GENDER DIFFERENCES IN NAVIGATION DIALOGUES WITH COMPUTER SYSTEMS

A thesis submitted for the degree of Doctor of Philosophy

by

Theodora Koulouri

School of Information Systems and Computing

Brunel University

January, 2013

Abstract

Gender is among the most influential of the factors underlying differences in spatial abilities, human communication and interactions with and through computers. Past research has offered important insights into gender differences in navigation and language use. Yet, given the multidimensionality of these domains, many issues remain contentious while others unexplored. Moreover, having been derived from non-interactive, and often artificial, studies, the generalisability of this research to interactive contexts of use, particularly in the practical domain of Human-Computer Interaction (HCI), may be problematic. At the same time, little is known about how gender strategies, behaviours and preferences interact with the features of technology in various domains of HCI, including collaborative systems and systems with natural language interfaces. Targeting these knowledge gaps, the thesis aims to address the central question of *how gender differences emerge and operate in spatial navigation dialogues with computer systems*.

To this end, an empirical study is undertaken, in which, mixed-gender and same-gender pairs communicate to complete an urban navigation task, with one of the participants being under the impression that he/she interacts with a robot. Performance and dialogue data were collected using a custom system that supported synchronous navigation and communication between the user and the robot.

Based on this empirical data, the thesis describes the key role of the interaction of gender in navigation performance and communication processes, which outweighed the effect of individual gender, moderating gender differences and reversing predicted patterns of performance and language use. This thesis has produced several contributions; theoretical, methodological and practical. From a theoretical perspective, it offers novel findings in gender differences in navigation and communication. The methodological contribution concerns the successful application of dialogue as a naturalistic, and yet experimentally sound, research paradigm to study gender and spatial language. The practical contributions include concrete design guidelines for natural language systems and implications for the development of gender-neutral interfaces in specific domains of HCI.

Contents

List of Figures

List of Tables

Dedication

I dedicate this work to my parents, Georgia and Michalis Koulouris, and brother, Dr. Theofrastos Koulouris, to whom I owe all the good things that I am and have done. Thank you for a lifetime of love.

Αφιερώνω τη διατριβή αυτή στους γονείς μου, Γεωργία και Μιχάλη Κουλούρη, και στον αδερφό μου, Δρ. Θεόφραστο Κουλούρη, στους οποίους χρωστάω όλα τα καλά πράγματα τα οποία είμαι και έχω κάνει. Σας αγαπάω με όλη μου την ψυχή.

Acknowledgements

First and foremost, my immense gratitude goes to my first supervisor, Dr. Stanislao Lauria, for his constant support and encouragement and his unwavering patience when I was holding unrealistic ideas. It has been a great pleasure.

I am also forever indebted to Professor Robert D. Macredie. It was a privilege to collaborate with him in several publications, and an unparalleled learning experience. Despite not being my supervisor, he was exceptionally generous with his time, expertise and thoughts, without which the outcome of this work would have been impoverished and its process a lot less enjoyable. His endless energy, empathy and genuine brilliance will serve as a source of inspiration for the rest of my academic career and, indeed, as a recipe for life. Thank you for the lessons.

I would like to thank the examiners of my thesis, Dr. Mark Perry and Professor Kenny Coventry. Their questions and approach during the viva transformed a notoriously terrifying experience to one that I could enjoy and benefit from – a true dialogue – and their feedback, detailed and insightful, led to greatly improving this thesis.

I would also like to express a special thanks to Ms. Ela Heaney for always having a solution to my administrative conundrums – her efficiency and sympathetic nature are extraordinary and make a difference to the Department of Information Systems and Computing.

I am also thankful to several other people from our Department; Dr. Sola Oni, Dr. Marije Kanis and Dr. Panagiotis Panagiotopoulos, for their words of encouragement and for being great examples of academics. Big thanks are also due to Dr. Gregory J. Mills, who helped me shape my initial ideas on dialogue, for being a great friend and a most ingenious interlocutor.

I am also grateful to all the researchers whose scientific efforts and contributions enabled me to develop my own work. These researchers are listed in the References section of this thesis. I extend my thanks to the editors and anonymous reviewers of my journal and conference papers, whose insightful comments improved this thesis. I'd also like to thank all people from the HCI, Dialogue Systems and Robotics communities, for sharing intelligent ideas, fun moments, food and drinks in Tokyo, Texas, London, Edinburgh and Derry.

There are also a few people who have made my every day more beautiful and every problem less serious: Konstantinos Iliadis, Galateia Iliadi, and Eftychia Asimaki.

I would particularly like to thank George Konstantinou for being the one to unwaveringly believe in me, to put up with me when I was impossible, to make me want to be the best I can, for being my one true source of companionship for twelve years. Thank you for all the fish.

Finally, I could not have done this without the unconditional love, kindness and support of my partner, Ricardo Canal. It would take more than 97,581 words to describe the ways that I love you

Publications

The research reported in this PhD thesis has produced the following publications:

Journals

Koulouri, T., Lauria, S., Macredie, R. D., and Chen, S. (2012). Are we there yet?: exploring the role of gender on the effectiveness and efficiency of user-robot communication in navigation tasks. *ACM Transactions on CHI*. *19*(1), ACM, New York, USA. 4:1 – 29

Koulouri, T., Lauria, S., and Macredie, R. D. Do (and say) what I say: linguistic adaptation in human-computer dialogues. *Human-Computer Interaction*. (accepted for publication).

Conferences

Koulouri, T., Lauria, S., Macredie, R. D., and Chen, S. (2012). Exploring the role of gender on the effectiveness and efficiency of user-robot communication in navigation tasks. In CHI2012. Austin, Texas, USA, 5-10 May 2012. (Invited TOCHI paper).

Koulouri, T., and Lauria, S. (2010). Route communication in dialogue: a matter of principles. In Proceedings of the SIGDIAL 2010 Conference: the 11th Annual Meeting of the Special interest Group on Discourse and Dialogue. Tokyo, Japan, 24-25 September 2010, Association for Computational Linguistics, Stroudsburg, PA, USA. 95-98.

Koulouri, T., and Lauria, S. (2009). A corpus-based analysis of route instructions in human-robot interaction. In Proceedings of Towards Autonomous Robotic Systems (TAROS09). Londonderry, UK, 31 August – 2 September 2009. 281-288.

Koulouri, T., and Lauria, S. (2009). Exploring miscommunication and collaborative behaviour in human-robot interaction. In Proceedings of the SIGDIAL 2009 Conference: the 10th Annual Meeting of the Special interest Group on Discourse and Dialogue. London, United Kingdom, 11-12 September 2009. M. Purver, (Ed). Association for Computational Linguistics, Stroudsburg, PA, USA. 111-119.

Koulouri, T., and Lauria, S. (2009). A WOz framework for exploring miscommunication in human-robot interaction. In Proceedings of the AISB Symposium on New Frontiers in Human-Robot Interaction. Edinburgh, UK. 8-9 April 2009. 73 – 80.

Workshops

Koulouri, T. (2010). Position Paper. In Proceedings of the 6th Young Researchers' Roundtable on Spoken Dialogue Systems. Tokyo, Japan, 22-23 September 20010. pp. 43-44.

Koulouri, T. (2010). Abstract. Where do we go from here?: an experimental investigation in route instruction dialogues. CLS 2010 Junior Researchers' Workshop. Zadar, Croatia, 28-29 August 2010.

Koulouri, T. (2009). Position Paper. In Proceedings of the 5th Young Researchers' Roundtable on Spoken Dialogue Systems. London, UK, 13-14 September 2009. pp. 49 -50.

1 Introduction

This chapter aims to lay the foundations of the thesis. It outlines the area within which this thesis is situated and indicates the research gaps, which motivate its central research question. Teasing apart the research problem, a number of specific research questions are articulated under three main themes. Then, the methodology is briefly described and justified, and the expected contributions are listed. The chapter concludes with an overview of the thesis, and a graphical outline that should serve as a preview of the ensuing chapters. On these foundations, this thesis proceeds with a detailed description of the theoretical and empirical research undertaken.

1.1 Navigating the domain of gender differences

Attesting our common phylogeny with the rest of roving animals, humans possess extraordinary abilities to locate targets in space, perceive distance, directional and spatial relations, perspectives and the structure of objects, and to use this information to traverse and manipulate objects in the world. What sets us apart, however, is the ability to talk about objects and places, and use this information to direct someone else's attention to them. Spatial language is among the first to emerge in children evidencing its central role in the development of human cognition (Casasola, 2008). Spatial language is simply defined as the words and phrases in human language used to encode objects and space, their location, motion and properties (Landau and Jackendoff, 1993; Landau, 1998).

How people traverse and talk about space has rich implications for theoretical fields and the design of practical systems. Research in these areas has resulted in the development of theories of human communication, cognition, and behaviour and informs the design of

computer applications and user interfaces, including Geographic Information Systems and dialogue systems for robot navigation. For instance, the capability to generate and process route instructions is essential for assistive systems, as exemplified by the robotic wheelchairs of the Diaspace program (Shi and Tenbrink, 2009) and the IBL project (Lauria et al., 2001), and robots used for rescue/exploratory purposes (for example, (Lemon et al. 2002)). The domains of spatial abilities and language afford great variation among individuals which strongly correlates with gender.

Mainstream media and virtually every scientific field share an interest about gender and sex differences. Yet, as Money (1987, p.13) states, 'the difference between male and female is something that everybody knows and nobody knows'. The realisation that 'gender' is something we do, rather than something we 'are' led researchers in psychology and relevant fields to focus on 'gender' differences, making a crucial distinction between 'sex' (the biological identity of men and women) and 'gender' (the behavioural identity). In effect, this distinction enabled external observation and understanding of how gender modifies performance in daily tasks (Chrisler and McCreary, 2010, p.2).

According to popular belief, enormous psychological and cognitive differences exist between males and females. Research from diverse fields such as psychology, neuroscience, education, marketing, economics and computing has confirmed that males and females communicate and process information differently (for instance, see Kucian et al., 2005; Beckwith and Burnett, 2004).

Halpern et al. (2007) present a comprehensive review and evaluation of scientific research in gender differences in cognitive abilities that are important for the fields of science and engineering. They conclude that while measuring cognitive (including visuospatial and verbal) abilities is a multifaceted (mediated by factors such as age, ethnicity, research methodology) process that involves difficulties and controversies, men appear to have superior visuospatial abilities, but with higher inter-individual variability, and women typically possess better verbal skills. It is argued that the most prominent differences between females and males are observed in the upper end of the ability distribution, whereas differences are attenuated for females and males with abilities around their population average. The paper discusses possible origins of gender differences – identifying biological, evolutionary and social/environmental contributions, with the aim to understand the factors

behind and offer solutions for the underrepresentation of women in science and engineering careers.

On the other hand, a recent analysis of forty-six meta-analyses on studies of the last two decades argues that claims of gender differences have been inflated (Hyde, 2005). It reveals that the magnitude of differences is close to zero or small for many parameters, including types of cognitive abilities, aspects of verbal and nonverbal communication, aggression, leadership, self-esteem, moral reasoning and motor behaviour. It is also argued that gender differences are diminished or amplified depending on the factors of context and age. Nonetheless, there are three areas in which gender differences have been consistently found.

First, Hyde's review corroborated that spatial abilities remain one of the few domains of cognition in which gender differences are reliably detected. *Prima facie*, results in terms of performance in spatial tasks generally favour males (Lawton, 2010, p.317); more precisely, while a strong male advantage has been observed in psychometric tests, results are less conclusive in real-world tasks, such as navigation (Coluccia and Losue, 2004). At the same time, spatial abilities is not a unitary field resulting in various definitions and methodologies being employed, which, in turn, has added to the controversy surrounding the existence and magnitude of gender differences in spatial abilities (Lawton, 2010, p. 318; Hausmann and Schober, 2012; Caplan et al., 1985).

Second, gender is argued to be a major factor underlying language abilities (Ullman et al., 2008). In particular, it is broadly agreed that women possess superior verbal skills, although differences do not uniformly arise in all language dimensions – females primarily excel in tasks involving verbal memory and word retrieval, with less consistent advantage in other tasks (Kimura, 1999, p.11). In addition to performance differences, research has identified qualitative differences in both domains of spatial and verbal abilities. In particular, males and females appear to rely on different strategies to navigate themselves and others. Men generally formulate instructions using the cardinal system and metric distance, whereas women use proximal landmarks when giving and following instructions (Lawton, 1994). Moreover, gender differences in communication style, use of linguistic elements and level of participation have been reported in both settings of face-to-face communication and Computer-Mediated Communication (CMC) (see, for example, Crowston and Kammerer, 1998).

Third, literature in the field of computing has recognised that there are important differences in the ways females and males interact through and with technology. These differences pertain to skills, performance outcomes, perceptions and attitudes across numerous domains of Human-Computer Interaction (HCI) (Chen and Macredie, 2010). Despite this realisation, our understanding of the ways that gender interacts with the characteristics of the technology and mediates its effectiveness and acceptance remains rudimentary (Burnett et al., 2011). As a result, the design of systems continues to exclude gender considerations (Bardzell, 2010). Ultimately, by making surface or ad-hoc decisions, developers are bound to create technology that is appropriate for some users, while marginalising the needs and preferences of another user group.

The complexity in the findings presented so far suggests that there is a need for further targeted and systematic investigation of gender-related differences in the domains of spatial abilities, language and HCI. The thesis seeks to contribute to the existing corpus of research through an empirical study of how females and males coordinate and communicate route information, with the focus on the interactions with spatially-aware dialogue systems. Central to the claims developed in this thesis is that insights gained through empirical investigation of task-oriented dyadic interactions can be of immediate relevance for the design of practical systems. This argument is justified in the next section.

1.2 Gender differences in dialogue with systems

There has been significant and sustained research over the last three decades into the design of natural language user interfaces, embedded in dialogue systems, robots and embodied conversational agents, to support goal-oriented use of computer systems (Jokinen and McTear, 2010, p.10). Despite widespread predictions of success, these systems have yet to enable effective, efficient and natural interactions with the user. This failure has been at least partly attributed to insufficient understanding about how users will address the system or, indeed, what people really do when they communicate (Fischer, 2006). Similarly, relatively little is known about the design and nature of the computer as an interlocutor itself (Porzel, 2006). It is argued that empirical studies of human communication that investigate inter-individual coordination in spontaneous task-oriented dialogue have the potential to provide important insight for the development of successful natural language user interfaces to

computer systems. There are several examples of implementations based on findings from such studies (such as the TRINDIKIT (Larsson and Traum, 2000), Galatea (Skantze, 2008) and RavenClaw (Bohus and Rudnicky, 2009)).

Most language, including spatial language, naturally occurs in dyadic interaction, that is to say, dialogue. For instance, we hardly ever produce route instructions without an intended recipient. Communicating route knowledge is a collaborative, goal-oriented process, anchored in a specific spatial and temporal context. This makes it a prototypically dialogic situation. Spatial language is a lively area of research, but, surprisingly, the overwhelming majority of studies have investigated language production and interpretation in isolation, in monologue and often in artificial settings (Coventry et al., 2009, p.3). Such isolated study of language fails to take into account that language is dynamic, adaptable to the context of use and emerges as a function of inter-individual processes. A normal dialogic situation is more than an information transfer between speakers. In empirical studies of human communication on which we can draw in understanding how to model and inform the design of user-system interaction, language is seen as a collaborative activity in which partners introduce, negotiate, and accept information (see the Interactive Alignment Model of Garrod and his colleagues (Pickering and Garrod, 2004), and the Collaborative Model of Clark and his colleagues (Clark, 1996)). There are four additional reasons of practical significance that motivate the study of language in dialogue.

When interlocutors introduce and accept information, they perform a coordination process known as ‘grounding’ (Clark and Marshall, 1981), that is, they mutually establish that what has been said has also been understood. The form and precision of grounding are determined by the affordances of the interaction condition. For example, in conditions of physical/visual co-presence, interlocutors share visual and auditory common ground and, as such, grounding and, as a result, the interaction becomes less effortful. These phenomena of human communication also emerge in HCI; indeed, studies have shown that collaborators who shared visual information in CMC and computer-supported cooperative work (CSCW) had more efficient interactions (Gergle et al., 2004; Kraut et al., 2003). Such findings exemplify that the theoretical understanding of human coordination processes can lead to awareness of how to better support interactions with collaborative systems.

Second, empirical research in dialogue foregrounds the phenomenon of linguistic alignment. In particular, people naturally align to each other's vocabulary, sentence structure and acoustic features in dialogue. Alignment is argued to be a basic interactive mechanism that takes place in dialogues at all levels – phonetic, phonologic, lexical, syntactic, semantic and pragmatic – and that makes communication between people 'easy', efficient and effective (Pickering and Garrod, 2004; Garrod and Pickering, 2004). From a practical perspective, alignment, as a mechanism that promotes language repetition, may be exploited in system design to predict and constrain user input as well as yield more natural and felicitous interactions. There has been a growing interest in alignment in the interaction between users and computer systems, initiated by Branigan and her colleagues (see Branigan and Pearson, (2006) and Branigan et al. (2010)). But, given the originality of this research, these studies have been confined to the investigation of lexical and syntactic alignment in simple picture-naming tasks. No attempt has been directed towards identifying alignment as it operates and develops in the course of the dialogue with the computer in a realistic task. Thus, little is known about how this mechanism operates in human-computer dialogues, let alone how it may be exploited to improve the efficiency of the interaction with the systems.

Third, viewing language production and interpretation as two autonomous processes has also precluded the observation of the natural and ubiquitous phenomenon of miscommunication. Yet, miscommunication (manifesting as system execution errors, non-understandings and incorrect user commands) is pervasive and possibly formative in the interactions with humans and systems.

Finally, the results of Hyde's review of meta-analyses in gender differences suggested that contextual factors influence the magnitude of gender differences. The author presents the theoretical argument that gender differences may be moderated, exacerbated or even reversed due to dyadic interactions between participants (Hyde, 2005). This argument appears stronger in light of the empirical research in dialogue presented above, which postulates that interaction success is dependent on the inter-individual processes of alignment and negotiation. Nevertheless, the vast majority of findings on gender differences in spatial abilities and navigation have also been derived from studies that have used non-interactive settings. As such, while this thesis does not question their theoretical value, it argues that their generalisability to interactive contexts of use, particularly practical settings of HCI, may be limited.

1.3 Research overview

The brief review of the literature illustrated that gender plays a major role in how people interact with each other and with their artefacts, with robust gender differences manifesting, *inter alia*, in the areas of spatial cognition and language use. Yet, given the multidimensionality of these domains, many issues are controversial while others remain unexplored. Moreover, empirical research has demonstrated that dialogue fundamentally changes performance and communication patterns. In this light, existing findings from non-interactive studies may provide incomplete accounts of gender differences, and, for many practical purposes, incorrect. As such, this thesis reframes the problem as an empirical question of *how gender differences emerge in spatial navigation dialogues with computer systems*. By addressing this question, the thesis aims to produce implications for communication theory development as well as ecologically-valid design guidelines for the development of collaborative systems and natural language interfaces.

1.3.1 Research questions

Decomposing the elements of the central research question, the thesis formulates a number of specific research questions. These research questions target the knowledge gaps that emerged from the detailed literature analysis presented in the next chapter. The research questions are grouped under three main themes:

A. Gender differences in performance and route communication in interaction.

This set of questions aims to draw an initial picture by testing predictions from non-interactive studies with regards to gender differences in **(i)** performance and **(ii)** route communication as well as **(iii)** user perceptions of the interaction. Do these predictions also apply in dialogue between humans and between humans and computer systems?

B. Effect of visual information on performance in HCI and communication and the effect of its absence by gender.

This set of questions seeks to **(i)** clarify the benefit of the availability of visual information on performance and communication in a novel HCI domain and, then, **(ii)** determine the effect of

its absence on females and males. Namely, some of these questions address whether the performance of females or males is more adversely affected and whether one gender adapts their strategies more drastically than the other gender in response to the visually-impoverished interaction condition.

C. Alignment in HCI and gender-related alignment in task-oriented interaction.

This set of question aims to **(i)** describe the phenomenon of linguistic alignment in human-computer dialogues and, then, **(ii)** establish whether its strength depends on gender.

1.3.2 Research methodology

The work presented in this thesis sets to investigate gender differences in navigation and communication in real-time dialogue with a system. The thesis is fundamentally data-driven and draws on experimental paradigms within the language-as-action tradition that investigate inter-individual coordination processes in task-oriented dialogue (Interactive Alignment Model and Collaborative Model, as presented above). The thesis developed a data collection and analysis approach oriented towards the research problem and the specific research questions.

The experimental study deployed a Human-Robot Interaction (HRI) navigation task. It employed a Wizard-of-Oz setup, motivated by similar HRI studies within the Diaspace and IBL projects and the previous studies in alignment in HCI (for example, Branigan et al., 2003; 2004). The study involved same-gender and mixed-gender pairs collaborating to complete the navigation task, with the user being under the impression that he/she was instructing a robot. This setup also served to observe inter-gender interactions, while inhibiting social elements that arise in human-human interactions and adversely affect spatial performance and communication (Picucci et al., 2011; Inzlicht and Ben-Zeev, 2000). A custom system was developed to support the simulation and enabled synchronous navigation and communication between a user and a robot, and the monitoring of the unfolding dialogue in context. The study included two experimental conditions, in which the participants could or could not monitor the ‘robot’s’ actions (this corresponded to the absence or availability of visual information as related to the second set of research questions).

The data analysis approach involved the fine-grained analysis of the performance and dialogue data. For the performance analysis, commonly-employed objective metrics complemented by user perceptions were used. For the dialogue data, it developed an analytic framework that integrated dialogue act analysis (following the HCRC Map Task scheme by Carletta et al., 1996), component-based analysis of the utterances and route instruction classification (following Tenbrink et al. (2010) and the CORK framework by Vanetti and Allen (1988)) and miscommunication analysis (following common definitions provided within previous models (Clark, 1996; Allwood, 1995; Hirst et al., 1994)).

The study implemented a between-subjects factorial design that investigated the simple and interaction effects of User Gender, Robot Gender and Visual Information on the performance-based and dialogue-based dependent variables.

1.3.3 Contributions

This thesis expects to produce several contributions with theoretical, methodological and practical implications.

From a theoretical perspective, the thesis aims to add to our understanding of gender differences in navigation performance outcomes and in qualitative differences in terms of language use and communication strategies. Second, original insight is expected to emerge with regards to previously unexplored interactive processes, that is to say, linguistic alignment and miscommunication, and their interplay with gender. Third, using models of human communication as its theoretical foundation, the thesis should help clarify relevant principles of coordination and communication, such as grounding. Fourth, the thesis aims to make specific contributions to route communication protocols, complementing them by detailing how people produce route instructions in real-time dialogue.

The methodological contributions result from the use of dialogue. The thesis aims to illustrate the validity and value of a dialogue paradigm that permits the collection of naturalistic data under experimentally controlled conditions. Second, the thesis involved the development of a complete data analysis framework based on established methods, classification schemes and theoretical models of communication, which is hoped to be of use to future studies in the same domain.

The findings of this thesis are expected to have practical significance for the field of HCI by identifying and describing gender differences in the novel domains of HRI and dialogue systems. Second, as it involves the concealed computer-mediated collaboration between two people, the thesis may provide insight in how gender differences arise in task-oriented CMC and CSCW. Most importantly, this thesis seeks to define the interactive mechanisms that enable users of both genders to coordinate and achieve their interaction goals. Third, having used visual information as an experimental manipulation, the findings of this work should also add to our knowledge of how it can be exploited to facilitate remote collaborations through technology. Finally, the thesis seeks to contribute to HCI by presenting an account of alignment in human-computer dialogue and enumerate design recommendations for systems with natural language interfaces that can leverage the potential of alignment towards more natural and efficient interactions.

These contributions are only a starting point in addressing the complexities and unresolved questions regarding individual differences in HCI. Their significance should lie in their ability to trigger further scientific research. Investigation of gender differences in conditions that approximate real interaction settings will not only help determine which interface features are suitable for users of both genders but also uncover the natural behaviours and strategic mechanisms that produce high performance outcomes and user experience. Designing features that promote such behaviours and mechanisms may hold the key to delivering systems that match the needs and preferences of all users.

1.4 Thesis overview

The overview of the thesis is provided below. It is also presented diagrammatically in Figure 1.1. This diagram presents the contents and inter-connections within each chapter and should serve to outline the work in this thesis. The structure of the thesis aims to narrate the research in a reflective way, illustrating the insights that motivated the choices made and the transitions from the theoretical background to the recommendations for future research.

Chapter 2 sets the theoretical foundations of this thesis by reviewing relevant literature in the two major themes of the thesis: *gender* and *dialogue*. First, it examines existing knowledge in gender differences in spatial abilities and navigation performance, verbal abilities and

communication styles, and HCI, with a special focus on the domains of spoken dialogue systems (SDS), virtual world navigation, HRI and CMC. Then, it continues, with the investigation of dialogue phenomena, drawing on the theories and principles within the Interactive Alignment Model and the Collaborative Model. In particular, it reviews literature in alignment in human communication and HCI; the effect of visual copresence in coordination mechanisms in CMC and task-oriented human communication; and, finally, miscommunication as it arises between humans, between humans and dialogue systems and robots and between males and females.

Chapter 3 builds on the analysis presented in Chapter 2 and identifies specific research gaps as they pertain to the research problem. This process leads to the framing of three sets of research questions, as outlined in section 1.3.1 above.

Chapter 4 discusses and justifies the experimental methodology developed in order to address the research problem and specific research questions. As outlined in section 1.3.2 above, the experimental study involved pairs of participants collaborating in a simulated robot navigation task, using a text-based CMC system. It, then, details the quantitative and qualitative techniques used for the analysis of the performance and dialogue data.

Chapter 5 reports the results of the statistical analyses performed on the performance and dialogue metrics, which investigated main and interaction effects of User Gender, Robot Gender and Visual Information condition on the measures of performance and communication, and their associations. Finally, these results are summarised and formulated as high-level 'answers' to each of the research questions.

Chapter 6 discusses the findings of the empirical study and the 'answers' to the research questions in light of existing literature. The chapter distils the results of the study into design recommendations for spoken dialogue systems and provides observations to inform the design of collaborative systems, robots and web navigation.

Chapter 7 concludes the thesis and lists its main contributions. It discusses the limitations of the research, which motivate future work, and outlines additional areas that merit further exploration.

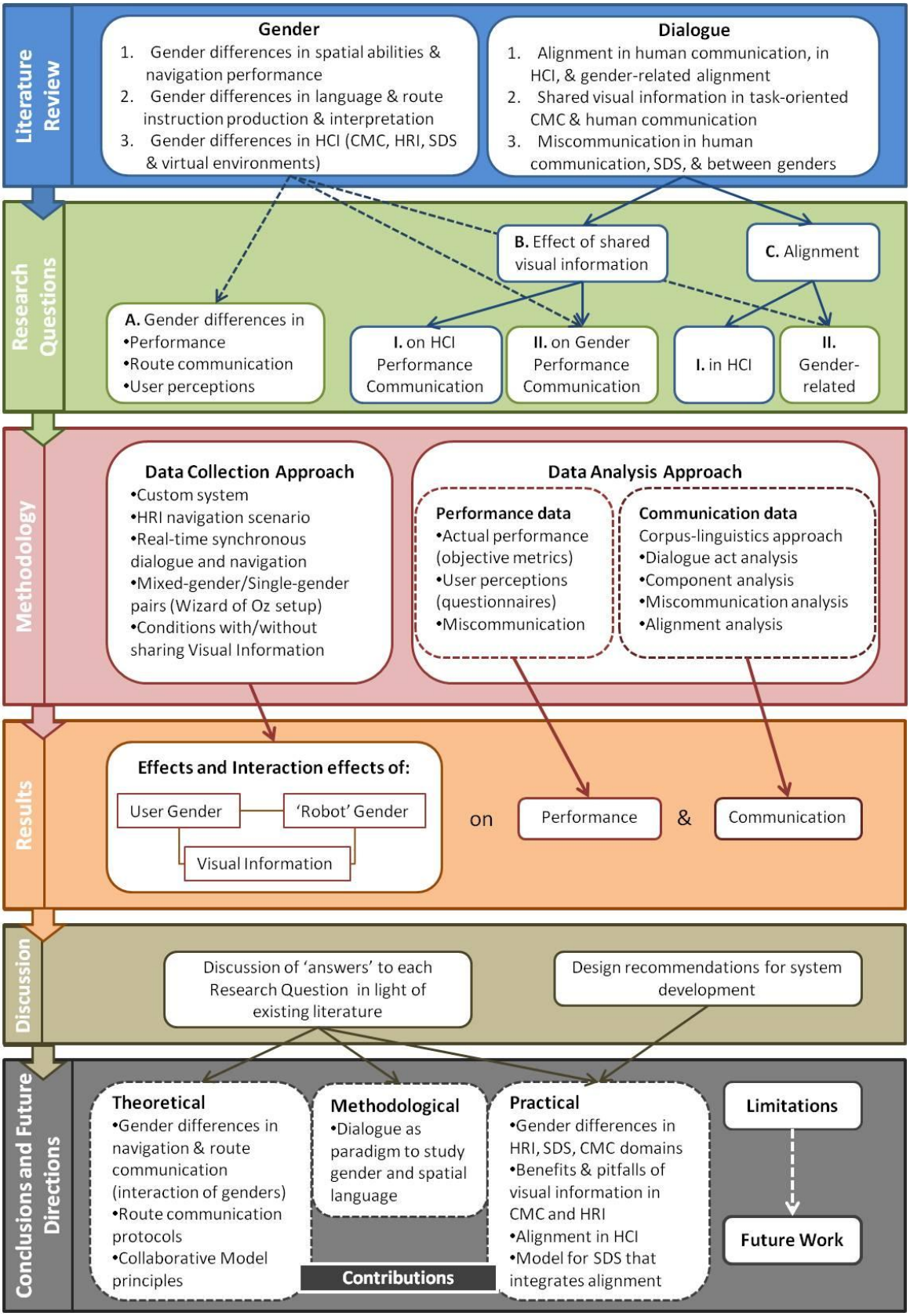


Figure 1.1: An outline of the thesis, showing contents, links and transitions.

2 Literature Review

2.1 Introduction

The previous chapter introduced the knowledge areas of gender and dialogue. Gender underlies performance and qualitative differences in the domains of spatial cognition, language and Human-Computer Interaction (HCI). Dialogue affects performance, coordination and communication patterns. This motivated the necessity of their joint investigation and led to the empirical question of *how gender differences arise in navigation dialogues with computer systems*. This chapter attempts to define in depth the area in which the research is situated and motivate the research questions of the next chapter through a critical analysis of the relevant literature. Due to the interdisciplinary nature of the research, the review draws from diverse domains and their subfields, including linguistics, cognitive and environmental psychology, and HCI.

The chapter dissects and integrates the two main themes of the thesis: (i) Gender differences occur in *navigation*, *language* and in the *interaction with computers*; and (ii) *Dialogue* – dyadic interaction – and the phenomena that emerge and develop over its course change performance and language patterns. As such, the following six sections of this chapter are dedicated to the discussion of existing work in these areas of gender differences. In the remaining sections, the focus is shifted to dialogue phenomena.

The chapter is organised as follows: section 2.2 defines and examines the types of spatial abilities with a view to disentangling the extent and magnitude of gender differences. Similarly, section 2.3 discusses gender differences in navigation, which is seen as the practical application of spatial abilities. It continues with a review of gender differences in the domains of language (section 2.4). Section 2.5 introduces the concepts and challenges of route instructions and discusses gender differences in their production and interpretation.

Section 2.6 outlines possible explanations with regards to the origins of gender differences. Section 2.7 discusses how gender becomes prominent in the interaction with systems. Section 2.8 examines existing knowledge in the mechanism of linguistic alignment in human communication and Human-Computer Interaction. Section 2.9 discusses visual co-presence and its effect on collaborative interactions. Section 2.10 explores miscommunication that naturally arises in the interactions with people, dialogue systems and between individuals of different genders. To facilitate reading, an interim summary is provided at the end of each section.

Research in gender involves debatable and controversial issues. As such, it is necessary to clarify the usage of term ‘gender’ of this study.

2.1.1 Distinction between ‘sex’ and ‘gender’

The distinction between the terms *sex* and *gender* in the English language has been encouraged by psychologists, sociologists and medical professionals since 1970. The APA offers the following guideline (APA, 2001, p.63): ‘*Gender* refers to culture and should be used when referring to men and women as social groups’, whereas *sex* ‘refers to biology and should be used when biological distinctions are emphasized’. This distinction is endorsed by World Health Organisation which states that ‘*sex* refers to the biological and physiological characteristics that define men and women. *Gender* refers to the socially constructed roles, behaviours, activities, and attributes that a given society considers for men and women’. Similarly, the Institute of Medicine (2001, p. 1) defines ‘*sex* as the classification of living things, generally as male or female according to their reproductive organs and functions assigned by the chromosomal complement, and *gender* as a person's self-representation as male or female, or how that person is responded to by social institutions on the basis of the individual's gender presentation. Gender is shaped by environment and experience’. In effect, the distinction made in the definitions above is that gender is a product of nurture and sex is associated with nature, leading to one of the fiercest debates across disciplines. However, as detailed above, spatial and verbal behaviour appear to be neither nature nor nature, but rather it becomes evident that it is an amalgamation of cultural and social factors, genetics and physiological and psychological adaptations. The complex interaction of these parameters may lead to differences in the observed behaviour of men and women. As such, scientists

have argued that emphasising the distinction between ‘gender’ and ‘sex’ in behavioural strands of science may be counter-productive. In particular, it is suggested that the distinction according to which sex is related to biology and gender is related to society/environment is no longer fruitful – and may, in fact, be artificial – because biological and environmental factors and their interdependencies determine the development of cognitive abilities (Halpern et al., 2007). A more useful approach to better understand differences between men and women is to pose simple questions such as ‘Is there a real difference? If so, is it associated with one’s sex? If so, how is it influenced by socialisation and culture?’ (Mills, 2011). In popular media, the terms gender and sex are often interchangeably used, so it is important that scientific publications ‘clarify the use of sex and gender’ (Institute of Medicine, 2001, p. 6). As such, while this thesis does not seek to make a distinction between gender and sex, it acknowledges that sex and gender are not interchangeable and equivalent. Following the majority of research in the field, it employs the term ‘gender differences’ as an umbrella term to signify differences in outcomes between men and women, with little focus on their biological or social origins. Echoing Halpern et al. (2007, p. 3) this is an arbitrary choice, made for the sake of clarity, which stems from the argument that the distinction between gender and sex relates to the unviable separation between what is biologically and what is socially determined. This decision is further examined in the last chapter of the thesis, as part of the discussion of limitations (section 7.4.1).

2.2 Gender differences in spatial abilities

Cognitive psychologists define three main classes of cognitive ability/intelligence: verbal, numerical and visuospatial (or simply spatial) ability (Halpern, 2000; Guttman, 1954). With implications for numerous theoretical and technical fields, research in spatial abilities began 100 years ago and the scientific interest has remained strong (Mohler, 2008). Since then, a multitude of definitions, spatial ability families and methods for measuring them have been proposed. Spatial ability refers to any ability involved in generating, representing, transforming and recalling spatial information (Linn and Petersen, 1985, p.1482). It practically manifests as these skills that allow us to process information about small and large-scale objects.; for instance, locating a pen under a 300-page thesis, deciding which way to turn to reach a destination and also being able to imagine what will be seen after

approaching a junction from a different direction (Ferrara et al., 2011). As previously noted, although gender differences in spatial ability are generally acknowledged, their locus and magnitude are still debated. Spatial ability is not a unitary process but encompasses a highly heterogeneous set of skills and, thus, its conceptualisation and experimental investigation vary between studies (Montello et al., 1999). The majority of research seems to agree on three broad components of spatial ability, *mental rotation*, *spatial orientation* (Montello et al., 1999) and *spatial visualisation* (Linn and Petersen, 1985; Voyer et al., 1995; Mohler, 2008; Colom et al., 2002). Other researchers distinguish spatial orientation and spatial visualisation as two categories of spatial ability, the latter incorporating mental rotation (Hegarty and Waller, 2004). Neither is there a consensus in the classification of measures of spatial ability, but several spatial, cognitive tasks have been developed to address each of these categories.

In the following paragraphs, gender differences in the three main areas of spatial ability, mental rotation, spatial orientation and visualisation, will be discussed. Navigation in real and virtual environments is often seen as the practical application of spatial abilities and will be addressed separately. While the main domains of spatial ability have been studied by cognitive psychologists and psychometricians, navigation and wayfinding have been the focus of interest of researchers in applied fields as diverse as geography, environmental sciences and engineering.

Mental rotation refers to the ability to mentally rotate two- or three- dimensional objects[1]. Typical tests in this category include the original Mental Rotation Test by Shephard and Metzler (1971) or by Vandenberg and Kuse (1978), the Primary Mental Abilities test by Thurstone and Thurstone (1958) and the Card Rotation Test (Ekstrom et al., 1976). The tests can be timed or untimed. Two extensive meta-analyses of gender differences in spatial task performance by Linn and Petersen (1985) and Voyer et al., (1995) conclude that the most robust gender differences are located in mental rotation tests, with moderate to large effect sizes between 0.48 – 0.90. Boys and men consistently perform faster and more accurately in this task. Studies with infants show differences emerging as early as five months (Moore and

[1] Several researchers refer to the mental rotation ability as spatial relations (Lohman, 1979; Colom et al., 2002).

Johnson, 2008). A recent study by Silverman et al. (2007) confirmed a male advantage in mental rotation in participants of thirty-five countries, providing support to the claim that it is 'universal across regions, classes, ethnic groups, ages, and virtually every other conceivable demographic variable' (Eals and Silverman, 1994, p.95).

Spatial orientation[2] refers to the ability to imagine how an object would look from different perspectives (orientations) of the observer. A widely used measure is the Spatial Orientation Test (Guildford and Zimmerman, 1948) and a number of studies have reported significantly better performances by males in this test (Moffat et al, 1998; Tan et al., 2003). Gender differences in this type of spatial ability have been moderate compared to mental rotation (Linn and Petersen, 1985; Voyer et al., 1995; Hyde, 2005), with no consistency across different types of spatial orientation tasks.

Spatial visualisation is the ability to recognise the parts of an object if they were moving or displaced from their original position (Mohler, 2008). Typical tests include the Embedded Figures Test, which involves discerning a figure within a complex pattern, Paper Folding, in which participants are asked to predict how a folded piece of paper would look when unfolded, and Differential Aptitude Test-Spatial Relations, in which blocks are used to construct a particular shape (Bennett et al., 1956). No significant gender differences have been detected in this category(Linn and Petersen, 1985).

Two additional categories of spatial ability have been only recently described in literature, yet they offer ground on which gender differences arise. First, *dynamic spatial ability* encompasses the skills to perceive and extrapolate real motion, predict trajectories of moving objects and estimate their time of arrival (Colom et al., 2002). Studies employing the relevant tests suggested that men are more accurate than women. However, a number studies have suggested that, in addition to differences in ability, strategic elements (how people choose to go about solving the task) contribute to the outcome. In particular, some people display an impulsive behaviour and follow a 'trial and error' strategy, while others are more conservative and wait until they are certain before making a decision about the situation at

[2] In other classifications, spatial orientation is equated to or incorporated into a category, called spatial perception (Mohler, 2008; Linn and Petersen, 1985).

hand (Contreras et al., 2007). The second area is *object location memory*. It is argued that this is the only spatial ability in which women outperform males (Honda and Nihei, 2009), and persists across different countries (Silverman et al., 2007) and age groups. The first study to make this observation was conducted in 1992 by Silverman and Eals, and the results have been replicated through various experimental setups (Spiers et al., 2008; De Goede and Postma, 2008). A meta-analysis of 36 studies by Voyer et al. (2007) confirmed gender differences favouring females in object identity memory tasks and object location memory tasks. Yet, differences disappear for tasks in which abstract objects were used, suggesting that the female advantage in previous studies tapped visual memory and not memory of location (spatial memory) (Rahman et al., 2011). In view of a general female superiority in linguistic tasks (discussed in detail in section 2.4), scientists suggest that women score higher than men because they are better in remembering object names. It is noteworthy that several studies failed to identify any differences in object relocation tasks (Iachini et al., 2005; Postma et al., 1998; James and Kimura, 1997). Moreover, the female advantage diminishes when the objects are not presented in front of the participants, but projected on a wall (Saucier et al., 2007). The authors associate the finding with the distinct abilities to perceive space close to or further away from one's body. It also relates to different navigational strategies employed by women and men; namely, women are reported to rely on proximal landmarks for navigation, whereas men use distant landmarks (Lawton, 1994) (discussed in the first part of section 2.3.3).

Differential research is extensive and growing. *Prima facie*, research from cognitive sciences appears to converge on a male advantage in psychometric spatial tasks (Kimura, 1996). However, this literature analysis revealed that the spatial domain is multifaceted, and a male advantage is not uniform across the domain. The review of the literature confirmed robust differences favouring males in mental rotation, moderate differences in orientation and dynamic spatial ability and minimal differences in spatial visualisation, whereas females appear to outperform males in object location memory tasks. The mental rotation test and the other psychometric tests mentioned above are highly popular among psychologists due to their consistency, simplicity to administer and cross-reference between other tests. However, although their advantages and value are undeniable, such tests have a rather abstract connection to real-world tasks and how spatial abilities are applied in every day contexts. In addition, there is a conspicuous lack of established definition and conceptual taxonomy for

spatial abilities and measures, which has led many scientists to question the validity of research findings (Caplan et al., 1985). However, even small differences between males and females in spatial aptitudes are argued to be of great practical consequence holding a predictive power for performance and experience in numerous activities and areas of life (Halpern, 2000).

2.2.1 Closing the gap in spatial abilities

There is a lively academic debate with regards to whether there are gender differences in cognitive abilities. Research is trying to address questions whether gender differences are 'innate' (due to brain organisation or hormonal differences), are the product of differences in socialisation, or a combination of both (discussed in section 2.6). Substantiating gender differences in spatial abilities is not a goal in its own right. The importance of these academic endeavours lie on the fact that visuospatial abilities are shown to underlie choice and achievement in science and engineering, mathematics and computing fields. Thus, research conclusions will be used to shape public policies (Halpern et al., 2007). In addition, there is a risk that such findings will be used to provide scientific ground to stereotypes and discourage women to select professional careers in these disciplines that depend on high spatial abilities (Brownlow et al., 2011; Chipman, 2005; Kinsey et al., 2008). Therefore, the outcomes of such research are not only of theoretical relevance but also of pressing practical significance. They should be used to develop methods that improve spatial abilities, which should serve in addressing the underrepresentation of women in scientific and technological fields.

Empirical evidence and results from longitudinal studies support that high spatial ability is a strong predictor of attainment in the fields of Science, Technology, Engineering, and Mathematics (Casey et al., 1995; Shea et al., 2001; Wai et al., 2009). Irrespective of the origins of gender differences, studies have shown that spatial ability can be enhanced through experience and training. In particular, it is established that training leads to improvements in spatial skills for men and women, children and adults. Moreover, training benefits do not fade over time, but improvement in performance persisted three months later. Most importantly, spatial skill training is transferrable, leading to improvements in novel tasks (Uttal et al., 2012). Studies also show that the gender gap in spatial task performance can be decreased or even eliminated. In particular, after several hours of playing action video games women were

able to surpass their previous scores and match men's performance in tasks of mental manipulations, attention and rotations (Feng et al., 2007; Subrahmanyam and Greenfield, 1994). A study by Saccuzzo et al. (1996) illustrated that, although practice in a particular set of psychometric tasks improved the performance of all participants, women improved at a faster rate and closed the previous performance gap with men. Similarly, in Terlecki et al (2007), improvement for females was overall greater than males but peaked much later in the training process. It should be noted that many studies did not conclude that women's post-training performance reached that of men. However, sufficiently closing the spatial task performance gap may also be sufficient to close the gap in the entrance, accomplishment and retention in mathematical and engineering fields. The feasibility of this proposition was exemplified by a ten-year project at Michigan Technological University. The project involved the development of multimedia software to improve the 3-D spatial visualisation skills of engineering students and its integration in an academic course. The results of the project included higher grades in follow-on courses and an increase in retention of female engineering students (Sorby, 2007).

Finally, understanding the true nature and underlying causes of gender differences in spatial task performance can not only invalidate stereotypes of female disadvantage but also lead to straightforward solutions. For instance, numerous studies have shown that women report low confidence and high anxiety when performing spatial tasks, such as wayfinding (Picucci et al, 2011; Lawton and Kallai, 2002; Malinowski and Gillespie, 2001). Recent findings confirm that increasing confidence and reducing the levels of anxiety in women hold the potential to improve their performance (Brownlow et al., 2011; Moè and Pazzaglia, 2006). The topic of spatial anxiety is revisited in the final part of section 2.3.3.

Halpern et al. (2007) reviewed and evaluated scientific evidence with regards to the magnitude and origins of gender differences, with the aim to explain and address the gap in achievement in math and science fields. The review revealed a complex network of variables ranging from early experience and biological constraints to educational policy and cultural context which interact and influence cognitive ability performance and career choices. These factors are discussed in section 2.5.

2.2.2 Spatial abilities beyond psychometric tests

The vast majority of studies explore spatial abilities and gender differences using lab-based tests, while much fewer are based on real-world, 'environmental' tasks (Malinowski and Gillespie, 2001; Hegarty et al., 2006). Environmental tasks include map learning, navigation in indoor or outdoor environments, and producing and interpreting route instructions (Hegarty et al., 2006). There are some difficulties associated with exploring spatial ability at this scale. Environmental tasks are much harder to run, control and cross-reference. Participants might have prior or various degrees of familiarity with the environment. All real-world tasks activate a variety of cognitive abilities, which is considered problematic by many researchers. On the other hand, important questions arise whether abstract tasks like mental rotation evoke spatial abilities applied in a real-world situation. These measures are too rigid to capture the richness of everyday, practical spatial activities and spatial behaviour (Montello et al., 1999). It may, thus, be misleading to assume that the magnitude and characteristics of gender differences that arise in 'paper-and-pencil' tests can be readily extrapolated to more realistic activities.

2.2.3 Interim summary

There appears to be a male advantage in spatial abilities as measured by psychometric tests. Due to the difficulty in demarcating the field of spatial abilities combined with the multiplicity of methodologies employed, this advantage is not undisputed. Psychometric tests have also been criticised as being too unidimensional to hold predictive power with regards to every day, practical spatial activities. Navigation is said to be 'the most prominent real-world application of spatial cognition' (Wiener et al., p. 152, 2009) and, is, thus, the spatial task selected to serve as the test bed for gender differences in this study. The following section focuses on literature that investigates gender differences in navigation.

2.3 Gender differences in navigation

Spatial *navigation* is generally described as the 'coordinated and goal-directed movement through the environment' (Montello, 2009, p.163). It consists of two components, locomotion

and wayfinding. *Locomotion* encompasses the real-time navigation tasks that are responses to current sensory-motor input of the immediate surroundings, like steering, avoiding obstacles, identifying surfaces of support and moving towards visible landmarks. Locomotion can take the form of running and walking or performed using vehicles like cars, bicycles and aeroplanes. *Wayfinding* refers to tasks that require decision making, planning (which route to take) and orientation, involve the use of a mental or physical, external map, and aim at reaching a destination that is not yet visible. The constituent of navigation that is the focus of this work and the related array of disciplines is wayfinding. Following the taxonomy by Wiener et al. (2009), a further distinction can be made between *aided* and *unaided* wayfinding. Aided wayfinding is conducted with the assistance of an external representation, namely, maps, signs, route instructions or route planning and navigation systems on portable devices, whereas these tools are not employed in unaided wayfinding. Such distinction between wayfinding types is significant because of the different cognitive processes underlying them.

As previously noted, successful wayfinding is equated to reaching a goal destination efficiently. It requires knowledge about the actions to perform in order to traverse the path between the point of origin and destination (Klippel et al., 2003). This knowledge is the function of visual and sensorimotor experiences. It is acquired by direct experience, physically navigating the path to the destination. It can also be constructed indirectly, by accessing external media like maps, route sketches or visualisations (as is the case with in-car navigation systems and web route planner applications). Most importantly to this study, another external source of spatial information to support wayfinding is route instructions. There are two types of spatial knowledge, route and survey knowledge. Route knowledge is the ‘knowledge of linear sequences of landmarks connected by travel patterns; routes are ordered and contain minimal metric scaling’ (Montello, 2009, p. 164). Survey knowledge is the ‘knowledge of two-dimensional layout from which spatial relations among places can be determined even if travel between them has never occurred’ (Montello, 2009, p. 164). The latter is considered to be a more advanced form of spatial knowledge and is argued to be the wayfinding strategy typically employed by males (see discussion in the first part of section 2.3.3).

A correlation between performance in psychometric and navigation tasks has been hypothesised and often confirmed in literature (Allen et al., 1996). In particular, superior

performance in mental rotation is related to navigational ability (Galea and Kimura, 1993; Moffat et al., 1998; Saucier et al., 2002), ability to use Euclidean information (Saucier et al., 2002) and orientation (Silverman et al., 2000). Yet, functional imaging studies indicate that different areas in the brain are activated for navigation or mental rotation tasks (Iaria et al., 2008). Apart from mental rotation, the correlations between navigation and performance in all other psychometric tests are low (Moffat et al., 1998). The findings also suggest that this correlation between psychometric and navigation tasks is stronger in simulated environments than in real environments (Hegarty et al., 2006).

Despite the abundance of popular accounts of gender differences in wayfinding, research outlines an intricate pattern of evidence (Coluccia and Louse, 2004). Once again, the often conflicting findings may be attributed to task-specific and methodological issues. Navigation is a complex activity that relies on more than one cognitive ability and could tap on all three components of spatial abilities. Studies may employ a variety of measures and settings. Typical tasks include actual wayfinding, recalling landmarks and routes, sketching routes, giving route instructions and orientation activities like indicating directions of landmarks, estimating distances, reading and interpreting a map and learning a map or route. These tasks can be conducted in various settings, such as maps, mazes and real, virtual or table-top scale models of indoor or outdoor environments. Many studies also examine people's experience and perceptions through surveys. Finally, probably the largest portion of differential research does not deal with performance but rather focuses on qualitative (that is, strategy or 'stylistic') preferences (Saucier et al., 2002; Lawton, 1994; Schmitz, 1997).

Coluccia and Louse (2004) conducted a meta-analysis of studies on gender differences in navigation published from 1983 to 2003. They found that in real-world navigation, males outperformed females in 58.8% of the studies, and equal performances were observed in 41.2% of cases. This pattern of results is replicated for navigation in virtual environments with 57.1% of studies supporting a male advantage, and 42.9% of them showing no differences. Yet, if only computer simulations are considered in which participants can move (omitting studies using video recordings and slide sequences), the number of studies favouring males climbs to 85.7%. In none of the reviewed studies, females performed better. The following sections detail findings of studies in gender differences in real and simulated environments.

2.3.1 Gender differences in navigation in real environments

Many studies replicate the results of psychometric tests in large-scale, real-world wayfinding tasks. Although gender differences are not as consistently found, when they are, they generally favour males. In a study by Malinowski and Gillespie (2001), participants oriented themselves in a forest using compass and map in order to locate ten points within four hours. The results indicated that men were more successful in locating points and quicker than women. Males were also more accurate in orienting and finding their way back to the point of origin within an unfamiliar building (Lawton et al, 1996). Silverman et al. (2000) reported the results from two similar wayfinding tasks, one in a building and the other in a wooded area, in which males had superior performance in all measures. In a study by Schmitz (1997), boys, aged 10 to 17, navigated a real-life maze more quickly than girls, but the author mentions that in a previous study, boys' speed compromised accuracy (Schmitz, 1995, as cited in Schmitz, 1997).

In 'off-the-field' wayfinding tasks, like map learning, reading, interpreting and sketching, some studies reported higher male accuracy and speed (Coluccia et al, 2007; Galea and Kimura, 1993[3]; Allen, 2000a; McGuiness and Sparks, 1983), whereas no quantitative differences were found in the reproduction of a previously traversed maze (Schmitz, 1997) and a furnished/unfurnished house (O'Laughlin and Brubaker, 1998). Moreover, tasks of pointing accuracy to landmarks either yield better scores for males (Lawton, 1996; Lawton and Morrin, 1999) or non-significant differences (Montello and Pick, 1993; Golledge et al., 1995). In Coluccia and Louse's (2004) review, studies favouring males or reporting similar performances corresponded to 64.3% and 35.7%, respectively.

2.3.2 Gender differences in navigation in simulated environments

Virtual environments refer to real-time graphical simulations with which the user can interact and control within a spatial frame of reference (Moshell and Hughes, 2002). They can vary in

[3] It should be noted that in the studies by Galea and Kimura (1993) and Coluccia et al. (2007), the subjects had to learn and reproduce a route or features on a map sketch, so no actual wayfinding was involved.

terms of their complexity, ranging from basic ‘desk-top’ interfaces to immersive displays that involve the operation of wearable or haptic controls. Simulated environments are increasingly being adopted for instruction, assessment and training purposes (Ross et al., 2006) and are invaluable research tools for testing hypotheses in applied and theoretical sciences. Navigation tasks set in simulated environments are an inexpensive and convenient alternative to real-world setups while being more ecologically valid than ‘paper-and-pencil’ spatial tests. A drawback of real-world experiments is that they allow for minimum experimental control. On the other hand, experimental conditions in a simulated environment can be clearly defined and reproduced, as the researcher has access to all overt sources of information available to the participant. Certainly, caution should be exercised when transferring conclusions from simulated to real-world spatial tasks. Yet, comparative studies noted similar performance and cognitive processes operating in navigation and orientation in real and virtual environments, and argue for the value of using virtual environments to measure environmental spatial ability (Ruddle et al., 1997; Richardson et al., 1999).

The male advantage in spatial tasks generally extends to simulated environments, although there are a few studies in which no differences were found (Tlauka et al., 2005). For instance, large differences between males and females have been found in maze navigation tasks (Moffat et al, 1998; Astur et al., 1998; Cánovas et al., 2008, Lövdén et al., 2007). Comparing the same tasks performed in real-world and simulated environment, it has been observed that male superiority is more pronounced in the latter (Waller, 2000; Coluccia and Louse, 2004). Once again, the attention is drawn to methodological and usability issues. That is, it is possible that the virtual environment itself and the environment-specific demands interfere with the performance. In this case, a poor score would not reflect differences in the ability to meet the task requirements, but lack of proficiency in navigating fluently the virtual environment. Navigation in a virtual environment adds layers of complexity to a normal navigation task by often requiring the coordinated operation of mouse, joystick and other sophisticated multimedia controls. Moreover, the real world contains objects and landmarks that can be easily perceived and facilitate navigation. This is not as straightforward in artificial environments which often results in users becoming disoriented and lost (Smith and Marsh, 2004; Dalgarno and Lee, 2010). In view of this, several studies also point to the fact that males have generally more extensive game playing experience so they are more likely to be familiar with virtual environments and interface control (Barnett et al, 1997; Coluccia and

Louse, 2004). Therefore, the amplification of gender differences may be attributed to prior experience. In many studies, attempts to mitigate this advantage were made by experimentally controlling this variable or by providing training to all participants (Castelli et al., 2008). Still, as pointed out by Martens and Antonenko (2012) and Waller (2000), the type and amount of training needed by each group of participants remain inconclusive.

Gender differences in virtual world navigation, as well as how they can be mitigated, are also considered in section 2.7.5, as part of the discussion of gender in the interaction with computer systems.

2.3.3 Interpretation of gender differences in wayfinding

Naturally, research has focused on specifying the factors that influence gender differences in wayfinding. These factors include cognitive factors, such as (i) the use of strategies that lead to a performance advantage and (ii) working memory capacity and demands, and (iii) psychological factors like anxiety. These three factors are discussed below. The topic of the origins of gender differences in spatial as well as verbal abilities is revisited in section 2.6 which explores socio-cultural and biological factors.

Navigation strategies

So far the literature review has shown that empirical findings obtained from 'paper-and-pencil' navigation tasks (for instance, Galea and Kimura, 1993), real-world navigation (Malinowski and Gillespie, 2001) and virtual environment navigation (Astur et al., 1998) generally oscillate from a male advantage to no differences. However, scientific focus appears to have shifted from measures of achievement to navigation strategies. Namely, differences in performance should not be attributed to superior spatial ability but to superior strategies, and, arguably, the competence (Saucier et al., 2002) or preference to use these strategies (Andersen et al., 2011). Converging findings from self-reports and experimentation show that females and males attend to different information and stimuli in the environment during navigation. In particular, women rely on route knowledge, that is, local landmarks and route turns. Their preferential use of landmarks has been correlated with their better object location memory (as described in section 2.2 above). On the other hand, men use survey

knowledge, that is, a global perspective based on spatial, geometric relations within the environment (Lawton 1994, 1996; Lawton and Kallai, 2002; Montello et al., 1999; Coluccia et al., 2007; Glück and Fitting, 2003; Lövdén et al., 2007). In the map-learning study by Galea and Kimura (1993), men were more accurate in recalling details about directions and distances on the map they reproduced, whereas women were better in recalling landmarks and street names. Similarly, in McGuinness and Sparks (1983), males excelled in recalling pathways while reproducing a map, whereas females recalled a larger number of landmarks and other map elements. Findings obtained from eye-tracking studies also showed that women fixate on landmarks longer than men (Andersen et al., 2011). The strategy employed by men accounts for their higher aptitude in configurational tasks, such as pointing and distance estimation (Ruggiero et al., 2008). In general, this wayfinding strategy is argued to be more efficient as well as more flexible and robust than the route perspective used by women (Saucier et al., 2002; Lawton, 1994; Hund and Padgitt, 2010). Moreover, since women do not exploit global cues, spending more time sampling landmark cues contributes to longer task completion times.

Landmarks are essential in women's navigation strategies, but also central in men's strategies. Sandstrom et al. (1998) found that men are able to use both landmark and geometric information, whereas females depend only on landmark cues. Their performance was similar when landmark information was available and relevant for task completion. However, male performance was not disrupted by the absence of landmarks and only transiently impacted when landmarks were available but misleading. In contrast, female performance was significantly impaired. Similarly, in a study in which participants used written instructions to navigate around and inside campus buildings, men performed best when they followed Euclidean-based route instructions, whereas women performed best when using landmark-based instructions. Yet, men and women performed equally well with landmark-based instructions (Saucier et al., 2002). After testing individuals in virtual maze navigation with or without landmarks, it was argued that an experimental setup devoid of landmarks would ultimately yield substandard performances by women (Andersen et al., 2011; Lövdén et al. 2007). The results supported that landmark-rich environments help eliminate gender differences, without impairing male performance. Furthermore, in wayfinding tasks aided by maps, performance of females is comparable to men's performance. Namely, in Coluccia and Louse's (2004) meta-analysis, the results were

balanced showing that 42.1% and 39.5% of the studies favoured males and females, respectively. According to Montello et al. (1999), maps offer the survey knowledge perspective that women lack, so viewing a map enables them to mitigate this disadvantage (see alternative interpretation of map aids in the next subsection below). Taken together, both genders utilise landmarks as a resource for navigation, but males appear to be capable to adapt better in topographies that lack landmarks.

Cognitive demands

Many studies attempt to explain why results from studies are inconsistent, either indicating no gender differences or a male advantage in navigation. It is argued that differences do not emerge unless the task is cognitively demanding (Piccardi et al., 2011; Chen et al., 2009). In Coluccia and Louse's review, this hypothesis was favoured to interpret gender differences in navigation, and it is concluded that gender differences are observed in unfamiliar and complex environments.

Another related factor is argued to be the working memory demands involved in a wayfinding task. Visuospatial working memory is found to be a predictor of navigation performance (Bosco et al., 2004). In Garden et al. (2002), introducing interference with visuospatial working memory hindered navigation performance more than verbal performance. Females appear to have a lower visuospatial working memory capacity than males (Lawton and Morrin, 1999), which may contribute to their inferior performance. Similarly, Allen (2000b, p.17) suggested that male superiority in map tasks that require symbolic representations to be reproduced in the real world should be attributed to male's more efficient visuospatial working memory. Indeed, in Brown et al. (1998), gender differences disappeared when participants had access to a map, which reduced memory demands. Taken together, it appears that the function of visuospatial working memory capacity and task demands are additional factors that influence male and female performances.

Psychological factors

Psychological factors like self-confidence (that is, the beliefs about one's ability) and stress are regularly demonstrated to affect cognitive performance, such as in the areas of mathematics (Shih et al., 2002) and spatial abilities. Low self-confidence, 'spatial anxiety' and the 'fear to get lost' are found to negatively correlate with navigation performance (Lawton, 1994; 1996; Picucci et al., 2011). In Schmitz (1997), participants who scored higher in fear and anxiety scales traversed a maze more slowly. In Lawton (1994, 1996) women reported feeling higher levels of anxiety when performing a spatial task, such as navigation in an unfamiliar environment (Lawton, 1994; 1996). In addition, these studies explore the interaction of wayfinding performance, strategy and anxiety and conclude that wayfinding strategy selection is mediated by levels of anxiety and fear. In particular, males outperformed females, reported lower anxiety levels and used the survey knowledge strategy. Female participants were more anxious and depended on a landmark-based strategy. Low confidence and feelings of uncertainty have also been reported by women in wayfinding while driving (Burns, 1998), in indoor wayfinding (Lawton et al., 1996) and map drawing (O'Laughlin and Brubaker, 1998). It is noteworthy that the actual task performance of females and males was similar in the aforementioned studies.

'Stereotype threat' is another obstacle that women face. Stereotype threat is the popular belief about the competencies/deficiencies of a group (ethnic, racial, gender). It leads to suboptimal performance among the individuals of the stigmatised groups, caused by excessive cognitive and memory demands, anxiety and fear (Schmader et al., 2003). These negative states lead to a 'self-handicapping', extending to all future tasks (Schmader et al., 2008) and ultimately distancing oneself from the area (Keller, 2002). Brownlow et al. (2011) performed an interesting experiment in which same-gender peers informed the participants about the task, conveying that the task can be accomplished. The results showed that the performance of both males and females improved when they believed that the task was feasible by people of the same profile as them. It was also confirmed that women who expected to do well in the task (high self-confidence), indeed, performed as well as men. Similarly, in Moè and Pazzaglia (2006), participants were divided into experimental groups, and either received neutral instructions or were told that 'males/females are better in this task'. The results showed that performance increased dramatically if participants believed

their gender had an advantage. It is important to note that these studies employed a mental rotation test, which is the only spatial task for which fifty years of research has reported robust differences favouring males (Voyer et al., 1995). Interestingly, a study, which explored the effects of gender and ethnicity stereotype threats on mathematical and spatial ability, found that the performance of women improved when they were told that their data would not be used for evaluating their ability but to understand other psychological processes (Gonzales et al., 2008).

2.3.4 Interim summary

The literature analysis presented in this section revealed that findings on gender differences in navigation are oscillating, between showing no performance difference and a male advantage. Yet, differences favouring males are pronounced in simulated environments. Wayfinding in simulated environments offers experimental control and reproducibility, but caution should be exercised so that the interface does not add to the complexity of the task and provide an advantage to males who generally have prior experience with similar devices and applications. While the diversity of measures, setups and environments may account for the discrepancies in findings, three factors are found to underlie the observed differences- or lack thereof; first, the choice of navigation strategy; second, task complexity; and third, spatial anxiety and self-confidence. Therefore, it is argued that the quest of identifying gender-related superiority in spatial tasks should not be a goal in itself but should lead to further research efforts that focus on how: an efficient strategy can be supported, the complexity of a spatial task can be alleviated, and negative psychological states can be prevented.

2.4 Gender differences in language

The previous section discussed gender differences in performance in spatial cognitive tasks and wayfinding, and the underlying reasons. This section provides an overview of gender differences in performance in verbal tasks and communication styles.

Most theoretical and applied studies assume that a language is uniformly used and processed by its speakers. Growing evidence points to large individual and group differences

with possible evolutionary, psychological and cognitive origins. Among the most prominent are gender differences (Ullman et al., 2008). Counteracting the male spatial ability superiority, women present a higher aptitude in verbal tasks from a very early age. Women outperform men in a range of tasks like lexical retrieval, verbal fluency and synonym finding (e.g., Maitland et al., 2004; Herlitz et al., 1999) and memory tasks querying sounds, digits, words and paragraph content (Kimura, 2000, p.11; Kaushanskaya et al., 2011). As reported in previous discussion on the female advantage in object memory and strategy preferences (see sections 2.2 and 2.3.3), women excel in recalling landmarks and street names and use them during navigation, which appears to stem from their ability to internally verbalise these elements (Galea and Kimura, 1993; Saucier et al., 2003).

Besides differences in achievement, qualitative differences are also well-documented. Beginning with Lakoff's (1975) seminal work, it is suggested that women and men form two distinct speech communities, with large differences in linguistic style, features and attitudes. Fitzpatrick et al. (1995) list 35 variables in which gender differences are consistently documented; these include linguistic and supra-linguistic features such as number of words, pauses, fillers, interruptions, sentence length, affirmation, questions, intensive adverbs and, hedges (hedges are words or phrases that mitigate the forcefulness or assertiveness of a statement, for example, 'I think', 'sort of', 'to my understanding') and question tags (in which a question is added after a statement, for example 'this is really exciting, *isn't it?*'), personal pronouns, emotion references etc. On the use of tags and hedges, Lakoff (1975, p.15) suggests that it is indicative of a lack of confidence on the speaker's part. In dyadic interactions, questions are more abundant in women's speech, whereas men seem to issue a larger number of directives (Mulac et al., 1988). In a professional context, however, the pattern was reversed (Mulac et al., 2000). Men also appear to produce more words overall and maintain a higher turn ratio in a conversation (Dovidio et al., 1988), while women provide lengthier sentences in writing and speech (Mulac and Lundell, 1994). Tentative and polite speech with high frequency of hedges is characteristic of 'female speech' (Lakoff, 1975; Mehl and Pennebaker, 2003; Leaper and Robnett, 2011). An effect of pair gender composition is also present, such that women spoke less and used hedges and tags more frequently when interacting with men compared to when addressing other women (Mulac et al., 1988; Dovidio et al., 1988). Relevant to our study, this shows that communication styles are affected by the speaker's gender relative to the gender of the addressee. The results of a

large meta-analysis of written and spoken texts show that females use higher numbers of pronouns, adverbs, emotion words, first-person references, expressions of politeness, hedges, negation, longer sentences and a lower number of directives (Newman et al., 2007). No significant differences were found in number of words or questions. In this study, women appear to refer to psychological and interpersonal topics more frequently, whereas men's language was more factual. Finally, they point out that these effects are expected to be weaker in task-oriented speech.

Tannen (1994) proposes that gender differences arise not only in language production but also in interpretation. In particular, empirical data shows that females and males understand specific linguistic features differently. Maltz and Borker (1982) and Mulac et al. (1998) found that men and women attach different meanings to backchannels ('uh-uh', 'ok'); males use them to show agreement, and interpret them as expressing uncertainty. On the other hand, females use them to signal continuing attention and view them as signs of being interested in the other person's opinion. Similarly, males use questions ('then, what did they do?') to elicit information, whereas females use questions to keep the conversation going. These discrepancies in function and interpretation are argued to lead to inter-gender miscommunication (discussed in section 2.10), and may also explain why females produce a considerably larger number of backchannels (Mulac and Bradac, 1995) and questions than men (Fitzpatrick et al., 1995). Interestingly, it was found that perceptions on backchannels and questions also depended on the gender of the addressee, such that backchannels spoken to a woman were seen as more controlling and questions asked by a woman to a man as being the most uncertain. Gender-related language differences do not only arise in face-to-face interaction but also in computer-mediated communication, such as email, instant messaging and virtual shared workspaces (e.g., Savicki et al., 2006). Differences have also been consistently found in anonymous online interactions where the participants' gender is not disclosed (e.g., Herring, 1996); this finding refutes earlier hypotheses postulating that the absence of overt gender clues will lead to parallel communication styles (Landow, 1992). Fox et al. (2007) found that women's communications in instant messaging were more expressive, containing emphatic terms, laughing, emoticons and adjectives. Moreover, an effect of recipient was identified; messages sent to men contained more words and turns but fewer emotion references. That study did not find any interaction effects between speaker and recipient gender, but in Thomson et al., (2001), it was found that women made more

references to emotion when speaking to women. In addition to linguistic differences, noted differences exist in participation and coordination patterns and software feature use in a CSCW system (Prinsen et al., 2007). Given that gender in computer-mediated communication is primarily an HCI issue, it is further discussed in relation to studies in gender and computer-mediated collaboration and computer-supported cooperative work in section 2.7.2.

2.4.1 Interim Summary

It is generally agreed that women perform better than men in most verbal tasks. Moreover, studies have demonstrated that certain linguistic and supralinguistic elements are characteristic of female or male speech. Yet, these findings are mostly sourced from social interactions, with less data on task-oriented interactions. Albeit limited, there is evidence that the gender of the recipient may modify the usage of these elements, which clearly demonstrates an interactive effect. The important issue of how interactive processes may lead to the adaptation of a gender-preferential style is revisited in section 2.8.4, which discusses gender in conversations. The following section continues the discussion of gender differences in language. It introduces concepts and challenges relating to route instructions in theoretical and applied research, before focusing on how males and females produce and interpret them.

2.5 Route instructions

Route instructions are written or spoken linguistic descriptions of the environment, designed to aid navigation (Montello, 2009, p.165). They are a form of spatial language and procedural discourse which has attracted scientific interest across diverse disciplines. On one hand, route instructions hold a rich theoretical value for linguists and cognitive scientists. Linguists study route instructions to account for how people talk about space. In cognitive science, spatial cognition is a primary area in the human cognitive system (Kaufman, 2007) as space is fundamental in the development of any roaming animal- and language is understood to be the window to cognition, externalising the ‘inner world of spatial concepts’ (Levinson, 2003, p.131). It is also scientifically intriguing that route instructions come about through a series of transformations and dependencies across different cognitive systems; people acquire, use,

communicate and interpret route knowledge relying on a variety of behavioural and sensory abilities. From a practical perspective, the ability to generate and process route instructions is of immediate concern for the development of natural language interfaces, such as those used for navigating robots.

2.5.1 Studies in route instructions

Route instructions are a unique type of spatial discourse that is not produced to describe static scenes of the environment. Rather, their fundamental function is to elicit a particular behaviour from an agent (human or robot) which will enable it to navigate in an unfamiliar environment (Daniel and Denis, 1998). Such communication episodes do not always lead to efficiently or successfully reaching the destination.

When giving route instructions, route knowledge (that is, the knowledge about the actions required to navigate from point A to point B) is externalised as a set of verbal route instructions. Route knowledge is acquired by direct experience, physically navigating the path to a destination, or indirectly, by accessing media like maps, route sketches or visualisations (as is the case with in-car navigation systems and web route planner applications) (Klippel et al., 2003). The recipient of the route instructions follows a reverse process: interpreting the linguistic content of the instructions, modelling and planning the actions and finally executing the actions in the environment[4].

There has been a wealth of research on the production and interpretation of route instructions. Influential work in this field includes studies by Gary Allen and his colleagues, Michel Denis and the Human Cognition Group in Paris, and Barbara Tversky and the STAR group. Vanetti and Allen (1988) developed the Communication of Route Knowledge (CORK) framework, which describes the structural organisation of a route communication episode and proposes a classification scheme for component analysis of route instructions (these frameworks are detailed in Chapter 3, section 4.4.5). The work also includes an

[4] This description of a route communication episode does not imply that communication is merely a coding/decoding process between a transmitter and a receiver.

account of individual differences. In particular, Vanetti and Allen (1988) investigated the impact of verbal and spatial abilities on the production and comprehension of route instructions in a large-scale environment and Allen (2000a, 2000b) explored gender-related differences in map-reading and when following route instructions.

Allen (2000a) also describes the principles for forming route instructions, which are shown to facilitate wayfinding. First, the instructions have to be presented in the correct temporo-spatial order. Second, descriptives (that present a static picture of spatial relations and provide the travellers with the opportunity to verify their position or reorient themselves) and delimiters (that provide specificity and distinguishing information about the environment) should be concentrated in choice points and at the end of the route. Finally, selection and placement of these components should depend on the characteristics of the environment and the perceived needs of the traveller. It is noteworthy, however, that the aforementioned studies typically include subjects navigating a real-world environment (like a town or a campus) while following a set of instructions prepared in advance by the experimenters (Allen, 2000a) or another group of participants (Vanetti and Allen, 1988).

Research by the Human Cognition Group in Paris has focused on protocols of how people generate route instructions. Analyses of spontaneously produced route instructions for navigation in an urban environment have provided the basis of a well-known classification scheme (Denis, 1997). As part of their corpus of research into route instructions, Michon and Denis (2001) have explored the special role of landmarks in two studies: the first on production of route instructions; and the second on following, evaluating and finally 'filling the gaps' in a stripped-down set of instructions. In Daniel and Denis (2003), the conditions in which people provide highly concise instructions are identified, along with what they see as constituting 'poor', 'good' and 'skeletal' instructions (Daniel and Denis, 1998; Denis et al., 1999) achieved by collecting and manipulating naturally-produced route instructions and then using them to guide navigation. Along the same lines, a study by Lovelace et al. (1999) explored the quality of route instructions, how components (references to turns and landmarks) contribute to this quality, and the effect of the familiarity of location on the configuration of route instructions.

Another area of research into route instructions focuses on the importance of different communication modalities (i.e., language, diagrams and gestures) in conveying information

about space and routes. For instance, in Tversky and Lee (1999), participants were asked to either write down route instructions or sketch a map to navigate someone to a particular location, whereas in Tversky et al. (2009) participants were given maps and asked to explain them using gestures or gestures with speech. The results from these studies suggest that all modalities communicated the route in similar ways, depicting the route as a sequence of turns or actions at landmarks while angles and distances were left under-specified. The structure of the route was also comparable across modalities, with beginnings acting to orient the follower, the middle parts of the route specifying a series of step-by-step actions at landmarks, and endings indicating arrival. These findings imply that the same cognitive processes underlie all communication modalities.

The approach of all aforementioned studies followed a monologue paradigm, in which a set of people independently formulated and subsequently executed or rated route instructions. However, it is expected that route instructions produced in isolation will be fundamentally different to route instructions produced in dialogue (Coventry et al., 2009, p.6). The theme of dialogue will be explored in the final sections of the chapter (section 2.8 – 2.10).

Finally, there are large individual and group differences in how people process and communicate route instructions, related to, amongst other things, core spatial and verbal abilities (Vanetti and Allen, 1988), age (Golding et al., 1996), education, and previous experience (Newcombe et al., 1983). Research has shown that observed individual differences may also be 'stylistic' – that is, relating to preference rather than ability (Barkowsky et al., 2007). Gender-related differences in route communication are discussed in section 2.5.3.

2.5.2 Challenges for practical systems

The capability to navigate outdoor environments through natural language instructions is essential for assistive systems, as exemplified by the robotic wheelchairs of the Diaspace program (Hui and Tenbrink, 2009) and the IBL project (Lauria et al., 2001), and robots used for rescue/exploratory purposes (for instance, the robot helicopter of the WITAS project (Lemon et al. 2002)). Research also focuses on characterising how route instructions should be formulated to optimise navigation. It is motivated by and informs the development of

geographic information/wayfinding systems (GIS). Important questions are explored such as which type of environmental features should be incorporated, which features should serve as landmarks, how they should be referred to, their right proportion in route instructions and whether the spatial layout or the landmarks should hold salience. These questions do not have simple answers, since different configurations of route instructions may be appropriate for different groups of individuals (Montello and Sas, 2006).

Applied research on route instructions faces several challenges. Route instructions inherit the basic properties of spatial language (Barkowsky et al., 2007). First, they are underspecified and non-quantitative. For instance, the route instruction, 'move forward for 12 metres and rotate 90 degrees to the right' is more accurate and efficient, but less natural and less likely to be uttered than 'move forward and turn right'. On the other hand, systems are capable of processing spatial information with a fine level of metric precision, which is incompatible to how people process and produce it. This adds to the complexity of designing effective interfaces between systems and their human users.

Second, route instructions are context-specific, where context encompasses multiple aspects associated with the interlocutors (e.g., their perceived characteristics and familiarity of the environment), and the physical properties of the situation. In particular, it is well established that speakers adapt to the perceived needs and abilities of the addressee. For instance, an individual will speak in different ways to a young child, a colleague or someone from a different country. However, in the context of human-machine communication, forming assumptions about what a system can do and understand is problematic for most people. In turn, forming assumptions of how users will talk to the system is also likely to be problematic for system developers. Moreover, situated dialogue refers to dynamic temporal and spatial events and thus, its interpretation encompasses far greater complexity compared to the typical 'unsituated' interactions with dialogue systems (that is, telephone-based information seeking applications).

Considering the aforementioned characteristics, natural communication about space with an agent (human or system) requires complex, flexible and rich interactive mechanisms. The potential for variability in how users will communicate with a system is enormous and has been dubbed 'The Vocabulary Problem'. The extent of the problem was measured in the well-known study by Furnas et al. (1987), in which participants were asked to name objects

for a computer to understand in five scenarios. The probability of two people using the same word to refer to an object was below 0.2. If the two people are assumed to be a system designer and end user, this suggests that there is an 80-90% chance that the word selected by the system designer to refer to an object or action will not be employed by the end user. There are myriad ways to formulate the same route instruction, with varying levels of granularity – defined by Tenbrink et al. (2010) as the 'level of specification in the representation of a particular situation, event or object'. For example, the route instruction 'Take the second turn on the right.' (which has a low level of granularity) is pragmatically identical to 'Go straight ahead until you pass a junction on your right; do not take this turn, go straight on until there is another junction on your right. Turn there.' (which has a high level of granularity). Different levels of granularity also yield differences in the effectiveness of instructions (Tenbrink et al., 2010; Allen, 2000a).

2.5.3 Gender differences in the production and interpretation of route instructions

In the domain of route instructions, clear patterns of gendered language use emerge which mirror the wayfinding strategies that women and men prefer to use. In particular, men formulate abstract instructions using cardinal directions, mileage estimates or metric distances. On the other hand, women give more concrete instructions and incorporate simple relational terms (left/right) and references to proximal landmarks and other visual objects encountered along the route (Ward et al., 1986; Lawton, 1994; Moffat et al., 1998; Galea and Kimura, 1993; Dabbs et al., 1998; Hund and Padgitt, 2010).

In addition to how language is used, it is necessary to look at how route instructions are interpreted. Findings are less clear with regards to which types of route instructions are more efficient for navigation or preferable by female and male addressees. In Anacta and Schwering (2010), groups of participants followed a combination of absolute (based on cardinality) and relative (left/right/straight) instructions in a real-world urban navigation task. The study found that the performance of both males and females was impaired when they received absolute instructions. It replicates the results of a similar study by Ishikawa and Kiyomoto (2008), which also showed that Japanese people prefer relative instructions. The measures used were number of stops, deviations, off-route distances and completion time.

These findings are surprising, given men's preferential use of cardinal information as a navigation strategy and for producing route instructions. A tentative explanation may be that, although men are capable of orienting and using cardinality when navigating, following cardinality-based instructions involves additional layers of processing complexity.

2.5.4 Interim summary

Due to the practical and theoretical consequences, significant research efforts have led to better understanding of route instructions. However, these findings are based on non-interactive studies that investigate either the production or following of route instructions; this significantly limits the value of these findings for more realistic and practical contexts of use. Mirroring the pattern of results with regards to gender-related navigation strategies (first part of section 2.3.3), females have been found to produce landmark-based instructions more than men, while cardinal directions (usually provided by male instructors) may be difficult to interpret for both genders. The next section briefly reviews our current knowledge with regards to the origins of gender differences in the domains of spatial and verbal abilities.

2.6 Origins of gender differences in spatial and verbal abilities

There are several theories regarding the factors contributing to gender differences. These theories can be roughly classified as environmental-based and biological-based. First, there is the evolutionary 'hunter vs. gatherer' explanation, according to which the division of labour in prehistoric societies promoted the gender differences in spatial vs. language skills; namely, males developed superior navigation skills, because they needed to hunt and track prey over large distances, probably rendering speech as a distraction rather than a necessary skill (Joseph, 2000). On the other hand, females reared children, manipulated domestic tools and gathered food from edible plants in groups around the home base, so they developed verbal skills and spatial skills necessary to remember nearby locations and the appearance of plants (Silverman and Eals, 1992). Second, experiential factors may also contribute to gender differences. Several studies argue that boys are highly encouraged and given more opportunities to explore outdoor environments compared to girls, either as part of chores or leisure time activities (Webley, 1981). Boys are also more likely to conduct these activities

without supervision. Moreover, boys are encouraged to play games that hold a spatial ability component, such as team sports and building structures with blocks. As previously noted (section 2.3.2), males are also more likely to have extensive exposure to computer and video games that depend on navigation skills (Terlecki and Newcombe, 2005). Evidently, greater experience in navigation tasks not only builds the necessary skills that help develop efficient strategies but also the confidence to carry out related tasks later in life. Finally, biological theories argue that hormone, genetic and brain organization differences may account for gender differences in navigation. In particular, research showed that higher levels of male hormones are associated with better navigation performance and efficient wayfinding strategies (also found in studies with rodents), while concentration of female hormones facilitate verbal abilities (Williams et al., 1990; Kimura, 1996; Moffat and Hampson, 1996). Another theory correlates spatial abilities with the presence of an X-linked recessive gene (Walker et al., 1981). Moreover, it is well-known that spatial processing is localised in the right hemisphere of the brain and people with right-hemisphere dominance have better spatial abilities (McGlone, 1980). Men have an early, more pronounced right-hemisphere dominance (Harris, 1978) and their brains are characterised by higher functional specialisation (Joseph, 2000). On the other hand, women develop a left-hemisphere dominance, which is favourable for language tasks but interferes with aptitude in spatial tasks (Annett, 1992), and their brains have lower lateralisation for cognitive functions.

Many scientists advocate that environmental and biological factors are not mutually exclusive, but their interaction better explains gender differences. According to interactionist theories, an individual with right-hemisphere dominance who is exposed to spatial activities will outperform an individual with the same cerebral pattern but with low experience. Moreover, the innate predisposition of males for spatial abilities encourages them to actively pursue relevant activities, which in turn further develops their skills (Casey et al., 1999).

2.7 Gender in the interaction with and through systems

In recent years, there has been an exponential growth of software applications with ever-increasing sophistication that support users in professional, educational, recreational and social activities. This has generated the necessity to develop effective user interfaces to cater for needs of a body of users of diverse expertise in diverse domains. Developers strive to

allow for customisation and personalisation of products and the ‘one size fits all’ paradigm is gradually depreciated, but, still, the underlying assumption in the majority of visual user interface designs is that users have similar perceptual and cognitive abilities. There are, at least, three arguments that underscore the necessity of taking gender into account in the development of interactive systems. First, the previous sections described the influence of gender on people’s abilities to perceive, interpret and manipulate visuospatial relations and objects. Interfaces of commercial systems typically depend on such elements, and, therefore, gender is expected to play a role in how people use these systems. Second, in addition to cognitive visuospatial differences, there are well-documented gender differences in linguistic and coordination patterns during face-to-face interaction which may cross over to computer-mediated communication and collaboration. Third, gender shapes many aspects of human interactions. Therefore, since human social norms persist in the interactions with artifacts (Nass and Reeves, 1996; Nass and Moon, 2000), gender-related effects should also be present in human-computer interaction. This section reviews literature that documented gender differences in various application areas of human-computer interaction, with a focus on those of most relevance to this thesis, computer-mediated communication, virtual world navigation and the specialised fields of human-robot interaction and dialogue systems.

2.7.1 Gender in Human-Computer Interaction

A fundamental principle behind user-centred design of interactive systems is to understand the user, with the aim to develop usable applications. It is argued that user differences are more likely to compromise the effectiveness and acceptance of a system than any factors that have to do with operation and training (Egan, 1988). Thus, early in the development lifecycle, analysts work towards defining user characteristics that may affect task performance and experience. However, user analysis often only involves the distinction between expert and non-expert users, and usually excludes gender considerations (Dillon and Watson, 1996). As a result, even today ‘the user’ remains genderless (Bardzell, 2010).

Gender and its effect in cognitive abilities and preferences have a broad impact on computer skills and hardware and software requirements. In particular, research has revealed gender differences in usage, preferences and perceptions in various application areas of human-computer interaction, ranging from online shopping and web applications (Bae and

Lee, 2011; Bimber, 2000) to computer games (Cassell, 1998; Hartmann and Klimmt, 2006), office software suites (Burnett et al., 2011) and decision support systems (Djamasbi and Loiacono, 2008). Overlooking gender differences has negatively impacted the adoption of these technologies by females, their judgments of 'self-efficacy' and attitudes towards computers (Busch, 1995). So, scientists were led to identify a gender gap in the use of computers and unify all related research into a new subfield of HCI, termed 'Gender HCI' (Beckwith et al., 2006a). This body of work focuses on the differences in how males and females interact with 'gender-neutral' systems and, by taking gender issues into account, how systems can be designed to be equally effective for both men and women (Fern et al., 2010).

Research within the area includes the pioneering work by Czerwinski and her colleagues (Czerwinski et al., 2002). Their approach was first to identify gender differences in Virtual Reality (VR) navigation and then to find solutions to offset these differences in display and VR world design (by provision of larger displays and wider views). Another example is the Gender HCI project of the EUSES consortium, which uncovered gender differences in end-user programming in terms of confidence and feature use and proposed solutions for the design of programming environments (Beckwith, 2007). Similarly, Fern et al. (2010) showed differences between male and female users, and the relation between their strategies and success in a debugging task. The position held in this research is that software design determines how well female problem solvers can make use of the software. Understanding how gender influences strategies, behaviours and success is the first step towards design that promotes successful behaviours and strategies by users of both genders.

Along the same lines, the work reported in this thesis seeks to contribute to 'Gender HCI' by detecting gender differences in the domain of robot navigation and dialogue systems, which are prime examples of collaborative/goal-oriented interaction between humans and computer systems. It also shares the aim to offer recommendations for gender-effective design.

2.7.2 Gender in Computer-Mediated Communication

The proliferation of computer and telecommunication technologies have enriched the ways we interact not only with computers but also with each other, such that we can communicate

across geographic distances and remotely work together on various tasks. To facilitate remote interactions, technologies may incorporate video or enable sharing visual perspective since visual information is found to facilitate collaboration. There has been considerable work on how visual information and other media affect coordination patterns and is presented in section 2.9. Moreover, as previously discussed in section 2.4, there are well-known gender differences in how females and males communicate, mainly in social settings. Yet, there is a relative paucity of research addressing the role of gender in shaping computer-mediated collaborative work and communication. As the popularity and prevalence of computer-mediated communication (CMC) and computer-supported cooperative work (CSCW) increase, it is important to clarify the role of gender in group dynamics and task-oriented interactions in order to avoid biases and inform the design of gender-neutral software and hardware for future collaborative systems.

Initially, some researchers projected that CMC (particularly text-based CMC) would make gender and other social differences less salient, because it could exclude physical cues of individuals (e.g., Hert, 1997; Lea and Spears, 1992). This, in turn, would benefit collaboration as people would attend more to the message than the social status of the speaker. Others have disputed this view, showing that power dynamics and biases found in face-to-face communication are transferred to CMC (Herring, 2000). Literature has shown gender differences in terms of discussion participation, communication styles and experience in CMC and computer-supported collaborative learning environments (Prinsen et al. 2007; Ding et al., 2011). Most of this research suggested that characteristics of gendered language in face-to-face communication are carried over to CMC. In particular, the majority of studies have suggested that males dominate the discussions, sending more messages or taking more turns (Carr et al., 2004). Differences in linguistic and communication styles have been observed, such that females appear more supportive, cooperative and agreeable (Sun, 2008). Females also ask more questions than males in collaborative learning situations (Ding et al., 2011; Prinsen et al., 2009). In a collaborative architectural task using a tabletop interface, females tended to explicitly question and state requirements, resources and plans for action (Richert et al., 2011), while males tended to execute more and negotiate the plan less. With regards to experience and satisfaction, comparative studies indicated that females prefer CMC than face-to-face interactions. Again, the majority of these studies has focused on the

behaviour of individuals with relatively fewer studies exploring how gender contributed to pair and group dynamics.

Among studies that also consider pair and group dynamics, Bernard et al. (2000) showed that males reported high levels of self-efficacy and satisfaction and low anxiety in a task involving synchronous three-way communication. Savicki et al. (1996) found that all-female groups produced more words and expressed greater satisfaction and confidence with group processes and communication than mixed and all-male groups. Females in mixed groups were, in fact, the least satisfied.

Again, fewer studies have addressed whether gender and pair/group composition is important for interaction efficiency and effectiveness; that is, whether mixed-gender or single-gender pairs perform better. A study in which physics students used a computer-supported collaborative learning environment found that females in all-female pairs outperformed females in mixed pairs, while males did equally well with female or male partners (Ding et al., 2011). In a day-trader collaborative game, all-female pairs were faster than all-male pairs (Sun, 2008). However, this study did not include mixed-gender pairs. Prinsen et al. (2009) argued that females performed better in a collaborative learning task because they have superior verbal skills and asked more questions. The authors concluded that computer-supported collaborative learning environments may have a greater utility for females, because females are able to show their potential and use their linguistic skills more easily than in face-to-face interactions in which females may face anxiety and the stereotype threat (as discussed in the final part of section 2.3.3). For instance, Inzlicht and Ben-Zeev (2000) pointed out that females in the presence of males tended to perform worse in a problem-solving task. Taken together, more focused research is necessary to add to these findings on the influence of gender and pair dynamics on performance, coordination, communication and user experience in CMC. It is also interesting to ascertain whether, the same patterns emerge when the gender of the collaborators is masked, thus reducing social preconceptions.

2.7.3 Gender in the interaction with robots

There is a limited body of research that demonstrates that gender affects user perceptions, experience and the processes of interaction with a robot. In particular, Schermerhorn et al. (2008) reported that males perceived robots as more human-like than women. In Nomura and Takagi (2011), males of a natural sciences educational background manifested a more positive impression of the robot, whereas in an earlier study (Nomura et al., 2008) the authors described a complex gender effect on negative attitudes and anxiety during short social interaction with a robot (that involved greetings and physical contact). In Woods et al. (2005), males and females interpreted the 'personality' of a social robot differently. The studies by Nomura and Takagi (2011) and Siegel et al. (2009) also considered the 'gender' of a robot. They found that males favour 'female' robots, developing stronger trust, engagement and positive attitudes compared to female users. These studies deployed a non-goal-oriented, social interaction setting. Instead, Mutlu et al. (2006) analysed users' perceptions developed before and after the completion of a competitive or cooperative task. The task involved playing a motion-based computer game against or in collaboration with the robot ASIMO and users were asked to rate the robot and their own affective state across certain scales. The findings demonstrated that user perceptions are influenced by the nature of the task and the effect is particularly pronounced for male users. In particular, male users rated the robot more desirable in the cooperative than in the competitive task. Men also reported feeling more excited when competing against the robot compared to when cooperating with it. In all conditions, women did not vary in their ratings of the robot or themselves. As the authors acknowledged, a limitation of the study was that the interactivity between users and ASIMO was minimal, involving only gaze and greeting.

This short review clearly shows that social responses arise in human-robot interaction, and are shaped by gender. Moreover, the effect of gender depends on the nature of the task. However, this research has mostly focused on non-task-oriented interaction and purely on aspects of user experience. It remains an open question how gender influences the performance and communication patterns of users in a task-oriented interaction with a robot.

2.7.4 Gender in the interaction with dialogue systems

As presented in section 2.4, empirical evidence from various human-human interaction settings reveals systematic differences in the way that women and men use language, in terms of what they say and how they choose to say it (Newman et al., 2008). Developers of systems with natural language interfaces have only considered gender differences in terms of acoustic features in order to adapt the acoustic models used by speech recognisers (for instance, Raux and Eskenazi, 2004). Thus, additional research is needed to develop dialogue systems that allow adaptation of the higher-level functions of dialogue, in which females and males differ, such as dialogue strategies, communication style, vocabulary choices and sentence structures.

2.7.5 Gender in virtual environments

The literature presented in section 2.3 has suggested that there is a male advantage in navigation tasks, which is exacerbated in virtual environments. There is a growing interest in isolating the parameters that can be manipulated in order to mitigate the performance gap between males and females in virtual world navigation. Examples of such research include the work by Czerwinski et al. (2002) and Tan et al. (2003; 2006). In the former, large displays that offered wide fields of view were found to benefit female navigation in virtual environments and enabled them to achieve similar results to men. The latter studies replicated this finding and showed that adding optical flow[5] cues helped females, without any negative effect on male performance. Hubona and Shirah (2004) argued for 'gender-neutral' visual interfaces. They maintain that research should not emphasise differences, that is, in which tasks women or men excel and which visual features are utilised by males and females, but the focus should be on creating visual interfaces that have features that are suitable for both. For instance, interfaces should be rich in static and dynamic visual cues (such as landmarks and optical flow), given that these cues are important for women and do not negatively affect the performance of males, while features involving mental rotation, for instance 3-D features,

[5] Optical flow refers to the apparent motion of objects, surfaces and edges in a scene caused by one's own motion in it.

should remain optional (since these impair female performance). Moreover, they challenge the trend of more tightly packed visual interfaces and recommend that some visually-presented information can be replaced by textual content. This recommendation echoes results from a study that investigated the selections of wayfinding aids by men and women (Devlin and Bernstein, 1995); it was found that men prefer visual and map information, while women choose to complement the visual and map aid configuration with written verbal instructions.

Such research provides evidence that the use and value of visual elements in a computer interface are influenced by the gender of the user. Most importantly, it suggests that there are design techniques that can benefit a user group, without disadvantaging another group.

Taken together, it is imperative that system designers take into account that individuals differ in their perceptual and processing abilities and preferences with regards to the information sources they use to find their way to a destination in real-world and virtual environments. Better awareness of gender differences in navigation and communication has implications for numerous fields ranging from natural language interfaces and computer-mediated communication tools to virtual worlds and interfaces that involve navigation (such as websites and online environments).

2.7.6 Interim summary

Taking gender into account in system design is of paramount importance, because interacting with computer systems draws on perceptual and communication abilities and preferences that have been found to be largely dependent on one's gender. Indeed, empirical studies show that gender differences emerge in numerous HCI domains, in terms of how people use and experience technologies. As such, dedicated research in gender in HCI has been gaining momentum, focusing on identifying and understanding gender differences and using this insight in the development of gender-neutral systems that can equally support users. For instance, research in virtual world navigation has led to identifying techniques that improve the performance of users of both genders. Existing literature has also hinted at a gender component in social CMC and provided some evidence that gender pair/group composition impacts performance outcomes and experience. Given the growing popularity of CMC and

CSCW systems, further systematic work is needed to clarify how gender mediates the use of such systems and influences group dynamics and collaborative work. At the same time, it remains largely unknown how gender mediates task-oriented interactions with dialogue and robotic systems.

Having reviewed existing knowledge in gender differences in spatial abilities, navigation language and the interaction with computer systems, the focus of the chapter is now shifted towards understanding dialogue and the phenomena that emerge during its course. The ensuing sections aim to build the case that dialogue offers a natural setting of interaction, where interactive mechanisms arise and fundamentally shape the ways people perform, coordinate and use language in collaborative tasks. Such mechanisms are argued to be of immense practical relevance for system development.

2.8 Dialogue and alignment with humans and systems

Empirical research in human communication has shown that the way in which people produce or interpret language is dependent on whether it takes place in dialogue or monologue. In studies of this kind, language is seen as a collaborative activity in which partners introduce, negotiate and accept information. In this sense, meanings do not exist *a priori* but are 'agreed upon' by the dialogue participants (Brennan and Clark, 1996). This explains why over-hearers that do not participate in the dialogue often have difficulty understanding what is being said (Schober and Clark, 1989). Moreover, it has been observed that dialogue is largely repetitive; that is, speakers in dyads progressively use the same expressions. This natural phenomenon has been referred to as 'adaptation', 'entrainment', 'accommodation', 'convergence' and 'alignment'. The thesis adopts the term 'alignment' as it is part of a complete framework of language use, the Interactive Alignment Model (Pickering and Garrod, 2004). According to this model, successful communication is the result of a process of alignment across all linguistic levels, such that speakers converge in how they understand and use sounds (phonetics), language structure (syntax), word meanings (semantics) and contextual information (pragmatics).

Although alignment is a prominent and well-documented phenomenon in human communication, it has received little attention in the context of human-computer interaction.

This is particularly surprising given that alignment, as a mechanism that promotes language re-use, can be of practical relevance to the 'Vocabulary Problem' (see section 2.5.2). That is, it is argued that alignment can be exploited not only to support successful and natural interaction but, more importantly, to predict and constrain the variability of user input.

Similarly, the most natural context of production and interpretation of route instructions is in dialogue: people hardly ever produce route instructions with no addressee. Still, the overview of studies in route communication (as reported in section 2.5.1) reveals that the majority of studies follow a monologue paradigm, in which a set of people either give or follow instructions produced independently and beforehand by another set of people or the experimenters. Such approaches offer invaluable insight, but do not provide a full account of how language is used in normal communication settings, where speakers respond to the needs of their addressees, the dialogue history and the context, which encompasses interlocutors, spatial relations, features and dynamic events. They neither allow for the observation of natural phenomena of interaction, like miscommunication and alignment. These phenomena are prevalent in communication and hold both theoretical and practical implications. Therefore, it is argued that studying production and comprehension of route instructions in isolation is an oversimplification, as dialogue fundamentally changes language through, amongst other things, alignment of communication between dialogue partners.

As such, this thesis uses the dialogue paradigm to explore route communication. It also attempts to identify and categorise the occurrence of alignment in users' interactions with computer systems, and elucidate its characteristics in problem-free communication as well as in cases of user error, system error and non-understanding. This section discusses existing knowledge in alignment in human-human interaction and in human-computer interaction. It, then, discusses gender-related alignment and studies in which females and males were found to adapt their communication styles.

2.8.1 Alignment in human communication

A recurring finding in human communication is that what speakers say is influenced by their perceptions about their addressees as well as the addressee's own linguistic behaviour. The latter often leads to 'linguistic alignment', a phenomenon in which there is convergence in

the linguistic behaviour of interlocutors, such as them both moving to the use of the same term to refer to an object. This is illustrated in the study by Garrod and Anderson (1987), which reported goal-oriented experiments in which people produced spatial descriptions guiding each other in a maze and found that, over time, the pairs converged on similar spatial descriptions. They proposed alignment operates as follows: interlocutors initially start by using different spatial descriptions and, as the dialogue progresses, the most frequently used words, syntactic structures and situation structures become increasingly likely to be reused, inhibiting the other competing expressions.

As such, alignment is argued to be a basic interactive mechanism that takes place in dialogues at all levels – phonetic, phonologic, lexical, syntactic, semantic and pragmatic – and that makes communication between people 'easy', efficient and effective (Garrod and Pickering 2004; Pickering and Garrod, 2004). The evidence for this comes from multiple data-driven studies which show that alignment occurs at the phonetic and phonological levels with participants converging in terms of pronunciation (Pardo, 2006). In respect of lexical alignment, dialogue is full of repetition of the same words (Tannen, 1989); interlocutors align in terms of vocabulary in the sense that they use the same referring expressions (Garrod and Anderson, 1987); and when interlocutors refer to the same object they tend to re-use a previously-used term, even when simpler terms are available (Brennan and Clark, 1996). In terms of syntax, speakers will select a specific syntactic structure (such as either 'give the apple to Jim' or 'give Jim the apple') based on that which their interlocutors have been using (Branigan et al, 2000). At the situational (or pragmatic) level, interlocutors align on reference frames, such that if one speaker uses an egocentric frame of reference (e.g., using 'to the left', signifying his/her own left), the other speaker will do the same (Schober, 1993). Alignment also occurs at similar rates in dialogues with native and non-native speakers (Bortfeld and Brennan, 1997).

The phenomenon has been described as part of the Interactive Alignment Model, proposed by Pickering and Garrod (2004), and develops through two processes. First, alignment occurs as a result of the local, between-speakers priming mechanism ('input-output matching') at the same linguistic level (for instance, lexical, where speakers repeat each

other's lexical choices). Subsequently, alignment at one level leads to further alignment at other levels, such that the re-use of a particular lexical item will activate a particular situation model (that is, the information relevant for the situation under discussion)[6]. From this perspective, since successful communication is seen as alignment of the interlocutors' situational models, communication success largely results from linguistic alignment (Pickering and Garrod, 2006). Eventually, the repeated use of the same syntactic, phonetic and lexical expression to refer to the same object results in the development of the chunking of those expressions into 'dialogue routines' which, over time, optimise and stabilise interaction. With respect to 'dialogue routinisation', the Collaborative Model (by Clark and colleagues, see Clark, (1996)) seems to coincide with the Interactive Alignment account; in particular, Brennan and Clark (1996) propose that when interlocutors use the same expression to refer to an object, they enter into a tacit 'conceptual pact' in which they agree to keep referring to the same object in the same way.

Finally, there have been several explanations about why this phenomenon occurs. These include the social explanation which argues that people that align linguistically with their partners are positively perceived (see, for example, Giles et al.'s (1991) 'Communication Accommodation Theory'), and the 'audience design' explanation of the Collaborative Model, which argues that by choosing the same referring expressions interlocutors maximise their chances of successful communication (Brennan and Clark, 1996). This also resonates with the Interactive Alignment Model (Pickering and Garrod, 2004) which holds that alignment will result in communicative success. Yet, the accounts diverge in terms of whether the mechanism of alignment is strategic or automatic. In particular, Pickering and Garrod (2004) argue that alignment is a process that invariably occurs owing to mechanisms within the human processing system and is the basis of communication success. Brennan and Clark (1996), however, assume that alignment is mediated by conscious modelling of the

[6] Pickering and Garrod (2006) make clear that knowledge that forms a situation model is separate from the interlocutors' general knowledge, suggesting that the distinction between situational and general knowledge is analogous to the distinction between working memory and long-term memory and adding that 'successful conversations occur between interlocutors whose general knowledge is quite different in respects that are irrelevant to the conversation at hand' (p. 215). They do, though, propose that alignment in situational models can lead to alignment in general knowledge.

interlocutor and context, which is updated on a turn-by-turn basis, in order to increase the likelihood of communication success. Critiquing these aspects of the Collaborative Model, Pickering and Garrod (2004) maintain that 'audience design' is an one-off, optional decision in the beginning of the dialogue, and interlocutors mechanistically repeat each other's linguistic changes without the 'computationally-intensive' task of dynamically modelling each other's mental states and common ground (the concept of common ground within the Collaborative Model is discussed in detail in section 2.9 in relation to the effect of visual co-presence).

2.8.2 Perspective alignment and spatial ability

Across all studies that employ interactive spatial tasks, findings appear consistent with the central notion of the 'language as action' tradition in linguistics; according to which, partners share a responsibility for mutual understanding and adapt their behaviour based on the needs of their partner and the situation in order to minimise the collective effort for themselves and their partner and ensure interaction success (see, for example, Clark, 1996). So for instance, the affordances of the interaction situation, such as visibility (that is, partners can see each other and what they are doing), influence the choice of perspective, granularity of spatial descriptions and efficiency of the coordination (for instance, see Clark and Krych, 2005). Relevant studies are discussed in detail in the following section (section 2.9). Similarly, as mentioned above, other factors such as familiarity play an important role in how people plan and describe routes; that is, more details, words and landmarks are used for addressees who are unfamiliar with the environment (Isaacs and Clark, 1987).

There is also evidence that speakers make even more tacit judgements about the abilities of the addressee, with a considerable effect on spatial language use. Schober (2009) has provided novel insight into how spatial ability affects the choice of perspective and the formulation of spatial descriptions in interactive situations. In this study, participants were categorised as 'high-' or 'low-' ability, based on their performance in a mental rotation task. It was found that high-ability individuals tended to use partner-centred descriptions (such as 'to your right' and 'in front of you'), whereas low-ability participants were more likely to use egocentric ones. More interestingly, as the dialogue unfolded, the high-ability speakers intensified their partner-centred strategy, when collaborating with low-ability partners.

Similarly, low-ability speakers increased the use of egocentric spatial descriptions when paired with high-ability people. This appears to fit within the Collaborative Model's concept of 'audience design'; speakers produce language based on a priori assumptions about the addressee's abilities, these assumptions are continuously adjusted based on feedback from the addressee, as the interaction unfolds. As expected, the study found that the efficiency and accuracy of the dyad depended on their ability, such that high-ability pairs outperformed mixed-ability pairs, who outperformed low-ability pairs.

2.8.3 Alignment in Human-Computer Interaction

Having argued that alignment is pervasive in human communication, there remains the question of whether this mechanism also operates in the communication between a human and a computer system and, if it does, in what ways. If alignment is an automatic process (following the Interactive Alignment Model), then it should present similar patterns as are seen in alignment in human-human interaction. If it is a strategic process (following the 'audience design' explanation of the Collaborative Model), it should manifest in different ways. If alignment has a 'social' dimension, it is less clear how, or indeed if, alignment will occur, since one party in the dialogue is non-human.

There is a corpus of research looking at aspects of human-computer alignment which may be relevant here. A large segment of this research is dedicated to the study of alignment at the phonological/acoustic level. This research shows that people tend to adjust their speech rate (Bell et al., 2004), amplitude and pause frequency (Oviatt et al., 2004; Suzuki and Katagiri, 2007) to that of the computer with which they interact. Moving beyond speech input as the focus, Branigan et al. (2003; 2006) and Cowan et al. (2011) have investigated syntactic alignment between a human and a real or simulated computer in a picture-naming task, demonstrating evidence of alignment beyond the phonological level. From the perspective of dialogue system development and the 'Vocabulary Problem', however, alignment in terms of vocabulary seems to have more practical significance.

Pioneering work on lexical alignment in Human-Computer Interaction was conducted by Brennan (1996) who aimed to address the question of whether people adopt the same lexical terms used by the computer to the same extent as they do when interacting with other

humans. Wizard-of-Oz experiments were performed in relation to a database query task, and the results showed that when the 'system' responded using a different term to that originally used by the human user, the user tended to accept and subsequently use the system's term. The rate of alignment with the computer (or 'convergence') was found to be comparable to the alignment rate with humans. This finding supports the hypothesis that alignment is a basic, automatic mechanism operating in all contexts of language use.

A series of studies by Branigan and her colleagues also focus on lexical alignment in HCI (see Branigan and Pearson (2006) and Branigan et al., (2010) for an overview). In Branigan et al. (2004), users were told that they would interact with a computer program or a human (via a computer) in an object-naming and selecting task, though the interlocutor was a computer program in both conditions. In the study, the users saw two objects on the screen (for instance, a bench and an apple). The objects could be referred to in two ways; for instance, the bench could be referred to accurately as 'bench' or less accurately as 'seat'. In both conditions, the computer would name one of the objects using the more or less accurate term and the user would select the named object. Subsequently, the roles were reversed; presented with the same pair of objects, the user named one of them and could see the computer's selection of it. The researchers measured whether the user would choose the less accurate term if the computer had done so. The same experimental setup was again deployed in Pearson et al. (2006). This study involved users completing the task with a computer but they were made to believe that they would interact with either the 'basic' or the 'advanced' version of the system, whereas in reality both versions were the same.

The findings of these studies show that alignment is prevalent in both human-computer and human-human interaction, with users in both studies using the less accurate term when it was used by their interlocutor (human or computer). On first consideration, this may suggest that alignment is an automatic process, a perspective supported by Branigan et al.'s (2003) study in syntactic alignment which observed similar rates of alignment for both computer and 'human' addressees, leading to the conclusion that alignment is an automatic imitation mechanism that does not involve any decision or strategic component. However, in Branigan et al.'s (2004) study, lexical alignment was considerably greater where the user was interacting with the computer compared to when their interlocutor was (what s/he thought was) a human, possibly because the former was perceived as being more 'error-prone'. The explanation for this is that speakers align their linguistic behaviour according to the

perceived, rather than actual, capabilities of the system. This is confirmed by observations from Pearson et al.'s (2006) follow-up study which showed that users aligned more to the 'basic' version of the system (than to the 'advanced' one). As the authors point out, this indicates that alignment has a strategic dimension, as users aligned more in order to maximise the likelihood of successful communication (Branigan and Pearson, 2006).

In summary, alignment between humans may be equally mediated by automatic priming processes, a social and a strategic component. Yet, human-computer interaction appears to involve a stronger strategic component, which is specifically clear in the case of lexical alignment.

The studies suggest that users align to systems in both automatic and strategic ways, raising interesting questions as to whether, through appropriate design, users could be made to interact with a dialogue system in a predictable and desirable way. However, caution has to be exercised in applying the results from these studies as the tasks and scenarios that they employed were not naturalistic, meaning that the results may not be readily extended to real-life applications.

2.8.4 Gender-related alignment

As discussed in section 2.4, females and males have different communication styles, indicated by the use and frequency of particular linguistic and supralinguistic features. In addition, there have been indications that the interaction of speaker/hearer gender may alter these patterns. Work by Mulac and his colleagues (Mulac et al., 1988; Fitzpatrick et al., 1995) has focused on understanding the role of gender composition of a dyad in language use. It was found that gender-specific differences in speech characteristics were prevalent in same-gender pairs, but scarce in mixed-gender pair interactions. As such, it was argued that both speakers converge on some aspects of their speech when interacting with the opposite gender and, as a result, gender-related differences are attenuated. The decrease in the use of the language linked to their own gender was reported to be higher between married couples (Fitzpatrick et al., 1995). However, Fitzpatrick et al. (1995) demonstrated that females tended to adjust their own gender-related language twice as much compared to male speakers. On the other hand, husbands aligned more to their wives' style than wives did, but males who

perceived themselves as ‘traditional’ did not accommodate their speech. The level of ‘traditionalism’ about marriage and family of a participant was determined by completing the Relational Dimensions Instrument developed by Fitzpatrick (1988). Generally, women were found to have a stronger tendency to moderate their own language during interactions, with husbands, unfamiliar men and other females. Bilous and Krauss (1988) also found that females converged more than males in interruption rates but less clear results were reported for other variables like pauses, laughter and backchannels (e.g., ‘uh-huh’, ‘ok’ etc). In an earlier literature review, Coates (1986) suggested that females frequently ‘masculinised’ their speech when interacting with males, but male alignment was much less common. Stupka (2011) also observed that females changed their linguistic patterns to accommodate the linguistic style of males, while minimal convergence by males was reported. Namy et al. (2002) presented a study that investigated phonetic accommodation in which participants repeated individual words spoken over headphones by females or males. It was found that female participants phonetically accommodated to a greater extent than male participants, and also converged more to male voices than to female voices. The stronger tendency for women to linguistically align more than men has been attributed to a variety of sociolinguistic reasons; namely, the model of Communication Accommodation Theory proposes that women converge more than men because of their greater need for affiliation and social approval, greater desire for communication effectiveness, ability to be more attentive to accommodation patterns and a lesser concern about not following the stereotypical gendered behaviour (Coupland et al., 1988).

These findings support the hypothesis that females align more to their interlocutor’s speech patterns. However, casual social conversations greatly vary from task-oriented interactions and the variables analysed were exclusively ‘non-functional’ linguistic elements (for instance, backchannels, laughter and intensive adverbs). Therefore, these findings cannot offer generalisable results, but may only serve as indicators.

Moreover, the observation that speakers depart from their own gender-specific style of speaking when interacting with the other gender contradicts the theory that men and women belong to different speech communities. In monologue settings or same-gender interactions, men and women manifest communication styles and strategies related to their gender. However, in mixed-gender dyadic or group interactions, people appear to moderate their linguistic patterns. This gives rise to the argument that language is not gendered but gender-

preferential (Fitzpatrick et al., 1995). Taken together, the findings point to an interaction effect between genders and dyad composition and further suggest that the strength of linguistic alignment is gender-related. Yet, as mentioned above, it remains unknown whether or how these effects occur in goal-oriented interactions.

2.8.5 Interim summary

Dialogue is the primary mode of human interaction and changes the way people coordinate and use language. Therefore, accounts that are based on monologues are of limited use, especially for practical purposes. An interesting phenomenon that occurs during dyadic interaction is the tendency of interlocutors to repeat each other's linguistic structures, a phenomenon which arguably underlies communication success. Alignment, as a mechanism that promotes language repetition, may be exploited in system design to predict and constrain user input as well as yield more natural and felicitous interactions. A number of studies have investigated the operation of alignment in HCI. However, a limitation of these studies is that they used a simple object-naming task, which is weakly related to real-life applications. Therefore, more research is needed to describe the occurrence and effects of alignment in HCI. With regards to the gender factor, studies in social conversation have demonstrated that while single-gender pairs exhibit a strong gender-specific communication style, the differences are attenuated in mixed-gender interactions, and that females align more than male speakers and to male hearers. Yet, there is no data with regards to task-oriented interactions. Still, by illustrating a pair composition effect, these findings provide initial support to the main argument of this thesis that dyadic interaction changes predicted patterns of gender differences.

The choice of perspective, the efficiency of spatial descriptions, and the success of the coordination has been found to depend on people's (combined) spatial abilities, and also on the affordances of the communication situation, such as the visibility between partners. The effect of sharing visual information is discussed in the next section.

2.9 The effect of visual information on communication and performance

In task-oriented interaction, a shared visual space is where the collaborators/interlocutors can see the same objects and environment at the same time. This section considers the ways that shared visual information influences goal-oriented interaction, through the examination of relevant literature. Such research is necessary for understanding phenomena in normal human communication. It also informs the development of computer systems that share the same visual or physical space with human users. Moreover, computer-mediated communication and collaboration technologies usually integrate video or support sharing visual perspective and, therefore, better awareness of the role of visual information as a conversational resource can lead to improved designs. Given the increasing use of these technologies in domains as diverse as education, medicine, and business, this section begins by outlining some concepts and challenges with regards to their development. It then discusses the effect of visual information on coordination processes drawing on empirical work and theoretical elements within the Collaborative Model.

2.9.1 Visual information in task-oriented Computer-Mediated Communication

As noted in section 2.7.2, with the advent and enhancement of networks and related telecommunication technologies, there has been a growing interest in communication and collaboration activities that can be remotely conducted. Computer-Mediated Communication (CMC) and Computer-Supported Cooperative Work (CSCW) takes place in everyday contexts, using speech (such as phone and VoIP), instant text messages, video, and in specialised domains, like videoconferencing, shared workspaces and collaborative virtual environments. These activities rely on a complex and delicate coordination of verbal communication and physical actions. It is well-established that performance in collaborative tasks deteriorates in conditions that lack collocated interaction. Current technologies have failed to compensate for this weakness, hindering the usability and pervasiveness of computer-mediated interactions (Olson and Olson, 2000; Whittaker, 2003a). This failure is

attributed to incomplete understanding of how people coordinate as well as how the mediation technology itself affects these coordination mechanisms (Gergle, 2006).

Communication-mediating technologies are typically classified in terms of two affordances: the mode which they support, that is, linguistic (for instance, phone,) or visual and linguistic (videoconference, shared workspace) and whether they are interactive (phone, instant text messaging) or non-interactive (email) (Whittaker, 2003b). The characterisation of the technologies and their affordances serve to understand how these affect communication behaviours, and, in turn, predict the content, processes and success of communication. In particular, the visual mode may involve physical co-presence, facial expressions, head nods, gaze, gesture and shared visual access to the environment. These elements are expected to have variable effects on attention, understanding, conveying attitudes and emotions, agreement, turn-taking and reference (Whittaker, 2003b). Developers may also use the 3C model to classify a CSCW and CMC application, understand its domain and guide its implementation (Ellis et al., 1991; Fuks et al., 2005); that is, whether it involves and how it combines the following three elements: communication (exchange of messages and information), coordination (managing people and their activities) and cooperation, with the latter referring to complex joint work within a shared space. The effect of the technology and the task are non-trivial, and, thus, developers should not rely on simple intuition or superficial characteristics when analysing requirements and features.

Many studies have focused on the effect of visual information on goal-oriented interaction (Clark and Krych, 2004; Gergle et al., 2004; Kraut et al., 2003; Brennan, 2005). The aforementioned studies compared the condition in which the instructors could monitor the followers' physical actions with speech-only communication and identified an effect of visual information on performance and conversational strategies. They converge on the finding that visual information leads to more efficient interactions. For computer-mediated communication, the implication is that multimodal technologies like videoconferencing that combines speech and vision could better support communication than speech- or text-only interfaces like telephone or instant messaging. Yet, comparable performances in video and audio-only communication are sometimes observed (Sellen, 1995). In fact, it is found that visual information may even disrupt communication, as in the case of unsynchronised or delayed visual feedback (Gergle et al. 2004; O'Malley et al., 1996). In fact, it is argued that the value of visual information depends on two factors; the nature of the information

provided and the task (Whittaker, 2003a). First, interlocutors do not benefit by being able to observe each other's bodies, but by sharing workspace, that is, by viewing physical actions and movements and relevant shared objects in the environment. However, a shared workspace does not necessarily ensure shared perspective, and, thus, misplaced assumptions of joint focus of attention and point of view ultimately hinder coordination (Whittaker, 2003a). Second, a study by Whittaker et al. (1993) contrasted speech-only communication with a condition of shared workspace and speech in three tasks. More successful and efficient interactions were observed for spatial and editing tasks but not for a brainstorming task.

The role of shared visual information in computer-mediated communication has been an active area of investigation, but visual information does not always lead to more efficient interactions, to the degree predicted by studies in human communication. At the same time, how interaction with systems (such as robots) is affected by the presence/absence of visual information remains largely unexplored. The following subsections illustrate the theoretical principles underlying the benefit of visual information to communication and performance in collaborative tasks.

2.9.2 The effect of visual information on grounding and situation awareness in task-oriented interactions

Studies in computer-mediated communication and prototypical human communication in collaborative tasks describe the effects of visual information in terms of *grounding* and *situation awareness*, largely relying on concepts from the Collaborative Model developed by Clark and his colleagues (Clark, 1996).

In dialogue, interlocutors engage in a process of *grounding* of their utterances, that is, they continuously seek and provide evidence that the utterances that have been presented have also been understood (Clark and Brennan, 1991). Grounding can be explicit, like backchannels or implicit like moving on to next utterance. The necessary form, strength and amount of grounding are determined by the interlocutors depending on, inter alia, the resources afforded by the communicative medium (Clark and Brennan, 1991; Clark and Marshall, 1981). A criterion underlying these decisions is the *principle of least collaborative effort*, according to which, people select the method of grounding that takes the least

collective effort for the interlocutors, in terms of time, errors, resources etc. Communication efficiency depends on the amount of common ground that the interlocutors secure. Common ground is the result of linguistic and physical/visual co-presence (Clark and Marshall, 1981). In conditions of physical/visual co-presence, interlocutors share visual and auditory common ground and, as such, the strongest type of evidence is afforded. That is, visual evidence of understanding is faster and more secure than spoken claims of understanding (Clark and Marshall, 1981). However, as previously mentioned, the nature of the shared visual information is important; namely, grounding is more efficient when speakers can monitor their addressee's workspace (Clark and Krych, 2004) compared to when viewing their faces (Whittaker, 2003a). So, for instance, interlocutors collaborating in a task and sharing workspace might opt for grounding methods like pointing and nodding rather than verbal contributions. When sharing visual information and viewing the same objects, those objects are part of the common ground and joint attention and reference can be easily established. As a result, the act of referring to elements in the environment becomes short and efficient, leading to utterances such as 'put that there'.

When visual feedback is not available, speakers compensate at a time cost having to verbally assert that something was understood or executed. Under such conditions, the addressees are in the best position to confirm their understanding and, as such, their partners rely on them (following the *principle of mutual responsibility*, described in Clark and Wilkes-Gibbs (1986)). But if the speaker can monitor the addressee's actions, interlocutors expect that the speaker should assume this responsibility and assess the perceptual evidence provided by the addressee, and thus saving the addressee from having to produce an utterance. Generally, the responsibility falls to whoever is judged to have the strongest evidence, so that collective effort is minimised.

Therefore, the structure of turn-taking in the interaction is adaptable; with visual evidence, grounding is not performed with discrete utterances, but with visible physical actions, and understanding is established continuously and instantaneously. Indeed, cross-timing participants' actions indicates that, when visual evidence is available, grounding overlaps with the planning and presentation of the subsequent utterance by the speaker (Brennan, 2005; Clark and Krych, 2004). As such, dialogue is said to be an 'artful orchestration' of vocal and gestural signals (Clark and Krych, 2004, p.79).

Shared visual information also enables speakers to maintain *situation awareness*; namely, they can monitor task status and their partner's activities. That is, speakers can assess the progress of the task, and the information necessary towards its accomplishment. Moreover, monitoring one's actions and their completion means that the next instruction will be provided precisely at the moment needed. Similarly, an incorrect execution is readily recognised by the partner and he/she can take immediate action towards repairing it.

In summary, visual information produces more efficient interactions because speakers are able to ensure mutual understanding immediately and reliably, establish joint reference to objects of interest, formulate shorter and fewer utterances and monitor task status and their partner's actions. It should be reiterated that these benefits of visual information in task-oriented interactions have been confirmed for shared workspaces, but are less pronounced for other situations like viewing partners' faces and bodies (Kraut et al., 2003, Whittaker, 2003a), and the latter remaining out of the scope of this thesis. Therefore, *visual information, shared workspace* and *visual feedback* will be the terms interchangeably used henceforth to refer to shared visual workspace.

2.9.3 Interim summary

Based on theoretical and empirical accounts, visual information clearly benefits performance and communication processes in task-oriented interactions, because collaborators can ground information more efficiently and maintain higher situation awareness. While there has been significant work on visually-supported CMC, the predicted benefit is not realised in all cases. As such, further research is needed to identify the techniques to better exploit visual information in CMC. Moreover, it would be interesting to explore how spatial language is produced and understood, given its numerous application areas that range from multi-player games to remote coordination of teams. Such research findings could also be relevant to the design of robots which operate under supervision or no supervision in collaborative tasks. Moreover, in connection to the findings discussed in section 2.7 (particularly, in section 2.7.5), it is argued that there are differences in how females and males process, use and benefit from visual elements on computer interfaces. Therefore, systematic research is required to determine the ways that visual information influences performance outcomes, communication processes and the strategies of users and how these are mediated by gender.

2.10 Miscommunication

This section provides an overview of miscommunication. After discussing miscommunication in human communication literature, it focuses on related issues in the field of spoken dialogue systems. It concludes by presenting miscommunication as it arises between people of different genders.

An error is a general term that denotes an action or decision that results in one or more unintended negative outcomes (Strauch, 2004, p.21). In dialogue studies, errors and other problems are referred to as miscommunication. Most research in linguistics emphasises a model of communication based on the 'conduit metaphor', according to which, the speaker encodes a message into a signal and transmits to a receiver who, in turn, decodes the signal and reconstructs the message. Such accounts are problematic because they presuppose a level of transparency between the information and mental states of the interlocutors that is rarely possible and imply that mishearings (because the signal was corrupted by 'noise' in the channel) are the only potential form of miscommunication. Miscommunication has been viewed as a pathological and marginal phenomenon of language, and, as a result, neglected up until recently (Coupland et al., 1991). Indeed, empirical studies demonstrate that communication is inherently imperfect and partial, and, miscommunication is now generally recognised as a natural and ubiquitous phenomenon.

The pervasiveness of miscommunication is said to relate to the principle of least collaborative effort (discussed in section 2.9.2) (Clark and Brennan, 1991). People try to complete a task putting the least effort possible to achieve a satisfactory result, and as Carletta and Mellish (1996, p. 71) maintain, 'in task-oriented dialogue, this produces a tension between conveying information carefully to the partner and leaving it to be inferred, risking a misunderstanding and the need for recovery.' Thus, when the interaction conditions are favourable (as in case of visual co-presence, see section 2.9.2), speakers typically opt to use deictic expressions, such as 'put that there', instead of 'put the book in the box', which is more economic but increases ambiguity and the risk of incorrect interpretation.

In many accounts, miscommunication is considered from the point of view of the addressee. Hirst et al., (1994) and McRoy (1998) distinguish two main types of miscommunication, misunderstandings and non-understandings. A misunderstanding occurs

when the addressee obtains an interpretation that he/she believes is correct and complete, but not the one that the speaker intended him/her to obtain. Misunderstandings go unnoticed and interlocutors may continue to converse at cross-purposes; they are only detected when the addressee acts upon them. A non-understanding occurs when the hearer obtains an uncertain interpretation of an utterance, no interpretation or more than one. Instances of non-understandings are immediately recognised, as the hearers are aware of them and articulate them. McRoy (1998) also distinguishes misinterpretations which occur when the most likely interpretation of a participant suggests that their beliefs about the world are out of alignment with those of their interlocutor. Other approaches shift the focus away from the addressee to the dyad, and refer to the concept of grounding (see section 2.9.2), according to which, successful communication is achieved through a process of mutual grounding between interlocutors. These approaches include the four-level hierarchy of communication, independently developed by Clark (1996) and Allwood (1995). According to this model, miscommunication can occur at any linguistic level, from failing to establish contact with the speaker to failing to recognise the function of the utterance in context. Interlocutors select repair initiations that indicate their current level of understanding and the source of the problem. The model was adapted by Mills and Healey (2006) and is shown in Table 2.1. Based on this model, Gabsdil (2003), Schlangen (2004) and Rodriguez and Schlangen (2004) proposed a classification of clarification requests in task-oriented dialogue and outlined the kinds of problems that may occur at each level of the hierarchy. Table 2.1 summarises the classification and the four-level model. The right-hand column in the table contains example of utterances that may give rise to these specific problems.

Table 2.1. The four-level model of communication (adapted by Mills and Healey (2006) from Clark (1996) and Allwood (1995)) and problems that can occur according to the classifications by Schlangen (2004) and Rodriguez and Schlangen (2004).

Level of communication		Kind of problem	Examples
4	Action Recognition	Problem with recognising or evaluating the intention	‘Do you have a pen?’ Question or request?
3	Meaning Recognition	Lexical problem, Parsing problem, Reference and contextual relevance resolution problem	Unknown vocabulary ‘I shot an elephant in my pyjamas’, ‘Put that there’

2	Utterance Recognition	Acoustic (speech recognition) problem	'Recognise speech' vs. 'Wreck a nice beach'
1	Securing Attention	Channel	-

In dialogue systems, speech recognition errors are the predominant source of miscommunication (Bohus and Rudnicky, 2005). Because of this, handling of miscommunication in dialogue systems is concentrated on misunderstandings and non-understandings by the system, with less attention given to other sources of miscommunication, such as problematic input by the user. In particular, miscommunication often arises due to false assumptions held by the user with regards to the linguistic and functional capabilities of a dialogue system or an embodied agent. Thus, out-of-grammar vocabulary and requests for unavailable functionalities are frequent. In human-robot interaction, physical co-presence may lead users to make misplaced assumptions of mutual knowledge (see section 2.9.2), increasing the use of underspecified reference and deictic expressions. Robots operate in and manipulate the same environments as humans, so failure to prevent and rectify errors has potentially severe consequences. From the domain of navigation, it is well-known that route instructions are structurally underspecified and arbitrary. Most importantly, since situated dialogue involves dynamic temporal and spatial events and giving route instructions can be a cognitively demanding task, users are liable to err, and issue incorrect instructions. For example, people often confuse 'right' and 'left'. Indeed, MacMahon (2004) observed that one third of route instruction protocols written by participants were so problematic that they resulted in the followers being lost. McTear (2008) has predicted that as the accuracy of speech recognition improves, miscommunication due to the aforementioned factors will be more prominent and important to handle. In conclusion, as miscommunication grows in scope, frequency and costs, the necessity to integrate it in the analysis and design process of interactive systems becomes imperative.

Research from social sciences has provided insight into miscommunication from diverse contexts; for instance, miscommunication arises between people of different nations, cultures, religions, generations, class and gender (Coupland et al., 1991). Studies found that miscommunication is common at the workplace, for example, between senior and junior managers, and is particularly prevalent between female and male colleagues and employees (Stubbe, 2010). A simple explanation is that same words (or paralinguistic features, like gestures) carry different meanings for members that belong to different gender, age, culture or organisational level groups. McTear (2008, p. 105) has argued that most of this research

has been deemed irrelevant and overlooked by developers of dialogue systems and interactive interfaces.

The review in section 2.4 showed that gender-related differences in language occur in how, why and about what men and women talk. Given the same communication task, males and females will consistently opt for particular linguistic and stylistic elements (for instance, use of adverbs and references to emotions). Such observations have led many researchers to argue that males and females essentially form two distinct linguistic communities (see, for instance, Tannen, 1990). Similarly, gender differences do not only concern language production but also interpretation (Tannen, 1994). These differences ultimately explain why miscommunication arises between females and males. For instance, Maltz and Borker (1982) and Mulac et al. (1998) suggest that men and women understand and use questions and backchannels differently. In particular, these studies observed that, to men, questions serve to elicit information and backchannels are used to indicate agreement. On the other hand, women were found to be more likely to use these features to maintain the conversation. Therefore, this dichotomy is bound to produce miscommunication between females and males. Inter-gender miscommunication has been predominantly examined in social terms, and particularly, those of dominance and power. For instance, Henley and Kramarae (1991) argued that miscommunication between males and females is not simply a by-product of different sociolinguistic community membership (as suggested so far, see, for example, Tannen, 1994; Maltz and Borker (1982)), but a tool to support male dominance of conversation.

2.10.1 Interim summary

Miscommunication is particularly interesting from the socio-linguistic perspective of gender and its practical relevance for interactive interfaces. Given miscommunication is a natural and ubiquitous phenomenon of communication, its occurrence and, most importantly, its effects should be part of any account of language use in both contexts of human-human interaction and HCI.

2.11 Chapter summary

How people generate and follow route instructions has implications for theoretical fields and the design of practical systems. The review of the relevant literature has illuminated two areas of concern. There are robust gender differences in navigation performance and language use. Gender is also a major factor that permeates skills, performance outcomes, attitudes and perceptions across numerous application domains of HCI. Yet, system design typically excludes gender considerations. Dialogue is the most basic and primary setting of communication, and has a fundamental effect on how people perform, coordinate and use language. Yet, the overwhelming majority of existing literature has studied navigation and route communication in monologue settings. As such, the generalisability of these findings to other contexts of use, including human-computer interaction, is problematic.

Targeting these knowledge gaps, the work presented in this thesis sets out to investigate gender differences in navigation and communication in real-time dialogue with a computer system. This approach enables the investigation of the operation and development of alignment in human-computer dialogue, in problem problem-free communication as well as in cases of user and system errors. The empirical study also considers how shared visual information is utilised as a communication resource in the accomplishment of the navigation task. The findings of this investigation could enhance awareness in how the user's gender influences behaviour, strategies and, ultimately, success. Such insight may be exploited in the design of effective system interfaces that support users of both genders by promoting successful behaviours and strategies.

This chapter described gender differences by drawing from research in the diverse fields of cognitive psychology, linguistics and human-computer interaction. The review of existing findings related to the themes of navigation performance, route instruction production and following, HCI, and the natural dialogue phenomena of alignment and miscommunication led to the identification of specific research gaps. These gaps merit further experimental consideration. Therefore, the following chapter examines the gaps and uses them to develop specific research questions.

3 Research Questions

3.1 Introduction

The review of the literature illustrated that gender is a major factor mediating performance in spatial tasks and language use, revealing an intricate pattern of research findings. Gender differences have also been documented in many domains of HCI, but our knowledge remains incomplete, especially in the areas of collaborative systems and natural language interfaces. The literature review also discussed empirical models of human communication – on which research in HCI often draws – which argue that how people produce and understand language and coordinate in task-oriented dialogue largely depends on inter-individual processes and the context of use. However, the vast majority of studies in gender and spatial language use non-interactive or artificial experimental settings. Therefore, their findings may be of limited use for the field of HCI and may not be extended to other contexts of interaction. These knowledge gaps give rise to the central research question of the thesis, of *how gender differences emerge in spatial navigation dialogues with computer systems*. By addressing this question, the thesis aims to produce implications for communication theory development as well as ecologically-valid design guidelines for the development of future collaborative systems and natural language interfaces.

Using an existing dialogue system in the empirical study would be of little value for future applications, thus, defeating the research objective of this thesis. Therefore, the experimental approach involved pairs of participants (dyads) collaborating in a robot navigation task, with one of the participants, the user, being under the impression that he/she was giving route instructions to a robot (which was simulated by the other participant). The dyads were seated in separate rooms and communicated using a custom tool that supported

synchronous text-based communication and execution of instructions. The experimental approach is described in detail in the following chapter.

The complexity of the central research question necessitates its analysis and decomposition. As such, the central research question will be addressed in a bottom-up process, through understanding its major components, that is, the concepts relating to navigation *performance* and *dialogue*. In particular, performance is interpreted through the analysis of *miscommunication* (that is, user errors and system errors and non-understandings), actual task performance (such as time and number of turns) as well as subjective user opinions. Similarly, dialogue is construed through the examination of linguistic *alignment* and the analysis of the dialogue acts of the interlocutors and their components (for instance, references to landmarks). Since miscommunication is a communication phenomenon, it is also viewed in conjunction with alignment. Finally, having established that shared *visual information*, or the lack thereof, has specific effects, it will be used as an experimental manipulation to understand the mechanisms that underlie gender differences in performance and communication. The concepts and their relations investigated in this thesis are encapsulated in the schematic diagram below (Figure 3.1). The solid-line boxes illustrate the concepts that play a major role in the development of the thesis and are thoroughly discussed, whereas the concepts in the dotted-line boxes were dealt as auxiliary but necessary to address the central research question.

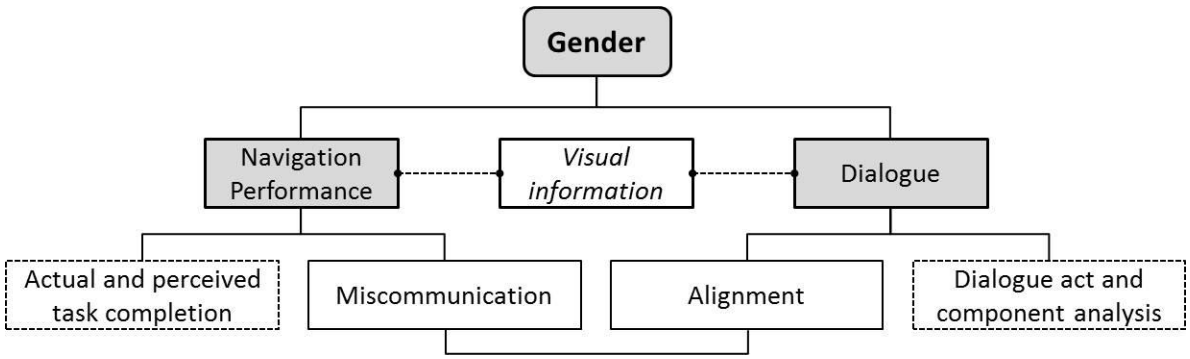


Figure 3.1: Diagram outlining the concepts analysed in this thesis and their relations.

This decomposition of the central research question is motivated by current literature as discussed in Chapter 2 and serves as the guide for the data analysis approach adopted in the thesis. This chapter revisits the main themes of the literature, the findings, knowledge gaps and unresolved arguments, which, in turn, generate specific research questions. These

research questions are grouped under three main themes, which are listed and discussed below.

A. Gender differences in performance and route communication in interaction.

B. Effect of visual information on performance and communication in HCI and the effect of its absence by gender.

C. Alignment in HCI and gender-related alignment in task-oriented interaction.

With regards to the first theme, the discussion of gender-specific performance and qualitative differences in the domains of navigation and communication may suggest particular directions for experimental hypotheses; for instance, females appear to be poorer navigators and rely on landmarks. Moreover, previous findings predict that females report lower task success perceptions and self-efficacy. Yet, because of lack of data from interactive studies, the ways that interaction will convolute the expected patterns are largely unknown. In fact, interesting permutations are anticipated as a result of the gender composition of dyads. With regards to the second theme, previous findings about the influence of shared visual information lead to specific predictions regarding task performance and the content and structure of communication. The expectation is that the availability of visual information will result in more successful interactions and produce major changes in language use and turn-taking patterns. If the expectation is validated, the question that naturally follows is whether it creates additional gender differences or moderates existing ones. With regards to the third theme, the mechanism of linguistic alignment is prevalent in human communication and is argued to be the basis of interactive success. Although it has important implications for the development of natural language interfaces, the occurrence, nature and role of alignment in HCI remain ill-defined. Therefore, after gaining a better understanding of alignment, the study will focus on the effect of gender on its development and operation.

The following sections define research questions that address gender differences in performance, communication and user perceptions in interaction (section 3.2) and how these are influenced by visual information (section 3.3) and influence the processes of alignment (section 3.4). To this end, section 3.3 begins by presenting a set of questions to clarify the effect of shared visual information on performance and communication patterns, while

section 3.4 begins by formulating research questions motivated by the analysis of current literature on alignment in HCI and human communication.

The following chapter provides an overview of the study design developed to answer the central question of the thesis through addressing the sub-questions. While several questions echo a particular prediction about outcomes (based on past research), others remain non-directional.

3.2 Gender differences in navigation performance and communication in dialogue

In Chapter 2, the review of literature revealed gender differences in spatial task performance. In particular, a male advantage consistently emerges in psychometric tasks such as mental rotation. A similar conclusion has also been reported by empirical studies that employ real-world route instruction tasks. Distilling the results of these studies, males are generally associated with more accurate wayfinding performance, superior strategies and are also able to provide more robust and efficient instructions. Previous findings show that females rely on landmarks when formulating route instructions and also as a sole strategy to navigate, such that their performance deteriorates when they receive route instructions deprived of landmark references. Yet, while males have a preference for directional instructions, their performance remains unaffected when following route instructions with or without landmark references. This research tends to be from a particular perspective/set of assumptions or simplifications that are problematic. The most important of these is that it sees language production and comprehension in isolation, lacking interactivity between the parties. In effect, there appears to be a gap in relation to how gender differences arise in wayfinding performance and strategies in interactive communication.

3.2.1 Gender differences in navigation performance

While the aforementioned findings are sourced from non-interactive studies on route instruction -giving or -following, this thesis investigates dialogue. If these findings were extrapolated to form hypotheses in dialogue, it would be predicted that all-male dyads[7] (that is, a male instructor/user interacting with a male follower/'robot') would outperform all other pairs, and pairs with female instructors/users (hereafter, users) and male followers/'robots' (hereafter, 'robots') would have lower or comparable performance. Pairs with female 'robots' would be anticipated to be the least successful. Moreover, it could be hypothesised that the performance of female 'robots' paired with male users would deteriorate because males provide purely directional instructions. However, such an approach would essentially lead to the same fallacy of the 'conduit' model of communication, having assumed that interaction is the sum of two separate processes of language production and interpretation. Therefore, these findings may serve to form inferential research questions. The first set of questions concerns whether males are associated with more efficient (faster and accurate) task performance.

Research Question 1(a)

Are all-male pairs (male users interacting with male 'robots) the fastest in completing the navigation task?

Research Question 1(b)

Are pairs with male 'robots' faster in completing the navigation task than pairs with female 'robots'?

Research Question 1(c)

Do all-male pairs produce the lowest miscommunication (execution errors, non-understandings and inaccurate instructions)?

Research Question 1(d)

[7] 'Dyad' or 'pair' in this study refer to two participants interacting in a Wizard-of-Oz set-up, with one participant, the 'user', giving real-time route instructions to a follower, whom he/she believes is a 'robot system'.

Do male 'robots' produce fewer execution errors and non-understandings than female 'robots'?

Research Question 1(e)

Do male users provide fewer inaccurate instructions than female users?

3.2.2 Gender differences in route communication

Accordingly, the second set of questions assesses previous research findings with regards to gender-related preferences and strategies in formulating spatial descriptions by users and 'robots' and, in particular, whether females in either role use landmark references more frequently than males.

Research Question 2(a)

Do female users include more landmark references in their utterances than male users?

Research Question 2(b)

Do female 'robots' include more landmark references in their utterances than male 'robots'?

A central element in the conceptualisation of this thesis is the inter-individual processes that permeate communication. As such, it is assumed that the dyads' gender composition, that is, the interaction of genders and roles, will impact the results.

3.2.3 Gender differences in user perceptions of performance

In addition to objective parameters for assessing task performance, the investigation looks at subjective aspects of gender differences. Research discussed in Chapter 2, section 2.2.3.3 highlighted that different affective and psychological processes operate when females and males perform spatial tasks. In particular, females regularly report feelings of high anxiety and fear. They also provide judgments of low self-efficacy and beliefs of poor performance before and after a wayfinding task, irrespective of actual results (see Lawton 1994; 1996; Lawton and Kallai, 2002). Moreover, as presented in Chapter 2, section 2.7.1, women were found to have less self-efficacy when completing complex computing tasks (Busch, 1995). Thus, it is interesting to probe whether there are differences between females and males in

perceptions of their performance and their experience with a computer system when instructing it on how to perform a navigation task.

Research Question 3

Do female users rate their performance lower than male users?

3.3 The effect of visual information on navigation performance and communication between user and system

Previous studies have provided substantial evidence regarding the effect of the absence/presence of visual information on performance and communication, allowing for the formulation of specific hypotheses. A shared visual workspace leads to more efficient task-oriented interactions as it offers several advantages for the accomplishment of the task, namely: it enables moment-by-moment direct observation of task status; it provides visible feedback on the addressee's understanding and activities, and aids in establishing a joint focus of attention and common reference frame (see Chapter 2, section 2.9). These observations are derived from task-oriented human-human dialogues in normal and computer-mediated communication and are accounted for by the Collaborative Model (Clark, 1996), mainly through the process of grounding. Yet, it may be only speculated that the common ground between user and computer is modelled similarly to common ground between humans, or, indeed, at all. It is, thus, necessary to ascertain whether comparable coordination and communication patterns occur when users share visual space and communicate with an artificial agent. Moreover, there are no comparative studies that focus on spatial tasks, such that it remains unclear how visual co-presence influences the configuration of route instructions, like, for instance, the use of environmental features.To this end, this study obtained experimental data from two interaction conditions, a condition in which users were able to observe the actions of 'robots' and a verbal-only condition. Two sets of inferential research questions are constructed and concern task performance and communication. Addressing these questions is a necessary step and primarily serves to expose aspects of gender differences in task-oriented HCI. In particular, the questions are used to inform the development of further research questions that essentially attempt to determine whether females and males are more sensitive to variations in interaction conditions. These questions will be enumerated in section 3.3.1.

Visual information allows for higher situation awareness through monitoring of task status and the actions of the addressee. According to previous findings, it is predicted that when visual information is available, task performance will be more efficient compared to a verbal-only condition, leading to the following specific research questions:

Research Question 4(a)

Are tasks completed more quickly when visual information is available?

Research Question 4(b)

Does the number of incorrect instructions by the user decrease when visual information is available?

Research Question 4(c)

Does the number of execution errors and non-understandings by the 'robot' decrease when visual information is available?

Research Question 4(d)

Do interlocutors use fewer words, turns and route instructions to complete a task when visual information is available?

From the perspective of communication processes, previous findings presented in Chapter 2, section 2.9.1 confirm that visual co-presence enables joint reference which allows for the use of shorter, unambiguous referring expressions. The principle of least collaborative effort predicts that a physical action renders a verbal turn redundant, eliminating the expectation for the addressee to verbally assert execution and understanding, and the responsibility for coordinating the interaction shifts to the user. Put simply, if 'robots' are aware that users can see what they are doing, then their action serves to demonstrate understanding and substitutes turns. As such, turn-taking is expected to be dominated by users. Therefore, the expectation is that *when visual information is shared, communication will be more economic compared to a verbal-only condition*. It is investigated through the following specific research questions:

Research Question 5(a)

Does the use of deictic pronouns and expressions (for example, 'turn there' and 'take this turn') increase when visual information is available?

Research Question 5(b)

Does the number of verbal acknowledgements by the 'robot' decrease when visual information is available?

Research Question 5(c)

Are route instructions less detailed, precise and explicit when visual information is available?

Research Question 5(d)

Are spatial descriptions by the 'robot' less detailed, precise and explicit when visual information is available?

Research Question 5(e)

Does the number of user-initiated queries decrease when visual information is available?

Research Question 5(f)

Does the number of user turns exceed 'robot' turns, when visual information is available?

Landmarks are central for navigation and their frequency correlates with gender. Yet, it remains unclear whether visual co-presence between instructor and follower, user and computer system, promotes or restricts the use of landmark references. On one hand, it might be plausible to assume that, in a shared visual space condition, references to landmarks will be prevalent, since they are objects of reference that belong to the common ground. On the other hand, incorporating a large number of landmark references is a resource-intensive activity and opposes the principle of least collaborative effort, and, thus, interlocutors may omit them or replace them with shorter expressions, like deictic pronouns. This frames the following research question:

Research Question 5(g)

Do interlocutors use fewer landmark references to complete a task when visual information is available?

3.3.1 Gender and visual information

The literature demonstrates that the availability of visual information facilitates grounding and situation awareness which improve task performance and alter coordination patterns. The validity of these findings for the application domain of this study is explored through the research questions presented above. The next step in the investigation is to determine whether the magnitude of the benefit in performance and the impact on language use due to the availability of visual information varies with gender. In other words, the study considers (i) whether the performance of females or males is more vulnerable to changes in interaction conditions, and, specifically, less optimal interaction conditions of no visual information, and (ii) whether females or males exhibit stronger tendencies to adapt their communication strategies in response to such changes.

It is expected that performance and communication efficiency will be comparatively lower in an interaction condition deprived of visual information. Results from studies in gender differences in navigation tasks oscillate between marked differences favouring males and no gender differences. The explanation proposed by researchers is that gender differences emerge only when the task is more difficult and disappear when the task is easy (Coluccia and Losue, 2004). This argument can lead to the hypothesis that the absence of shared visual information will be more detrimental to women's performance than it is to men's. Yet, this may be a precarious assumption, unless the role of landmarks is taken into account. In particular, it was proposed that the availability of a shared visual workspace will have a direct effect on the frequency of landmark references (explored in Research Question 5(g)). One possibility is that visual feedback will reduce the necessity for explicit spatial descriptions and the use of landmarks will decline. An alternative scenario is that landmarks will be mutually perceived, which will encourage interlocutors to refer to them. These two possibilities have specific implications for females, given females' total reliance on landmarks compared to males; that is, the lack of landmarks will impair females' performance and the abundance of landmarks will improve it. Thus, more empirical evidence is required to clarify whether the performance of female users and 'robots' will be impaired in the less optimal interaction condition of no visual information. This leads to the following question:

Research Question 6(a)

Is task performance of females more negatively affected by absence of visual information than males' performance?

As previously noted, previous research supports the expectation that, in addition to an impact on performance, the presence/absence of visual evidence will shape the content and structure of communication. However, there is no empirical data with regards to whether the effect of visual information on communication processes will be stronger, weaker or altogether different depending on gender. This gap can be addressed in the following question:

Research Question 6(b)

Do females adapt their communication strategies more than males in response to lack of visual information?

3.4 Alignment in human-computer dialogue

Speakers tend to repeat their own and each other's linguistic choices in dialogue, leading to alignment across linguistic and situational levels, a phenomenon which arguably underlies communication success. Linguistic alignment, as a mechanism that promotes language re-use, can be exploited not only to support natural interaction but, more importantly, to predict and constrain the variability of user input. Yet, development of interactive systems has overlooked this natural tendency. Having identified the importance of better understanding of alignment not only to human communication but also to the field of HCI, this investigation aims to identify and categorise the occurrence of alignment in users' interactions with computer systems. It ultimately aims to determine whether alignment correlates with gender; namely, whether female or male speakers have stronger tendencies to align to their (human or artificial) partners. The related research questions are detailed in the following subsection (3.4.1).

The studies discussed in Chapter 2, section 2.8.3 provided strong evidence regarding the presence of alignment in HCI. However, four possible limitations have been identified. First, the studies employed tasks and scenarios (e.g., object-naming) that were restricted and only weakly related to real-life applications. Second, they failed to assess the fundamental characteristic of alignment; in particular, that alignment is mutual. Instead, they focused on

the ‘one-way’ alignment of user to system. It would be interesting to see whether user alignment varies depending on whether the system is also primed to repeat user’s expressions. Third, alignment was measured in interaction with a system that was completed in two utterances. Yet, alignment operates and develops over the full course of a dialogue (as shown from the original ‘maze game’ experiments by Garrod and his colleagues), during which, other natural phenomena, like miscommunication, arise. Fourth, these studies provide evidence of the local priming mechanism of alignment (‘input/output matching’), with less scope for the global, longer-lasting alignment that persists throughout the dialogue (relating to ‘dialogue routines’). As a result, questions remain with regards to whether and how alignment occurs and develops in human-computer dialogues.

Motivated by these studies and in an attempt to address the noted limitations, the thesis uses experimental data from human-robot dialogues to address the following inferential questions:

Research Question 7(a)

Does alignment occur in the interaction between a human user and a computer system?

Research Question 7(b)

If alignment does occur in this context, is it a mutual phenomenon?

As outlined in section 3.3 above, previous studies showed that the availability of visual information has a prolific effect on task accomplishment and communication structure. These findings give rise to rich questions with regards to how visual feedback affects the interaction with a computer system, leading to Research Questions 4(a) – 4(d) and 5(a) – 5(g). Given the focus of this work on alignment, it would be interesting to identify how visual feedback influences the coordination mechanism of alignment between a human and a computer system. In particular, the following research question seeks to identify whether the strength of alignment is different across two conditions of (i) absence and (ii) presence of visual information.

Research Question 7(c)

Does visual information influence alignment between a user and a system?

In addition to the practical importance, exploring whether alignment is stronger or weaker depending on the interaction condition may have implications for theoretical models of communication. As shown in section 2.8, findings remain inconclusive regarding whether alignment is an automatic, 'post-conscious' (Bargh, 1989)[8] process or an optional strategy that interlocutors employ to maximise the probability for communication success. Therefore, if it is found that alignment is consistent across both conditions of presence and absence of visual information, it may suggest that it is an automatic mechanism that ordinarily occurs irrespective of situation. On the other hand, if alignment is stronger or weaker in one condition, it could hint at the existence of a strategic component.

The next research question is concerned with miscommunication. As discussed in Chapter 2, section 2.10, instances in which the hearer fails to correctly interpret an utterance are natural and ubiquitous in goal-oriented human communication. Similarly, speakers commonly produce not only underspecified and vague utterances, but also inaccurate ones. For systems with natural language interfaces, miscommunication is more prevalent owing to natural language understanding errors, out-of-grammar words and out-of-functionality commands. Thus, there is considerable scientific interest distributed in areas like error prevention, detection, prediction and recovery.

Relevant to the objectives of this thesis, in addition to using the frequency of miscommunication to quantify task performance, it is important to understand the behaviour of users when miscommunication is detected. Miscommunication appears to be the basis of linguistic change, as it is at this point when speakers need to consciously reformulate their utterances – to be more compatible with what the 'hearer' can understand. Therefore, it is expected that miscommunication will disrupt lexical alignment, leading to the next research question below. Within the same problem domain, it is practically relevant to continue the

[8] According to Bargh (1989), 'post-conscious' processes are unconscious processes based on information that have first been encoded consciously. Empirical research in social priming by Bargh and colleagues (e.g., Bargh, 1989) and Higgins and colleagues (Higgins et al., 1977) have suggested that priming is post-conscious. At the same time, pre-conscious and post-conscious processes are usually lumped together as they are functionally equivalent (Bargh, 1989), and, as such, this thesis does not make any attempt to differentiate between pre-conscious and post-conscious processes. The key distinction made here and in relevant literature lies between *automatic* and *conscious* processes.

investigation to find out whether users will attempt to recover from an error by using vocabulary that 'worked' earlier in the dialogue, or they will use an entirely novel expression.

Research Question 7(d)

Does miscommunication locally disrupt the process of alignment in human-computer communication?

As noted in section 2.8.1, the main premise of studies adopting the Interactive Alignment Model is that alignment underlies successful communication. Moreover, there is evidence that alignment has a social dimension, leading people to align their verbal and non-verbal behaviour to express affiliation (Giles et al., 1991), and that this behaviour is perceived favourably by peers. Although it is a contentious issue whether the same social norms persist in people's interactions with computers (see Nass et al., 1999), research has shown that users rated more positively systems that imitated their head movements (Bailenson and Yee, 2005), personality attributes (Moon and Nass, 2001) and acoustic and prosodic features (Nass and Lee, 2001; Ward and Nakawaga, 2002). Therefore, Research Question 7(e) deals with the relationship between alignment and user evaluation of interaction success:

Research Question 7(e)

Does lack of alignment also compromise user perception of interaction success?

While the studies discussed in Chapter 2, section 2.8.3 explored an unknown territory and provided original ideas and novel data on the operation of alignment in HCI, there was no focused attempt to draw specific recommendations for interactive systems. Thus, tapping the findings from the research questions outlined above (Research Questions 7(a) – 7(e)), the thesis aims to distil guidelines relevant to the development of practical, goal-oriented dialogue with systems. Moreover, there has been limited work in developing formal or computational models that leverage the effects of this mechanism. Therefore, this thesis aims to describe a theoretically- and empirically-motivated dialogue model that supports and exploits alignment.

3.4.1 Gender and alignment

The frameworks that describe linguistic alignment do not make any claims about individuals being stronger aligners than others. There are a few studies, however, that show that certain personality traits facilitate or inhibit priming effects (Gill et al., 2004; Brockmann et al., 2005). The question that naturally arises is whether gender is a factor that arbitrates the strength of linguistic alignment.

As discussed in Chapter 2, section 2.8.3, empirical data from social interactions suggest that females and males uphold a strong gender-preferential communication style (that is, exhibiting all stylistic features attributed to their gender), when conversing with a person of the same gender. However, in mixed-gender interactions speakers align to their interlocutors' gender-related style and, thus, differences are moderated. Although the results are largely inconclusive with regards to whether men align to female interlocutors, it is consistently reported that women strongly align to men. Existing research has focused on alignment in terms of stylistic elements from social conversations and phonetic alignment while repeating recorded words. Thus, it remains unknown whether similar patterns emerge for alignment at other linguistic levels in goal-oriented interaction. Therefore, building on previous findings and in order to address this knowledge gap, the following questions are framed:

Research Question 8(a)

Do female speakers align more strongly than male speakers in task-oriented interaction?

Research Question 8(b)

Do female speakers align more strongly to male addressees than to female addressees in task-oriented interaction?

Research Question 8(c)

Do speakers in same-gender pairs align more strongly than mixed-gender pairs in task-oriented interaction?

Interlocutors in mixed-gender pairs generally appear to depart from their own gender-preferential register and accommodate to their partner's style. Thus, if gender-preferential language can be extended to encompass preferences in forming route instructions, it gives rise to the following question:

Research Question 8(d)

Do users in mixed-gender pairs moderate the use of their own gender-preferential strategies and provide instructions as preferred by their addressees (landmark-based and purely directional route instructions to females and males, respectively)?

In effect, the issue of gender and navigation strategy choice addressed in Research Questions 2(a) and 2(b) is reframed as a question of whether, and to what degree, a speaker's linguistic choices is influenced by the needs associated with the gender of the recipient. Given that the experimental approach of the thesis involves a remote Wizard-of-Oz setup, thus, masking the interlocutors' gender, the findings related to Research Questions 8(b) – 8(d) could provide interesting findings.

Research Question 7(d) focuses on the effect of miscommunication on the operation of alignment, with the aim to identify changes in user behaviour in terms of linguistic input after the occurrence of user and system errors and problematic understanding. This question is extended to consider whether gender determines how users respond to and handle miscommunication.

Research Question 8(e)

Does miscommunication have different effect on male and female users in terms of communication strategies?

3.5 Chapter summary

Little is understood with regards to gender differences in the interactions with collaborative and dialogue systems. Moreover, the relegation of dialogue as a research paradigm has led to serious knowledge gaps with regards to gender differences in spatial task performance and communication strategies. To address these issues, this chapter began by reframing past arguments and conclusions as questions of how females and males perform and communicate in a robot navigation task in interaction. Through the second set of questions, the thesis sought to clarify the effect of visual co-presence in spatial tasks and in the interaction with artificial agents, with the primary aim to determine whether the magnitude and characteristics of this effect varies with gender. Next, it was argued that alignment, an interactive

mechanism of practical relevance for HCI, is not well-understood, which motivated a series of related questions. Finally, the chapter enumerated research questions that aimed to examine whether the tendency and strength to align to one's interlocutor is mediated by gender. The research questions presented in this chapter guide the practical work of the study and are summarised in Table 3.1 below. A diagram categorising the questions in terms of whether they address the performance, dialogue, or both, components of the central research question is also presented in Figure 3.2. The next chapter will discuss the development of a suitable methodology to provide 'answers' to these research questions.

Table 3.1: The research questions of the study. The left-hand side column refers to the research question number.

	Research Questions
	A. Gender differences in navigation performance, route communication and user perceptions in interaction
1(a)	*Are all-male pairs (male users interacting with male 'robots) the fastest in completing the navigation task?*
1(b)	*Are pairs with male 'robots' faster in completing the navigation task than pairs with female 'robots'?*
1(c)	*Do all-male pairs produce the lowest miscommunication (execution errors, non-understandings and inaccurate instructions)?*
1(d)	*Do male 'robots' produce fewer execution errors and non-understandings than female 'robots'?*
1(e)	*Do male users provide fewer inaccurate instructions than female users?*
2(a)	*Do female users include more landmark references in their utterances than male users?*
2(b)	*Do female 'robots' include more landmark references in their utterances than male 'robots'?*
3	*Do female users rate their performance lower than male users?*
	B(i). The effect of visual information on navigation performance and communication between user and system
4(a)	*Are tasks completed more quickly when visual information is available?*
4(b)	*Does the number of incorrect instructions by the user decrease when visual information is available?*

4(c)	*Does the number of execution errors and non-understandings by the 'robot' decrease when visual information is available?*
4(d)	*Do interlocutors use fewer words, turns and route instructions to complete a task when visual information is available?*
5(a)	*Does the use of deictic pronouns and expressions (for example, 'turn there' and 'take this turn') increase when visual information is available?*
5(b)	*Does the number of verbal acknowledgements by the 'robot' decrease when visual information is available?*
5(c)	*Are route instructions less detailed, precise and explicit when visual information is available?*
5(d)	*Are spatial descriptions by the 'robot' less detailed, precise and explicit when visual information is available?*
5(e)	*Does the number of user-initiated queries decrease when visual information is available?*
5(f)	*Does the number of user turns exceed 'robot' turns, when visual information is available?*
5(g)	*Do interlocutors use fewer landmark references to complete a task when visual information is available?*
	B(ii). Gender and the effect of visual information
6(a)	*Is task performance of females more negatively affected by absence of visual information than males' performance?*
6(b)	*Do females adapt their communication strategies more than males in response to lack of visual information?*
	C(i). Alignment in human-computer dialogue
7(a)	*Does alignment occur in the interaction between a human user and a computer system?*

7(b)	*If alignment does occur in this context, is it a mutual phenomenon?*
7(c)	*Does visual information influence alignment between a user and a system?*
7(d)	*Does miscommunication locally disrupt the process of alignment in human-computer communication?*
7(e)	*Does lack of alignment compromise user perception of interaction success?*
	C(ii). Gender-related alignment in task-oriented interaction
8(a)	*Do female speakers align more strongly than male speakers in task-oriented interaction?*
8(b)	*Do female speakers align more strongly to male addressees than to female addressees in task-oriented interaction?*
8(c)	*Do speakers in same-gender pairs align more strongly than mixed-gender pairs in task-oriented interaction?*
8(d)	*Do users in mixed-gender pairs moderate the use of their own gender-preferential strategies and provide instructions as preferred by their addressees (landmark-based and purely directional route instructions to females and males, respectively)?*
8(e)	*Does miscommunication have different effect on male and female users in terms of communication strategies?*

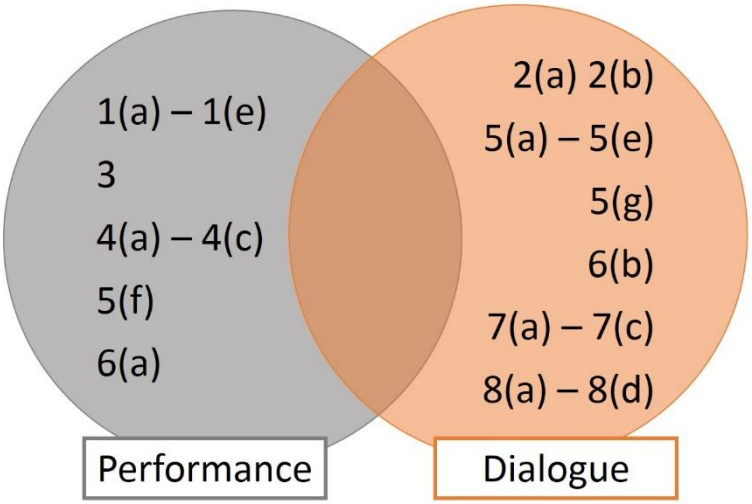


Figure 3.2: Diagram showing the number of the research questions that address performance, dialogue elements, or both.

4 Methodology

4.1 Introduction

Chapter 2 discussed the literature on gender differences across the domains of navigation, communication and human-computer interaction. The literature analysis also described the rich and diverse interaction phenomena observed in dialogue. It was revealed that existing knowledge in gender differences is largely obtained from non-interactive or artificial studies. As such, it was argued that such studies may provide incomplete accounts of gender differences and their findings may have limited value for practical HCI applications. This brought forward the central research question of how gender differences arise in navigation performance and dialogue with a computer system. In Chapter 3, drawing on existing literature in gender differences and studies within the 'language-as-action' tradition (the Collaborative Model and the Interactive Alignment Model), the analysis of the central research question identified relevant knowledge gaps leading to the formulation of specific sub-questions. This chapter describes the development of, and rationale behind, the methodology of the thesis to address these research questions. The thesis is fundamentally data-driven and is motivated by experimental paradigms that investigate spontaneous task-oriented dialogue. This chapter details the experimental technique and setup adopted in this thesis, which includes the re-enactment of an HCI navigation scenario, and discusses the data analysis approach. The data analysis approach combines objective measures of performance with a corpus-linguistics methodology that integrates existing classification frameworks to analyse dialogue moves and their components, instances of miscommunication and linguistic alignment.

The chapter is structured as follows: section 4.2 identifies the experimental approach appropriate for the purposes of this study and describes the task domain, the system developed for this study, the sample and procedure. Section 4.3 briefly discusses previous studies that followed comparable methodologies and the merits and limitations of the experimental setup. Section 4.4 details the analysis approach of the performance and dialogue data. Section 4.5 presents some conclusions.

4.2 The experimental method

Human-human interaction differs from human-computer interaction (Amalberti et al., 1993; Fraser and Gilbert, 1991) and human-robot interaction (Tenbrink et al., 2010). Therefore, data and ideas to inform the design of computer-based dialogue systems should be derived from interactions with such systems, rather than from studies of human-human interaction. This, of course, requires that a dialogue system already exists or that one is simulated. A commonly-employed approach that uses a simulated system is the Wizard of Oz (WOz) method (Fraser and Gilbert, 1991) where two people interact, one of whom is made to believe that he/she is interacting with a system rather than a person. The 'wizard' in a WOz experiment is the experimenter or a single, trained confederate. However, this approach will inevitably offer one (expert and possibly biased) interpretation of the instructions, inhibiting effects of interaction and individual differences in language interpretation and strategy. To address this, in the WOz experiment employed in this study, the wizards were also naive participants who were given no dialogue script or guidelines on what to say.

The study was designed to elicit spontaneously generated route instructions as they emerge in real-time dialogue within a controlled spatial network. The experimental technique involved dyads of participants (instructors and followers) collaborating in an urban navigation scenario, with the instructors being under the impression that they converse with a software agent (a robot follower). A system was developed to enable synchronous text communication and execution of route instructions between the paired participants. To implement the experimental conditions aiming to assess the effect of presence/absence of visual feedback on performance and communication patterns of males and females (relevant

to Research Questions 4(a) – 6(b) and Research Question 7(c)), the system could enable or restrict visual access to the actions of the robot.

Given the focus of the research on inter-individual processes involved in the production and interpretation of route instructions, both instructors and 'robots'/followers were subjects in the study. To allow gender differences in route-giving and -following tasks as they emerge from interaction to be explored, pairs were formed with all possible combinations of roles and gender:

1. Female user/instructor – Female 'robot'/follower (henceforth referred to as F_uF_r)
2. Female user/instructor – Male 'robot'/follower (henceforth referred to as F_uM_r)
3. Male user/instructor – Female 'robot'/follower (henceforth referred to as M_uF_r)
4. Male user/instructor – Male 'robot'/follower (henceforth referred to as M_uM_r)

4.2.1 The task domain

The domain used in the experiment was pedestrian navigation in a simulated town. The user had to guide the robot to six designated locations in the town. The destination location was shown in red and the tasks that had been completed were shown in blue. The environment consisted of highly salient landmarks such as buildings and landmarks of lower salience such as pathways, which aimed to approximate a realistic urban environment. At the same time, environments containing a fair number of landmarks have been shown to be appropriate for users of both genders (as discussed in Chapter 2, section 2.7.5). The cooperative nature of the task lies in two additional characteristics. First, in each pairing, only the user/instructor (hereafter the user) knew the destinations and had a global view of the environment, so the 'robot'/follower (hereafter the 'robot') had to rely on the user's instructions and location descriptions. Second, the user needed the 'robot's' descriptions to determine its exact position and perspective. Participants were able to freely interact and develop their own strategies to carry out the experimental and discourse task. As opposed to experimental setups with have involved real-world urban navigation (see Chapter 2, section 2.5.1), in this study each participant had two overt sources of information: what was on his/her map; and

what their partner said. Thus, the participants were given the opportunity to interact with each other in a relatively natural manner, while the information available to them was finite and controlled at any point in the dialogue. Data was captured on each participant's actions and utterances, to support analysis and understanding of how the participants approached the task and any problems that arose. All actions and utterances also had time and position data associated with them. Moreover, as discussed in the literature review (section 2.3.3.2), a factor underlying gender differences in navigation tasks is argued to be efficiency of visuospatial working memory. Thus, unlike navigation and direction-giving in real-world urban environments, the tasks of this study did not require learning the route through navigation or recalling the map or instructions from memory, which give rise to different cognitive demands and errors.

4.2.2 The system

The experiment relied on a custom-built system which supported the interactive simulation and enabled real-time direct text communication between the user and 'robot' in a pair. The system connected two interfaces over a Local Area Network using the TCP/IP as the communication protocol, kept a log of the dialogues and also recorded the coordinates of the current position of the robot at the moment messages were transmitted. Thus, it was possible to analyse the descriptions against a matching record of the robot's position and reproduce its path with temporal and spatial accuracy. The interfaces consisted of a graphical display and an instant messaging facility (the dialogue box). The dialogue box displayed each participant's messages (in green) in the upper part of the dialogue box; the messages sent by the other participant in the pair were displayed (in magenta) in the lower part of the dialogue box. The desktop PCs used by the participants were equipped with 17-inch LCD monitors with 1024 × 768 pixel resolution.

The interface seen by the user displayed the full map of the simulated town. In order to explore the effect of the provision of visual information ((relevant to Research Questions 4(a) – 6(b) and Research Question 7(c)), there were two variants of the user's screen. In the first, called the 'Monitor condition', a small 'monitor' was displayed in the upper right corner of the screen showing the 'robot's' immediate locality, but not the robot itself (see Figure 4.1).

This meant that the user shared the same visual space as the ‘robot’. This experimental decision follows the relevant literature investigating the effects of visual information (for example, Whittaker, 2000b; Kraut et al., 2003), in which the instructor can see what the follower is seeing and doing, but not the follower himself/herself. This is traced back to the ‘What You See Is What I see’ paradigm in the design of groupware and CSCW systems, developed by Xerox PARC (Stefik et al., 1987). The size of the ‘monitor window’ on the user’s computer screen was approximately 7.2 × 7.2 cm. It was considered appropriate, given that it displayed a scaled-down, high-fidelity image of a relatively uncluttered environment, which was also part of the user’s own map. Users stated that the visual information provided by the ‘monitor’ was sufficient, when probed during the short post-task interviews. No delays were noted in the display of the messages or the visual feedback.

Displaying the map and the robot’s visual space on one screen was considered more usable and less distracting for users than requiring to view two different media (for instance, paper and computer monitor or two separate monitors). This, however, resulted in a compromise in the size of the monitor window showing the ‘robot’s’ visual space. At the same time, similar interfaces have been used by the related studies of Kraut, Gergle and Fussell discussed in section 2.9. For example, in Kraut et al. (2003), the helper’s display consisted of the repair manual and schematics of the bicycle, and a small rectangle window on the right bottom corner showing the view from the head-mounted camera of the worker. In the ‘No Monitor condition’, this feature was disabled so that the user had no direct visual information relating to the ‘robot’s’ position and actions in the environment (see Figure 4.2).

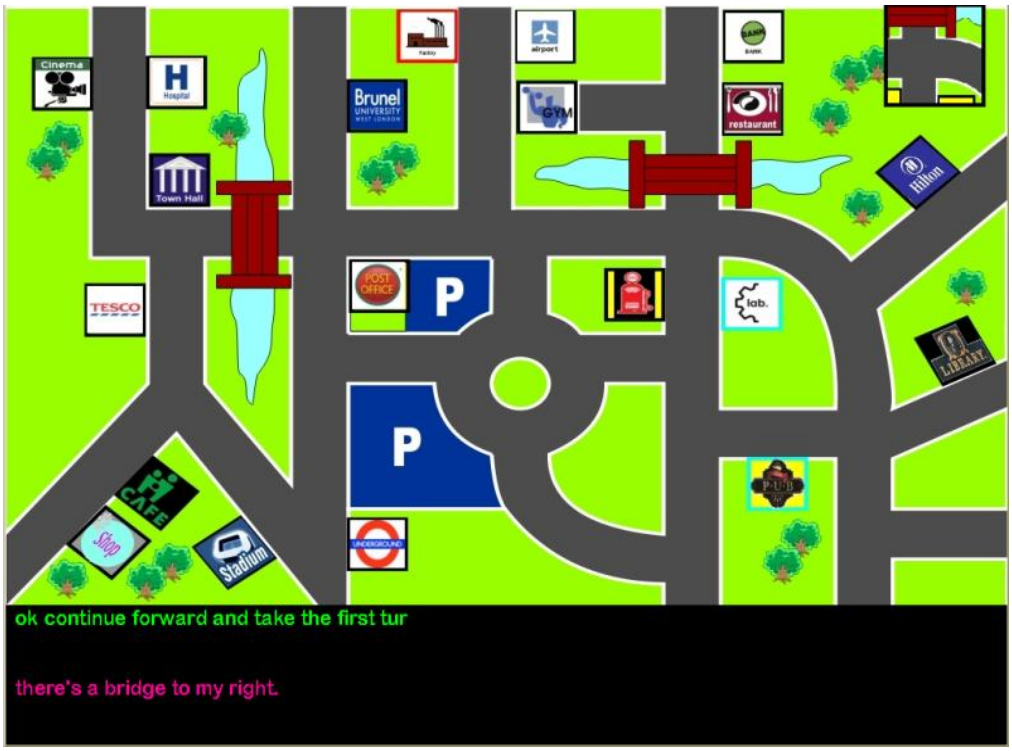


Figure 4.1: The interface of the user/instructor as presented in the Monitor condition. The monitor window can be seen in the upper right corner.

Figure 4.2: The interface of the user/instructor as presented in the No Monitor condition.

The 'robot's' interface displayed a fraction of the overall environment map, showing only the surroundings of the robot's current position (see Figure 4.3). The 'robot' (signified by a red circle with a yellow 'face') was operated by the follower using the arrow keys on the keyboard. The dialogue box also displayed a history of the user's previous messages to the 'robot'. To simulate the ability of the 'robot' to learn routes, after each task was completed a button for the completed route appeared on the 'robot's' screen. If the 'robot' was then instructed to go to a previously visited destination, the follower could press the corresponding button and the 'robot' would automatically execute the move. In the example provided in Figure 4.3, the 'robot' has 'learnt' two routes: (i) from the 'start' to the 'pub'; and (i) from the 'pub' to the 'lab'.

In addition to maintaining the illusion of an automatic interlocutor, the buttons served another purpose; there is empirical evidence that interlocutors re-use (contracted forms of) the same referring expressions when describing the same objects (see section 2.8). As such, given that each new route was incrementally more complex and could, thus, contain a previously-described route, participants would most likely include the same content and dialogue moves. Such tendency would have resulted in a fallacious increase in aligned responses over time. On the other hand, by using the buttons/learnt routes, participants could avoid having to describe the previous route and repeat language.

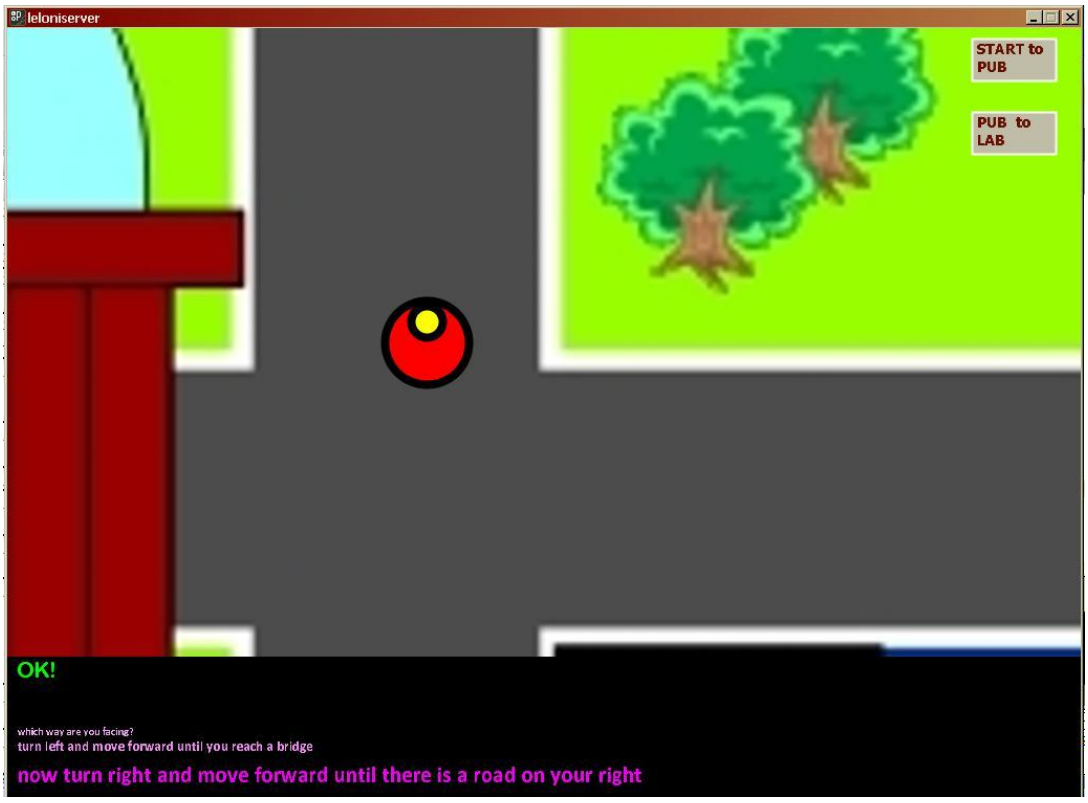


Figure 4.3: The interface for the 'robot'/follower.

4.2.3 Participants

A total of 64 participants (32 males and 32 females) were recruited from undergraduate and postgraduate students of various departments at a UK university. The participants were randomly allocated to the two roles (user or 'robot') and to each of the experimental conditions (Monitor or No Monitor). Each participant was paid £10 for participating in the experiment. Previous experience in using computers was necessary, and familiarity with instant messaging applications. No other specific computer expertise or other skill was required to take part in the experiment. Participants were native or near-native speakers of English.

4.2.4 Procedure

Users and 'robots' were seated in separate rooms equipped with desktop PCs, on which the respective interfaces were displayed. 'Robots' were scheduled to arrive 20 minutes before users, since additional time was needed to explain the Wizard-of-Oz setup and familiarise with the interface features that were not present in the user's interface (moving the robot and buttons). It was ensured that 'robots' and users never met before or after the experiment. Participants received verbal and written instructions related to the task from their role perspective (the documents with the written instructions for the 'robots' and users can be found in Appendix I and II, respectively). They were told that the experiment aimed to explore how people interact with robots. It was also made clear that it did not aim to measure their abilities to follow or give instructions. They were informed that their interaction will be recorded anonymously for subsequent analysis. Participants were also advised that they can request the dialogue logs to be deleted and are free to leave at any time and still receive full payment. Written consent from the participants was obtained (the form of consent can be found in Appendix IV).

The participants that were assigned to be 'robots' were fully informed about the experimental setup and that they were to pretend to be robots. However, they were not aware of the experimental conditions, that is to say, whether their actions could be monitored by the user. No examples or instructions were provided on how to communicate or complete the task. The 'robots' were also given a brief demonstration of, and time to familiarise themselves, with the operation of the interface. The training of the 'robots' in terms of communication style followed the guidelines set in Amalberti et al. (1993): natural language should be used, there were no constraints in comprehension and production and no dialogue script, but 'robots' could only produce task-related utterances, and the use of slang words was not permitted (abbreviations and misspellings were automatically corrected).

The users were told that they would interact directly with a robot, which for practical reasons was a computer-based, simulated version of the actual robot. The users were given minimal information about the 'robot'. They were informed that the 'robot' had advanced capacity to understand and produce spatial language and learn previous routes. This aimed to reduce the likelihood of users inferring during the interaction that the 'robot' was actually a

person. Users were asked to open each interaction with 'hello' (which actually initialised the application used by the 'robot') and end it with 'goodbye' (which closed both of the applications used by the pair). Users were asked not to employ cardinal reference systems (such as 'North', 'South', 'up', 'down'), since use of reference systems was not a focus of the study and it was thought that it may lead to confusion/ambiguity since no reference system was provided within the map. Instead 'forward', 'backward', 'right' and 'left' were to be used as directional statements. The users were told that 'robots' could only see its surrounding area. In addition, users in the Monitor condition were told they would be able to see what the 'robot' sees. The users were given no other examples of, or instructions about, how to interact with the robot. The pairs attempted six tasks presented to each pair in the same order; the user navigated the 'robot' from the starting point (bottom right of the map) to six designated locations (pub, lab, factory, tube, Tesco, shop). The users were free to plan and modify the route as they wished. The destinations were selected to require either incrementally more instructions or the use of previously taught routes. Dialogues ran until the task was completed or the user chose to end them. After each task (completed or abandoned), the users filled in a short questionnaire, which consisted of seven Likert-scale statements for which the users stated their level of agreement. The questionnaire is discussed in section 4.4.3, and a copy can be found in Appendix III.

At the end of the experiment, the users were debriefed and the full nature of the experimental setup was disclosed and explained. Before this disclosure, questioning was used to determine whether users had become aware that the experiment was a simulation. Though previous research has shown that participants can be misled (Fraser and Gilbert, 1991), giving confidence that the experimental setup would be successful, the experimenters were prepared to discard relevant data if any user expressed that s/he was not convinced by the simulation. However, all users confirmed their belief in the setup and expressed surprise on being told during the debriefing that they had been interacting with a human acting as a 'robot'. This gives confidence that any effects identified in the results are not a result of language adaptation by the users arising from them believing that they were instructing another person. It was also anticipated that a user (or 'robot') might reveal his/her gender, for example, by greeting ('Hello, I am Bob'). In such cases, the logs would be deleted. Inspection of the data confirmed that no dialogue contained any overt clues of gender. It should also be

noted that there was no evidence that users assigned a ‘gender’ to ‘robots’. Interestingly, however, there was an instance in which a male user referred to his interlocutor as ‘Mr Robot’, but it did not appear to be a general phenomenon, and may have to do with images of robots in popular media which almost exclusively have ‘male’ appearance or names.

The procedure of the experiment is diagrammatically shown in Figure 4.4 below.

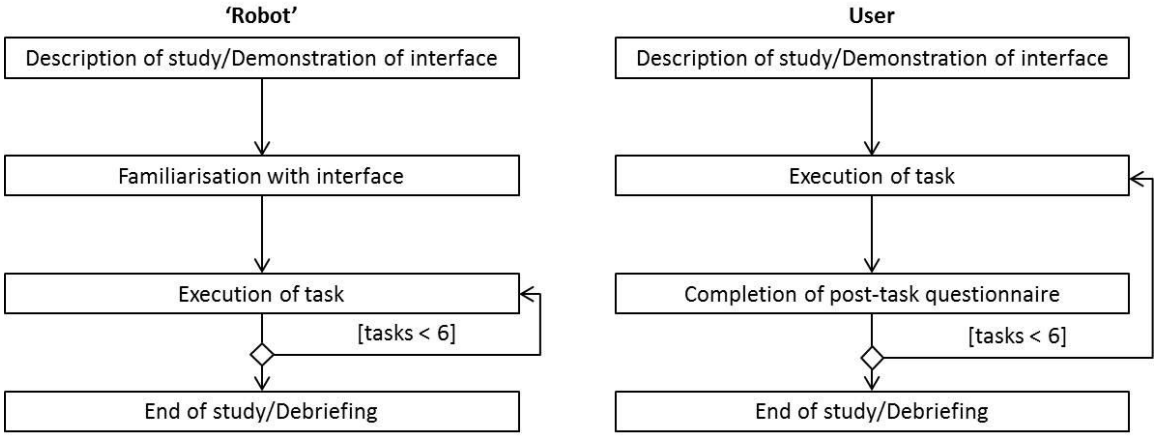


Figure 4.4: Sequence of main activities of the experiment for the participants assigned as ‘robots’ and users.

4.2.5 Pilot study

A pilot study was performed in which three pairs of expert participants (research students in HCI) and three pairs of naive participants (students from various departments of the university) completed the tasks. In addition to dialogue and robot action system logs, data from post-task interviews and walkthroughs were collected and analysed. The pilot study confirmed the viability and credibility of the setup, but it also revealed features that helped improve aspects of the interface and experimental procedure; for instance, it was observed that the most effective and efficient way for ‘robots’ to familiarise themselves with the operation of the interface was a short demonstration video which did not provide dialogue examples. ‘Out-of-sync’ or overlapping messages were very frequent during the pilot study interactions, so ‘generating message’ was displayed to the users while their partners were typing. In the first round of the pilot study (with the expert participants) the town map did not include recognisable landmarks, and, as a result, the tasks were extremely difficult to

complete – this was rectified in the next version of the interface. Finally, the pilot study optimised the planning and timing of the procedure, and wording of instructions, which were important for maintaining the WOz setup.

4.3 Discussion of the experimental method

This section reviews a number of studies in HCI that followed similar dialogue-based language and task corpus methods. It, then, represents possible merits of the experimental approach. Finally, it discusses possible limitations, which relate to the communication modality, the dimensionality of the navigation environment and the influence of the interface.

4.3.1 Related studies

Language and task corpus collection is an integral part of system development as it defines the linguistic and task requirements for a particular application domain. Corpora oriented towards robot systems include a few but prominent examples. Pioneering work within the IBL project resulted in a robot able to learn and execute verbal route instructions (Lauria et al., 2001). Similar to this thesis, the IBL project used the domain of urban navigation and was based on a corpus collected through a WOz study. Green et al. (2006) also employed the WOz methodology to explore how users interact with a robot to refer to objects in the environment using speech and gesture. Unlike the study reported in this thesis, however, minimal contributions by the 'robots' were allowed in both these studies.

A dialogue-based methodology yielded the SCARE corpus, which was collected as part of the development of multimodal dialogue systems. This corpus contains dialogues produced by two human partners that perform a treasure-hunt task in a virtual reality indoor environment (Stoia et al., 2006). A corpus of human dialogues in an outdoor navigation tasks to explore error handling for dialogue systems were used to inform the design of a dialogue system (Skantze, 2005). In both cases the corpora were collected in overt human-human interaction conditions (unlike the covert WOz setups adopted in the present study and the human-robot interaction studies mentioned above).

Closely related to the work reported in this thesis, the ongoing Diaspace Program in the University of Bremen explores dialogic interaction for spatially-aware robots (see, for example, Hui et al., 2010). In particular, experimental scenarios and manipulations have been designed in order to collect dialogues to navigate a robotic wheelchair within an indoor environment (real system or in a WOz experiment).

4.3.2 Merits of experimental setup

Placing a 'robot' (rather than making explicit that it was another person) at the other end of the communication channel holds three advantages. First, the obvious merit of this approach is that the results are relevant and directly transferable to the design of future interactive and collaborative systems. Second, in human-human interaction, interlocutors rely on assumptions of shared knowledge and general linguistic conventions (Grice, 1975). But the ways in which these elements shape the interaction are neither relevant nor transparent to a system designer (Dahlback et al., 1993) and are likely to be confounding in terms of the aims of the study. On the other hand, when talking to a non-human agent, users are expected to avoid using this knowledge and to depend on assumptions and conventions set up within the course of the particular dialogue only, allowing clearer insights into their patterns of interaction. The third advantage of this approach is traced to the fact that using natural language is primarily a social activity (Clark, 1996; Nass and Moon, 2000), such that relationship behaviours arise as an artefact of using language. In a comparative study, Schechtman and Horowitz (2003) point to two 'drawbacks' in using human-human interaction experimental setups and applying their findings in social technology design. People put more time and effort into interactions with people than computers. Such behaviour is a hindrance to task goals, when time or efficiency is important. Second, emotions like fear, shame, anxiety and embarrassment are very likely to occur in interactions between people (Schechtman and Horowitz, 2003). As detailed in Chapter 2 (final part of section 2.3.3), intense 'spatial anxiety' and 'fear to get lost' (Lawton, 1994; 1996) reported by females adversely affect performance and navigation strategies (Schmitz, 1997) and have been related to the 'stereotype threat' experienced by women. Thus, in a study in which the person interacts with a non-human agent, such emotional and psychological states that interfere with performance and communication strategies can be alleviated. Masking the gender of the

interlocutor may also help avoid other gender stereotype issues, such as men being less likely to listen to instructions from female voices (see, for example, Jonsson et al., 2008).

4.3.3 Discussion of limitations

Valid questions emerge regarding whether the results from text-based interaction in a 2D environment are extensible to spoken dialogue taking place in 3D or real-world environments, and whether navigation differences can be attributed to different prior proficiency in using computer interfaces. Issues with regards to the credibility and generalisability of the WOz approach are also discussed.

Communication modalities

A system simulation that involves speech is certainly harder to design and perform than text-based experiments. A concern, however, is whether the validity and extensibility of the results from the experiment are limited due to the differences between text and spoken utterances as modalities. Experiments comparing the modalities outline several differences. Hauptmann and Rudnicky (1988) compared three modes: typing to a computer, speaking to a computer directly, and speaking through a human intermediary. Their findings are consistent with the well-known studies by Chapanis et al. (1972) exploring human-to-human typed or oral interaction. Their data reveal that the spoken utterances were lengthier containing a higher number of words and task-unrelated words and 'noisier' (ungrammatical sentences containing 'fillers' like 'uhms'). No differences were found with regard to the frequency of questions, commands and statements. On the other hand, the differences were exacerbated in the speech conditions, when participants interacted with either computers or people. Therefore, it is argued that these differences are not likely to interfere with the basic and particular interactive mechanisms which are explored in this thesis. Furthermore, Moratz and Tenbrink (2003) conducted a study in which users navigated a robot using either typed or spoken instructions and reported that similar instructions were employed in both modalities. Finally, the study (described in Chapter 2, section 2.8.3) that explored lexical alignment in HCI reported similar patterns of results for speech- and text-based interaction with the system (Brennan, 1996). Even if this were not the case, the study would be useful given the

immediate practical relevance that any findings would offer for text-based interaction with a computer system or computer-mediated communication between people.

Dimensionality of environment

Another potential criticism of the study is that the interface displayed a plan view of the environment, whereas in a real-world situation the instructor and follower face three-dimensional objects. However, Tenbrink (2007) compared the spatial descriptions used in a computer-, picture-based 2D scenario and a real-world human-robot interaction study and found no differences with respect to conceptual and linguistic strategies. The interpretation provided was that three-dimensional concepts do not influence linguistic representations given that the objects that form spatial descriptions exist on the horizontal plane. This provides confidence in the use of a plan view in the study.

Landmarks in the environment

People rely on landmarks to organise information, orient themselves, navigate and locate objects. It is necessary to define the term landmark, and explain how it is used in this study. Presson and Montello (1988) minimally define a landmark as an element or feature in space that can serve as point of reference. Along the same lines, Sorrows and Hirtle (1999) state that a landmark is a visually or cognitively salient object that people can remember to help them orient themselves and locate other objects. According to Pick et al. (1988) and Benyon and Hook (1997), landmarks do not necessarily have to be static; they can be moving or change over time.

It is acknowledged that the 2D icons used as landmarks in the simulated environment of the experiment are substantially different to the versatile 3D objects that exist in abundance in the real world. However, based on the adopted definition, their function (visually salient objects to serve as point of reference) is essentially the same for their respective environments. At the same time, it is argued that this issue relates to the broader debate of whether navigation in simulated environments reflects navigation behaviour in the real world, as discussed in section 2.3.2; while there is empirical evidence suggesting that people rely on

similar cognitive processes in both environments, caution should be exercised when transferring conclusions from one domain to the other.

The effect of interface operation

Previous research in virtual environment navigation has identified a mediating effect of interface proficiency (Waller, 2000). Navigation in a virtual environment often involves operating a mouse, joystick or multimodal, haptic interface devices, which adds to attentional and cognitive demands of the task. The mediating effect of interface controls becomes even more critical in studies exploring gender differences. In particular, males hold an advantage over females as they are more likely to have prior experience in operating such interface devices, through playing computer games (Castelli et al., 2008). Consequently, relevant studies usually involve extensive interface training to reduce the effect of the interface controls. However, it remains unspecified how much and what type of training any individual should receive for this effect to be eliminated. Taking these issues into consideration, the interface of this study was designed to offer intuitive navigation that required minimal skills and no previous experience. In particular, moving the robot (a red circle) only involved pressing the arrow keys on the keyboard. Moreover, the experimental procedure allowed time for familiarisation with the operation of the system, and was confirmed that no participant had difficulty in learning to use the interface.

Generalisability and credibility of WOz

Valid questions may emerge with regards to the WOz variant deployed in the study; that is, whether the lack of trained confederate(s) and dialogue script limit the generalisability of the results to HCI. The study was exploratory in nature and its primary aim was to measure lexical alignment, so it was felt that the use of training materials, predefined messages or action sequences would have influenced the content of the responses of the ‘system’ and, by extension, of the user. The ‘typical’ WOz approach is also based on the assumption that all user behaviours can be predefined and this might not accommodate unanticipated interaction patterns. Instead, the study aimed to constrain the content of ‘robot’ responses by using a specific domain of interaction, and by designing an environment with a limited set of spatial

relationships and landmarks and predefined destinations about which the participants could converse. Nevertheless, it would be interesting to replicate this study using a 'typical' WOz setup, in which 'robots' are trained to use either the same or a different lexical item at given points in the developing dialogue.

A concern could also be raised with regards to the credibility of the setup. As described above, 'robots' were instructed to use natural speech, but only interact about the task and not to use slang. At the same time, users were told that the robot was proficient in understanding and producing spatial language. The post-task interviews with the users confirmed that users believed that they had been interacting with a robot throughout the session. During the interviews, no user expressed that they were surprised with the (linguistic and functional) capabilities of the robot. This may be due to the fact that users have no experience of interacting with real robotic systems, which may lead to inflated or no *a priori* assumptions about what a robot can do. There is an interesting body of research focusing on users' perceptions of systems' capabilities. The study by Amalberti, et al. (1993) presented an experiment in which two groups of users interacted with the same human experimenter; one group was told that they would talk to a human, and the other group that they would interact with a dialogue system. The human experimenter followed the same guidelines as the 'robots' in the study reported in this paper. The results showed that users approached the roles in the interaction differently, and tended to rely less on the problem-solving capacity of the 'computer' compared to the human interlocutor. Interestingly, any linguistic differences tended to disappear as subjects gained familiarity with the system. Along the same lines, research by Levin and colleagues (Levin et al., 2013; Levin et al., 2008) demonstrates that people are willing to attribute human-like cognitive characteristics such as intentionality to robots more than they do with computers, but only when users are given time to observe intentional behaviour by the robot. However, robots (even future ones) cannot be perceived as fully intentional.

4.3.4 Social responses and the anthropomorphism explanation

Another interesting finding emerged through inspection of the corpus; while users almost exclusively produced task-related utterances, some users would commend or thank the robot

for successfully completing a route (for example, 'well done', 'thank you'). While this appeared to be a subject-specific behaviour, it resonates with a recurring finding in HCI literature that 'individuals mindlessly apply social rules and expectations to computers' (Nass and Moon, 2000, p. 82). In a series of experiments by Nass and colleagues, people were found to engage in *overlearned social behaviours* such as *politeness* when interacting with artifacts, and, even *overapply human social categories*, such as *gender*. In fact, a user in the present study once referred to the 'robot' as Mr. Robot. In addition, a large number of users in this study would end each successfully completed interaction, by thanking or praising the robot. In this section, the issues of politeness, gender and anthropomorphism are discussed.

In a study by Nass et al. (1999), users interacted with a system using text-based communication. After the completion of the task, the system directly asked the users about its performance. Two different groups of users also answered the same questions by either another computer or in a typical post-session questionnaire. The users who were directly asked by the same system with which they had interacted provided extremely positive evaluations. The authors conclude that people are polite to computers and liken this behaviour to face-to-face human interactions, in which we are less prone to give negative comments (and prefer being dishonest) in order to avoid hurting the other person's feelings. This tendency towards computers is attributed to *overlearning*; people maintain some social rules so deeply established that they tend to apply them to all situations.

Gender is one of the most prominent social categories (Bem, 1981), and, as such, gender was also investigated by Nass and colleagues (Nass et al., 1997). In their study, participants had to provide assessments for systems with either female or male voices. Their findings suggested that stereotypes associated with human females and males were extended to the systems.

The question that naturally arises is whether these tendencies can be attributed to anthropomorphism, that is, people apply social rules to artifacts because they perceive them as humans. Anthropomorphism is defined as 'a *sincere, conscious* belief that computers are human and/or deserving of human attributions' (Kim and Sundar, 2012, p.1). As discussed above, there is extensive evidence that people provide social responses to computers, which may suggest that people anthropomorphised computers. Research by Nass and colleagues

(for example, see Nass and Moon, 2000) provide evidence against the anthropomorphism explanation. They argue that during debriefing of the hundreds of the participants in their studies, all of them appeared to believe that computers do not warrant human-appropriate treatment since they are not human. Reviewing related research, they conclude that people do not anthropomorphise computers (that is, they do not sincerely and consciously believe they have human characteristics), but, they are capable of developing emotional or social responses towards objects (such as talking to or giving a name to one's car or swearing to a printer). They call this behaviour 'ethopoia', directly responding to an entity as human, while knowing it is not. Their conclusion is that mindlessness qualifies as a better explanation compared to anthropomorphism; individuals mindlessly apply social rules and expectations to computers. This is because humans will not (put the effort to) consider the all cues and differences to create new categories for computers, so they would oversimplifyingly apply the social rules they already use in human-human interactions.

4.4 Data analysis approach

The thesis employed a fundamentally quantitative analysis framework. The first part of this section reports and discusses the measures that were used to calculate performance (interaction efficiency and effectiveness) and user perceptions. This analysis concerns the performance elements of the central research question, and related sub-questions (see Figure 3.1 in Chapter 3). The analysis of performance data is presented in subsections 4.4.1 – 4.4.3.

To address the sub-questions that probe the dialogue elements, a corpus-linguistic methodology was adopted (Biber et al., 1998, p.4). In particular, it follows existing annotation and analysis schemes to evaluate the spontaneous utterances of the participants and their components and to identify the degrees of alignment. Then, it measures the associated frequencies, which are used in the statistical analysis (detailed in the next section). This analysis is described in subsections 4.4.4 – 4.4.6. Following the corpus-linguistics methodology, the approach in this thesis also integrates qualitative components (McEnery and Hardie, 2011). As such, while the outputs are generated based purely on statistical processing of the corpus data, qualitative statements are formulated in light of these

quantitative results. Moreover, where appropriate, the quantitative findings will be reinforced by dialogue examples drawn from the corpus.

4.4.1 Analysis of performance

The analysis of actual performance targeted Research Questions 1(a) – 1(e), 4(a) – 4(d), 5(f), 6(a), 7(d), 7(e) and 8(e) (see Figure 3.2 in Chapter 3). It used objective measures that have been commonly equated with interaction effectiveness and efficiency in research in human communication (e.g., Clark and Krych, 2004), navigation (e.g., Ishikawa and Kiyomoto, 2008), computer-mediated communication (e.g., Gergle et al, 2004) and spoken dialogue systems (e.g., Walker et al., 2000a; Litman and Pan, 2002). These will be broadly referred to as performance-based measures, and are listed in Table 4.1 below. It should be noted that the 'task' served as the basis of measurement (each pair normally completed six tasks).

Table 4.1: List of performance-based measures.

Performance-based Measures
Time
User turns
'Robot' turns
User words
'Robot' words
User word-ratio
User turn-ratio
Route instructions
Incorrect instructions
Execution errors
Non-understandings

Time was the time in seconds elapsed from the beginning until the completion of a task as automatically logged and computed by the system. The number of 'robot'/user turns and 'robot'/user words were also part of the system logs. Word and turn ratios were calculated from these data. The number of instructions was derived from the dialogue logs and manual coding, which will be explained in section 4.4.4. Dialogue efficiency was also assessed through the metric of miscommunication, which comprised of number of incorrect instructions, execution errors and non-understandings, and was obtained through labelling

based on system logs. The analysis of miscommunication is discussed in detail in the following section.

The objective measures of performance outlined above were complemented by subjective user ratings of interaction efficiency and success (relevant to Research Questions 3 and 7(e)), gathered through a post-task questionnaire described in section 4.4.3.

4.4.2 Analysis and annotation of miscommunication

The logged interactions were annotated in order to detect and classify interaction problems. In order to obtain 'hard' data, the analysis is formalised and described below.

As detailed in Chapter 2 (section 2.10), in dialogue studies, errors and other problems are referred to as miscommunication. Miscommunication encompasses two forms of problems, misunderstandings and non-understandings (Hirst et al., 1994). Misunderstandings are only noticed, when the addressee acts upon them. Thus, this analysis measures execution errors, which refer to deviations from the described route. On the other hand, non-understandings are immediately recognised, as the hearers are aware of them and articulate them (for example, in the form of clarification requests). These two forms of miscommunication are normally attributed to the recipient, who, in this scenario, is the 'robot'. The source of execution errors was not only incorrect interpretations of an instruction, but they also occurred as a result of inaccurate instructions. Therefore, the analysis of miscommunication extends to consider 'user errors'. The identification and annotation of execution errors, non-understandings and incorrect instructions are described in detail in the next subsections.

Execution errors

As mentioned above, misunderstandings corresponded to execution errors, which refer to instances in which the 'robot' failed to understand the instruction and deviated from the described route. The coordinates (x, y) of the 'robot's' position were recorded for each exchanged message and placed on the map of the town (which was defined as 1024 by 600 pixels), allowing the movements of the robot to be retraced when undertaking analysis of the

dialogues. Execution errors were determined by matching the coordinates corresponding to each of the user's utterances with those returned as a result of their execution by the 'robot'. An excerpt of a dialogue containing an execution error is shown in Table 4.2. Figure 4.5 illustrates the route which the user described and the 'robot' followed during the interaction presented in Table 4.2. The 'robot' accurately executed the instructions in utterances 5, 6 and 7. However, the 'robot' misunderstood the next instruction (utterance number 8) and ended up in an unintended location.

Table 4.2: An excerpt of a dialogue containing an execution error [NMF5_TE54-62][9]. The columns denote (from left to right): the speaker (User or 'Robot'), the utterance number, the utterance, and the 'robot' coordinates and time that the utterance was sent.

Speaker	Utterance Number	Utterance	Coordinates and Time Stamp
U	1	Hello	1000,530 @13:37:32
R	2	Hello	1000,530 @13:37:36
U	3	We are going to Tesco	1000,530 @13:37:42
R	4	Ok. Directions please.	1000,530 @13:38:5
U	5	Go straight ahead and turn right at the junction	1000,530 @13:38:20
U	6	Then go straight and follow the road round the bend to the left	909,464 @13:38:47
U	7	You will pass a bridge on your right, continue going straight	902,358 @13:39:12
U	8	Then cross the bridge and turn left	675,259 @13:39:35
U	9	Tesco will be on the right hand side and that is the destination	561,117 @13:40:8

[9] The codes in the square brackets contain the dialogue ID in the corpus. The letters before the underscore character stand for the condition (Monitor or No Monitor) and the arbitrary pair ID. The letter and numbers after the underscore denote the task (Pub, Lab, Factory, Tube, Tesco and Shop) and turn number.

Figure 4.5: The 'robot's' execution of the instructions given in the dialogue presented in Table 4.2: the solid blue line illustrates the accurately executed route; the blue dashed line represents the route that the instructor described but the 'robot' failed to execute; the red line shows the deviation from the intended route; the numbers in brackets along the executed route indicate the utterances communicated at that point.

Non-understandings

The second type of miscommunication considered in the analysis includes the utterances by the 'robot' that expressed non-understanding. These responses could be formed explicitly as in statements like 'I don't understand' or clarification requests (Gabsdil, 2003). The annotation of non-understandings follows the definition provided by Hirst et al. (1994) and Gabsdil (2003). Non-understandings occur when: (1) the 'robot' forms no interpretation of the user's utterance; (2) the 'robot' is uncertain about the interpretation he/she obtained; or (3) the utterance is ambiguous to the 'robot', leading to more than one interpretation of the instruction. Table 4.3 contains examples of utterances corresponding to these different sources of non-understanding. However, it should be made clear that the analysis did not consider each source separately.

Clark (1996) and Allwood (1996) independently developed a four-level model of communication (see Chapter 2, section 2.10). According to the model, non-understandings can occur at any of the four levels of communication (securing attention level, utterance recognition level, meaning recognition level and action recognition level); from failing to establish contact with the speaker to questioning the function of the utterance in context. Interlocutors formulate their responses showing in which level the non-understanding occurred. Following the model, the analysis of non-understandings also included cases in which the 'robot' understood the meaning of the instruction but had a problem with its execution. An example of this final type of non-understanding is where the user is telling the 'robot' to move forward, but the instruction cannot be executed given the 'robot's' current location on a t-junction, as in example (4i) in Table 4.3 below. Clarification requests termed task-level reformulations also fall into this category, as in example (4ii). Task-level reformulations rephrase the instruction, independently of its form, in terms of the practical effects of its execution, and are widespread in task-oriented dialogues (Gabsdil, 2003).

Table 4.3: Examples of non-understandings produced by the 'robot' in response to a user instruction.

Examples of Non-understanding	Utterance
1	*U*: Turn left. *U*: There is a pub. The building next to you. *R*: Please instruct which way exactly.
2	*U*: You must turn to your left and go to the end of the junction. Then you turn right. *R*: Turn right when I can see the tree?
3	*U*: Go back to last location. *R*: Back to the bridge or back to the factory?
4i	*U*: Go forward. *R*: There is a fork in the road.
4ii	*U*: Turn left. *R*: Ok; do I go over the bridge?

Incorrect instructions

The analysis considered cases in which users provided incorrect instructions. Oulasvirta et al. (2006) proposed a classification for 'user errors', according to which, 'user errors' can occur

at the goal level (for instance, false assumptions with regards to the system's general capabilities), task level (for instance, the user issues a command which is incorrect in relation to the present state of the dialogue), command level (that is, vocabulary and grammar errors), or modelling level (the user issues a command which clashes with the 'world' of the system). As with the annotation of non-understandings, classifying the source of incorrect instructions was out of the scope of the analysis. However, this classification helped in providing a formal framework for distinguishing the cases of 'user errors'. In the present dialogue corpus, incorrect instructions occurred mainly because of unintended mistakes or misconceptions regarding the position and orientation of the 'robot'.

Figure 4.6 shows a screenshot of an interaction and serves to exemplify an incorrect instruction due to a mistake in the spatial direction (at the command level, following the scheme by Oulasvirta et al. (2006)). The destination of the particular interaction was the Tube. As can be seen from the small window in the top right corner of the user's monitor and the 'robot's' message in the dialogue box ('There is a fork in the road'), the 'robot' is on the y-junction beside the Lab. The next instruction from the user is 'Ok, turn left here and then take the third *right*' which is false, having confused 'left' with 'right'. The 'robot' accurately executes the incorrect instruction and arrives at Brunel University. As such, this miscommunication incident was tagged as 'incorrect instruction' and not 'execution error'.

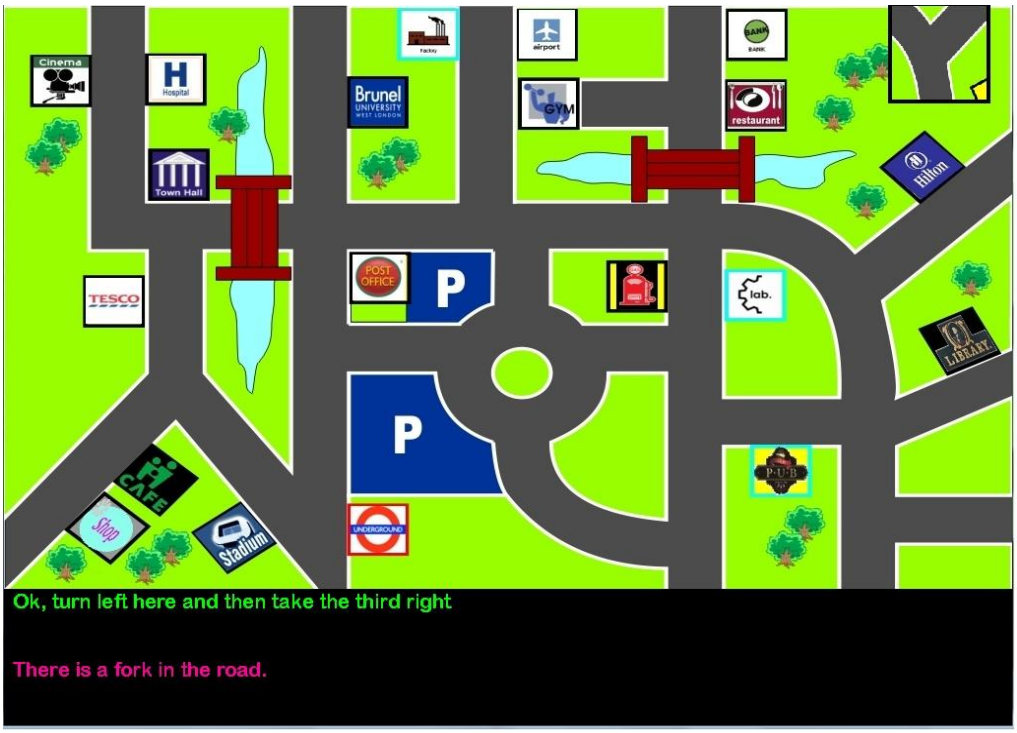


Figure 4.6. Screenshot of the user's interface during an interaction. The destination is the Tube station. The 'robot's' position is displayed in the small window on the top right of the user's interface. The dialogue box shows the user's message in green (on the top of the dialogue box) and the 'robot's' message in magenta (on the bottom of the dialogue box).

4.4.3 User perceptions of the interaction

Objective evaluation of the interaction needs to be complemented by subjective judgements by human users. In the field of spoken dialogue systems, common methods are interviews, focus groups and various forms of questionnaires. Questionnaires that consist of a number of rating scales are particularly useful as they produce 'hard', quantifiable data (Hone and Graham, 2000). These questionnaires are typically distributed to the users directly after the interaction with the system and contain declarative statements or questions with which participants are asked to rate their agreement. The questionnaires can be simple, consisting of a few statements assessing different dimensions of satisfaction. For example, the studies within the well-known PARADISE project (Walker et al., 1997; 1998; Litman and Pan, 2002) used eight to ten questions, either Yes/No or on a five-point Likert scale. Then, the

values of the responses were summed to give a single User Satisfaction measure. A more sophisticated questionnaire was developed by Hone and Graham (2000), called SASSI. It was mainly designed for and validated using speech input systems. It contained 50 seven-point Likert scale statements (for instance, 'the system is accurate') which were mapped across six dimensions; that is, System Response Accuracy, Likeability, Cognitive Demand, Annoyance, Habitability and Speed.

A decision was taken to design a simple questionnaire to collect data on user perceptions, based on the related studies by Williams and Young (2004) and Skantze (2005). After the completion of each of the six tasks, the users were asked to complete a questionnaire in which they rated their agreement with five declarative statements of opinion (shown in Table 4.4). The questionnaire used a Likert scale with seven levels of agreement: *strongly disagree; disagree; slightly disagree; neutral; slightly agree; agree; and strongly agree.* The items probed five different aspects of the user's experience of their interaction with the 'robot': perceived task completion (item 1); execution accuracy (item 2); ease of use (item 3); helpfulness of the system (item 4); and overall satisfaction (item 5). The responses were mapped to integer values between one and seven (with seven representing the highest level of agreement). The scores associated with each statement were summed for all six tasks, which resulted in a cumulative score for each statement ranging from 6 to 42.

Table 4.4: Statements in the questionnaire completed by users after the completion of each of the six tasks.

1. I did well in completing the task
2. The system was easy to use
3. The system was accurate
4. The system was helpful
5. I am generally satisfied with this interaction

4.4.4 The dialogue act annotation scheme

The first step in the analysis of the dialogue data was the classification of each utterance in the corpus. The utterances were annotated using a simplified version of the HCRC Map Task move coding scheme (HCRC Dialogue Structure Coding Manual by Carletta et al., 1996).

The HCRC coding scheme divides the participants' dialogue moves into *initiations* (expecting a response) or *responses.* The dialogue moves were classified according to a set of dialogue act categories; namely, *Instruct, Explain, Clarification Request, Query, Acknowledge, Reject, Reply* and *Clarify*. Definitions and examples of the dialogue acts are provided below. In the following examples, 'U' stands for user and 'R' denotes the 'robot'.

The *Instruct* dialogue act commands the follower to execute one or more actions, as in example 1 below.

Example 1 [MF3_P6]

U: You must turn to your left and go to the end of the junction. Then you turn right.

The *Explain* dialogue act states information that has not been requested by the partner (that is, it is not a reply to a previous query). Accordingly, the 'robot's' utterance in example 2 below is classified as 'Explain'.

Example 2 [NMF_F51-52]

U: Then, go forward and turn right.

R: I can see a car park now.

The *Clarification Request* dialogue act is close to the 'Check' category in the HCRC coding scheme but follows the definition adopted in the specialised clarification request taxonomies by Gabsdil (2003), Schlangen (2004), Rodriguez and Schlangen (2004) and Purver (2004), which are based on the four-level model of communication (see sections 4.4.2 above or 2.10, in Chapter 2). In particular, a clarification request negotiates a previous proposal (typically an instruction), with regards to problems in perception, vocabulary, reference and executing an action. The latter type of clarification request, a 'task reformulation' (Gabsdil, 2003), is illustrated in the 'robot's' turn in example 3 below. As discussed in the context of miscommunication, clarification requests signal lack of full understanding.

Example 3 [NMC3_T58-59]

U: Go right.

R: Ok; do I go over the bridge?

The *Query* dialogue act covers all questions addressed to the partner, which are not clarification requests, as in the user's turn in example 4 below.

Example 4 [NMC3_T56-57]

U: Where are you?

R: I am standing facing the Post Office, with the car park on my left.

The *Acknowledge* dialogue act is a minimal sign of positive feedback. It demonstrates that a previous utterance or action was received, understood or accepted. Acknowledgements can be formulated simply, as 'Ok', 'Yes' or as the 'robot's' response in example 5 below.

Example 5 [NMC7_T35-36]

U: go to lab and walk ahead, when you see two roads take left and then keep walking for a while and take second left

R: second left taken.

The *Reject* dialogue act is the opposite of an 'Acknowledge'. It minimally provides negative feedback, rejecting an utterance or action completely. The user's utterance in example 6 below was tagged as 'Reject'. This dialogue act was incorporated following Muller and Prevot's practice for annotating dialogue acts in route communication dialogues, arguing that rejections cannot be simply considered as replies since they hold a different communicative function in task-oriented dialogues (Muller and Prevot, 2009).

Example 6 [MF1_TE48-49]

R: Have I reached the desired destination?

U: No.

The *Reply* dialogue act is any reply to a query that contains only the information requested, as in the 'robot's' response in example 7 below.

Example 7 [NMC8_TE140-141]

U: What do you see?

R: Gas station and car park.

The *Clarify* dialogue act is a reply to a question that contributes with more information over and above what was strictly asked. The difference between 'Reply' and 'Clarify' is illustrated in the example 8 below; the first response by the 'robot' is classified as 'Reply' and the second is a 'Clarify' dialogue act. The 'robot's' response in example 4 above was also a 'Clarify'.

Example 8 [NMF2_F62-64]

U: Is the gym on your left or right?

R: The gym is on the right.

R: Brunel is on the left.

Figure 4.7 shows the decision tree used in the annotation of the dialogue acts.

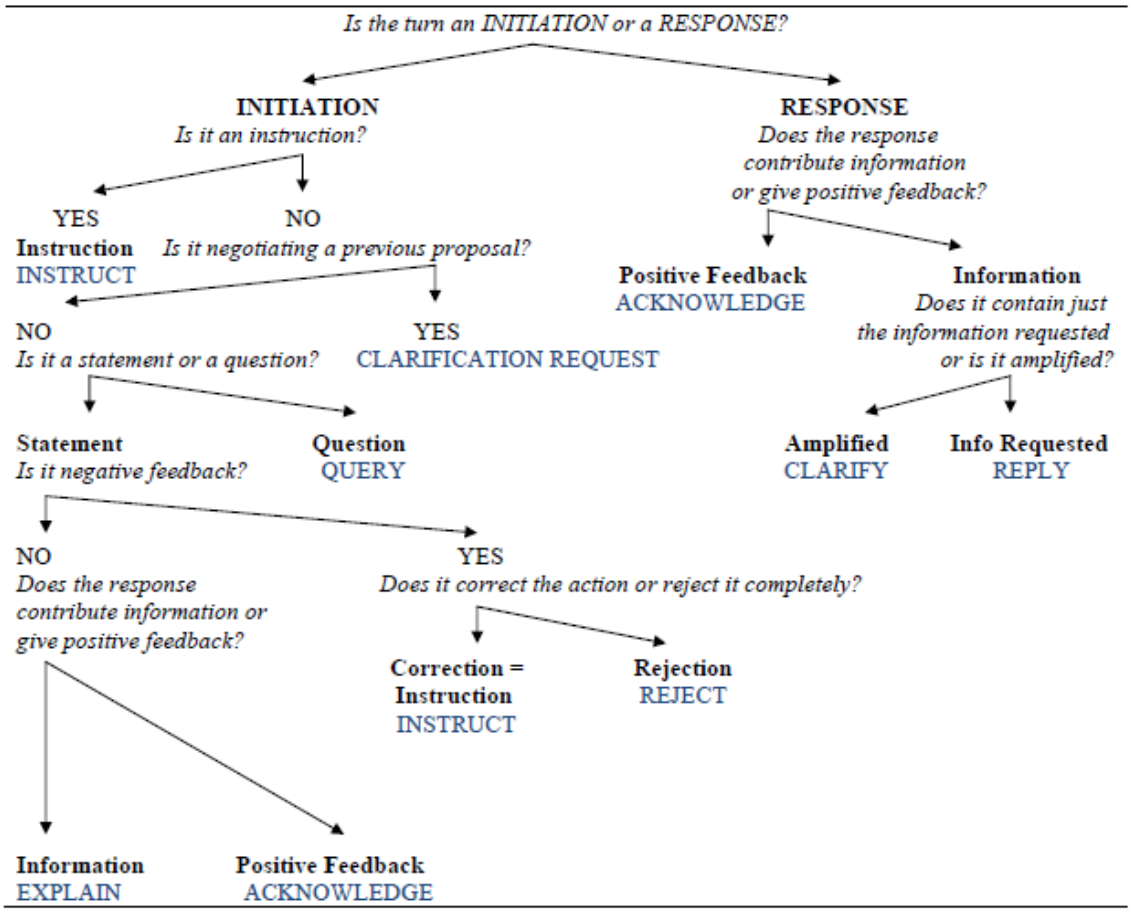


Figure 4.7: Decision tree for the annotation of dialogue acts. The dialogue act categories are shown in blue.

4.4.5 Component-based analysis of instructions and utterances

The utterances that were coded as 'Instruct' dialogue acts were segmented into 1,660 main clauses, termed instruction units. As in examples 1 and 2 below, the 'Instruct' dialogue acts by the user contain two instruction units each.

Example 1 [MF9_P12]

U: Go forward, then turn right

Example 2 [MF3_F47]

U: Now turn right and move forward until there is a road on your right.

One of the most widely used classification schemes for route instructions was developed by Denis (1997) after extensive empirical research (as discussed in Chapter 2, section 2.5.1). The focus of the classification is on whether the instructions contain references to landmarks. In particular, the instructions were divided into the following classes[10]:

Class 1: prescriptions of an action without any references to a landmark. For example, 'go forward', 'turn right'.

Class 2: prescription of an action with landmark references. For example, 'move forward until you find a bridge', 'turn right at the junction'.

Class 3: Introduction/description of landmarks with descriptive verbs such as 'is', 'see', or 'find'. The landmark is mentioned without any reference to an action to be executed. For example, 'you will see the pub on your right'.

Class 4: Description of a landmark in terms of its characteristics. For example, 'it is a big grey building'.

Tenbrink and Hui (2007) employs a taxonomy that includes Denis' classes 1 and 2 but suggests that landmark references should be further divided into: references to three-dimensional landmarks (referred to as *spatial locations*, such as buildings and bridges) and two-dimensional landmarks (referred to as *path entities*, such as streets and junctions). An example of the annotation according to these schemes is provided below.

---

[10] In Denis (1997), there is also a class 5 that contains commentaries, such as "good luck", "be careful, the path is not well paved".

Example 3 [NMF4_T104]

U: Please walk ahead and turn left at the end of the road. Continue straight until you reach the underground. It is on your left.

The turn in example 3 is segmented into four instruction units, as shown in Table 4.5 below. 'DIR' denotes prescription of action, 'DES' denotes introduction/description of landmarks, and 'L' and 'P' stand for references to spatial location and path entity respectively.

Table 4.5: Example of annotation of instructions based on the schemes by Denis (1997) and Tenbrink and Hui (2007).

Instruction Unit	Annotation
Please walk ahead	DIR
Turn left at the end of the road	DIR P
Continue straight until you reach the underground	DIR L
It is on your right	DES L

These schemes were considered too coarse for the purposes of this thesis, as they do not categorise smaller constituents, like prepositions, and terms that specify actions, spatial relations and environmental features. Therefore, the present analysis employed the Communication of Route Knowledge (CORK) framework developed by Vanetti and Allen (1988), which is described in the following subsection.

The CORK framework

The CORK framework is comparable to the aforementioned schemes, but complements them in two respects. Firstly, it further divides the *path entity* category into '*choice points*', which include junctions, intersections and crossroads, and '*pathways*', which include channels of movement (streets, roads, etc.). Secondly, it introduces delimiters – features that define the instructions and provide differentiating information about landmarks.

In the CORK framework, instructions are divided into directive and descriptive communicative statements. Similar to classes 1 and 2 in Denis' scheme, directives are action-based commands that contain verbs of movement, like 'go' and 'turn'. Equivalent to class 3, descriptives contain 'state of being' verbs like 'see', 'be' and 'find'. Descriptives provide the followers with information about relations between the follower and a landmark or between landmarks (for instance, 'the shop is next to the café'). Moreover, a descriptive offers perceptual experience (as in 'you will see the pub on your right'). Directive and descriptive communicative statements may also contain references to environmental features, that is, locations, pathways and choice points. A location is defined as an environmental feature that can function as point of reference. Pathways refer to channels of movement, such as streets. Choice points are places that afford options with regard to pathways, for instance, junctions and crossroads; choice points allows for errors to be made. Finally, communicative statements can contain delimiters. Delimiters constrain the instruction or provide discriminatory information about the environment. There are four categories of delimiters, listed below.

1. *Distance designations* specify action boundary information or the space that separates points of reference. For example, 'move forward *until* you see a car park', '*from* the bridge continue straight *to* the university'.

2. *Direction designations* specify spatial relations in terms of an intrinsic body-based frame of reference (left, right) or cardinal directions (north, south, up, down, forward, backward). For example, 'turn *left*', 'go *back to* the lab'.

3. *Relational terms* are prepositions used to specify the spatial relationship between the follower and the environmental feature (on your left), or between environmental features (on the left of, toward, away from, between, in front of, beside, behind, across from etc.). For example, 'the lab will be *on your right*', 'go to shop *next to* the café'.

4. *Modifiers* are adjectives to differentiate features. For example, 'turn left at the *big red* bridge'. Modifiers include ordering expressions (Tenbrink et al., 2007). For example 'take the *first/second/last* road on the left'.

The classification scheme used in this study ultimately brings together Denis' (1997), Tenbrink and Hui's (2007) and Vanetti and Allen's (1988) taxonomies, and is summarised in Table 4.6. It should be noted that a distinction between references to the destination location and other locations along the route was made in the present scheme.

Table 4.6: Framework for the analysis of instructions and utterances in the corpus. The second column includes the tags used in the analysis.

Communicative Statement Type	Tag
Directive statement	DIR
Descriptive statement	DES
References to Landmarks	Tag
Location	L
Pathways	P
Choice points	C
Destination	D
Delimiter Category	Tag
Distance designations	1
Direction designations	2
Relational terms	3
Modifiers	4

Following this framework, the example presented in the previous section was annotated as shown in Table 4.7 below.

Example 1 [NMF4_T104]

U: Please walk ahead and turn left at the end of the road. Continue straight until you reach the underground. It is on your right.

Table 4.7: Component-based annotation of a dialogue example.

Instruction Unit	Annotation
Please walk ahead	DIR 2
Turn left at the end of the road	DIR 2 C
Continue straight until you reach the underground	DIR 2 1 D
It is on your right	DES D 3

Finally, the analysis using the CORK framework was extended to all user and 'robot' utterances, as illustrated in the examples below. The robot turn in Table 4.8 was tagged as an 'Explain' dialogue act and contained a reference to a choice point and a location and a relational term (category 3 delimiter).

Table 4.8: Example of component-based annotation of a 'robot' utterance

Utterance	Annotation
[MF9_L46] *R*: I have reached the junction by the bridge.	C 3 L

Table 4.9 includes the annotation of an instruction, as well as of utterances by the 'robot' and user. The turn by the 'robot' is an 'Acknowledge' dialogue act and consists of a location reference. The second turn by the user is a 'Query' and contains a location reference and a relational term (category 3 delimiter).

Table 4.9: Example of component-based annotation of a dialogue excerpt.

Utterance	Annotation
[MF3_L17-19] *U*: First, go to the pub	DIR L
R: I have walked to the pub	L
U: Is the pub in front of you?	3 L

Annotation of instruction granularity

Granularity is argued to be a decisive factor of route communication efficiency (see Chapter 2, section 2.5.2). The analysis of granularity in this thesis follows the definition and specifications provided by previous literature (Tenbrink et al., 2010; Klippel at al., 2009). Granularity refers to the level of specification used by a person to describe a particular situation, event or object. In particular, instructions with low granularity are simple in form, turn-by-turn directions to the destination and only contain spatial directions. Instructions of high granularity were defined as including location references, which could also be anchored spatially.

The classification of the instruction in the corpus to simple and compound (*low* and *high* granularity, respectively) was straightforward based on the annotation using the CORK framework. In particular, category 2 delimiters (such as left, right, down, forward) are the basic constituents of a route instruction since they indicate the direction of movement. Complementing the directional instructions with action boundary information (provided by category 1 delimiters), and/or terms that clarify the frame of reference (category 3 delimiters) and specify the target landmark (category 4 delimiters) increases the instruction's level of granularity and reduces referential ambiguity (Allen, 2000a; Tenbrink and Hui, 2007). Similarly, based on the interpretation of granularity, references to spatial entities (like locations, pathways and choice points) contribute to the specificity of the instructions. Thus, a simple instruction can only have two components (verb of movement and direction of movement, as in 'turn right') or just one (as in 'stop'). Any other component is a location reference and/or category 1, 3 and 4 delimiters. Based on this, a simple granularity metric was derived. Namely, the number of components was calculated and the instructions with more than 2 components were considered compound (high granularity), whereas instructions with one or two components were simple (low granularity). Examples of this annotation are given in Table 4.10.

Table 4.10: Examples of component-based annotation of user instructions (the tags refer to the CORK framework categories as summarized in Table 4.6: DIR: directive statement based on verb of movement; C: reference to choice point; L: reference to location; the numbers signify delimiter types 1, 2, 3 and 4).

Instruction Unit	CORK tag	Number of components	Granularity
Move forward	DIR 2	2	Low
Move forward until you get to the first junction on your right	DIR 2 1 4 C 3	6	High
Move forward until you reach a bridge	DIR 2 1 L	4	High

The annotation of granularity is illustrated by considering the second instruction in the example captured in Table 4.10 ('Move forward until you get to the first junction on your right'): the instruction is a directive statement (DIR) based on the verb of movement, 'move'; 'forward' is a category 2 delimiter designating direction; 'until' is a category 1 delimiter, providing boundary information for the action, 'move forward'; 'first' is a category 4

delimiter specifying the target landmark, ‘junction’; the ‘junction’ is a choice point; and the choice point is further complemented by the category 3 delimiter, ‘on your right’, stating its position in relation to the frame of reference. This results in six components in the instruction, classifying it as a compound instruction (high granularity).

Deictic and anaphoric pronouns

The analysis considered the use of deixis and anaphora by users and ‘robots’. In particular, the frequencies of deictic forms (in particular, ‘this’, ‘that’, ‘here’, ‘there’ and ‘now’) and anaphoric references (such as ‘it’) were measured. Deictic expressions are used for indexing entities in the local surroundings. They are generally preferred by speakers, as they substitute for longer referring expressions that are based on spatial relations like ‘left’, ‘right’, ‘front’ etc. Anaphoric pronouns refer to antecedent objects in the situation in place of nouns. However, they require both conversational partners to establish that these entities are in their joint attention. The studies in human communication discussed in section 2.9 of Chapter 2 predicted that a shared visual space increases the use of these expressions. As such, Research Question 5(a) explores whether users in the Monitor condition are more likely to provide instructions such as ‘take this road’ or ‘turn left now’ and ‘robots’ to refer to the destination by asking ‘is this it?’ (as in the example in Table 4.11 below).

Table 4.11: Dialogue excerpt containing deictic expressions

Utterance
[MF2_L23-25] *R*: turn left **now**.
U: Is this **it**?
U: **it** most certainly is.

4.4.6 Annotation of lexical alignment

As described in Chapter 2 (section 2.8), alignment manifests in various aspects of linguistic behaviour, ranging from alignment on the phonetic level to the sociolinguistic level of formality of language. The present study focuses on lexical alignment. As such, the analysis

basically investigated whether speakers use the same words as their partner. Following the Interactive Alignment Model of human communication (see Chapter 2, section 2.8.1) and addressing the limitations of related work in HCI (see Chapter 2, section 2.8.3), it was necessary to capture alignment both 'locally', as priming, and 'globally', as lexical innovation, in the dialogue.

First, alignment was measured by looking at the adjacency pairs in the dialogue and comparing the two utterances (what the Interactive Alignment Model terms 'input/output matching'). An adjacency pair is a sequence of two *related* utterances by two *different* speakers, such that the second utterance is a response to the first – for instance, paired responses like a question followed by an answer, or an offer followed by acceptance or rejection (Levinson, 1983, p.303). So, a turn was a 'match', if it contained the same type of component as the turn to which it was a response. For each matching component in an utterance, a score of 1 was given. If no component matched, the turn was a 'mismatch' and a score of 0 was given. The sum of matching components was noted. The annotation of alignment at the adjacency pair level is exemplified through three dialogue excerpts, shown in Tables 4.12 – 4.14.

In the first example, the user's utterance matches the previous utterance by the 'robot', repeating the modifier, 'bendy', and the pathway reference, 'road'. Thus, it is marked as '2' matches. The aligned components are in bold in Table 4.12.

Table 4.12: First dialogue example of alignment

Utterance	Match
[MF9_P29-30] *R*: I am at the junction by the bridge, facing the ***bendy road***.	
U: Go into the ***bendy road***.	2

In the second example (see Table 4.13), the first 'robot's' utterance repeats the choice point reference, 'crossroad' acknowledging the execution of the action. Hence, it is marked as containing '1' match. The user's subsequent utterance does not match any component of the utterance by his/her partner and is therefore annotated as '0' match (i.e., a mismatch). In contrast, the 'robot's' final utterance reiterates the destination reference, 'shop', and the relational term 'on my left' and is therefore marked as containing '2' matches.

Table 4.13: Second dialogue example of alignment

Utterance	Match
[MF7_S71-74] *U*: Turn right, go along the road until the ***crossroad***.	
R: I am at three-way ***crossroad***.	1
U: Turn left, go straight, the second building ***on your left*** is the ***shop***.	0
R: ***Shop on my left***, unknown building behind on my left.	2

In the third example (see Table 4.14), the user first produces an instruction which does not match the previous utterance. This is immediately reformulated to repeat the exact expression used by the 'robot', 'at y-shaped junction', containing '2' matches.

Table 4.14: Third dialogue example of alignment

Utterance	Match
[MC7_S142-143] *R*: I am at y-shaped junction.	
U: make a right.	0
U: make a right at ***y-shaped junction***.	2

Second, lexical innovation, the rate of unique words introduced over the course of the dialogue, was used as an indicator of global alignment (following the approach of Mills, 2007). That is, when interlocutors introduce new expressions instead of re-using those that have already occurred in the dialogue (as the Interactive Alignment Model postulates), alignment is low. Lexical innovation was calculated by comparing every constituent word in an utterance to the previous words in the dialogue. For example, an utterance such as 'turn left' leads to a backwards search in the dialogue for the previous occurrence of 'turn', adding '1' to the alignment score if not found and '0' if found, before moving on to the next word. Lexical innovation was also used to capture alignment achieved by the end of the dialogue and was measured by the ratio of unique words produced in undertaking the final task of the session. Simply put, the lower the ratio of unique words towards the end of the dialogue, the higher the level of alignment ultimately achieved.

The Interactive Alignment Model states that speakers tend to repeat their own and each other's linguistic expressions. In effect, this process is the cumulative effect of both 'self-alignment' and 'other-alignment'. For example, in the dialogue excerpt in Table 4.12, the

user repeated ‘bendy road’ and, hence, a ‘user match’ was added. Let’s assume that the ‘robot’ had subsequently signalled a problem by saying “the bendy road is closed”. The question that naturally arises is whether the ‘robot’ self-aligns or aligns to the user. Indeed, it is impossible to know, in such cases, whether a speaker aligns to himself/herself, to the other, or *both*. The annotation would not discount such cases or classify them differently, but would analyse them as all other responses to initiations; whether they matched or not the partner’s utterance.

4.4.7 Reliability of annotation

Lexical innovation was automatically calculated. The rest of the measures were manually annotated. The manual annotation was largely performed by cross-referencing the utterances with the system logs of the robot actions and position at the time each message was sent or received. The annotation process was performed in two stages. During the first stage, 25% of the corpus (48 dialogues, 933 turns, from both conditions) was coded by two annotators: an expert annotator and an annotator with no prior knowledge of discourse analysis or experience in dialogue data annotation, who received a training session before undertaking the analysis. The annotators coded the same 25% of the corpus, and worked independently. The consistency of the annotation was calculated by a series of Cohen’s Kappa. As explained above, the annotation of alignment and miscommunication involved little subjective judgement. The annotation of components based on CORK was also relatively mechanical. For the classification of dialogue acts, the scheme had excluded all categories that had proved problematic and ambiguous for annotators in the original evaluations (Carletta et al. 1996), and this ensured good inter-annotator agreement. The Kappa values are provided in Table 4.15 and show a generally high level of agreement between the annotators (values above 0.70 are normally considered satisfactory (Lazar et al., 2010, p. 298)). The few items where disagreement occurred were discussed between the annotators. In the second stage of the annotation, only the expert annotator annotated the remaining 75% of the corpus, because of the high level of inter-annotator agreement from stage 1.

Table 4.15: Agreement between the annotators expressed by Cohen’s Kappa.

Category	Cohen’s Kappa

Match/Mismatch/None	0.961
Execution Error/None	0.842
Non-understanding/None	0.886
Incorrect Instruction/None	0.816
Components (CORK)	0.978
Dialogue acts (simplified HCRC moves)	0.878

The annotation of dialogue acts, components of instructions and utterances, and miscommunication was undertaken by one main annotator. To ensure reliability, 1/4 of the corpus was annotated by three other people. These annotators had no knowledge of discourse analysis and previous experience in dialogue data annotation. They attended a tutorial with the main annotator and were supplied with written instructions, annotation examples and copies of the coding schemes provided in Figure 4.7 and Table 4.6, and the definitions of miscommunication, as described in section 4.4.2. The annotators coded the same dialogues, but worked independently. The results showed that for the annotation of dialogue acts and components, the level of agreement reached an average of 98%. For the miscommunication annotation, the mean agreement between the novices and main annotator was 91%. These percentages of inter-annotator agreement were considered satisfactory. After reviewing the annotation items in which disagreement occurred, it was decided that the analysis could be based on the main annotation.

4.5 Chapter summary

This chapter provided a detailed account of a methodology for the collection and analysis of performance and dialogue data, which was developed following the research paradigm, objectives and specific research questions of this thesis. Central in the experimental approach was the design of a system to elicit spontaneously generated spatial descriptions and natural interaction phenomena within a controlled spatial setting. The system reenacted a robot navigation scenario and enabled and recorded the interaction and synchronous execution of route instructions, while allowing for fine manipulation of experimental conditions and detailed examination of the unfolding interaction in context. Using multiple 'robots' instead of a single confederate and masking the gender of the interlocutors ensured that the data was not 'contaminated' by experimental bias and social preconceptions. The data analysis approach involved the examination of performance and dialogue data. For the analysis of

performance, established metrics were used to assess interaction efficiency and effectiveness and user experience. For the dialogue analysis, existing classification schemes grounded on empirical data and theoretical and linguistic models were evaluated, adapted and unified, leading to a fine-grained framework that integrated the analysis of dialogue acts, their components, miscommunication and linguistic alignment. Taken together, a complete evaluation framework was developed oriented towards the research questions of this thesis, but which is also hoped to be useful to future studies in the domain of interactive systems.

5 Results

5.1 Introduction

The analysis of past research in Chapter 2 brought forth a series of research questions which were listed in Chapter 3. The fourth chapter described the data collection and analysis methodology developed to address these questions. This chapter provides a justification of the experimental design and statistical analysis approach and reports the results of the analysis of the interaction data of the dyads. These results should guide the 'answers' to the research questions of this thesis.

The central research question is *how gender differences arise in navigation and route instruction dialogues with computer systems*, thus, performance- and dialogue-based measures were used. As detailed in Chapter 4, performance results were produced by the analysis on measures that are commonly employed in the evaluation of interaction efficiency and effectiveness: time taken and number of words, turns by user and 'robot' and instructions per task. In addition, post-task questionnaires were used to capture users' subjective evaluation of the dialogue with the system. Miscommunication, user- and 'robot'- induced, also served as an objective measure of performance. Dialogue-related findings were derived by two analyses: first, dialogue act analysis and fine-grained component analysis of user and 'robot' utterances were performed; in particular, the frequencies of types of dialogue acts (acknowledgements, queries and clarifications), landmark references and delimiters, directive and descriptive instructions and the level of granularity of instructions were measured. Second, alignment was measured 'locally' and 'globally' in the dialogue; as match scores between user and 'robot' utterances at the adjacency pair level, and as lexical innovation, that is, the rate of unique words introduced over the course of the dialogue.

The structure of the chapter parallels the structure of the previous chapter and presents the results grouped under the main themes of the central research question. For quick reference, the diagram illustrating the concepts and their relations investigated in this thesis is reproduced in Figure 5.1 below. Section 5.2 provides details of the experimental design and statistical tests performed on the data. Section 5.3 presents some basic information about the corpus. Sections 5.4 and 5.5 report the results relating to performance measures, miscommunication and user perceptions of the interaction. Sections 5.6 and 5.7 detail the findings produced by the dialogue act and component-based analyses and section 5.8 describes the results from the analysis of alignment. The chapter concludes by giving an overview of the findings, and also lists the research questions addressed in the study and distils the findings into high-level 'answers'.

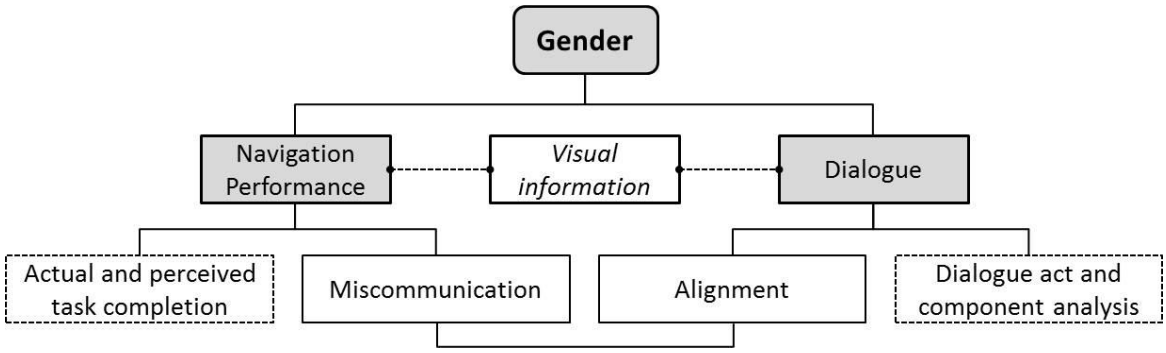


Figure 5.1: Diagram outlining the concepts analysed in this thesis and their relations

5.2 Statistical analysis approach

The premise of this thesis is that communication is mainly a function of intra-, inter-individual and contextual parameters. Therefore, the factorial experimental design is a natural choice because it elucidates and disentangles interaction effects (of role and gender in different visual information conditions) from simple main effects. The visual information conditions are implemented as 'Monitor' and 'No Monitor' conditions (as explained in Chapter 4, section 4.2.2).

A 2×2×2 Analysis of Variance (ANOVA) for independent groups was performed. The between-participants factors, each with two levels, were: (i) Monitoring (Monitor and No

Monitor), (ii) User Gender (Female users and Male users) and (iii) Robot Gender (Female 'robots' and Male 'robots'). The 2×2×2 factorial design is illustrated in Table 5.1 below.

Table 5.1: 2×2×2 factorial design: Factor 1: Monitoring (2 levels: Monitor/No Monitor), Factor 2: User Gender (2 levels: Female/Male), Factor 3: Robot Gender (2 levels: Female/Male)

Monitoring	User Gender	Robot Gender
Monitor	Female	Female
		Male
	Male	Female
		Male
No Monitor	Female	Female
		Male
	Male	Female
		Male

Correlational analyses were undertaken in addition to the ANOVAs to pinpoint significant relationships between dependent variables. Moreover, for categorical data, relationships were investigated through chi-square tests of independence. Given the limitations of the chi-square test compared to parametric ones[11], statistically significant results were clarified through additional measures; namely, odds ratios and the phi coefficient (φ). The odds ratio is extremely useful as it compares the two groups (levels) of a variable and represents the differences between them. The phi coefficient provides the magnitude of the relationship between two variables. Measures of correlation, like φ, have been criticised because they are hard to interpret meaningfully and tend to be small even when the effect is important (Howell, 2009, p.300). This is because φ decreases dramatically when the marginal totals (the totals in a contingency table) are variable (Zysno, 1997). However, non-uniform marginal totals have no effect on odds ratios. When appropriate, the analysis followed a 'top-down' approach in order to identify the locus of a significant effect. Namely, the data were

[11] Chi-square analysis shows that there is a statistically significant association between two independent variables or groups, but not the nature of the association (for instance, how two groups differ) and the strength of the association (how much they differ).

initially analysed as a whole. Subsequently, the analysis controlled for the Monitoring independent variable and separate chi-square tests were performed on the Monitor and No Monitor data. The User Gender and Robot Gender independent variables were explored with three tests: (i) a *2 × 2* chi-square test for different User Gender (Female/Male) and (ii) a *2 × 2* chi-square test for Robot Gender (Female/Male) and (iii) a *4 × 2* chi-square test for the four different pair configurations (as illustrated in Table 5.2 below). While Pearson's chi-square is the standard test for determining associations between categorical variable in this study, additional chi-square analysis was performed, as required by the nature of the data and design. In particular, results from Linear, Cochran-Mantel-Haenszel and McNemar's chi-square tests are reported, where appropriate.

Table 5.2: Pair configurations and the abbreviations, henceforth used.

Pair Configuration	Abbreviation
Female user – Female 'robot'	F_uF_r
Female user – Male 'robot'	F_uM_r
Male user – Female 'robot'	M_uF_r
Male user – Male 'robot'	M_uM_r

Statistically significant results indicate that there is a real difference between groups or association between variables, but they need to be supplemented by measures of how different these groups are or how strong this association is. In other words, a complete analysis needs to report the effect size and state not only that a result is significant but also 'important'. Thus, as part of the analyses of variance, the eta-squared (η^2) and Cohen's *d* are provided and the chi-square analyses include odds ratios and the phi coefficient (as detailed above). The phi coefficient and eta-squared estimates belong to the *r*-family of measures of effect size. They indicate the proportion of variability in the dependent variable scores that are explained or predicted by the independent variable. Odds ratios and Cohen's *d* show the differences among the groups and are of the *d*-family of measures. Many scientists regard the *d*-family of measures as more informative (see McGrath and Meyer, 2006). It should be noted that, in this analysis, eta-squared is computed by the original formula; that is, sums of squares

of the effect under consideration divided by the total sums of squares[12]. Cohen's *d* is calculated as the difference between the means of the two groups in comparison, divided by the pooled standard deviation. For instance, a $d = 1$ denotes that the two groups' means differ by one standard deviation. While noting the risk of setting generally applicable guidelines, Cohen (1988) originally proposed interpretations for the effect size based on the eta-squared and Cohen's *d* values, as illustrated in Table 5.3. Thus, if *d* is less than 0.2 – that is, the two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it was found to be statistically significant.

Table 5.3: Interpretation of effect sizes based on the values of eta-squared and Cohen's *d* measures.

Effect size	Eta-squared	Cohen's d
Small	0.01	0.2
Medium	0.059	0.5
Large	0.138	0.8

The design was balanced but the sample size of each group was small. Therefore, particular caution was exercised with respect to the assumptions for parametric tests (that is, normality and homogeneity of variance). For all dependent variables tested, the shape of the distributions was examined. Although some of the histograms did not look particularly 'normal' (because of the small sample size), the assumptions were not grossly violated, as there were no 'lumps' or large gaps in the distributions. In some cases, outliers were identified and removed. As such, the data are not inconsistent with being drawn from a normally distributed population. Bonferroni corrections were also performed to deal with the problem of multiple comparisons. Moreover, Levene's test was used to ascertain equal variances between groups. The main effects that were found significant were verified through t-tests for independent groups and inspection of error bar graphs. The significance level accepted was $p <= 0.05$ for all statistical tests. Non-significant results were not reported, unless this was deemed relevant to the discussion. The graphs typically show the 95% confidence interval (CI) of the mean of the variable of interest.

[12] SPSS (versions up to 18.0) produces a partial eta-squared estimate instead of eta-squared, which has misled researchers to report inaccurate effect sizes (Levine and Hullett, 2002).

Studies addressing the issue of the violation of assumptions have reported that, in practice, they are commonly violated (see Grissom, 2000, for a review). For instance, Grissom and Kim (2005, p.10) cite that only 3% of the data in behavioural research have the appearance of a normal distribution. Luckily, ANOVAs and t-tests are robust statistical procedures, and departures from the assumptions have minor effects, particularly when sample sizes are equal (Howell, 2009, p. 334). For ease of reference, all measures used in this analysis and their corresponding symbols are provided in alphabetical order in Table 5.4 below.

Table 5.4: Symbols of statistical measures used in the analysis.

Statistical Measure	Symbol
Cohen's d	d
Degrees of freedom	df
Eta-squared	η^2
Mean	M
Linear chi-square	M^2
Pearson's chi-square	χ^2
Pearson's product-moment correlation coefficient	r
Phi coefficient	φ
Standard deviation	SD

The following sections detail the results of the ANOVAs, correlational and chi-square analyses performed on the dependent variables corresponding to actual and perceived task performance, miscommunication, alignment and dialogue content. The analyses are expected to elucidate the main and interaction effects of the user and 'robot' gender and visual information factors on these variables as well as the statistical relationships between them.

5.3 Basic dialogue statistics

The experiments yielded a corpus of 184 dialogues[13], which comprised 3,875 turns by the participants (2,125 user turns and 1,750 'robot' turns). 15,471 words were produced; 9,971 by

[13] The corpus is freely available for academic research. If interested in using it, contact the author at: theodora.koulouri@brunel.ac.uk.

the user and 5,770 by the 'robot'. To provide a general picture of the corpus, the averages of basic dialogue elements of each pair are included in the table below (Table 5.5).

Table 5.5: Pair averages for several elements in the dialogues.

Variable	Mean
#Turns	121.09
#User Turns	66.41
#'Robot' Turns	54.69
#Instructions	52.38
#Words	491.91
#User Words	311.59
#'Robot' Words	180.31
#Unique Words	90.06
#Turns per Task	21.06
#Words per Turn	4.06

5.4 Performance

Studies on task-oriented dialogues with humans or computer systems link efficiency to task completion time, number of words, turns and instructions and turn length (for example, Walker et al., 2000a; Gergle at al., 2004, Clark and Krych, 2004; Brennan, 2005). In addition, fewer execution and understanding failures (by the 'robot') and incorrect route instructions (by the user) are taken as indicators of superior performance.

5.4.1 Time per task

The three-way factorial ANOVA revealed a main effect of Robot Gender ($F_{(1,23)} = 9.208$, $p = 0.006$, $\eta^2 = 0.241$, $d = 1.13$), which indicated that male 'robots' required less time to complete a task (M = 335.6 seconds, SD = 73.63) than female 'robots' (M = 411.9 seconds, SD = 64.11). However, from inspection of the error bar charts, it became apparent that only groups with male 'robots' paired with female users were significantly different from the other groups (see Figure 5.2), supported by a significant interaction effect ($F_{(1,24)} = 6.197$, $p = 0.02$). To isolate the locus of the effect, one-way analysis of variance and post-hoc Tukey tests were performed and confirmed that F_uM_r pairs (M = 301, SD = 58.59) are faster to complete each

task than F_uF_r (M = 415, SD = 58.81, d = 1.94) and M_uF_r pairs (M = 409, SD = 65.99, d = 1.73) by almost two standard deviations ($F_{(3,30)}$ = 5.106, p = 0.006).

It should be noted that visual information failed to provide a significant improvement in completion time. In particular, average completion times per task were 356 seconds and 379 seconds in the Monitor and No Monitor conditions, respectively.

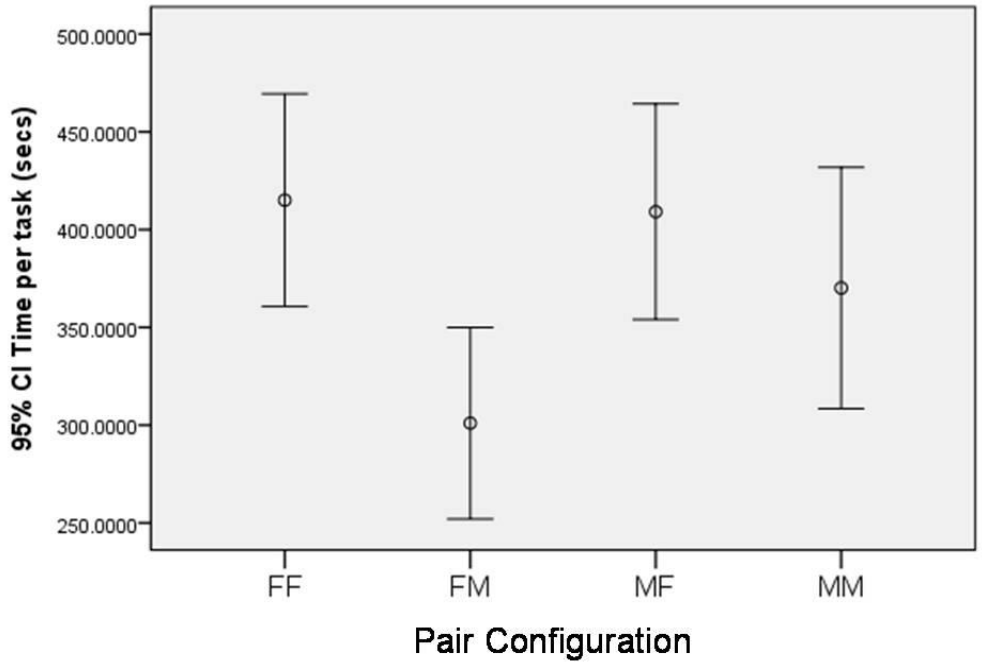


Figure 5.2: Means (and standard deviations) of time per task for all pair configurations: F_uF_r, F_uM_r, M_uF_r and M_uM_r.

5.4.2 Words, turns, turn length and instructions per task

The analysis revealed a main effect of Monitoring on number of words per task ($F_{(1,24)}$ = 6.904, p = 0.015, η^2 = 0.191, d = 0.94). Pairs in the No Monitor condition (Mean = 99.5, SD = 34.57) required a larger number of words to complete each task than the Monitor pairs (Mean = 72.46, SD = 21.27). Both users and 'robots', individually, used a larger number of words under the No Monitor condition. In fact, an additional significant effect of Robot Gender showed that users were 'wordier' when interacting with female 'robots' ($F_{(1,24)}$ = 4.393, p = 0.047).

The analysis of the proportion of turns by the user contributed with interesting results. A significant main effect of Monitoring initially indicated that users in the Monitor condition produced 57% of all turns, which dropped by 6% in the No Monitor condition ($F_{(1,23)} = 5.5$, $p = 0.028$, $\eta^2 = 0.131$, $d = 0.84$). This result was refined as a significant interaction effect of Monitoring by User Gender was found ($F_{(1,23)} = 5.548$, $p = 0.027$, $\eta^2 = 0.137$). As illustrated by the error bar graph below (Figure 5.3), female users dominated the conversational floor in the Monitor condition, having produced over 61.9% of turns. However, when monitoring was disabled, female users' turn-possession was balanced, dropping to 50.5%. Comparisons between the groups verified the difference between female users in the Monitor condition and No Monitor condition ($t_{(14)} = 3.211$, $p = 0.006$, $d = 1.6$). On the other hand, the turn ratio of male users remained consistent across conditions. The dependent variables, turn length and number of instructions, did not produce reliable differences. However, as described in a later section on instruction granularity (section 5.6.6), instructions issued in the No Monitor condition tended to be longer, consisting of more than two components.

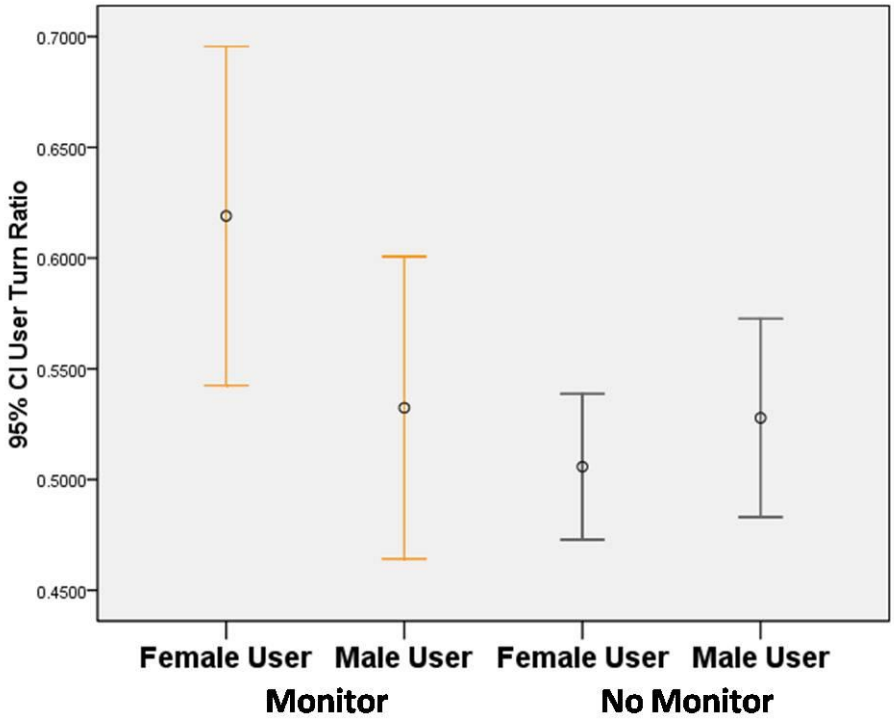
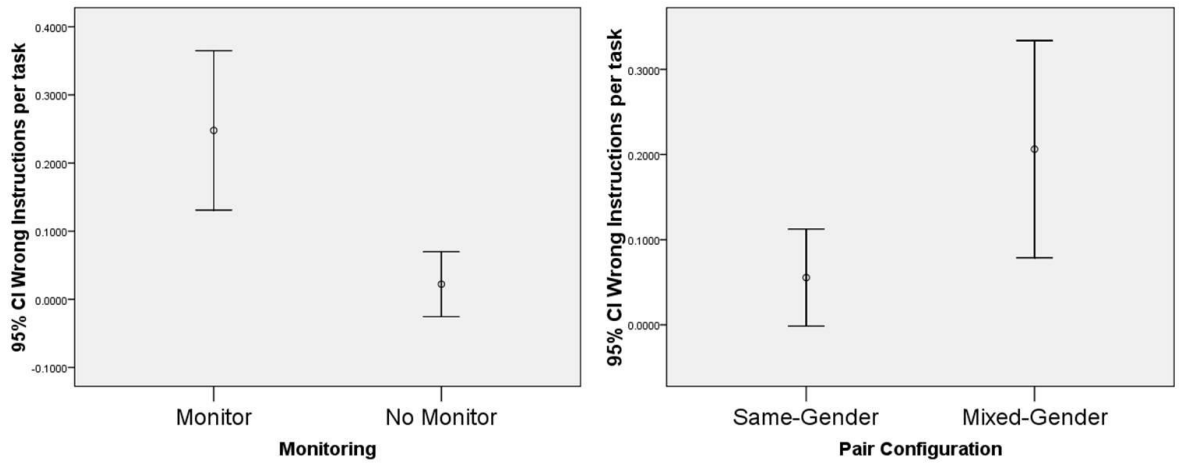


Figure 5.3: Ratio of user turns for male and female users in the Monitor and No Monitor conditions.

5.4.3 Miscommunication

As detailed in Chapter 4 (section 4.4.2), 'robot'-attributed miscommunication encompasses two measures: number of (i) execution errors and (ii) 'robot' turns that were tagged as expressing non-understanding. Miscommunication induced by the user was estimated by the number of incorrect instructions[14].

The three-way ANOVA revealed a strong significant main effect of Monitoring on the number of incorrect instructions per task. Surprisingly, the number of incorrect instructions per task was close to zero in the No Monitor condition and high in the conditions in which the user could confirm at all times the actions and understanding of the 'robot' ($F_{(1,23)}$ = 13.784, p = 0.001, η^2 = 0.304, d = 1.35). The User Gender × Robot Gender interaction was found to be significant ($F_{(1,23)}$ = 4.797, p = 0.039, η^2 = 0.106) indicating that users in mixed-gender pairs (F_uM_r and M_uF_r) tended to be less accurate compared to users speaking to 'robots' of the same gender (F_uF_r and M_uM_r). The contrast between same-gender and mixed-gender pairs also confirmed the finding, ($t_{(29)}$ = -2.251, p = 0.032, d = 0.81). The effects are shown in the graphs in Figure 5.4 below.



[14] In previous publications by the author (e.g., Koulouri et al. (2012)), miscommunication was calculated by combining number of execution errors and non-understandings only. In many instances in the data, execution errors were caused by incorrect instructions. The present analysis also considers incorrect instructions, which are measured separately. Thus, execution errors due to incorrect instructions are not tagged as execution errors.

Figure 5.4: Incorrect instructions per task in the Monitor and No Monitor conditions (graph on the left-hand side) and for same-gender (F_uF_r and M_uM_r) and mixed-gender pairs (M_uF_r and F_uM_r) (graph on the right-hand side).

Similarly, the ANOVA conducted on number of 'robot' turns expressing non-understandings yielded a significant main effect of Monitoring. Interestingly, when participants shared visual information, 'robots' produced a greater number of non-understandings ($F_{(1,24)} = 4.324$, $p = 0.048$, $\eta^2 = 0.134$, $d = 0.76$). Finally, for execution errors as dependent variable, no differences were found among the groups. The results are summarised in Figure 5.5, which shows the distributions of incorrect instructions, non-understandings and execution errors across the two conditions.

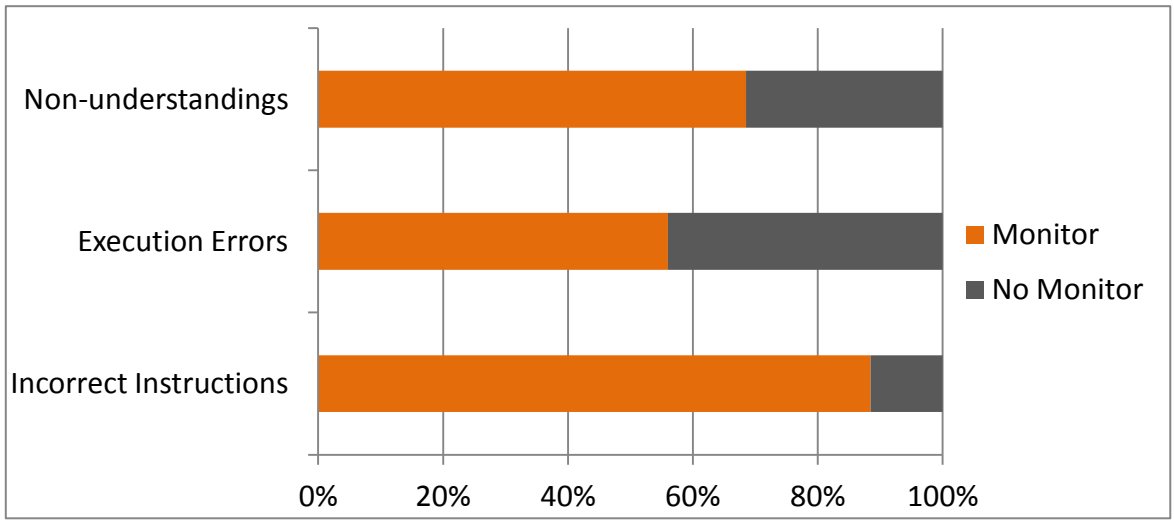


Figure 5.5: Distribution of miscommunication in the Monitor and No Monitor conditions.

5.5 User perceptions of the interaction

After each task, the users completed a seven-point Likert-scale questionnaire in which they rated their agreement with five statements. These statements covered ease of use (item 2), accuracy (item 3) and helpfulness (item 4) of the system, perceived task completion (item 1) and overall satisfaction (item 5). The levels of agreements were: *strongly disagree, disagree, slightly disagree, neutral, slightly agree, agree, and strongly agree*. They were mapped to integer values ranging from 1 to 7 (with 7 representing the optimal score). The values for

each statement were summed for all six tasks. Thus, the summed score for a statement could be in the range of 6 to 42. The set of statements is provided in Table 5.6 below for quick reference.

Table 5.6: Statements in the questionnaire completed by users after the completion of each of the six tasks.

1. I did well in completing the task
2. The system was easy to use
3. The system was accurate
4. The system was helpful
5. I am generally satisfied with this interaction

ANOVA and correlational analysis were performed. Though the use of parametric or non-parametric tests on rating scores has been a controversial issue, Likert scale data are commonly and legitimately treated as if they were interval (Gravetter & Forzano, 2012, p. 92; Norman, 2010). Employing such an approach has been recommended by HCI practitioners and applied statisticians (Sauro & Lewis, 2012, pp. 243-246; Lewis, 1993) and was therefore adopted in this study.

As expected, all statements were negatively correlated with frequency of non-understandings and execution errors, such that users rated their performance, the system and interaction less favourably. The results of the analysis are shown in Table 5-7 below.

Table 5.7: Correlation matrix showing significant correlations between execution errors and non-understanding and statements.

		Statement 1	Statement 2	Statement 3	Statement 4	Statement 5
Execution Errors and Non-understandings	Pearson Correlation	-0.438	-0.617	-0.523	-0.506	-0.721
	p value	.015	.000	.003	.004	.000
	N	30	30	30	30	30

A mixed ANOVA design was employed to explore the effect of gender. The within-subjects factor was Statement, with five levels corresponding to the statements in the questionnaire. The between-subjects factors were Monitoring, User Gender and Robot Gender. A significant difference between statements was found ($F_{(3.626,79.768)}$ = 3.080, p =

0.024, $\eta^2 = 0.094$)[15]. More importantly, the analysis determined a significant interaction effect of User Gender and Statement ($F_{(3.626,79.768)} = 2.750$, $p = 0.038$, $\eta^2 = 0.084$). In particular, the results indicated that male users perceived higher task success than females (item 1). On the contrary, system accuracy was rated more favourably by females (item 3). User satisfaction (item 5) was similar for both genders. System ease of use and helpfulness were also not significantly different. The interaction is shown in Figure 5.6.

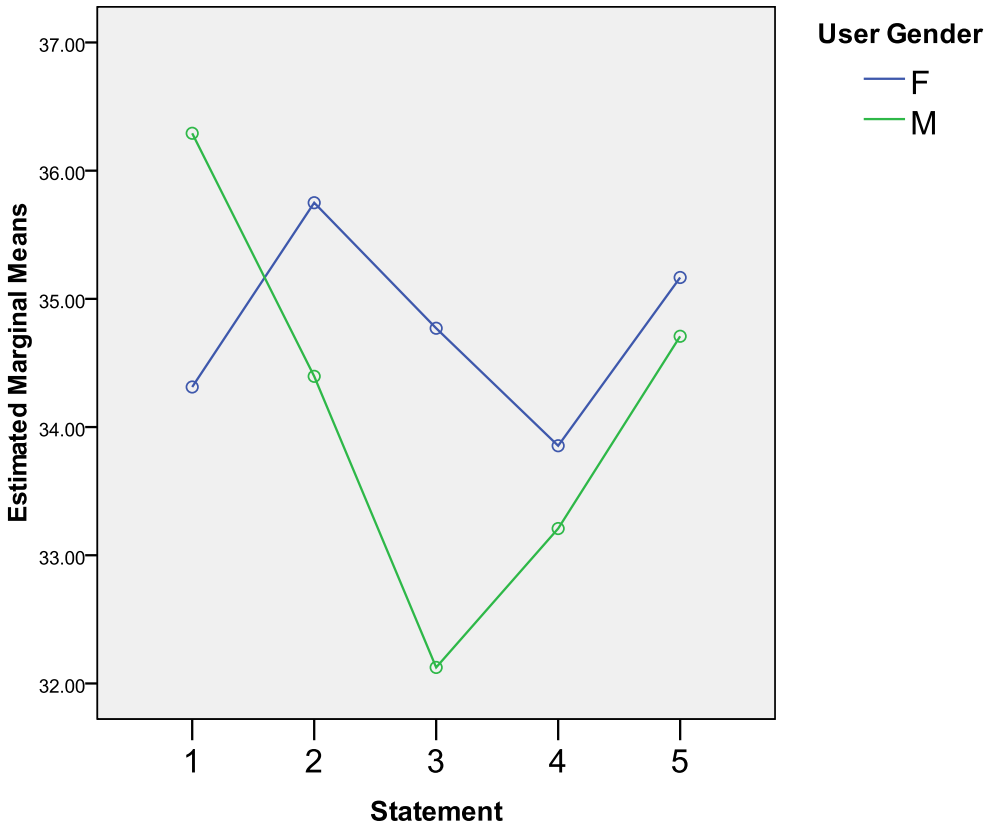


Figure 5.6: Mean summed scores of each statement for female and male users. The statements were the following: 1: I did well in completing the task; 2: The system was easy to use; 3: The system was accurate; 4: The system was helpful; 5: I am generally satisfied with this interaction.

[15] The analysis was adjusted using the Huynh-Feldt correction. The Greenhouse-Geisser and Huynh-Feldt corrections are the most widely used procedures to combat the violation of the sphericity assumption (which increases the type I error rate). Many statisticians recommend using the latter, as it is more powerful (Abdi,

Correlational analysis also revealed a significant negative correlation between user experience of task success ('*I did well in completing the task*') and lexical innovation (r = -0.473, p = 0.013), suggesting that users rated the interaction as less successful when alignment was weaker. This finding is discussed in relation to Research Question 7(e) in section 5.8 which investigates alignment.

5.6 Dialogue acts

Chapter 4 (section 4.4.4) described the primary annotation of the utterances, which was based on the HCRC dialogue act coding scheme. The analysis presented in this section considered the frequencies of certain dialogue acts: the number of queries (questions), acknowledgements (positive feedback to show that the utterance to which it responds has been understood and accepted) and clarifications (responses to questions that give information over and beyond what was asked) issued by the user and 'robot'. The frequencies of these dialogue acts were directly relevant to Research Questions 5(b), 5(d), 5(e) and 6(b).

5.6.1 Queries

The three-way ANOVA performed on the number of user queries yielded a significant main effect of the Monitoring factor ($F_{(1,22)}$ = 14.710, p = 0.001, η^2 = 0.251, d = 1.2). In particular, the user queries showed a dramatic increase when monitoring was disabled. The analysis also revealed an interaction effect between Monitoring and User Gender ($F_{(1,22)}$ = 7.247, p = 0.013, η^2 = 0.124). T-tests and inspection of the error bar charts (shown in Figure 5.7) confirmed that the greatest number of queries was given by female users in the No Monitor condition. Finally, a significant three-way interaction of Monitoring by User Gender by Robot Gender was detected ($F_{(1,22)}$ = 4.203, p = 0.05, η^2 = 0.072). It refined the effects and indicated that although in the Monitor condition, female users paired with female 'robots'

2010). The general recommendation is to use the Huynh-Feldt correction when the epsilons are around 0.75 (Howell, 2010, p.476). Here, the epsilon values were 0.608 (Greenhouse-Geisser) and 0.906 (Huynh-Feldt).

rarely asked questions (M = 0.25, SD = 0.29), when visual information was not shared, the number of their queries exploded, increasing by 2.36 standard deviations (M = 3.45, SD = 1.89). The effect is illustrated in Figure 5.8.

Looking at the other side of the communication, the analysis on the number of 'robot' queries per task also produced a main effect of Monitoring ($F_{(1,23)}$ = 11.014, p = 0.003, η^2 = 0.274, d = 1.17), but inversely: the 'robots' issued a larger number of queries when their partners were able to monitor their actions.

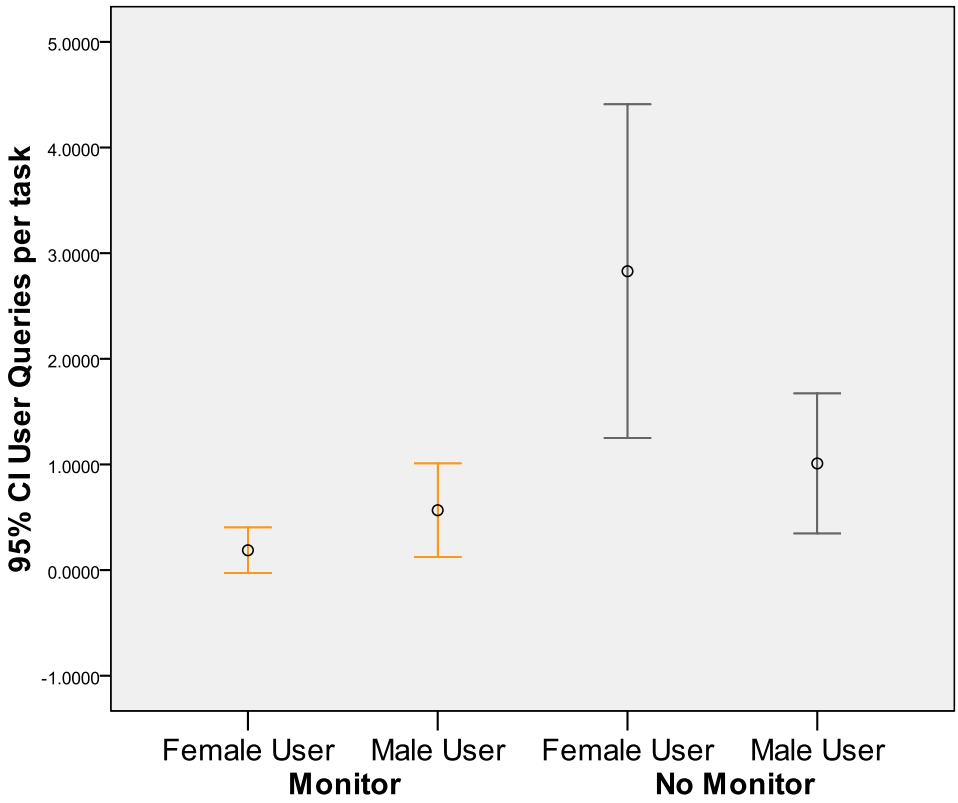


Figure 5.7: Queries by male and female users in the Monitor and No Monitor conditions.

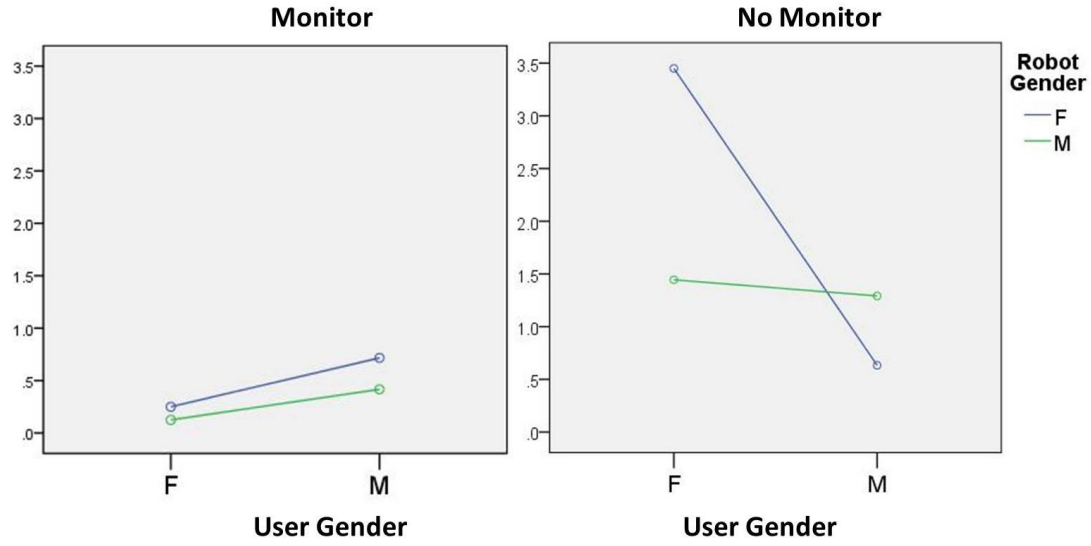


Figure 5.8: Plots of the interaction of User Gender and Robot Gender for each level of Monitoring. The Y axis represents the means of queries by Female or Male users.

5.6.2 Acknowledgements

The analysis on acknowledgements per task revealed an analogous pattern of significant effects. Namely, in the absence of shared workspace, participants produced a larger number of acknowledgements signalling understanding and acceptance of previous statements ($F_{(1,22)}$ = 4.459, p = 0.046, η^2 = 0.102, d = 0.74). This effect was overshadowed by a significant effect of Monitoring by User Gender ($F_{(1,22)}$ = 6.786, p = 0.016, η^2 = 0.155). Inspection of the error bar charts and t-tests showed that pairs of Female users in the No Monitor condition provided a significantly higher number of acknowledgements compared to the other groups (see Figure 5.9). A conclusive result was reached through the second-order interaction effect of Monitoring by User Gender by Robot Gender ($F_{(1,22)}$ = 4.195, p = 0.05, η^2 = 0.096, d = 2.23). In the Monitor condition, F_uF_r pairs exchanged very few acknowledgements (M = 1.625, SD = 1.5). By contrast, in the No Monitor condition, the number of acknowledgements for F_uF_r pairs quadrupled (M = 6.725, SD = 2.86), which translates to a difference of 2.23 standard deviations. Figure 5.10 illustrates the result by showing the interaction of User Gender by Robot Gender for each level of Monitoring.

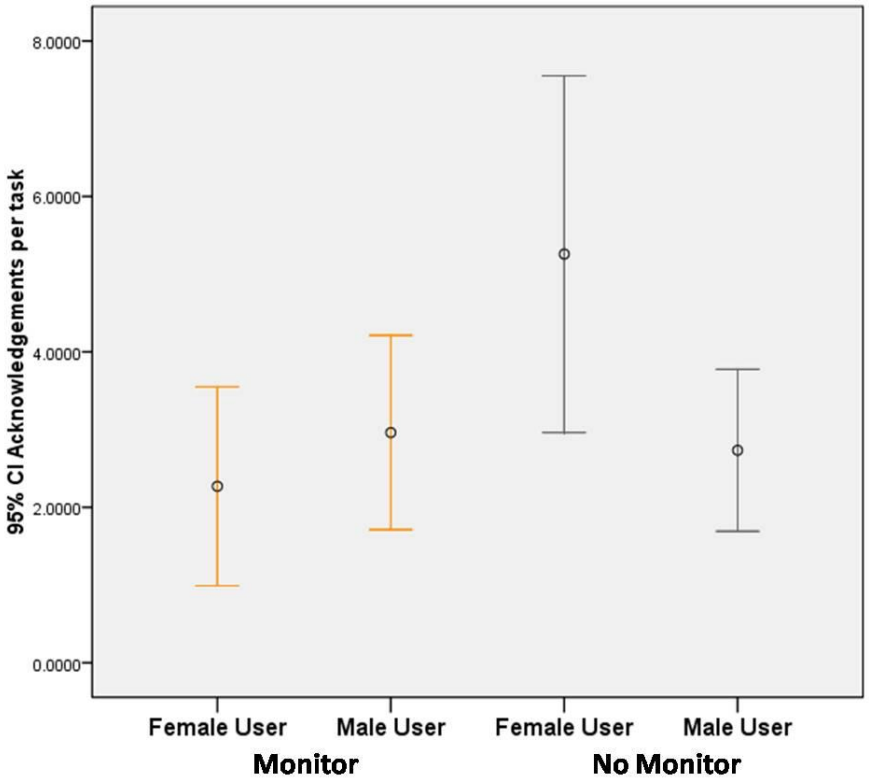


Figure 5.9: Acknowledgements given by pairs with female and male users in the Monitor and No Monitor conditions.

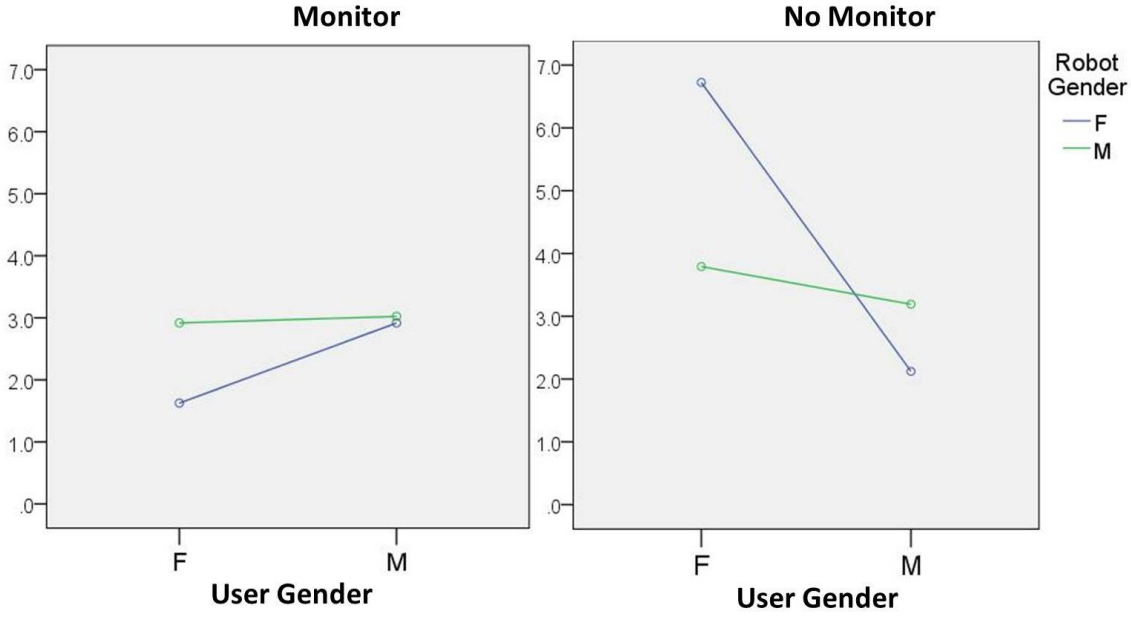


Figure 5.10: Plots of the interaction of User Gender and Robot Gender for each level of Monitoring. The Y axis represents the means of acknowledgements given by pairs of Female and Male users.

The analysis on the number of acknowledgements by the 'robot' also showed that when visual information is not shared, 'robots' provide evidence of positive understanding more frequently ($F_{(1,23)} = 9.629$, $p = 0.005$, $\eta^2 = 0.22$, $d = 1.04$).

5.6.3 Clarifications

The analysis on dialogue acts investigated the number of clarifications per task provided by the pairs. Inspection of the dialogue data showed that clarifications were provided exclusively by 'robots'. There was a significant effect of Monitoring ($F_{(1,24)} = 6.405$, $p = 0.018$, $\eta^2 = 0.173$, $d = 0.89$). In particular, 'robots' gave a higher number of replies that were richer in information, in the absence of shared visual space.

5.7 Utterance components

As detailed in Chapter 4 (section 4.4.5), the analysis of the corpus of user and 'robot' utterances followed the CORK framework (Vanetti and Allen, 1988). Communicative statements were classified as **Directives** or **Descriptives**. These communicative statements could contain references to environmental features (that is, landmarks). The types of environmental feature considered were: **Locations** (e.g., buildings or bridges), **Pathways** (e.g., streets), **Choice Points** (e.g., junctions) and the **Destination**. Last, utterances can be composed of delimiters, which fall into four categories:

1. **Distance designations**: e.g., '...*until* you see a car park'.
2. **Direction designations**: e.g., 'go *left*'.
3. **Relational terms**: e.g., 'on *your* left'.
4. **Modifiers**: e.g., '*big red* bridge', 'take the *first/second/last* road'.

This section presents the results of the analysis on the frequencies of landmark references, types of delimiters, communicative statements, deictic and anaphoric expressions and simple/compound instructions. It begins by providing the composition profile of the route instruction corpus and, then, juxtaposes the dialogue data of users and 'robots' with regards to each of these components.

5.7.1 The route instruction corpus

The corpus of utterances contained 1,676 route instructions. According to the CORK framework, instructions can either be directive or descriptive communicative statements that contain elements like references to environmental features and delimiters. Component analysis of the instructions revealed that users generally gave directive statements, and only 8% of their instructions were descriptive statements (Figure 5.11).

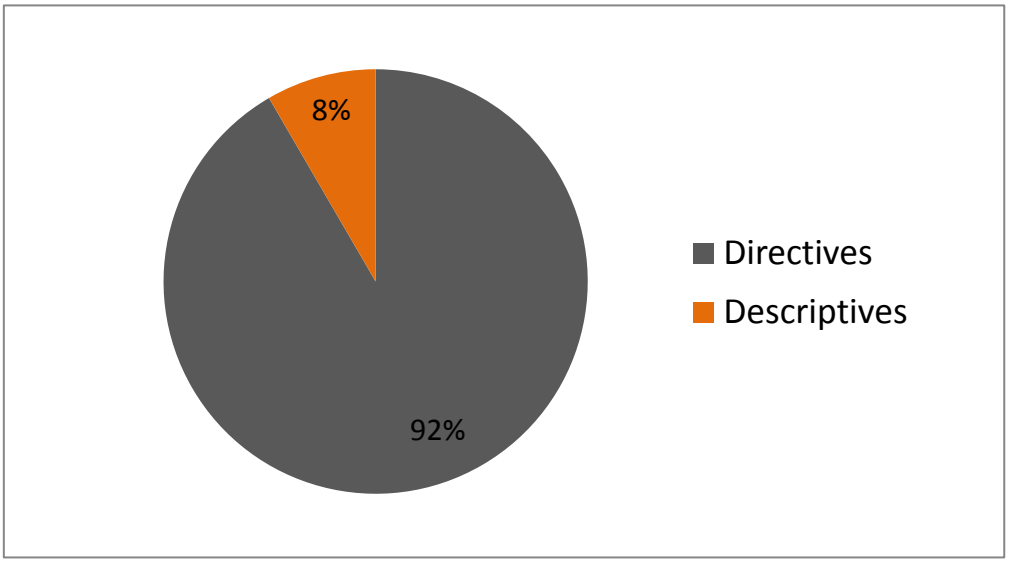


Figure 5.11: Distribution of instruction types in the corpus.

As seen in Figure 5.12, the majority of instructions lacked any type of reference to environmental features. On the other hand, 47% of instructions contained landmark references. In particular, users employed references to locations in 19.8% of their instructions. 5% of instructions included pathway references. Choice points were found in

11.4% of instructions. Finally, destination references were encountered in 10.9% of instructions.

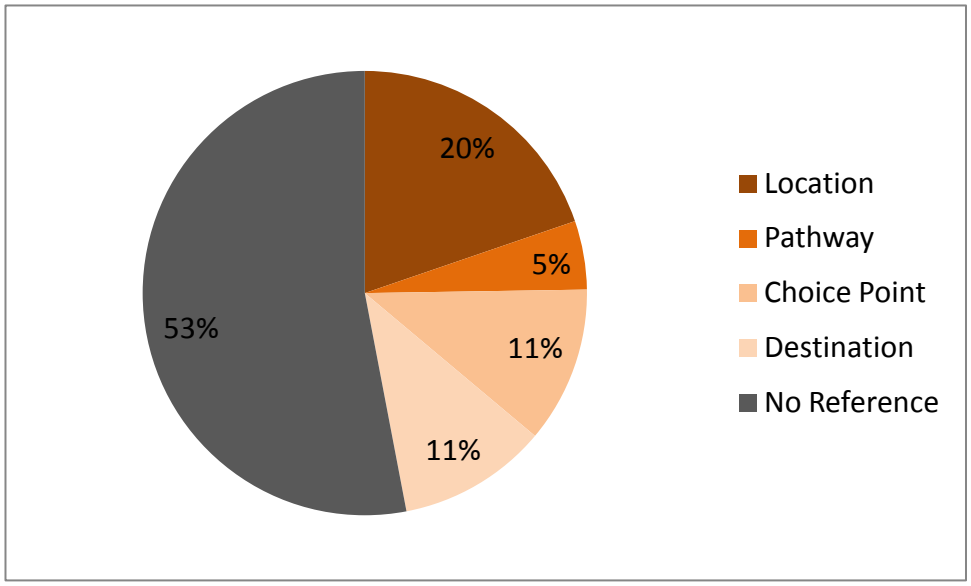


Figure 5.12: Proportion of route instructions with different types of landmark references and no references.

The following graphs presented in Figure 5.13 show how the different types of components were incorporated in directive and descriptive instructions. References to destinations were generally found inside descriptive statements and often accompanied by relational terms (category 3 delimiters) that state the frame of reference, as in the utterance 'you will see the lab on your right'. Such utterances were exclusively reserved for final instructions. Directive instructions mainly included a basic directional term (79%), for instance, 'forward' and 'left' (category 2 delimiters). Location references were commonly incorporated in directives (49%). Directives also incorporated a larger number of references to junctions and crossroads than references to pathways.

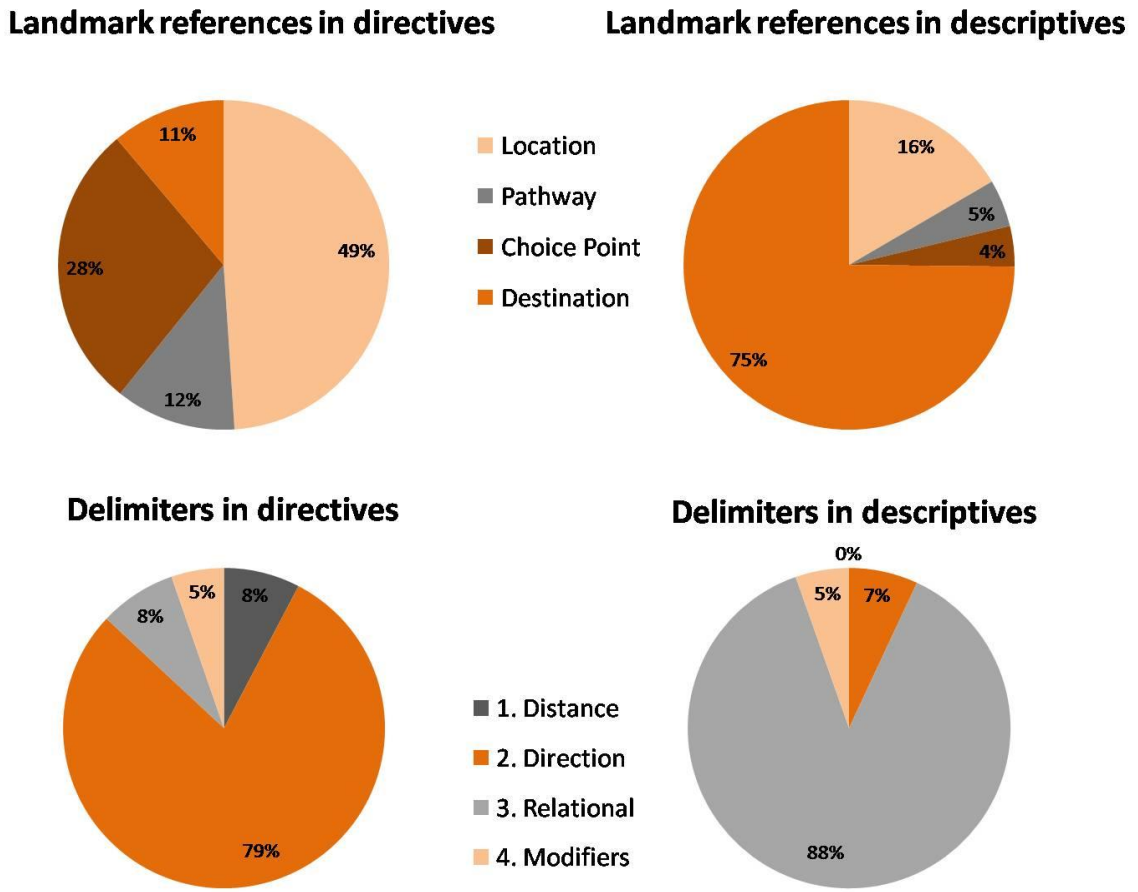


Figure 5.13: Configuration of directive and descriptive instructions in terms of landmark references and delimiters.

5.7.2 Landmark references in user utterances

The three-way ANOVA on number of landmark references in user instructions revealed a significant main effect of Robot Gender ($F_{(1,24)} = 6.454$, $p = 0.018$, $\eta^2 = 0.177$, $d = 0.9$). In particular, the findings suggested that when addressing females, users employed a larger number of landmark references. The distribution of landmark references in route instructions in the Monitor and No Monitor conditions is shown in Figure 5.14.

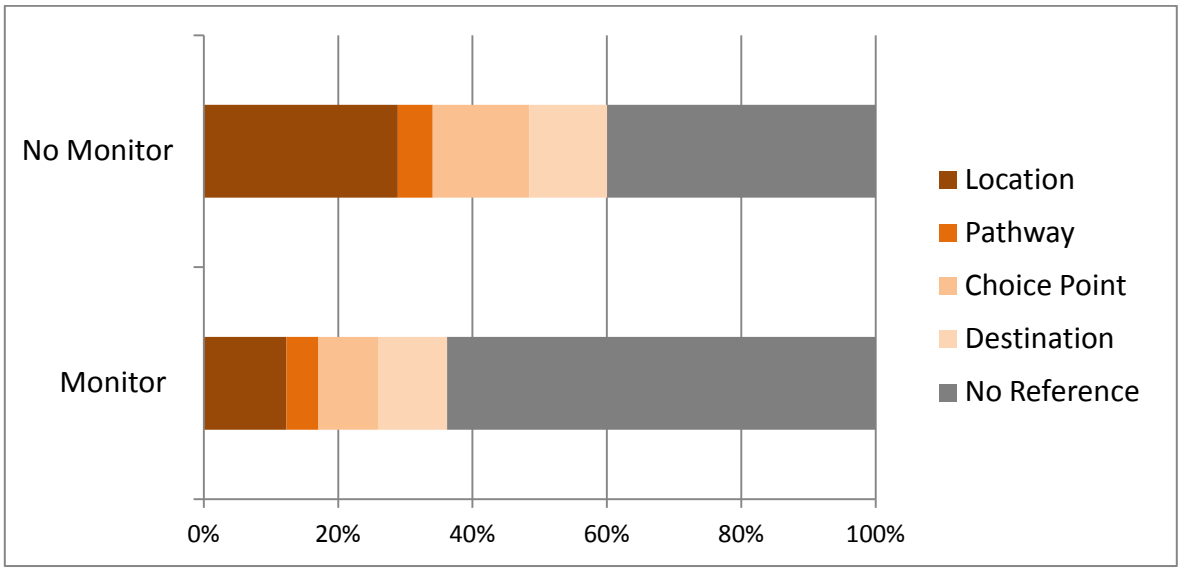


Figure 5.14: Inclusion of landmark references in instructions in the Monitor and No Monitor conditions.

Users also provided landmark references in replies to queries and clarification requests by their partner. Therefore, the analysis considered the number of landmark references in all user utterances. It showed that landmark references are most prevalent in the No Monitor condition ($F_{(1,24)}$ = 6.512, p = 0.017, η^2 = 0.163, d = 0.85). The initial finding that female 'robots' received a larger number of landmark references was also replicated ($F_{(1,24)}$ = 6.063, p = 0.021, η^2 = 0.151, d = 0.82).

Since the different types of landmark references (that is, references to locations, choice points, pathways and the destination) vary in function and information value, the analysis investigated the effects of each type in isolation. When location references were considered separately, the differences were accentuated; the analysis showed that users incorporated a larger number of location references in their instructions under the No Monitor condition ($F_{(1,24)}$ = 10.236, p = 0.004, η^2 = 0.233, d = 1.07). The proportion of instructions with location references increased from 12.3% in the Monitor condition to 28.9% in the No Monitor condition. The result is graphically presented in Figure 5.14 above. A main effect of Robot Gender was also observed, suggesting that a higher number of location references was provided to female 'robots' ($F_{(1,24)}$ = 7.469, p = 0.012, η^2 = 0.17, d = 0.88). Moreover, the analysis on all user utterances reiterated that location references were most frequently used in the No Monitor condition ($F_{(1,23)}$ = 10.893, p = 0.003, η^2 = 0.24, d = 1.11) and when the addressees were female ($F_{(1,23)}$ = 4.286, p = 0.05, η^2 = 0.094, d = 0.63). In addition, the

analysis revealed an interaction effect of Monitoring and User Gender ($F_{(1,23)}$ = 4.598, p = 0.043, η^2 = 0.101). The interaction was explored through t-tests which showed differences of almost two standard deviations between female users under the Monitor and No Monitor conditions ($t_{(14)}$ = -3.424, p = 0.004, d = 1.71). In particular, female users included three times as many location references when the 'robot's' actions were not visible. The findings are illustrated in Figure 5.15.

Analysis was also performed on normalised data (number of user location references over number of user words). The ANOVA confirmed the effect of Monitoring ($F_{(1,24)}$ = 6.754, p = 0.016) and the interaction effect of Monitoring × User Gender ($F_{(1,24)}$ = 4.262, p = 0.050). As such, it is ensured that the differences did not emerge because users *generally* produced a larger number of words. For completeness, the results of the analysis using normalised data are reported, where appropriate. However, the explanation that the number of words increased simply because participants needed to use more landmark references (and delimiters) appears more likely than the opposite, namely, that landmark references increased because participants needed to use more words. Thus, 'task' continues to be the common measure used in the data analysis detailed in this chapter.

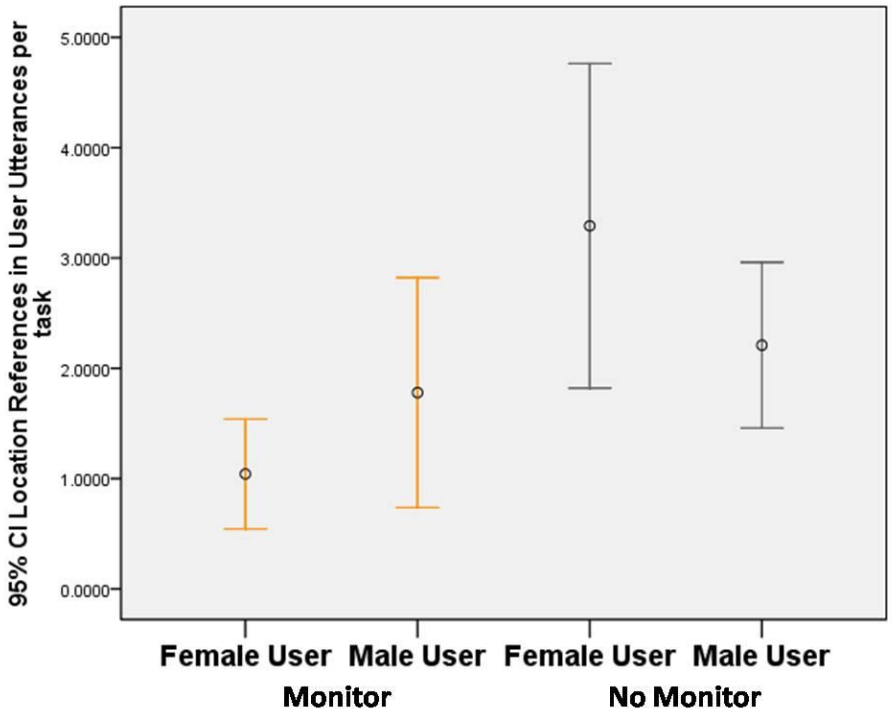


Figure 5.15: Location references in utterances by female and male users in the Monitor and No Monitor conditions.

There was a main effect of Robot Gender for number of references to pathways ($F_{(1,24)}$ = 4.274, p = 0.05, η^2 = 0.143, d = 0.79). In particular, female 'robots' received a greater number of references to pathways than male 'robots'. No effect of Monitoring was detected for number of references to choice points and pathways. This suggests that in the absence of shared visual space, users resort to location references, since these hold higher information value compared to two-dimensional landmarks, such as junctions and roads (Denis et al., 1999).

Finally, the analysis on frequency of destination references in user instructions found no reliable differences. This is because users commonly stated the destination in the beginning of the task, as in 'Your destination is the pub.', which is not an instruction. Thus, the three-way ANOVA on destination references in all user utterances produced a reliable interaction effect of Monitor by User Gender, echoing the previous finding with regards to location references ($F_{(1,24)}$ = 5.579, p = 0.027, η^2 = 0.174). The interaction was examined using t-tests which confirmed that female users stated the destination twice as frequently when visual information was not shared, compared to females in the Monitor condition ($t_{(13)}$ = 3.198, p = 0.007, d = 1.64). The differences are shown in Figure 5.16. This finding was also supported by the analysis using normalised data ($F_{(1,24)}$ = 4.100, p = 0.005).

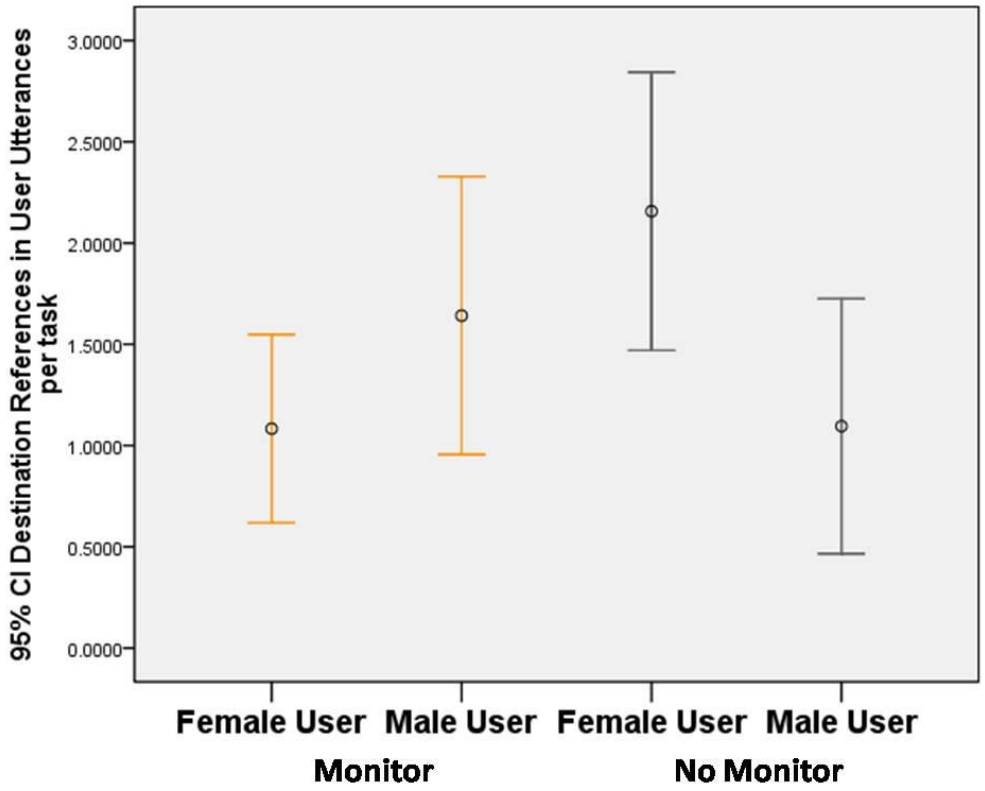


Figure 5.16: References to destinations in utterances by female and male users in the Monitor and No Monitor conditions.

5.7.3 Landmark references in 'robot' utterances

Figure 5.17 below shows the overall distribution of reference types in 'robot' utterances. Mirroring their partners' behaviour (discussed in the previous section), 'robots' in the No Monitor condition were also found to use more landmark references ($F_{(1,24)}$ = 4.892, p = 0.037, η^2 = 0.096, d = 0.64). In the Monitor condition, 65% of the 'robot' utterances contained landmark references, which climbed up to 82.5% when visual information was withheld.

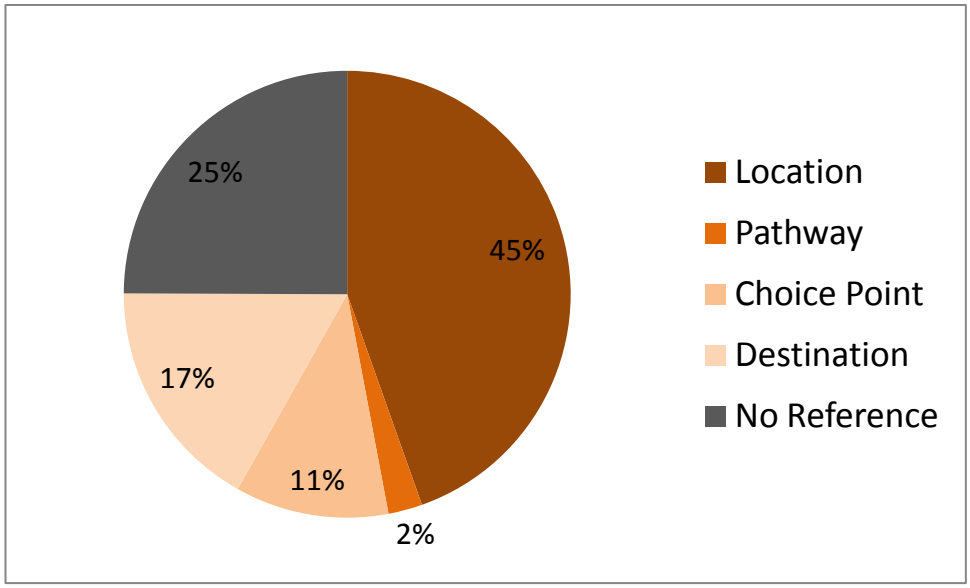


Figure 5.17: Proportion of 'robot' utterances (queries and statements) with references to different types of landmarks and without reference.

The increase was pronounced for location references (contained in 33% of 'robot' utterances in the Monitor condition compared to 53% of utterances in the No Monitor condition), which parallels their partner's preference for location references in the No Monitor condition ($F_{(1,24)}$ = 8.818, p = 0.007, η^2 = 0.247, d = 0.97). The User and Robot Gender factors were not significant.

The number of references to destinations was also significantly higher in the No Monitor condition ($F_{(1,23)}$ = 8.145, p = 0.009, η^2 = 0.222, d = 0.99). The observed differences are illustrated in the figure below (Figure 5.18). No differences were found with regards to references to choice points, while pathway references were quite rare (2% of the 'robot' utterances as shown in Figure 5.17 above).

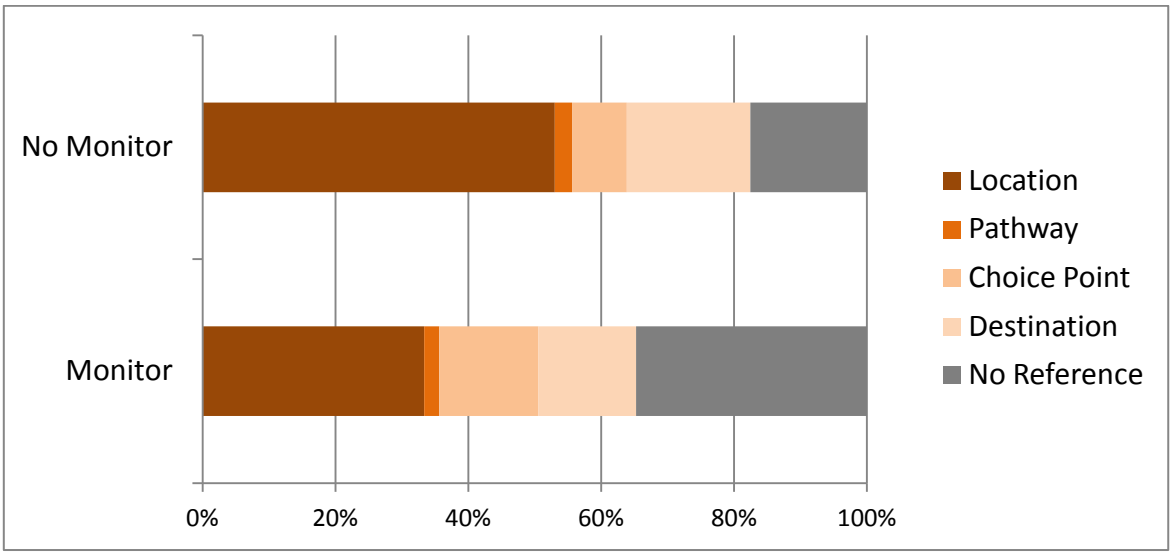


Figure 5.18: Inclusion of landmark references in 'robot' utterances in the Monitor and No Monitor conditions.

Finally, there was a significant relationship between the number of location references in the utterances produced by user and 'robots' (r = 0.616, p = 0.001). This finding shows that there is a matching tendency between partners in the use of location references, providing an initial indication of the presence of alignment. The relationship is illustrated in the scattergram in Figure 5.19.

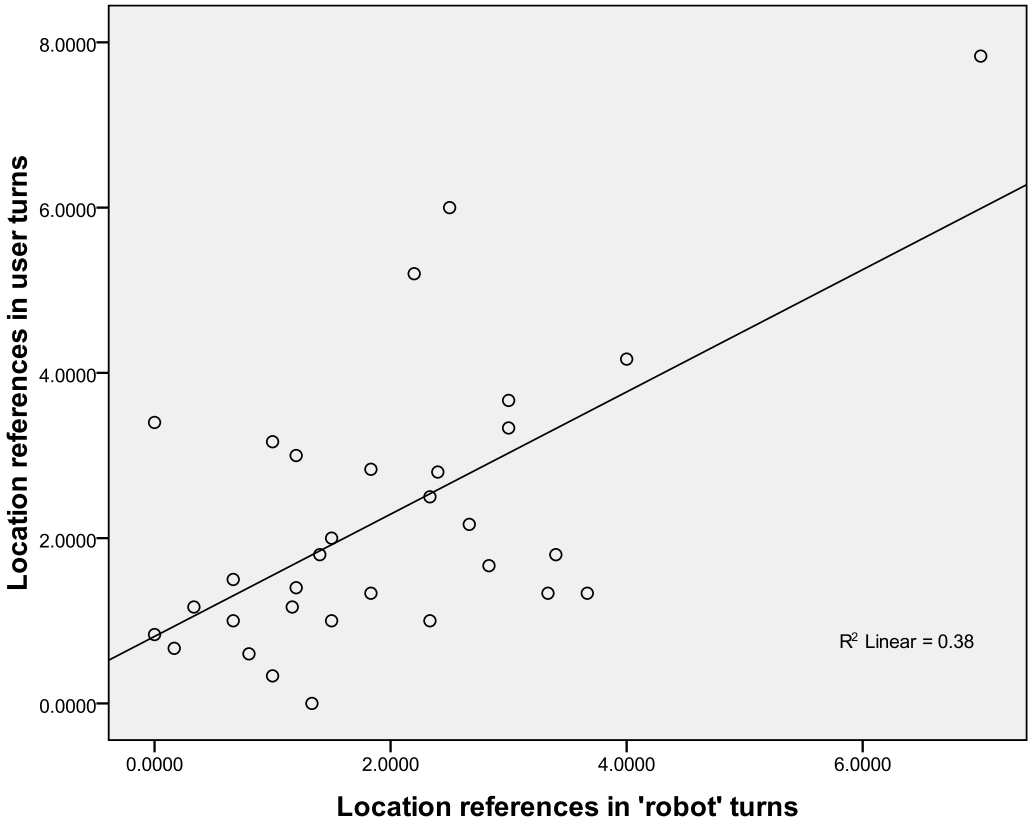


Figure 5.19: Relationship between number of location references per task in user and 'robot' turns.

5.7.4 Delimiters in user utterances

The analysis considered the frequency of delimiters[16] in route instructions (presented in Figure 5.20). The ANOVA revealed a significant main effect of Monitoring for distance designations (category 1 delimiters). These delimiters, which specify the boundary of the route, were scarcely used in the Monitor condition ($F_{(1,23)} = 4.539$, $p = 0.044$, $\eta^2 = 0.136$, $d = 0.77$). In addition, there was a marginal interaction effect of User Gender by Robot Gender ($F_{(1,23)} = 4.208$, $p = 0.052$, $\eta^2 = 0.126$). The comparisons showed that when users address

[16] Directional delimiters, such as 'left', 'right', 'forward', are the basic constituents of all route instructions; therefore, their frequency was not used as a dependent variable in the analysis.

'robots' of the same gender, they tend to clarify boundary and distance information more frequently than mixed-gender pairs ($t_{(28)}$ = 2.330, p = 0.027, d = 0.88). This finding is presented in Figure 5.21. Finally, the analysis performed on the frequencies of relational terms and category 4 delimiters (this category includes modifiers and ordering expressions) failed to yield any significant effect. It should be noted that that the occurrence of category 4 delimiters was generally low, which might be attributed to the particular navigation environment.

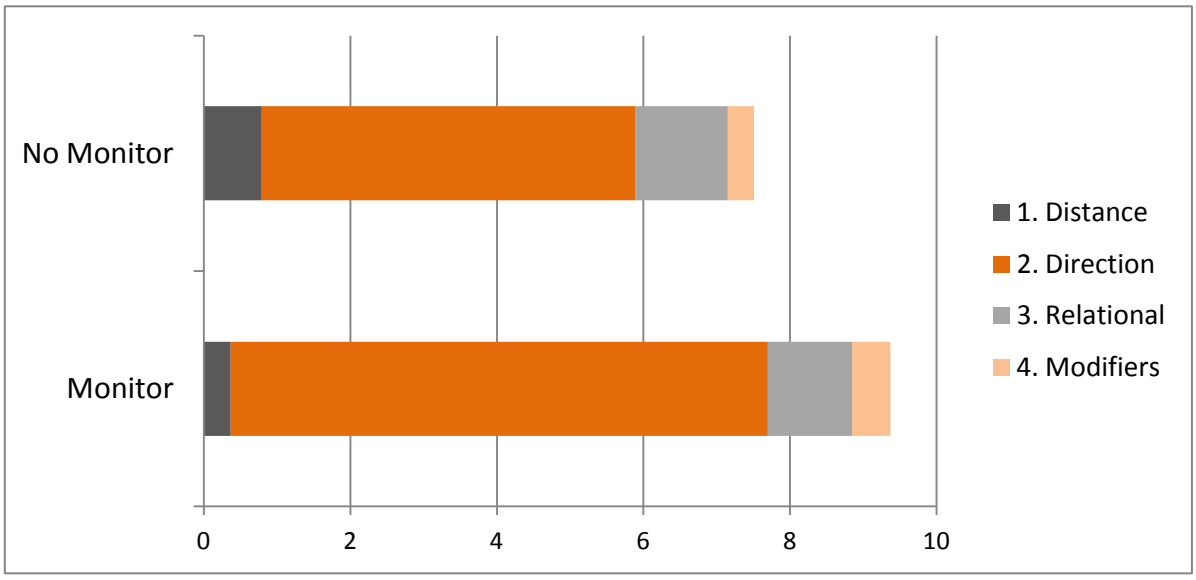


Figure 5.20: The frequencies of delimiters (per task) in user instructions in the Monitor and No Monitor condition.

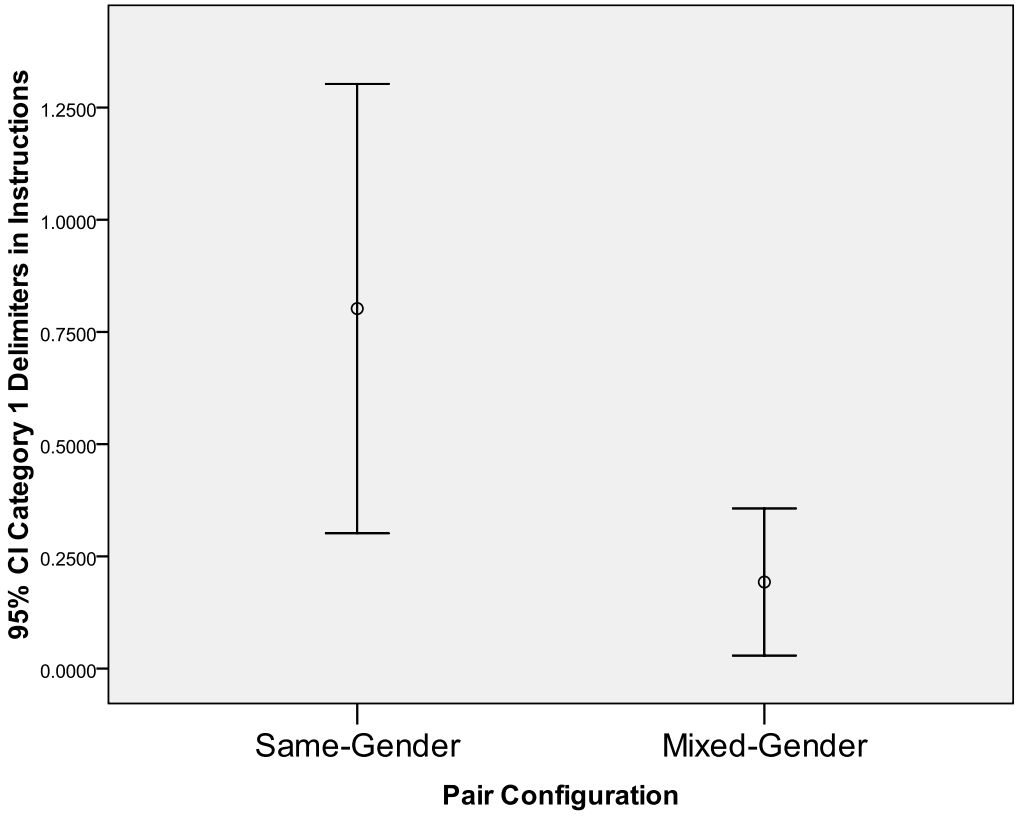


Figure 5.21: Use of distance designations (category 1 delimiters) for users in same-gender and mixed-gender pairs.

5.7.5 Delimiters in 'robot' utterances

The third category of delimiters includes terms that specify the relation between traveller and an environmental feature ('on your left') or between environmental features. 'Robots' were found to incorporate a larger number of these terms in their utterances in the No Monitor condition ($F_{(1,23)}$ = 5.332, p = 0.03, η^2 = 0.182, d = 0.90); that is, when not being monitored, 'robots' tended to be explicit about the frame of reference.

5.7.6 Directive and Descriptive Instructions

A series of chi-square tests were performed to assert whether the use of directive and descriptive instructions depended on User Gender, Robot Gender and Pair Configuration. The results revealed a significant relationship between User Gender and type of instruction (χ^2 =

3.940, *df* = 1, *p* = 0.047, φ = 0.071) in the No Monitor condition (shown in Table 5.7). Descriptive instructions give information about relations among features in the environment and tap perceptual experience (for instance, 'you will see a bridge') and were generally scarce in the corpus. Nevertheless, compared to female users, male users produced a larger number of descriptives. The odds ratio indicates that female users were 1.65 more likely to form their instruction as directive statements. In the Monitor data, no significant association was found.

Table 5.8: User Gender × Instruction type crosstabulation for the No Monitor data.

User Gender * Instruction Type Crosstabulation

User Gender			InstructionType		Total
			Directive	Descriptive	
	Female	Count	392	31	423
		Expected Count	384.0	39.0	423.0
	Male	Count	317	41	358
		Expected Count	325.0	33.0	358.0
Total		Count	709	72	781
		Expected Count	709.0	72.0	781.0

5.7.7 Granularity of route instructions

As described in Chapter 4 (second part of section 4.4.5), the level of granularity of instructions was determined by their number of components (that is, delimiters and landmarks). In particular, instructions that consisted of one or two components, verb or verb and direction, were considered simple (for instance, 'walk straight ahead'). Instructions with more than two components were compound (for instance, 'walk straight ahead until you reach the road junction', which has four components).

The three-way ANOVA revealed that the frequency of simple instructions that contained only the verb and the direction of movement was lower in the No Monitor condition ($F_{(1,24)}$ = 4.769, *p* = 0.039, η^2 = 0.144, *d* = 0.77). Figure 5.22 illustrates the differences between the conditions. A three-way interaction effect of Monitoring by User Gender by Robot Gender

was also detected, which initially suggested that the difference is more pronounced for same-gender pairs between the Monitor and No Monitor conditions ($F_{(1,24)} = 4.381$, $p = 0.047$, $\eta^2 = 0.126$). The difference is statistically significant only for F_uF_r pairs between conditions, showing that simple instructions by female users paired with females dramatically decreased in the No Monitor condition. The interaction is plotted for each level of Monitoring in Figures 5.23 below.

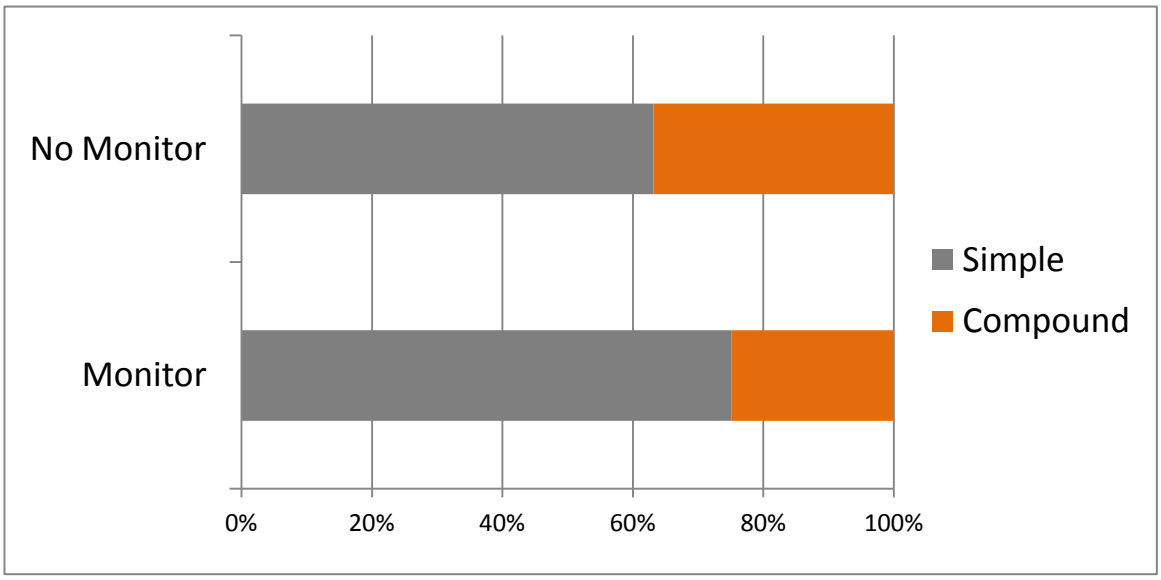


Figure 5.22: Proportion of simple and compound instructions in Monitor and No Monitor conditions.

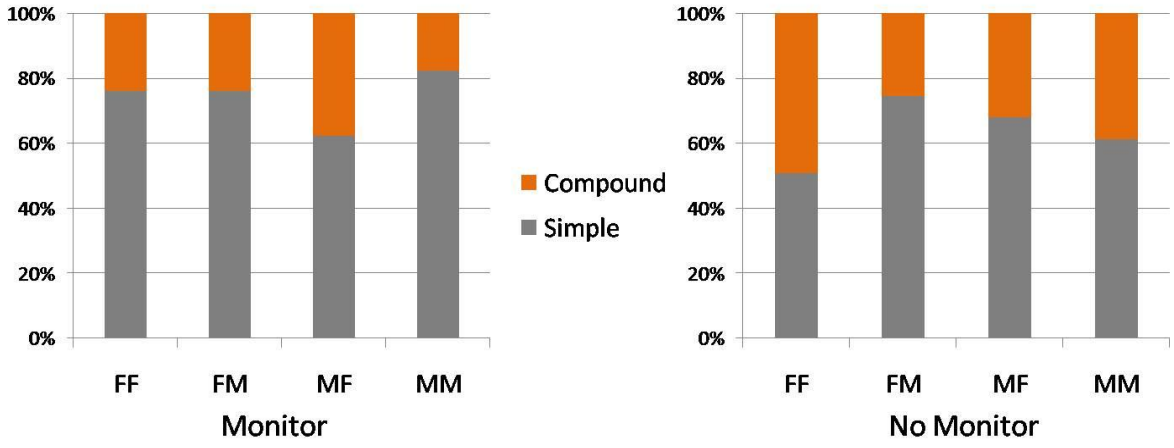


Figure 5.23: The proportion of simple and compound instructions by users in all pair configurations in Monitor and No Monitor conditions.

Finally, there was a marginally significant effect of Robot Gender on the use of compound instructions ($F_{(1,24)}$ = 3.865, p = 0.06, η^2 = 0.123, d = 0.69). Users appeared to provide instructions of higher granularity to female 'robots'. Further investigation is needed to assess the validity of the result.

Chi-square tests were used to detect relationships between the production of simple and compound instructions and User Gender, Robot Gender and Pair Configuration. Inspection of the contingency tables showed an association between granularity and Robot Gender for both Monitor (χ^2 = 10.314, df = 1, p = 0.001, φ = -0.106) and No Monitor data (χ^2 = 6.800, df = 1, p = 0.009, φ = -0.095) (see Tables 5.8 and 5.9). Confirming the tentative ANOVA outcome above, these results show that users are more likely to provide detailed and explicit instructions to female 'robots'. The odds ratios are 1.64 and 1.49 for the Monitor and No Monitor data, respectively. By contrast, male 'robots' mostly received simple, directional instructions.

Table 5.9: Robot Gender × Granularity of instructions crosstabulation for the Monitor data.

Robot Gender * Granularity Crosstabulation

			Granularity		
			Simple	Compound	Total
Robot Gender	Female	Count	336	141	477
		Expected Count	357.1	119.9	477.0
	Male	Count	352	90	442
		Expected Count	330.9	111.1	442.0
Total		Count	688	231	919
		Expected Count	688.0	231.0	919.0

Table 5.10: Robot Gender × Granularity of instructions crosstabulation for the No Monitor data.

Robot Gender * Granularity Crosstabulation

			Granularity		
			Simple	Compound	Total
Robot Gender	Female	Count	243	171	414
		Expected Count	260.1	153.9	414.0
	Male	Count	227	107	334
		Expected Count	209.9	124.1	334.0
Total		Count	470	278	748
		Expected Count	470.0	278.0	748.0

5.7.8 Deictic and anaphoric pronouns

The analysis of the dialogue data in this corpus failed to confirm the expectation that participants sharing visual space make extensive use of deictic and anaphoric pronouns. In fact, the use of these elements was very rare: spatial deictic terms, temporal deictic terms and anaphoric references accounted for only 1.2%, 1.5% and 0.4% of lexical items, respectively. Therefore, due to the low occurrence of these elements, it is not possible to infer that visual information had an effect on the use of deixis and anaphora. At the same time, this may imply that users are less likely to opt for underspecified deictic and anaphoric expressions to interact with computer systems compared to human communication.

5.8 Linguistic alignment

The analysis on lexical alignment essentially investigated whether speakers use the same words as their partner. The degree of alignment achieved by the dyads was assessed on the adjacency pair level (as 'match' scores between user and 'robot' utterances) and by measuring lexical innovation. Lexical innovation was determined as the ratio of unique words in the final task and as rate of new words introduced over time. This section reports the results of the analysis targeting each of the research questions with regards to alignment.

5.8.1 Alignment as lexical innovation

To address Research Question 7(a), evidence of alignment between user and 'robot' was sought. The rate of lexical innovation was determined by the number of new words introduced as the dialogue progressed. Figure 5.24 shows the number of new words against the utterance number (averaged for all pairs). The graph demonstrates a decrease of innovation over time and shows that the vocabulary utilised by the participants becomes relatively stable after approximately 70 turns. This finding fits the basic predictions by the Interactive Alignment Model which suggests that participants will come to rely on previously used expressions as dialogues progress. Addressing Research Question 7(a), the decrease in the rate of lexical innovation that occurs early in the dialogue hints at a rapid development of alignment.

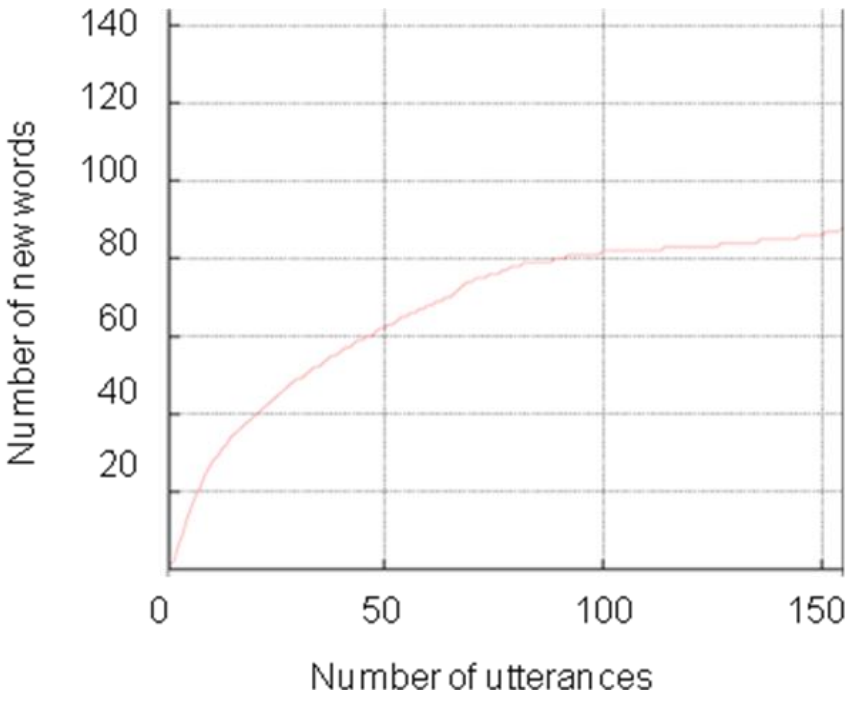


Figure 5.24: Lexical innovation over time

Lexical innovation was also measured by the ratio of unique words produced in undertaking the final task of the session. Not surprisingly, there was a significant negative correlation between match scores ('local' alignment) for users and 'robots' and the ratio of unique words in the final task (r = -0.529, n = 32, p = 0.002). That is, 'robots' and users that were aligning

to each other on the adjacency pair level were also more likely to conclude the dialogue with a more concise vocabulary. This finding also serves to validate the fitness of lexical innovation as a measure of alignment.

5.8.2 Alignment as 'matches' between user and 'robot' responses

The analysis in relation to lexical innovation pointed to the existence of alignment. Additional evidence was required to determine whether both interlocutors coordinate their lexical choices, and therefore whether, as Research Question 7(b) asked, alignment is a mutual phenomenon.

In addition to lexical innovation, alignment was measured 'locally', at the adjacency pair level. That is, a turn was a 'match', if it contained the same lexical item as the turn it responded to. If no component was matching, the turn was a 'mismatch' and the score 0 was given. Correlational analysis showed that user match scores and 'robot' match scores were positively and strongly related (r = 0.824, p = 0.001). The computation of r-squared[17] indicated that 68% of the variability in the user match scores could be directly predicted by the variability in 'robot' match scores. Therefore, as the 'robot' match scores increased the user match scores were also very likely to increase. This finding provides evidence that alignment is not merely present but also mutual and conditional: if one speaker uses aligned responses, their partners are more likely to do so at similar rates. The scattergram in Figure 5.25 illustrates that the data points are reasonably well-distributed along the regression line, in a linear relationship with no outliers. Similarly, there is a positive correlation between the mismatch scores of users and 'robots', with the mismatch scores of users rising when the mismatch scores for 'robots' rise (r = 0.419, p = 0.017).

[17] R-squared (r^2) belongs to the r-family of measures of effect magnitude (see Section 5.2) and provides the percentage of the variation in a variable is attributable to the variation in the other variable.

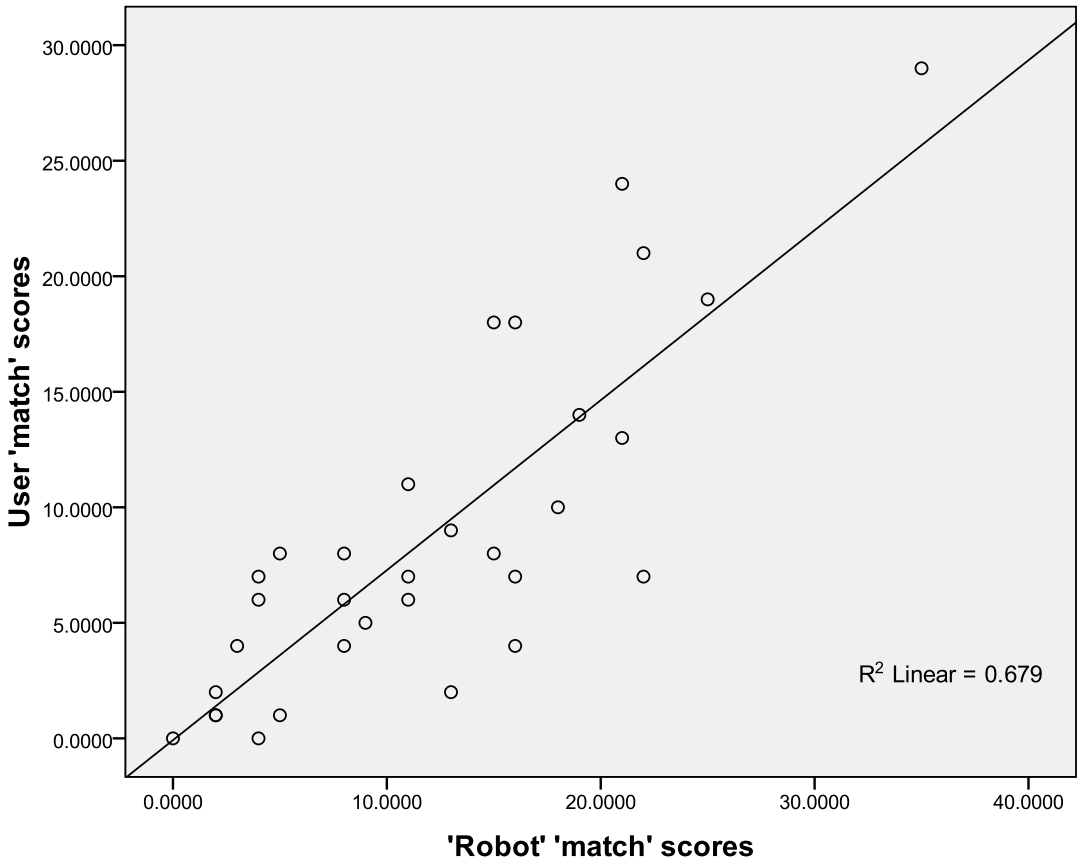


Figure 5.25: Scattergram showing the relationship between 'match' scores by users and 'robots'.

5.8.3 The effect of monitoring on alignment

Relevant to Research Question 7(c), the analysis sought to discover whether the levels of alignment varied with the absence of visual feedback.

A three-way factorial ANOVA performed on the match scores per task as dependent variable confirmed significant differences between the Monitor and No Monitor groups ($F_{(1,22)} = 9.354$, $p = 0.006$, $\eta^2 = 0.263$, $d = 1.17$). The match scores of pairs were significantly higher in the No Monitor condition (M = 4.333, SD = 1.784) compared to the Monitor condition (M = 2.14, SD = 1.953). User Gender and Robot Gender failed to yield significant effects. This result suggests that in the absence of visual feedback participants relied more heavily on alignment as a mechanism/strategy to ensure dialogue success.

In particular, the match scores by the user tripled in the No Monitor condition (M = 2.033, SD = 1.214) compared to the Monitor condition (M = 0.636, SD = 0.516), ($F_{(1,22)}$ = 12.885, p = 0.002, η^2 = 0.365, d = 1.5). Similarly, the match scores of the 'robot' responses were higher in the No Monitor condition (M = 2.696, SD = 1.251) than in the Monitor condition (M = 1.471, SD = 1.337), ($F_{(1,24)}$ = 6.507, p = 0.018, η^2 = 0.192, d = 0.94). This demonstrates that it is not the scores of *one* of the participants that account for the observation; rather, both 'robots' and users aligned more when visual information was not available.

The results in section 5.4.2 confirmed that when no visual feedback was available, more words were exchanged. Thus, it may have been the case that there were more matches in the No Monitor condition simply because there were more words in the dialogues. To eliminate this as a possible explanation of the observed alignment, the ratio of number of 'matches' to the total number of words per task was used to compare the Monitor and No Monitor conditions. This analysis reiterated the previous results: alignment was considerably higher in the No Monitor than in the Monitor condition ($F_{(1,24)}$ = 4.970, p = 0.035, η^2 = 0.187, d = 0.83).

Finally, the analysis considered lexical innovation in the final task to assess alignment. The three-way factorial ANOVA revealed reliable differences between the Monitor and No Monitor conditions ($F_{(1,24)}$ = 8.424, p = 0.008, η^2 = 0.217, d = 1) as well as an interaction effect of User Gender by Robot Gender (which is discussed in section 5.8.7). In particular, in the Monitor condition, the final task contained 21.1% new words (SD = 0.049), which dropped to 17.1% in the No Monitor condition (SD = 0.027). This finding provides further evidence that alignment is higher when users do not have access to visual information.

5.8.4 Miscommunication and alignment

This subsection presents the analysis related to Research Question 7(d); the effect of miscommunication on alignment was explored through lexical innovation.

First, lexical innovation in the final task was considered using the measure of the ratio of unique words. The analysis revealed that there was a positive relationship between the number of incorrect instructions and the ratio of new words, suggesting that pairs concluded

the dialogue being less aligned when more incorrect instructions had been given (r = 0.405, n = 32, p = 0.021).

As a result, a chi-square analysis was performed to clarify the link between lexical innovation and miscommunication. This analysis considered the number of new words contained in an utterance immediately after a i) non-problematic and ii) problematic utterance (that is, a dialogue turn marked as a non-understanding, an incorrect instruction or in which an execution error occurred; a combined measure was used since the nature and cause of miscommunication was not the focus of this analysis). All utterances were grouped based on whether or not they contained new words and whether or not they followed a problematic utterance.

The chi-square test is used for categorical (that is, nominal and ordinal) data. In reality, however, the standard Pearson's chi-square test does not take into account ordinal information, which could lead to elevated p values. There are several methods to address this issue. In the approach discussed in Howell (2009, p.306) and Agresti (2007), Pearson's correlation (r) is used to calculate the chi-square instead of the standard Pearson's chi-square statistic. In the present analysis, it is the case that the 'number of new words' category represents an ordered variable[18]. The analysis will report both linear and standard Pearson's chi-square for completeness. Finally, the statistical significant results will be refined through odds ratios and the phi coefficient.

A chi-square test was performed on the contingency table below (see Table 5.10) showed that an association exists between the number of new words in an utterance and the occurrence of miscommunication (χ^2 = 18.522, df = 1, p = 0.001). The linear-by-linear association (calculated using Pearson's r, as explained above) confirmed the result (M^2= 18.518, p = 0.001) and the phi coefficient was equal to 0.068[19]. The odds ratio was 1.78,

[18] The Miscommunication variable is not, of course, ordinal, but dichotomous variables can be treated as ordinal with no effect on the analysis (Howell, 2009, p.309).

[19] As explained in 4.1, the phi coefficient depends on the equivalence of the marginal totals. As can be seen from the Total column in the table (see Table 5.10), the marginal totals are 3744 and 233 which dramatically decreases the φ value.

indicating that the odds of novel words being used were 1.78 times higher after miscommunication than after a non-problematic utterance. Figure 5.26 illustrates the probability of new words being introduced after miscommunication and problem-free communication [20]. In order to provide a comprehensive account of the data, Figure 5.27 shows the probabilities for the occurrence of 0, 1, 2, 3, and 4 or more words.

Strictly speaking, the use of Pearson's chi-square analysis or Pearson's *r* on this data is incorrect, since the utterances were not independent from each other, being produced by the same thirty-two pairs. However, it appears that it is the most popular choice for researcher conducting analysis of similar experimental data. Acknowledging that the common Pearson's chi-square test is inappropriate, the analysis also included two alternative types of chi-squares; first, the Cochran–Mantel–Haenszel (CMH) test has been proposed as a method to strengthen the reliability of the chi-square (Cochran, 1954). This test allows to control for one variable (in this case, pair), while comparing the levels of the other two variables (in this case, (i) problematic/non-problematic and (ii) no new words/1 or more new words). The results of the CMH test confirmed the previous findings (χ^2 = 17.772, *df* = 1, *p* = 0.001). Second, statisticians also strongly advise the use of McNemar's test, which is a form of chi-square for within-subjects design. The association was also found significant at *p* = 0.001.

Table 5.11: Number of utterances containing zero and one or more new words after preceding problematic and non-problematic utterances.

Miscommunication * Number of new words in the next turn Crosstabulation

			Number of new words in the next turn		
			No new words	1 or more new words	Total
After a:	Non-problematic utterance	Count	2478	1266	3744
		Expected Count	2447.7	1296.3	3744.0
	Problematic utterance	Count	122	111	233

[20] The probability is calculated by the number of turns with 0 and 1 or more new words after a (problematic or non-problematic) turn, divided by the total number of (problematic or non-problematic) turns.

	Expected Count	152.3	80.7	233.0
Total	Count	2600	1377	3977
	Expected Count	2600.0	1377.0	3977.0

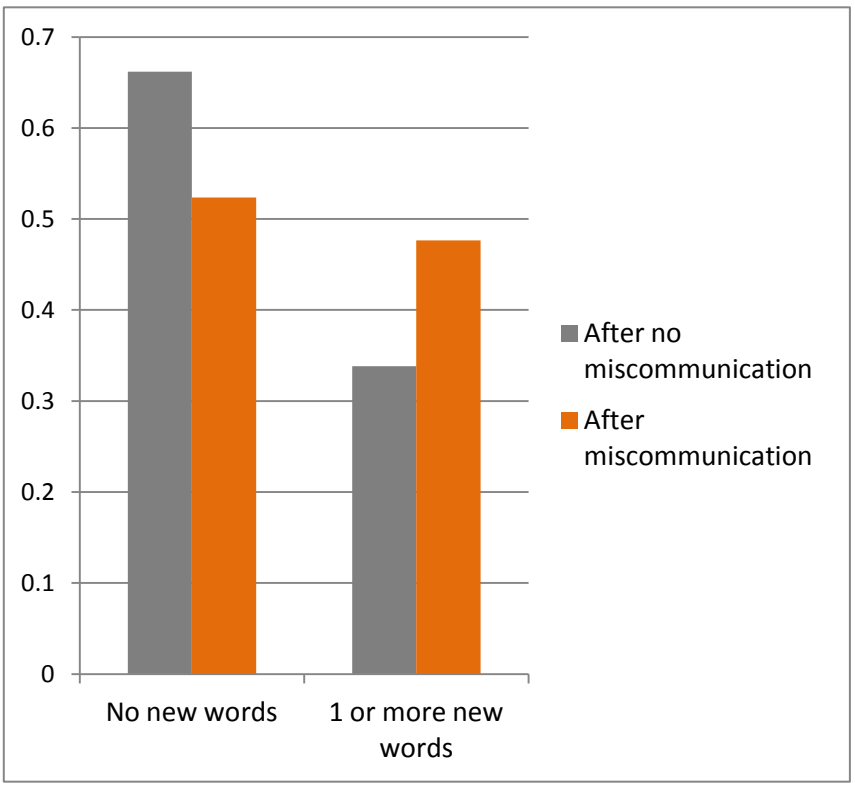


Figure 5.26: Probability of occurrence of new words after problematic and non-problematic utterances. Probabilities are calculated as the ratio of actual count over total number of utterances.

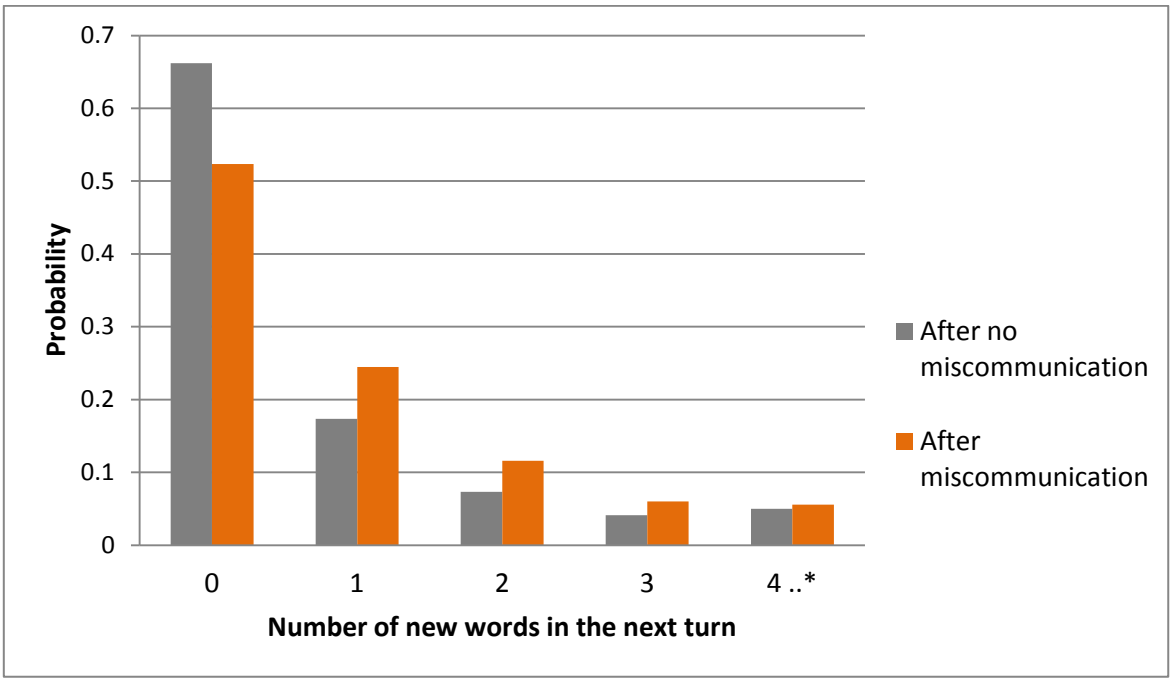


Figure 5.27: Probability of occurrence of new words (0, 1, 2, 3 and 4 or more) after problematic and non-problematic utterances.

The effect of monitoring on miscommunication and alignment

The findings presented in sections 5.4, 5.6 and 5.7 harmonise with previous literature indicating that visual feedback has a major effect on task performance and communication. Moreover, the results in section 5.8.3 suggested that alignment increased (high 'match' scores and low lexical innovation) when users did not have visual access to the 'robot's' actions. So far, this section has shown that novel vocabulary is more likely to be input by the user when s/he detects miscommunication, whereas in problem-free communication, vocabulary from the preceding dialogue is reiterated. Therefore, it was necessary to tease apart the effect of visual information, and refine our observations on how miscommunication shapes the development of alignment.

Again, chi-square analysis was carried out (see Table 5.11) to discover whether there was a significant relationship between the three variables: number of new words in an utterance (0 or 1 to many), type of previous utterance (non-problematic or problematic) and visual information (Monitor or No Monitor condition). The resulting test indicated a significant association between occurrence of miscommunication and lexical innovation, but only in the

No Monitor condition (χ^2 = 15.711, df = 1, p = 0.001), and was confirmed by the linear (M^2 = 15.704) and McNemar's chi-square (significant at p = 0.001). As discussed above, the CMH method was applied to compensate for the flouted assumption of independence between utterances. The test also verified the association (χ^2 = 14.670, df = 1, p = 0.001). Under both conditions, only around 34% of the utterances contained new words when communication was smooth. However, when a problem occurred, this figure climbed to 54% in the No Monitor condition. The odds ratio indicated that, if visual information was withheld, new words were 2.33 times more likely to be introduced after miscommunication. The graph in Figure 5.28 illustrates that the probability of introducing new words is elevated after miscommunication, whereas it is most likely that users draw their vocabulary from the preceding dialogue in cases where the communication is problem-free. Figure 5.29 provides the utterance data broken down in more than two classes.

The number of utterances with new words also rose, to 44%, in the Monitor condition, but failed to yield a significant result.

Table 5.12: Number of utterances with no and one or more new words after problematic and non-problematic utterance in the Monitor and No Monitor conditions.

Miscommunication * Number of new words in the next turn * Monitoring Crosstabulation

After:			Number of new words in the next turn		
			No new words	1 or more new words	Total
Monitor	Non-problematic utterance	Count	1114	598	1712
		Expected Count	1101.7	610.3	1712.0
	Problematic utterance	Count	81	64	145
		Expected Count	93.3	51.7	145.0
No Monitor	Non-problematic utterance	Count	1365	672	2037
		Expected Count	1347.8	689.2	2037.0
	Problematic utterance	Count	41	47	88
		Expected Count	58.2	29.8	88.0

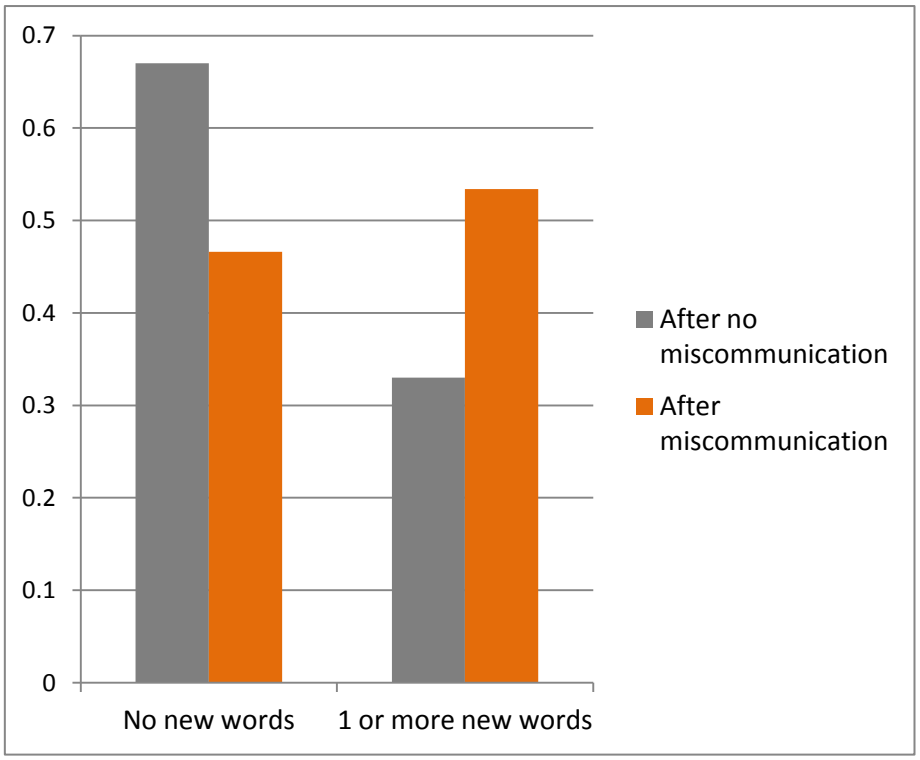


Figure 5.28: Probability of occurrence of new words after problematic and non-problematic utterances in the No Monitor condition.

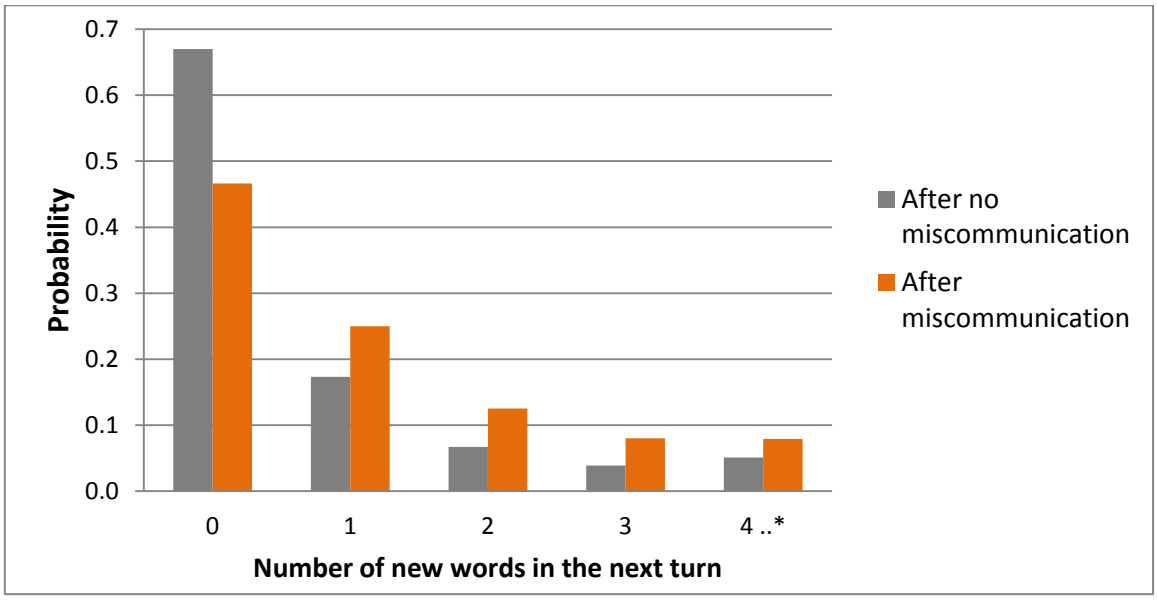


Figure 5.29: Probability of 0 to any number of new words (0, 1, 2, 3, and 4 or more) after problematic and non-problematic utterances in the No Monitor condition.

5.8.5 Alignment and user perceptions of interaction success

Research Question 7(e) looked at user perceived interaction success. As already revealed in section 5.5, the analysis found a significant negative correlation between user experience of task success ("*I did well in completing the task*") and lexical innovation (as ratio of unique words in the final task) (r = - 0.473, p = 0.013). This may indicate that users perceived that the interaction was less successful when alignment was weaker. The analysis failed to reveal significant relationships between the other statements.

5.8.6 Gender and alignment

To address Research Questions 8(a) – 8(c), the analysis investigated whether alignment is mediated by gender; that is, whether User Gender and Robot Gender have an effect on alignment as (i) lexical innovation and (ii) match and mismatch scores. The ratio of unique words produced in the final task was used to determine the degree of alignment ultimately achieved; namely, the lower the ratio of new words, the higher the level of final alignment.

As mentioned in section 5.8.3, the three-way factorial ANOVA on lexical innovation revealed a main effect of the Monitoring factor ($F_{(1,24)}$ = 8.424, p = 0.008, η^2 = 0.217, d = 1) as well as an interaction effect of User Gender by Robot Gender ($F_{(1,24)}$ = 4.431, p = 0.046, η^2 = 0.117). In particular, in the Monitor condition, the final task contained 21.1% new words (SD = 0.049), which dropped to 17.1% in the No Monitor condition (SD = 0.027).

Investigation of the User Gender × Robot Gender interaction effect through t-tests revealed that the ratio of unique words by the end of the dialogue is lower for same-gender pairs compared to mixed-gender pairs[21]. The findings are illustrated by the error bar graphs

[21] All t-tests were found significant. But the Levene's test showed heterogeneous variance.

The two assumptions of the t-test are homogeneity of variance and that the data are drawn from a normally distributed population. The t-test, however, is a robust test. Howell (2009, p.215) cites three studies that explored the effects of violating these assumptions. They found that for equal sample sizes, violating the assumption of homogeneity of variance has very small effects (±0.02 from the true value of α). As Howell (2009) notes, this level of inaccuracy is acceptable.

below (see Figure 5.31). In particular, F_uF_r and M_uM_r groups concluded the dialogue with 15.3% and 16.6% unique words (SD = 0.007 and SD = 0.056), respectively. By contrast, the ratios of unique words for the F_uM_r and M_uF_r groups were 19.5% and 21.6% (SD = 0.035 and SD = 0.011), respectively. To reinforce this finding, the contrast of mixed-gender pairs (F_uM_r and M_uF_r) and same-gender pairs (M_uM_r and F_uF_r) produced a clearly significant effect ($t_{(28)}$ = -3.404, p = 0.004, d = 1.17).

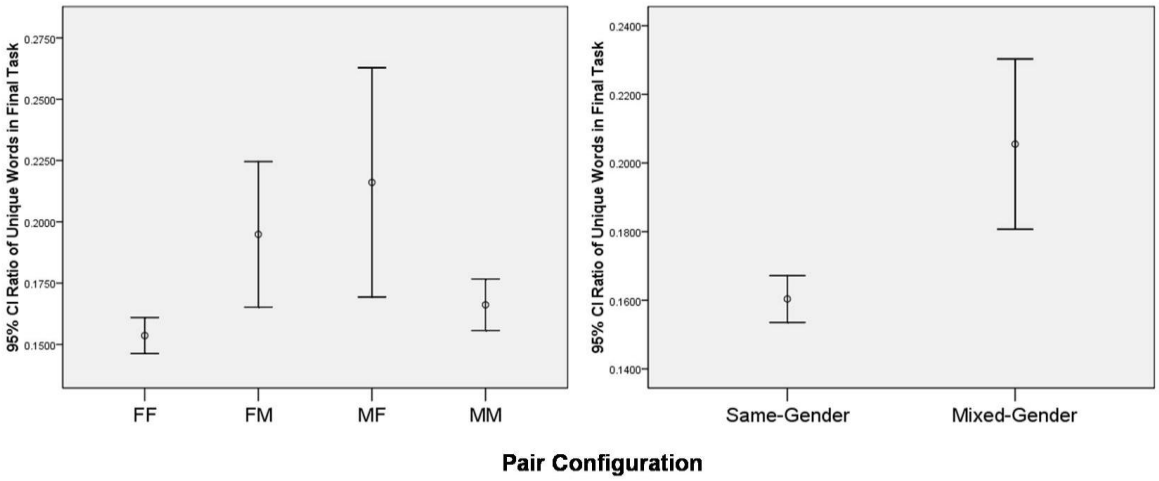


Figure 5.30: The ratio of unique words in all pair configurations (left-hand side graph) and same-gender and mixed-gender pairs (right-hand side graph).

Next, the research questions were further examined through chi-square analysis which looked at the association between the gender of user and 'robot' and the tendency to match or not the previous utterance of their partner.

The data was cross-tabulated with respect to whether a response was a match or a mismatch (see Table 5.12). As explained in section 5.2, chi-square tests were performed with User Gender, Robot Gender or Pair Configuration (that is, F_uF_r, F_uM_r, M_uF_r, M_uM_r) as the second classification variable. Subsequently, the analysis controlled for the Monitoring factor and chi-square tests were performed for the Monitor and the No Monitor data separately.

Preliminary analysis using the match and mismatch scores by both participants showed that female users gave more non-matching responses than male users (χ^2 = 12.380, df = 1, p = 0.001, Odds Ratio = 1.2) in the Monitor data. The next step of the analysis with Pair Configuration as the classification variable clarified the results and showed that pairs with male users and 'robots' produced a higher number of matched responses compared to F_uM_r

pairs (χ^2 = 17.565, *df* = 3, *p* = 0.001, Cramer's *V* = 0.2[22]). In order to make sure that the F_uM_r and M_uM_r data alone contributed to the effect, they were deleted from the table and chi-square analysis was performed again. As expected, the remaining data did not produce a significant effect. This result seems to resonate (at least, partly) with the findings presented previously in this section, which indicated higher alignment between same-gender pairs. The bar chart in Figure 5.32 helps illustrate the observed differences between M_uM_r and F_uM_r pairs. However, no differences were found between pairs in the No Monitor data, as evidenced in the right-hand side graph in Figure 5.32. This could link to the finding reported in section 5.8.3, which indicated that when visual information was not shared, participants generally resorted to alignment. Finally, the separate tests on the match/mismatch scores by the user (χ^2 = 4.155, *df* = 3, *p* = 0.042, φ = -0.158) and 'robot' (χ^2 = 15.034, *df* = 3, *p* = 0.002, φ = 0.234) produced similar results for the Monitor condition.

Table 5.13: Pair configuration × Matching crosstabulation for the Monitor data.

Pair configuration		Matching		
		Match	Mismatch	Total
F_uF_r	Count	44	59	103
	Expected Count	47.9	55.1	103.0
F_uM_r	Count	32	68	100
	Expected Count	46.5	53.5	100.0
M_uF_r	Count	57	60	117
	Expected Count	54.4	62.6	117.0
M_uM_r	Count	71	48	119
	Expected Count	55.3	63.7	119.0
Total	Count	204	235	439
	Expected Count	204.0	235.0	439.0

[22] Cramer's *V* is equivalent to the phi coefficient for larger than 2 × 2 tables and its interpretation is similar.

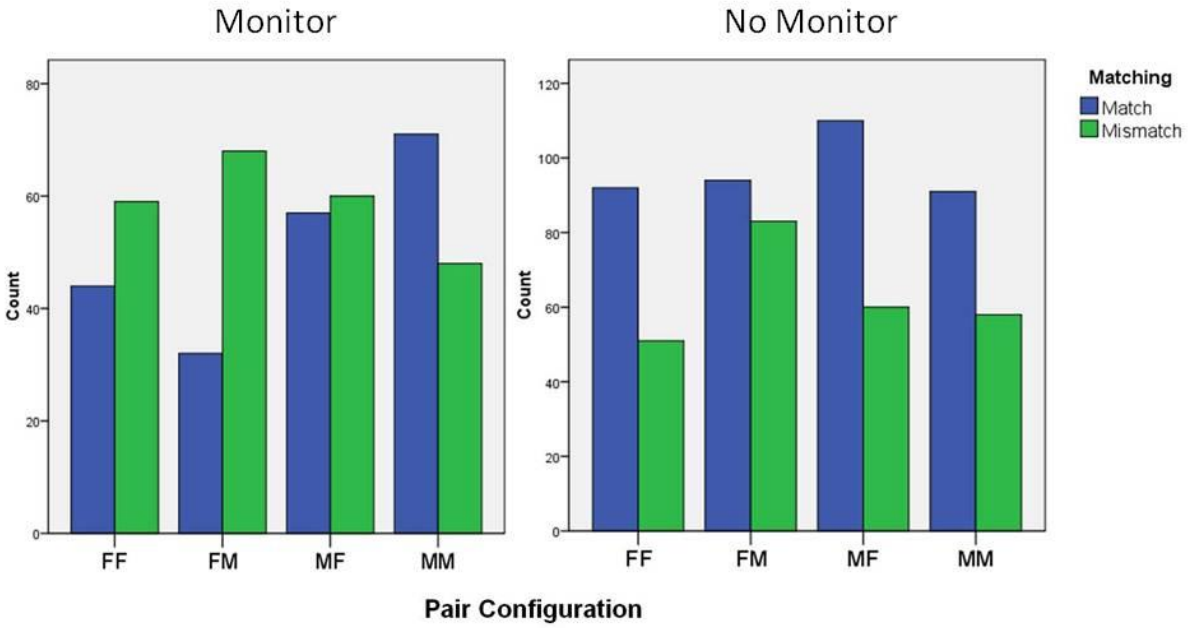


Figure 5.31: The graphs show the frequencies of matching and non-matching responses for F_uF_r, F_uM_r, M_uF_r and M_uM_r pairs in the Monitor and No Monitor conditions.

5.8.7 The effect of miscommunication on gender and alignment

The findings from section 5.8.4 led to the argument that when miscommunication occurs, users tend to introduce new words in the dialogue. The question that naturally follows is whether this behaviour depends on gender (addressing Research Question 8(e)).

The variables corresponding to the question are the gender of the user, occurrence of miscommunication and the number of new words present in the next turn (that is, no new words or one or more new words). From the analysis in the previous section (section 5.8.4), the relationship between the variables was only found significant in the No Monitor condition. Thus, the No Monitor data were used on this analysis to explore the gender factor (included in Table 5.13 below).

Table 5.14: Number of utterances with no or one or more new words after problematic or non-problematic utterances crosstabulated by User Gender.

User Gender * Number of new words in the next turn * Miscommunication Crosstabulation

			Number of new words in the next turn		
			0 New Words	1 or more New Words	Total
After non-problematic utterance	Female	Count	764	327	1091
		Expected Count	734.2	356.8	1091.0
	Male	Count	600	336	936
		Expected Count	629.8	306.2	936.0
After problematic utterance	Female	Count	27	20	47
		Expected Count	21.9	25.1	47.0
	Male	Count	14	27	41
		Expected Count	19.1	21.9	41.0

A significant relationship was found between gender and the number of new words category after problematic and non-problematic utterances. After problem-free communication, both female and male users tended to re-use old vocabulary. 70% of utterances (for female users) and 64% (for males) contained only previously used words. The Pearson's chi-square on the relationship yielded χ^2 = 8.035, confirmed by the linear relationship of 8.031 (df = 1, p = 0.005, φ = 0.063). However, there was a difference between male and female users. The odds ratio of adhering to the old vocabulary was 1.3 higher for female users than for male users after a non-problematic utterance.

As suggested by the findings reported in section 5.8.4, new words are most likely to follow miscommunication. But the analysis revealed that female and male users employed different approaches when miscommunication is detected. Male users responded by introducing new words (66% of the utterances). On the other hand, female users appeared to continue adhering to the old vocabulary. The linear and Pearson's chi-square of 4.779 and 4.723, respectively, supported the existence of the relationship, statistically significant at p = 0.029 (the phi coefficient was 0.233, explaining 5.42% of the variance). The odds ratio climbed to 2.6, indicating that female users were now 2.6 times less likely to try new words after miscommunication compared to males. In brief, the analysis suggests that females preferred to re-use vocabulary, even after miscommunication, whereas males were more

inclined to introduce new words. The results are summarised in the graph in Figure 5.33 below.

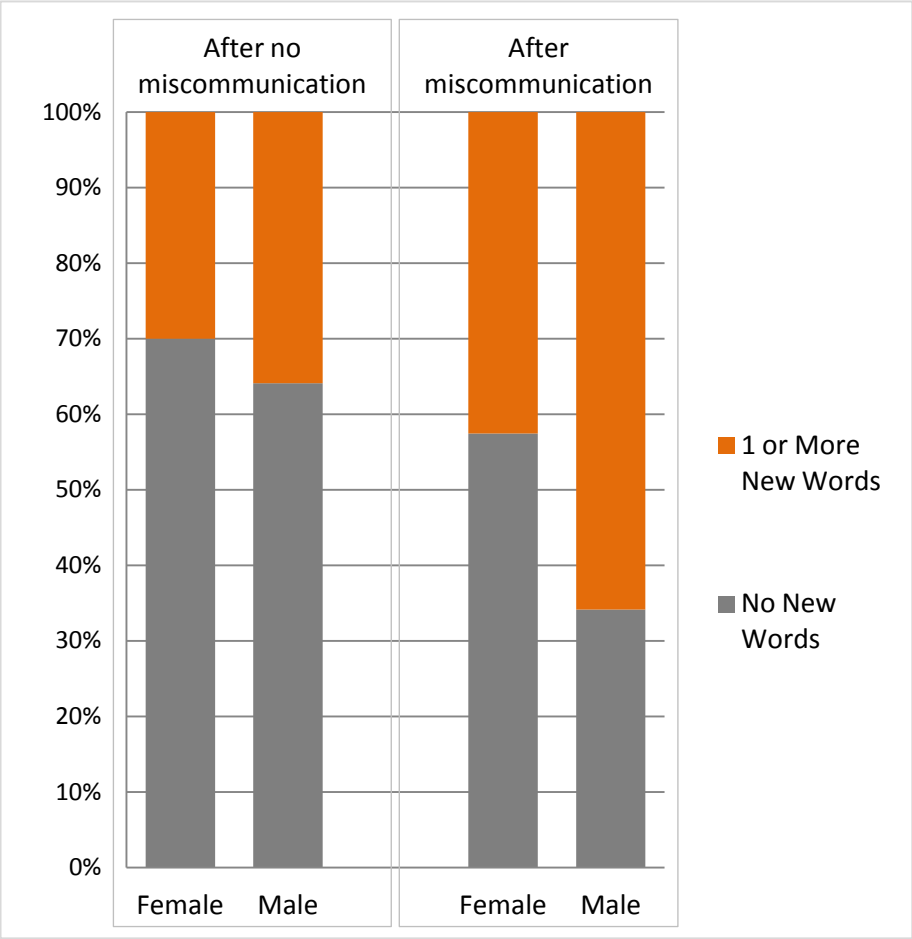


Figure 5.32: Probability of occurrence of new words after problematic and non-problematic utterances for pairs with female and male users in the No Monitor condition.

The analysis continued with CMH tests. It compared the levels of the two variables (miscommunication and new words), controlling for pairs with male and female users, separately. The results strengthened the previous interpretation. On one hand, it was verified that male users tend to produce considerably more utterances containing new words after miscommunication ($\chi^2 = 15.203$, $df = 1$, $p = 0.001$). On the other hand, the CMH test did not find a significant association between miscommunication and new words for pairs with female users. This finding appears to corroborate that females employ a conservative strategy of using previous vocabulary in *both* situations of problematic and problem-free communication.

5.9 Summary of results

The study was designed to assess the effect of gender on task performance and language use in dialogue through addressing 31 research questions. As anticipated by past research, visual information (i.e., the Monitoring factor) had a strong effect on several experimental variables. Notably, it was found that females were more reactive to the lack of visual information than males, in terms of dialogue strategies rather than performance. The results also provide insights into the local and global processes of alignment in a user's dialogue with a system. More importantly, the tendency and magnitude of alignment appear to correlate with the gender of the individual as well as the dyad. The following sections recapitulate the results of this study. To facilitate reference, three tables that sum up the results are provided in the end of the chapter. Table 5.14 lists the sub-questions and hypotheses and summarises the respective answers to them. Table 5.15 enumerates the significant main and interaction effects of Monitoring, User Gender and Robot Gender revealed by the chi-square analyses and ANOVAs. Table 5.16 provides the corresponding results of the statistical tests. Table 5.17 lists the significant positive correlations between the dependent variables. The right-hand columns on Tables 5.15, 5.16 and 5.17 give the number of the sections where the relevant results were presented.

5.9.1 The effect of shared visual information on navigation performance and communication between user and system

Although the Monitoring factor was primarily used as an experimental manipulation to clarify aspects of gender differences, interesting results emerged regarding the impact of visual information on performance and dialogue.

The analysis identified a robust effect of visual feedback on performance. In particular, when visual information was withheld, the number of words produced by both participants increased, and turn possession was balanced between interlocutors. On the other hand, users dominated the conversational floor when they could monitor the 'robot's' actions. Interestingly, the lack of visual feedback also brought a decline in miscommunication.

The analysis of linguistic content of utterances also illuminated significant differences between the Monitoring conditions. The lack of shared visual information led users to incorporate a larger number of references to three-dimensional landmarks in the environment and specify the boundary of movement, and purely directional route instructions were less frequent. In reciprocity, 'robots' under this condition regularly referred to locations, named the destinations, and clarified frame of reference. Measuring the frequency of dialogue acts indicated that users asked a large number of questions and 'robots' provided information-rich replies (clarifications), as well as explicitly stated that an instruction was understood and executed (acknowledgements). On the other hand, when visual information was available, 'robots' issued more queries. Finally, in the incidence of miscommunication, users in the No Monitor condition were more likely to introduce new expressions to the dialogue.

5.9.2 The operation and development of alignment in human-computer dialogue

The study addressed five questions that aimed to identify and categorise the phenomenon of alignment in the dialogue between a user and a computer system. First, the stabilisation of working vocabulary early in the interaction reveals the operation of alignment between speakers that settle on a set of grounded expressions for dealing with the ensuing dialogue. Second, the analysis of the experimental data confirmed that the magnitude of alignment is reciprocal, with interlocutors aligning to each other at similar rates. Third, analysis of data from two different visual co-presence conditions produced evidence that may also indicate that alignment in human-computer dialogues has a strategic component. That is, in the absence of visual evidence of understanding, correct execution and joint reference, speakers tended to adapt their linguistic choices more strongly, possibly in an effort to compensate for the lack of this resource and in an attempt to enhance (the impoverished) communication. Fourth, the development of alignment is locally disrupted by the occurrence of miscommunication such that novel words are introduced instead of falling back on previously used vocabulary. Users and 'robots' converged in shorter vocabularies when user errors were lower. Yet, while the lack of visual feedback promoted alignment, when miscommunication occurred users were considerably less likely to draw from the grounded expressions. Finally,

analysis of the user perception data revealed that users rated their performance less favourably when alignment was weaker.

5.9.3 The effect of user and 'robot' gender on performance, dialogue and alignment with and without visual information

The analysis outlined an intricate picture of differences in the communication behaviour of males and females and a complex pattern of results produced by the interactions of genders, roles, and visual information. Yet, the findings can be described under three common themes.

First, it was found that users of either gender produced wordier instructions and utterances when addressing female 'robots', which typically incorporated location and pathway references. Male 'robots', on the other hand, mostly received simple directional instructions.

Secondly, female participants and, specifically, female users appeared to be more susceptible to context and interaction conditions. Thus, the most pronounced contrasts were observed between female users, or all-female pairs in the Monitor and No Monitor conditions. In particular, female users dominated the dialogue, accounting for the majority of turns, when they could monitor the 'robot's' actions. But, when visual feedback was not available, female users gave way to their partners, and were less inclined to provide descriptive instructions compared to male users. Similarly, acknowledgements and user requests for location and status information were abundant between female partners only when visual information was not shared. Finally, utterances by female users in the Monitor condition mostly lacked location and destination references, whereas in the No Monitor condition, female users employed a large number of landmark references. Again, the locus of difference in the ratio of simple and compound instructions was identified between female users in the Monitor and No Monitor conditions. On the other hand, males maintained a consistent behaviour across conditions with regards to these variables.

The third theme brings together the results of alignment. Same-gender pairs appear to contrast with mixed-gender pairs in terms of lexical innovation rate and match scores. In particular, same-gender pairs achieved a significantly lower ratio of novel words compared to mixed-gender pairs. Moreover, male participants were more likely to match their partner's

utterances than F_uM_r pairs. Finally, users in same-gender pairs provided fewer erroneous instructions, which explicitly conveyed boundary of action information. As mentioned in section 5.9.1 above, users tended to introduce new words when miscommunication occurred, while utterances were generally composed of already used words in problem-free communication. This effect was found to be pronounced when users did not have access to visual feedback. Most interestingly, extending the chi-square analysis to include the gender factor showed that females and males responded differently to miscommunication. That is, males were most inclined to try new words with the 'robot', while females appeared to fall back to old vocabulary.

Finally, the analysis of the questionnaire items revealed two differences in user perceptions of task success attributed to gender. Female users rated their performance significantly lower than males, while rating the system's accuracy more favourably than males. In reality, it was not confirmed that male users performed better, since female users paired with male 'robots' were the fastest and, as mentioned above, the rate of miscommunication was comparable for female and male users.

These results will be reviewed in the next chapter, in relation to converging and conflicting evidence from previous literature. The following chapter will also discuss the implications of these results for the development of natural language interfaces to computer systems.

Table 5.15: List of research questions and respective high-level results.

	Research Question	Answer and High-level Result
	A. Gender differences in navigation performance, route communication and user perceptions in interaction	
1(a)	*Are all-male pairs (male users interacting with male 'robots) the fastest in completing the navigation task?*	No; F_uM_r pairs were the fastest, particularly compared to M_uF_r and F_uF_r pairs.
1(b)	*Are pairs with male 'robots' faster in completing the navigation task than pairs with female 'robots'?*	Yes; pairs with male 'robots' (F_uM_r and M_uM_r) were faster than pairs with female 'robots' (M_uF_r and F_uF_r) pairs.
1(c)	*Do all-male pairs produce the lowest miscommunication (execution errors, non-understandings and inaccurate instructions)?*	No; users in same-gender pairs (M_uM_r and F_uF_r) issued fewer incorrect instructions than mixed-gender pairs (M_uF_r and F_uM_r).
1(d)	*Do male 'robots' produce fewer execution errors and non-understandings than female 'robots'?*	Inconclusive; no significant differences were found between female and male 'robots'.
1(e)	*Do male users provide fewer inaccurate instructions than female users?*	No; users in same-gender pairs (M_uM_r and F_uF_r) issued less incorrect instructions than mixed-gender pairs (M_uF_r and F_uM_r).
2(a)	*Do female users include more landmark references in their utterances than male users?*	No; both female and male users employed more landmark references when addressing a female 'robot'.
2(b)	*Do female 'robots' include more landmark references in their utterances than male 'robots'?*	Inconclusive; no significant differences were found between female and male 'robots'.
3	*Do female users rate their performance lower than male users?*	Yes; females reported lower perceived task success than males, while rating the system more favourably.
	B(i). The effect of visual information on navigation performance and communication between user and system	
4(a)	*Are tasks completed more quickly when visual information is available?*	Inconclusive; no significant differences were found; task completion times were comparable across conditions.
4(b)	*Does the number of incorrect instructions by the user decrease when visual information is available?*	No; incorrect instructions increased when visual feedback on the actions of 'robots' was available.
4(c)	*Does the number of execution errors and non-understandings by the 'robot' decrease when visual information is available?*	No; non-understandings increased when users could monitor the actions of 'robots'.

4(d)	*Do interlocutors use fewer words, turns and route instructions to complete a task when visual information is available?*	Yes; 'robots' and users required fewer words to complete the task when visual feedback was available.
5(a)	*Does the use of deictic pronouns and expressions (for example, 'turn there' and 'take this turn') increase when visual information is available?*	Inconclusive; no significant differences were found; the frequency of deictic expressions was equally low across conditions
5(b)	*Does the number of verbal acknowledgements by the 'robot' decrease when visual information is available?*	Yes, 'robots' issued fewer acknowledgements when users could monitor their actions.
5(c)	*Are route instructions less detailed, precise and explicit when visual information is available?*	Yes; users generally omitted action boundary information and provided simple instructions (verb with or without direction of movement) when they could monitor the actions of 'robots'.
5(d)	*Are spatial descriptions by the 'robot' less detailed, precise and explicit when visual information is available?*	Yes; 'robots' regularly omitted stating their frame of reference when sharing their visual space with users.
5(e)	*Does the number of user-initiated queries decrease when visual information is available?*	Yes; users issued fewer queries when they could monitor 'robot' actions.
5(f)	*Does the number of user turns exceed 'robot' turns, when visual information is available?*	Yes; the proportion of utterances provided by users was higher when they could monitor 'robot' actions.
5(g)	*Do interlocutors use fewer landmark references to complete a task when visual information is available?*	Yes; the use of location and destination references by 'robots' and users dropped when sharing visual space.
	B(ii). Gender and the effect of visual information	
6(a)	*Is task performance of females more negatively affected by absence of visual information than males' performance?*	Inconclusive; no significant differences were found in task completion time, miscommunication rates and number of turns, words and instructions by both interlocutors.
6(b)	*Do females adapt their communication strategies more than males in response to lack of visual information?*	Yes; marked differences were found across conditions in terms of proportion of user turns and use of location and destination references, queries and acknowledgements of female users, while male users' behaviour remained consistent.
	C(i). Alignment in human-computer dialogue	
7(a)	*Does alignment occur in the interaction between a human user and a computer system?*	Yes; vocabulary stabilised early in the dialogue suggesting the operation of alignment.
7(b)	*If alignment does occur in this context, is it a mutual*	Yes; 'robots' and users aligned to each other at similar rates.

	phenomenon?	
7(c)	*Does visual information influence alignment between a user and a system?*	Yes; 'robots' and users aligned more strongly in the absence of visual information.
7(d)	*Does miscommunication locally disrupt the process of alignment in human-computer communication?*	Yes; the development of alignment is locally disrupted; new vocabulary was introduced after miscommunication.
7(e)	*Does lack of alignment compromise user perception of interaction success?*	Yes; lower task success perceptions were associated with higher final lexical innovation.
	C(ii). Gender-related alignment in task-oriented interaction	
8(a)	*Do female speakers align more strongly than male speakers in task-oriented interaction?*	No; refer to question 8(c) below.
8(b)	*Do female speakers align more strongly to male addressees than to female addressees in task-oriented interaction?*	No; refer to question 8(c) below.
8(c)	*Do speakers in same-gender pairs align more strongly than mixed-gender pairs in task-oriented interaction?*	Yes; alignment was stronger between same-gender pairs than mixed-gender pairs.
8(d)	*Do users in mixed-gender pairs moderate the use of their own gender-preferential strategies and provide instructions as preferred by their addressees (landmark-based and purely directional route instructions to females and males, respectively)?*	Yes; male users employed more landmark references when addressing a female 'robot'.
8(e)	*Does miscommunication have different effect on male and female users in terms of communication strategies?*	Yes; when visual information was withheld, males were more likely to introduce new words after miscommunication than did females.

Table 5.16: List of significant main and interaction effects of User Gender, Robot Gender and Monitoring. Lower case m and nm denote the Monitor and No Monitor conditions. Upper case F and M denote user and 'robot' gender. For instance, nm-F_uF_r stands for pairs with female user/female 'robot' in the No Monitor condition. The results for variables with an asterisk are derived from chi-square analysis. Parametric tests were performed on all other dependent variables.

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	Time			F > M			F_uF_r, M_uF_r > F_uM_r		5.4.1
	# Words	nm > m							5.4.2
	# User words	nm > m		F > M					5.4.2
	# 'Robot' words	nm > m							5.4.2
	% User turns	m > nm			m-F > nm-F				5.4.2
	# Incorrect instructions	m > nm					F_uM_r, M_uF_r > $F_UF_uF_r$, M_uM_r		5.4.3
	# Non-understandings	m > nm							5.4.3
	# Location references by User	nm > m		F > M	nm-F > m-F				5.7.2
	# Pathway references by User			F > M					5.7.2
	# Destination references by User				nm-F > m-F				5.7.2

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	# Location references by 'Robot'	nm > m							5.7.3
	# Destination references by 'Robot'	nm > m							5.7.3
	# Distance designations by User (Delimiter category 1)	nm > m					F_uF_r, M_uM_r > F_uM_r, M_uF_r		5.7.4
	# Relational terms by 'Robot' (Delimiter category 3)	nm > m							5.7.5
	Directive vs. Descriptive instructions *	nm only			nm-F > nm-M				5.7.6
	% Simple instructions (low granularity)	m > nm						m-F_uF_r > nm-F_uF_r	5.7.7
	Simple vs. Compound instructions *	m and nm		M > F					5.7.7
	# User Queries	nm > m			nm-F > m-F ..			nm-F_uF_r> m-F_uF_r	5.6.1
	# 'Robot' Queries	m > nm							5.6.1
	# Acknowledgements	nm > m			nm-F > m-F ..			nm-F_uF_r > m-F_uF_r	5.6.2
	# 'Robot' acknowledgements	nm > m							5.6.2
	# 'Robot' clarifications	nm > m							5.6.3
	# 'Matches'	nm > m							5.8.3
	# User 'Matches'	nm > m							5.8.3
	# 'Robot' 'Matches'	nm > m							5.8.3

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	'Matches' vs. 'Mismatches' *	m only						m-M_uM_r > m- F_uM_r	5.8.7
	User 'Matches' vs. User 'Mismatches' *	m only						m- M_uM_r > m- F_uM_r	5.8.7
	Robot' 'Matches' vs. 'Robot' 'Mismatches' *	m only						m- M_uM_r > m- F_uM_r	5.8.7
	% Unique words in final task	m > nm					F_uM_r, M_uF_r > F_uF_r, M_uM_r		5.8.3 & 5.8.7
	0 new words/ 1 or many new words in next utterance after no miscommunication *	nm only			nm-F > nm-M				5.8.4 & 5.8.8
	0 new words/ 1 or many new words in next utterance after miscommunication *	nm only			nm-F > nm-M				5.8.4 & 5.8.8
	Perceived task success		M > F						5.5
	System accuracy		F > M						5.5

Table 5.17: The results of the ANOVA or chi-square analysis (*p, F* and χ^2 values) for all significant main and interaction effects. The table does not include the results of post-hoc tests and effect sizes, but these can be found in the text.

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	Time			F(1,23) = 9.208, p = 0.006			F(1,24) = 6.197, p = 0.02		5.4.1
	# Words	F(1,24) = 6.904, p = 0.015							5.4.2
	# User words	F(1,24) = 4.138, p = 0.053		F(1,24) = 4.393, p = 0.047					5.4.2
	# 'Robot' words	F(1,22) = 6.654, p = 0.017							5.4.2
	% User turns	F(1,23) = 5.490, p = 0.028			F(1,23) = 5.548, p = 0.027				5.4.2
	# Incorrect instructions	F(1,23) = 13.784, p = 0.001					F(1,23) = 4.797, p = 0.039,		5.4.3
	# Non-understandings	F(1,24) = 4.324, p =							5.4.3

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
		0.048							
	# Location references by User	F(1,23) = 10.893, p = 0.003		F(1,23) = 4.286, p = 0.05	F(1,23) = 4.598, p = 0.043				5.7.2
	# Pathway references by User			F(1,24) = 4.274, p = 0.05					5.7.2
	# Destination references by User				F(1,24) = 5.579, p = 0.027				5.7.2
	# Location references by 'Robot'	F(1,24) = 8.818, p = 0.007							5.7.3
	# Destination references by 'Robot'	F(1,23) = 8.145, p = 0.009							5.7.3
	# Distance designations by User (Delimiter category 1)	F(1,23) = 4.539, p = 0.044					F(1,23) = 4.208, p = 0.052		5.7.4
	# Relational terms by 'Robot' (Delimiter category 3)	F(1,23) = 5.332, p = 0.03							5.7.5

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	Directive vs. Descriptive instructions *	nm only			$\chi 2$ = 3.940, df = 1, p = 0.047				5.7.6
	% Simple instructions (low granularity)	F(1,24) = 4.769, p = 0.039						F(1,24) = 4.381, p = 0.047	5.7.7
	Simple vs. Compound instructions *	m and nm		m: $\chi 2$ = 10.314, df = 1, p = 0.001 nm: $\chi 2$ = 6.800, df = 1, p = 0.009					5.7.7
	# User Queries	F(1,22) = 14.710, p = 0.001			F(1,22) = 7.247, p = 0.013			F(1,22) = 4.203, p = 0.05	5.6.1
	# 'Robot' Queries	F(1,23) = 11.014, p = 0.003							5.6.1
	# Acknowledgements	F(1,22) = 4.459, p = 0.046			F(1,22) = 6.786, p = 0.016			F(1,22) = 4.195, p = 0.05	5.6.2

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
	# 'Robot' acknowledgements	F(1,23) = 9.629, p = 0.005							5.6.2
	# 'Robot' clarifications	F(1,24) = 6.405, p = 0.018							5.6.3
	# 'Matches'	F(1,22) = 9.354, p = 0.006							5.8.3
	# User 'Matches'	F(1,22) = 12.885, p = 0.002							5.8.3
	# 'Robot' 'Matches'	F(1,24) = 6.507, p = 0.018							5.8.3
	'Matches' vs. 'Mismatches' *	m only						$\chi 2$ = 17.565, df = 3, p = 0.001	5.8.7
	User 'Matches' vs. User 'Mismatches' *	m only						$\chi 2$ = 4.155, df = 3, p = 0.042	5.8.7
	Robot' 'Matches' vs. 'Robot' 'Mismatches' *	m only						$\chi 2$ = 15.034, df = 3, p =	5.8.7

Dependent Variables		Factors			Interactions				Index
		Monitoring	User Gender	Robot Gender	Monitoring × User Gender	Monitoring × Robot Gender	User Gender × Robot Gender	Monitoring × User Gender × Robot Gender	
								0.002	
	% Unique words in final task	F(1,24) = 8.424, p = 0.008					F(1,24) = 4.431, p = 0.046		5.8.3 & 5.8.7
	0 new words/ 1 or many new words in next utterance after no miscommunication *	nm only			$\chi 2$ = 8.031, df = 1, p = 0.005				5.8.4 & 5.8.8
	0 new words/ 1 or many new words in next utterance after miscommunication *	nm only			$\chi 2$ = 4.779, df = 1, p = 0.029				5.8.4 & 5.8.8
	Perceived task success		F(3.626, 79.768) = 2.750, p = 0.038						5.5
	System accuracy		F(3.626, 79.768) = 2.750, p = 0.038						5.5

Table 5.18: List of significant correlations between dependent variables.

Positive Correlations		Index
# Location references by User	# Location references by 'Robot'	5.7.3
# User 'Matches'	# 'Robot' 'Matches'	5.8.2
# User 'Mismatches'	# 'Robot' 'Mismatches'	5.8.2
# 'Matches'	% Final words in final task	5.8.1
# Execution errors/Non-understandings	# Robot 'Mismatches'	5.8.5
# Incorrect instructions	# Robot 'Mismatches'	5.8.5
# Incorrect instructions	% Final words in final task	5.8.4
Score in perceived task success	% Final words in final task	5.5 & 5.8.6

6 Discussion

6.1 Introduction

While there is an extensive body of literature in how gender differences occur in spatial tasks, how females and males use language and how they interact with artefacts, little is known about how the interactive communication itself will shape patterns of performance and language use, either exacerbating or moderating differences. Chapter 3 enumerated specific research questions that were motivated by existing gaps in literature with the aim to provide clearer understanding in how gender differences occur in navigation and dialogue with a computer system. The first set of research questions (see Chapter 3, section 3.1) reframed the dominant notions that males are more efficient and accurate in following and giving instructions, females use landmarks more than males, and they rate their performance less favourably than males; do these findings also apply in dialogue? The second set of research questions (see section 3.2) utilised the experimental manipulation of visual information and clarified its impact on performance and communication in the novel domain of robot navigation and, then, explored whether females are more affected by a – less optimal – interaction condition without visual cues. The third set of research questions (see section 3.3) focused on alignment, a fundamental phenomenon of interactive communication. It examined the occurrence and operation of alignment in HCI and whether its strength depends on gender and pair composition. Chapter 5 reported the results of the empirical study which yielded data relevant to these research questions.

This chapter builds on the experimental work to provide a critical examination of the findings and reflect on the answers to each of the research questions in the light of existing literature (as discussed in Chapter 2). Theoretical and practical implications of the findings are presented, while specific results are distilled into design recommendations and guidelines.

The chapter is organised as follows: section 6.2 discusses the findings in relation to research questions 1(a) – 3, which outline an initial picture of how gender differences arise in dialogue. Section 6.3 assesses the effect of visual information on performance and communication through discussion of the findings from research questions 4(a) – 5(g) and identifies theoretical and practical implications. Section 6.4 extends the analysis of this effect to how females and males coordinate, drawing on the findings in relation to research questions 6(a) and 6(b). Section 6.5 interprets the findings from research questions 7(a) – 8(e) on alignment, discusses how gender and pair composition regulate alignment, and presents design recommendations for dialogue systems that integrate alignment.

6.2 Gender differences in navigation and route instruction dialogues

The central argument developed in this thesis is that interactive communication will influence gender differences in language use and performance. This argument was assessed through research questions 1(a) – 3, which examined the effects and interaction effects of user and 'robot' gender on the dependent variables relating to performance, route communication and user perceptions. In particular:

1. Are males more efficient and accurate (Research Questions 1(a) – 1(e))?
2. Do females use more landmarks (Research Questions 2(a) – 2(b))?
3. Do females rate their performance lower (Research Question 3)?

The high-level findings that emerged from these questions are discussed in relation to previous research.

6.2.1 Task performance

Most existing research has identified male superiority in a range of spatial activities and domains, leading to the prediction that all-male pairs would outperform all other groups and that all-female pairs would be the least successful. Similarly, it might be expected that pairs with a male in either user or 'robot' role (that is, M_uF_r or F_uM_r pairs) would show more

efficient interactions than F_uF_r pairs. The study presented in this thesis, however, revealed a more complex pattern of results.

As anticipated, gender influences navigation performance and communication. However, the findings suggested that it is the interaction – the combination of gender and role – that has the most significant impact. In particular, in this study, pairs of female users and male 'robots' (i.e., F_uM_r pairs) are associated with the fastest completion of tasks. While pairs with female 'robots' appeared to need more time and words, the analysis failed to confirm the expectation that female 'robots' will produce higher execution errors and non-understandings. In fact, users interacting with 'robots' of the same gender provided the most accurate instructions. Research in sociolinguistics has suggested that females and males belong to two distinct linguistic communities, such that they use and interpret linguistic and stylistic elements in different ways (for example, Tannen, 1994; Mulac et al., 1998); these differences ultimately lead to inter-gender miscommunication. These observations from social research may relate to the finding of this study that users in mixed-gender pairs produced a higher number of incorrect instructions than users interacting with same-gender followers. Taken together, these results do not comply with previous findings of male superiority, neither do they argue for a female advantage. In effect, the results support that dyadic interaction can moderate differences that typically emerge in monologue settings, and suggest that females can achieve comparable performances in interactive route communication episodes. This argument has been proposed by some scientists (for example, Hyde, 2005), but has not been evaluated through empirical investigation to date.

While the analysis of performance-related measures outlined a picture in which female users produced efficient and accurate route information in interactive communication, the dialogue-based analysis presented in the following section refines this view illustrating how interlocutors, especially users, prevented miscommunication by adapting their language use.

6.2.2 Communication processes

In non-interactive tasks, females use landmarks as their default strategy to find and describe a route, while males prefer purely spatial instructions. The results of this study did not confirm that female users employed a larger number of landmarks (references to environmental

features, such as locations, the destination, choice points and pathways). In fact, users of *both* genders included significantly more landmark references only when interacting with a female as 'robot'. The explanation proposed here is that male users adapted their own linguistic preferences to match the needs of the female 'robots', by incorporating more landmark references compared to when they were interacting with male 'robots'. This interpretation fits with empirical studies that show that speakers adapt their utterances according to the perceived needs, characteristics and spatial capabilities of their partners (see the studies within the Collaborative Model relating to the 'audience design' concept, for example, Isaacs and Clark (1987); Sacks et al., 1974; Schober, 1993; 2009). Similarly, male 'robots' received mostly simple instructions from their partners; that is, instructions composed of a verb of movement and a directional spatial term. Such instructions are more economic, and hence, in principle, more appealing for communication purposes. Yet, they are underspecified and ambiguous (Tenbrink and Hui, 2007), lacking landmark references that are used to provide cues for (re-) orientation and solve and prevent navigation problems (Michon and Denis, 2001). Female navigators were found to be particularly vulnerable to instructions and environments devoid of landmarks, while male performance remained unaffected (Andersen et al., 2011; Lövdén et al. 2007). As such, it may be argued that females and male users, by not relying as frequently on purely spatial instructions when guiding females, ensured that successful communication was achieved. If this interpretation of the findings is correct, it raises questions around how users were able to perceive their partner's needs within a very unusual communication situation of (albeit simulated) human-robot interaction. In conclusion, this finding exemplifies the abilities of human interlocutors to 'tune' to the interaction conditions and presents opportunities for further experimental investigation.

6.2.3 The effect of spatial ability

Section 2.8.2 of the Literature Review chapter discussed research by Schober (2009), which revealed a complex effect of spatial abilities on spatial language use; for example, it was found that individuals that had superior mental rotation abilities were also able to provide more accurate spatial descriptions and use an allocentric frame of reference. Similarly, studies have found that spatial ability as measured by mental rotation tests appears to correlate with virtual maze navigation (Moffat et al., 1998), ability to follow Euclidean-based route instructions (Saucier et al., 2002) and orientation (Silverman et al., 2000). However,

imaging studies in humans report that navigation tasks involve areas in the brain (hippocampal region) that are never activated during mental rotation tasks (Iaria et al., 2008).

In light of this research, valid questions arise with regards to whether the findings of the present study constitute gender differences or whether they essentially reflect differences in spatial ability (related to the fact that females usually have low mental rotation ability, whereas males have high mental rotation ability). At the same time, the review of existing research revealed that participants' gender or spatial ability is rarely controlled for when one of them is the independent variable. Yet, the aforementioned study by Saucier et al. (2002) offers insight into the issue. Their findings confirmed a male advantage in mental rotation. Males were also able to navigate more accurately than women when following Euclidean instructions, but no performance difference was found when males and females used landmarks to navigate. A significant correlation was also identified between mental rotation ability and navigation accuracy when following Euclidean instructions. But there was no correlation between mental rotation and accuracy of people following landmark instructions. Most importantly, it was also found that gender differences in mental rotation ability did not explain differences in navigation using either type of instructions. The authors concluded that the lower accuracy of women that followed Euclidean instructions did not result from their lower spatial ability (as measured by mental rotation), but, rather from their weaker aptitude in processing Euclidean instructions accurately and efficiently. As such, it is argued that because the navigation task of the present study did not involve Euclidean information or cardinality, the performance of participants was not mediated significantly by their low or high spatial ability.

Taken together, it is acknowledged that the validity of the results of the present study may be limited by the fact that spatial ability was not tested, and was not controlled for in the experimental design. However, the study of Saucier et al. (2002) provides empirical evidence showing that spatial ability cannot explain the differences observed between males and females. Moreover, it is clear that interactive navigation amalgamates a range of abilities (verbal and communication) beyond the domain of spatial abilities. It is argued that the combination of these abilities and their interaction with variables relating to the communicative situation, as discussed in sections 2.7 and 2.8, appear to better qualify as explanations of the complex pattern of results of the present study, and that attributing them to differences in spatial ability may be an oversimplification.

6.2.4 User perceptions of the interaction

It is a well-documented phenomenon that females experience lower confidence and judgments with regards to their abilities (termed self-efficacy) and higher anxiety than males in both the domains of spatial abilities (for example, Lawton, 1996;1994) and HCI (for example, Beckwith et al., 2007), which generalises across nationalities and levels of expertise. It was, thus, anticipated that females that performed a spatial task interacting with a computing system would provide poorer self-assessments compared to male users, and this was explored through research question 3. The results confirmed that males provided higher scores responding to the question 'I feel I did well in this task'. On one hand, several studies have found that self-efficacy affects usage and performance (Beckwith, 2007; Burnett et al., 2011). On the other hand, other studies have shown that although females' perception of ability and performance are lower, their actual performance is comparable to males' (McCoy et al., 2001; Hargittai and Shafer, 2006; Lawton et al., 1996). The latter observation is more in line with the results of the present study. In addition, while females rated their own performance more poorly than male users, they tended to rate the system's performance more favourably than males. Indeed, males tended to judge that the system was less accurate. Previous studies have showed that girls attributed failures in math exams to their own lack of ability, whereas boys offered various other reasons like luck or task difficulty (Dickhauser and Meyer, 2006; Stipek and Gralinski, 1991). In this light, the results of the study provide original empirical support to the proposition that females are more likely to blame themselves and their lack of skill if they face difficulties in performing a task using a computer system, while males are expected to view such situations as system failures (Beckwith and Burnett, 2004; Boiano et al., 2007).

The questionnaire used in this study was designed to be relatively short and simple to be completed after each task. Therefore, the resulting observations may not be conclusive. Post-study overall satisfaction questionnaires, which assess system usability characteristics, such as ease of learning, simplicity, effectiveness and user interface (as in the PSSUQ and CSUQ questionnaires of IBM; see Lewis, 1995) and affective factors, such as likeability, cognitive demand and annoyance (as in SASSI by Hone and Graham, 2000), could be additionally employed in a replication study in order to provide insight into the role of gender in all dimensions of user experience.

In conclusion, profound gender differences were identified in the novel domain of dialogue-based navigation of a robot system. The findings presented so far corroborate the claim that any female 'disadvantage' is moderated through interactive mechanisms, hinting at an adaptation mechanism that has benefited interactions with female 'robots'. The remainder of this discussion seeks to shed more light onto these processes.

6.3 The effect of visual information on navigation performance and communication between user and system

In task-oriented interaction, a shared visual space is the interaction condition in which the collaborators/interlocutors can see the same objects and environment at the same time. There has been significant research in task-oriented CMC and human communication exemplifying how shared visual information (particularly of the work area) increases awareness of the current state of the task and facilitates conversation and grounding, such that interlocutors can form more efficient utterances and monitor understanding. On the basis of this literature, Research Questions 4(a) – 5(g) were formed for the task domain of human-robot navigation. The questions explored whether visual information will allow collaborators to perform the task faster and more accurately and whether it will change the structure and content of communication. The investigation of visual information was not a primary research aim, but was utilised as a means to clarify gender-specific issues. Yet, interesting results emerged for this study's domain, some of which confirmed previous findings, while others contributed with new insight. The results of the present study were also largely consistent with conceptual work within the Collaborative Model.

6.3.1 Task performance

As addressed in Research Questions 4(b) and 4(c), it was expected that miscommunication will be less frequent when users and 'robots' shared visual information. This prediction was guided by empirical evidence, as detailed in the review of the literature (see Chapter 2, section 2.9). For instance, Karsenty (1999) observed that visual information about the workspace prevented misunderstandings and facilitated error recovery through the provision of appropriate information. Along the same lines, research by Kraut, Fussell and Gergle (see

for instance, Kraut et al., 2003; Gergle et al., 2013; Kraut et al., 2002) identified a benefit for task performance in computer-mediated collaboration, explaining that visual co-presence enhances situation awareness and grounding mechanisms. However, the results of the present study contradicted this prediction, indicating that the frequency of non-understandings and incorrect instructions increased when users and 'robots' shared visual information. Therefore, although counter-intuitive, it is argued that when the available evidence of understanding is less solid and reliable (that is, only speech, no visual feedback), the criteria to ensure that understanding is being achieved become stricter, forcing interlocutors to be more accurate, persistent and detailed (and, consequently, less efficient in terms of word and turn usage). On the other hand, visual evidence relaxes the criteria and causes interlocutors to be less precise, which, in turn, results in higher miscommunication.

There are theoretical and empirical findings in the literature that support this observation; the Collaborative Model postulates that interlocutors do not seek perfect and complete mutual understanding, but to sufficiently understand each other for current interaction purposes. Thus, people set their grounding criteria to be only as precise as they need to be. In an empirical study, Brennan (2005) reported that followers collaborating in a spatial task reached the target more closely when instructors did not receive visual feedback. Closer inspection of the dialogue data of this corpus reiterates Brennan's observation; when users could supervise the 'robot's' actions, 'almost there was good enough'. Two complete dialogues from pairs from the Monitor and No Monitor conditions provided in Table 6.1 below are indicative of this tendency. The destination was the Lab. The user in the No Monitor condition required that the 'robot' not only reaches but also goes inside the location before asserting that task was accomplished, whereas the user in the Monitor condition provided directions that led the 'robot' about 100 pixels off the target and ended the task (see image in Figure 6.1 below). It is also interesting to note that users in the Monitor condition did not usually state that this building was the destination, as in the dialogue example in Table 6.1 below. This suggests that visual co-presence may lead to inflated assumptions of what is mutually known or perceived. This observation has implications for the interaction with collocated and tele-operated systems, discussed in the next section (second part of section 6.3.4). Finally, similar to the findings presented above, Brennan (2004) also observed that execution error rates were not higher without visual information than with visual information.

Table 6.1: Dialogue examples from the Monitor (left-hand side) and No Monitor conditions (right-hand side).

Monitor condition *[MF9_L36-47]*	*No Monitor condition* *[NMF8_L18-29]*
U: hello	*U:* hello
R: hello	*R:* hello
U: go straight ahead	*U:* walk straight then turn right
U: turn right	*R:* now where do I go?
U: now, turn left then right	*U:* where are you now?
R: I have reached the junction	*R:* the pub is on my right
R: ok	*U:* walk straight past the pub and stop at the lab
R: straight ahead or turn left?	*R:* I am at the lab now
U: keep going straight	*U:* go into the lab
U: goodbye	*R:* I am inside the lab now
R: I have reached the junction by the bridge	*U:* goodbye
R: goodbye	*R:* goodbye

Figure 6.1: The execution of the instructions provided in the dialogues in Table 6.1. The thick yellow line represents the path taken by both 'robots'. The red

dashed line and blue solid line show the finishing execution of the 'robots' in the Monitor condition and No Monitor conditions, respectively.

The finding (in response to Research Question 4(a)) that visual information failed to improve performance in terms of task completion time is another point of departure from past research. Thus, questions arise with regards to why visual information did not provide a performance benefit, as robust as previously identified in various experimental conditions and setups in CMC. In those studies, it was argued that visual information improves performance by facilitating situation awareness and grounding. Hence, it is possible that these basic coordination mechanisms that naturally occur in communication between humans do not operate as strongly, or in comparable ways, in the interaction with an artificial agent. The following section focuses on communication processes and language usage, as addressed in Research Questions 5(a) – 5(g). Drawing on concepts within the Collaborative Model, it seeks to provide some insight into how situation awareness and grounding function in this study's HCI domain and develop an interpretation of the novel results described in this section.

6.3.2 Communication processes

Visual information has a dual effect on the mechanisms of situation awareness and grounding. Grounding is the process in which interlocutors establish common ground, that is, what has been said has also been understood (see Chapter 2, section 2.9.2). A communication episode consists of contribution cycles of two parts; the presentation and the acceptance stages. In the presentation stage, assumptions about what the addressee knows help the speaker plan his/her utterance. During the acceptance stage, interlocutors mutually ensure that the addressee has sufficiently understood the original utterance. The assumptions of a speaker with regards to what the addressee knows are based on three factors: community co-membership, linguistic and physical co-presence (Clark and Marshall, 1981). Community membership requires that both interlocutors believe that they are members of the same community and share particular universal knowledge. The second kind of evidence is linguistic co-presence such that a speaker can use an anaphoric expression in dialogue, for instance 'it' in the utterance 'The destination is the pub. It is on the right'. The third and strongest kind of evidence is based on (direct or indirect) physical co-presence. As such, in

the case of visual co-presence, the speaker may produce ‘turn here’ or ‘take this road’. In turn, the ‘robot’ may demonstrate understanding without having to explicitly state it but through performing the action (since visual evidence is stronger than linguistic). Such behaviour also relates to the principle of least collaborative effort which proposes that interlocutors adapt their communication patterns so that they will put the least effort in terms of language production and processing while not jeopardising the communication.

This theoretical framework helps interpret the findings in response to research question 5(g), which showed that visual co-presence does not encourage the use of reference to commonly perceived objects (landmarks), but actually renders it redundant.

As predicted by the Collaborative Model and similar to past studies, visual feedback reduced the necessity for verbally acknowledging that a message was understood (relating to the process of grounding) and that an action was executed (relating to situation awareness). The actions of the ‘robot’ served as an immediate, accurate and effortless indicator of his/her understanding. The users could monitor the progress of their partners and provide further instructions and corrections at the moment needed. This had a profound effect on the structure of the dialogue; the interaction was a cycle of the user giving instructions and the ‘robot’ executing them, with additional communication reserved to resolve ambiguities and misunderstandings. In particular, the results showed a significant decrease in word usage by both interlocutors and in explicit verbal confirmations of understanding and action completion by ‘robots’. In that condition, users dominated the interaction with user turns accounting for more than 60% of all turns. On the other hand, when visual feedback was withheld, ‘robots’ participated in coordinating the interaction providing information-rich responses more actively; as such, contributions between the interlocutors were balanced.

It is noteworthy that the interface environment of ‘robots’ did not change between conditions. Still, they also adapted their language use to compensate for the changes in the users’ interfaces. Such behaviour is predicted by the Collaborative Model, which stresses the cooperative nature of communication and proposes that, although both interlocutors are responsible to maintain understanding, responsibilities change and become unequally divided in order to reduce the *combined* effort. This shift in responsibilities is evident, when considering query initiations; ‘robot’ queries increase when their actions are monitored, and user queries increase when they cannot monitor the ‘robot’s’ actions. Similar to the

aforementioned studies by Kraut, Fussell and Gergle, these observations are based on basic dialogue act analysis; yet, in the present study, fine-grained component analysis complements these results. As indicated by the relative frequencies of the types of components that are known to contribute to the information value of instructions (that is, delimiters and landmark references), when sharing their visual space, users could readily confirm their assumptions about the information requirements of 'robots' and use linguistics shortcuts and simpler language. On the other hand, when visual information was withheld, uncertainty about the position and movement of the 'robot' created the need for elaborate and explicit instructions; as such, the level of instruction granularity was higher, corresponding to higher concentration of location references and distance designation delimiters. On the other hand, users that could supervise their partner's actions relied on underspecified purely directional instructions. Analysis of 'robot' utterances revealed parallel communication strategies; in the absence of visual information, 'robot' replies to user queries contributed with more information over and above what was asked and they explicitly stated frame of reference, locations and the destination. These differences are illustrated by juxtaposing dialogue examples from the Monitor and No Monitor conditions in Table 6.2 below.

Table 6.2: Dialogue examples from the Monitor (left-hand side) and No Monitor (right-hand side) conditions.

Monitor condition *[MF4_L18-27]*	*No Monitor condition* *[NMF6_F47-57]*
U: I am going to give you directions to the lab	*U:* go to the bridge and turn left before the bridge
R: okay	*R:* Done
U: go straight	*U:* is the gas station on your left?
U: then turn left	*R:* Yes it is
U: go back, take right instead	*U:* now keep going down the road until you see a car park
R: I am at an intersection	*R:* I am in front of the car park
U: keep straight	*U:* turn right and walk till the end, along the road you will see a gym on your right
U: go left	*R:* Yes, gym to my right side
U: turn left	*U:* good, keep going straight and you will see a factory on your left
U: there you are	*R:* Yes, factory to my left side
	U: well done, goodbye

While this study replicates previous findings and extends them to a novel domain of HCI, it also presents some discrepancies. Following the predictions of grounding theory and based on the empirical evidence of the aforementioned previous studies, it was expected that when collaborators shared visual space, they would be able to refer more quickly and efficiently to objects of the environment by using 'short-hand expressions', that is, deictic and anaphoric pronouns (formulated as Research Question 5(a)). However, this analysis revealed that the occurrence of these elements was extremely low in both conditions. An early comparative study by Guindon et al. (1987) corroborates this observation reporting that the use of pronouns was less common when people interacted with a system than when they addressed another person. While they serve to substitute longer expressions, deictic and anaphoric terms are often ambiguous referring to more than one point in time, space or precedent dialogue. Therefore, it may be concluded that users do not generally opt for underspecified deictic and anaphoric expressions to navigate a computer system. As discussed above, speakers formulate their messages based on the assumptions of what is in the common ground, reinforced by linguistic and physical co-presence and community membership. Thus, it can be argued that users were less certain about the common ground with the computer system, which inhibited the use of anaphoric expressions and deixis. Previous research has focused on the combined effect of visual information on both situation awareness and grounding with less attention to their individual contributions (as identified in the most recent work of Kraut and his colleagues (Gergle et al., 2013)). As such, the finding of this study may have several explanations. First, while the aforementioned significant effects demonstrated the development and benefit of situation awareness, the lack of significant amount of deixis suggested that the operation of grounding between participants was disrupted. This interpretation may add to the theoretical differentiation of the influence of visual co-presence on situation awareness and grounding. A second explanation lies on the distinction between visual and physical co-presence. Visually co-present collaborators maintain a common visual space, whereas physically co-present interlocutors are able to attend to spatial relations between objects of relevance, the wider environment and themselves. Thus, in a fully situated interaction, it could be argued that deictic references would be pervasive. Third, grounding

may have been weaker because of the nature of visual information offered; that is, although the perspective of the 'robot' could be easily established while the 'robot' was moving, when the 'robot' was static, its current perspective was less clear for the user[23]. As such, it is plausible that since shared perspective was not unequivocally ensured between interlocutors, it interfered with their ability to agree upon and use shorter referring expressions like anaphora and deictic terms. As discussed in section 2.9.1, the characteristics of the visual information available to users define its value and effectiveness (Whittaker, 2003a). Therefore, it should be stressed that the results derived from this experiment with regards to the benefit of visual information are only valid and reproducible, if the aspects of what and how visual information was offered are taken into consideration.

As mentioned above, previous studies have not shown large differences in terms of the content of dialogues. This was attributed to a coarse coding scheme which considered dialogue acts such as queries and replies (Kraut et al., 2003; Gergle et al., 2004). In this study, a detailed component-based scheme complemented the dialogue act analysis and illustrated a refined picture of how visual information influences the choice and distribution of utterance constituents.

To sum up, the analysis tested the proposal that visual information serves as a resource in collaborative spatial tasks with a computer system. The results are consistent with the Collaborative Model of human communication and added original insight into how visual co-presence supports situation awareness and grounding. This section continues by presenting theoretical and practical implications.

6.3.3 Theoretical implications

The results of this study provide corroborating evidence for concepts of the Collaborative Model: (i) the contribution model (Clark and Schaefer, 1987; 1989), (ii) Clark and Brennan's

[23] This experimental element follows relevant literature (see Chapter 2, section 2.9.1 and Chapter 4, section 4.2.2).

(1991) framework on the costs and tradeoffs of grounding in different communication media, and (iii) the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986).

First, although the 'robot's' interface remained the same in both experimental conditions, 'robots' adapted their language based on what the user could see. When users could not see the workspace, 'robots' provided lengthier descriptions, replies that contained more information than asked for and elaborate verbal acknowledgements. If language production was indeed an independent process, 'robots' would not modify their language use in response to what their partners could see. Similarly, although what the 'robots' knew did not change across conditions, users assumed that 'robots' needed more explicit instructions that were anchored on landmarks and specified action boundary. While this behaviour appears unnecessary, it confirms the basic argument of the Collaborative Model that speakers design their utterances based on what is assumed to be mutually known, in the common ground.

Second, the results verify that the strength and preciseness of grounding is determined by the affordances of the interaction, such that interlocutors adjust quite flexibly to the degree of perceptual co-presence. The results diverged from previous research that argued that communication is faster and more accurate when interlocutors share visual space (Gergle et al., 2004; Gergle, 2006; Kraut et al., 2003). On one hand, visual information facilitated situation awareness because pairs did not need to rely on lengthy linguistic descriptions to determine the state of the task and how the action was performed. On the other hand, the scarcity of deictic and anaphoric expressions when pairs shared visual space may suggest that visual information did not facilitate grounding as much as expected. It either indicates that grounding is weaker in a plausible HCI scenario, or the lack of visual access to the robot itself disrupted the operation of grounding (see discussion in the next section).

Moreover, the rise of miscommunication when visual information was available illustrated that stronger (visual) evidence does not necessarily lead to better performance; rather, better performance depends on how well people are able to adjust their grounding criteria. Grounding involves satisficing (Simon, 1981). That is, interlocutors ground as precisely as they need to for current purposes. Without visual information, they needed to set a higher criterion in order to reach equivalent performance levels. This meant that they had to select more effortful means of communication, but this safeguarded against errors by both partners. Closer examination of the performance and dialogue data supported this

explanation and showed that when speakers did not share visual space, they stipulated a stricter criterion to declare the successful completion of each task, while for users receiving visual information, close enough was good enough (as illustrated in the dialogue examples in Table 6.1 and Figure 6.1 above).

Third, the results also relate to Clark and Wilkes-Gibb's (1986) principle of mutual responsibility. That is, the responsibility for grounding fell on whoever had the strongest evidence at their disposal. So, when visual evidence of the 'robot's's movements were available, the user was responsible to signal that understanding failed or succeeded. This reduced the 'robot's' and collective efforts. It was also shown that physical actions by the 'robot' replaced or possibly functioned as verbal turns. In fact, verbal turns by 'robots' were typically perceived as redundant and were often ignored by the users, as exemplified by the dialogue excerpt in the Table 6.3 below.

Table 6.3: Dialogue from the Monitor condition [MF8_S66-82]

U: hello
R: hello
R: which way should I go?
U: go to the shop
[movement]
U: move forwards
[movement]
R: is this right?
[movement]
U: stop
U: turn around and then move forwards
[movement]
R: do I go the other way?
[movement]
U: take the road on the right
[movement]
U: stop
U: move forwards a little bit
[movement]
R: am I here yet?
U: move forwards
[movement]
U: stop
U: you're at your destination, goodbye

R: goodbye

In addition to corroborating the principles of the Collaborative Model, inspection of the data provided observations with regards to the cross-timing of verbal and physical actions. The time-stamped data of the corpus showed that when visual evidence was available, users could plan and present new information concurrently with 'robots' grounding the previous information by moving. So grounding was done continuously and not through individual turns. This overlap was accepted as natural, such that when 'robots' did not receive the next instruction by the end of the execution of the previous one, they took it as indication of error. In the interchange below (see Table 6.4), the robot received an instruction which was executed immediately (by pressing a button). Yet, several seconds of 'silence' triggered the 'robot' to question the accuracy of the execution. Clark and Krych (2004, p.73) observed a similar phenomenon in task-oriented human communication and termed it the 'immediacy constraint'. They noticed that when the instructors did not respond immediately after an action was performed by the follower, the follower would implicate that the instructor's answer is a rejection of the correctness of the action. They found the 'time window' for an expected response to be within 0.3 and 1.0 seconds. The 'immediacy constraint' raises interesting questions for computer system interfaces: what is the specific time limit of a user waiting for a system response in a particular application, before assuming that his/her action was incorrect. The observation of this study also relates to the phenomenon that participants in conversations or other forms of joint activity have expectations about how and when the other party will act, such that absence of activity is seen as meaningful and, as in this example, alarming. This has been termed 'interpredictability' and is reinforced when partners can take on the perspective of others, as in the visual co-presence condition of this experiment (Klein et al., 2005).

Table 6.4: Dialogue excerpt from the Monitor condition [MF8_T39-40].

Time stamp	Utterance
14:22:42	*R:* Go to the pub
14:22:44	[robot appears outside pub]
14:23:10	*R:* Is that ok?

Following most existing theories of human communication, the coding scheme employed in this study was based solely on verbal utterances. But, as illustrated in the examples above, grounding is also carried out through non-verbal means. Therefore, it is important that any

theory of discourse should be able to account for how actions ground meaning and how they are combined and integrated with linguistic utterances in task-oriented interaction.

Finally, the fine-grained component analysis of the route instructions data adds insight to research in route communication. Past empirical research has focused on defining the best practices of producing route directions and what makes them 'good' or 'poor' (for instance, Allen, 2000a; Denis, 1997; Lovelace et al., 1999). These accounts converge on the importance of landmarks that provide cues for (re-)orientation and delimiters that give specificity and distinguishing information about actions and environmental features. The validity of the findings and suggested principles has been derived through scenarios in which the directions were produced beforehand by either the experimenters or a separate group of subjects. Therefore, while this study extends the applicability of the framework using dialogue methods, it also demonstrates that the reliance on these principles is dramatically reduced as soon as the navigation task becomes less demanding (that is, by the availability of visual information).

6.3.4 Practical implications

Research in psycholinguistics and cognitive and social psychology is highly interconnected to the field of HCI. On one hand, such research describes the principles that underlie human communication – how information is processed and communicated – which can serve as the framework for interface designers to understand and improve interactions with computers. 'Ad-hoc' decisions, which are not motivated by a sound framework of principles, are unlikely to lead to successful applications, especially those based on natural language. For example, human communication is analogous to HCI in that they both involve coordinated action. Similar to human communication, users also expect that they should provide and receive evidence that they have been understood and that the task is still on track. In many cases, user and system errors occur because of inappropriate or insufficient feedback and impoverished context (Brennan, 1998). Thus, as suggested in the previous section, understanding how principles such as the 'immediacy constraint' and grounding operate may be useful for interface design. On the other hand, HCI offers a unique test-bed for cognitive psychologists and linguists to clarify and generalize their theories, models and principles, as it involves different communication media, interaction contexts and interpersonal dynamics.

The previous section discussed the theoretical implications of the results in relation to existing models of human communication leading to a better understanding of how visual information affects the processes and products of communication and collaboration. This understanding can be of practical relevance for dialogue systems and robots. Moreover, since it involved a computer-mediated (albeit concealed) collaboration between people, this study may also have practical implications for the development of Computer-Supported Cooperative Work (CSCW) and Computer-Mediated Communication (CMC) systems. Interaction with these systems may involve physical or perceptual co-presence, joint or directed activities and virtual or real manipulation of objects. This section briefly describes these technologies and analyses the ways that the theoretical findings of this study can be extended to provide design guidelines for collaborative systems.

Computer-mediated communication and collaboration

Computer-mediated communication and collaboration have become a normal setting of human activity. Most software tools provide visual information and, as reiterated by the present study, there is ample evidence that sharing visual information improves communication and performance in cooperative work. However, understanding of the mechanisms and the factors underlying this benefit remains incomplete (Gergle, 2006). It is also essential to take a step back and attempt to define the processes involved in CSCW. These processes vary on several levels, which are largely dictated by the nature of the collaborative task. That is, CSCW and CMC applications are typically classified based on whether they involve one or the combination of the following three elements: communication (exchange of messages and information), coordination (managing people and their activities) and cooperation, with the latter referring to complex joint work within a shared space (Ellis et al., 1991; Fuks et al., 2005). The effect of the technology and the task are non-trivial, and, thus, developers should not rely on simple intuition or superficial characteristics when analysing requirements and features. As argued above and throughout this thesis, insight offered by empirical models of human communication can be of great value. But it should be also complemented by focused experimental research in order to clarify how the particular technology – and the tasks users intend to accomplish by using it – change, interoperate with and rely on language. This approach will enable developers to identify how the design of existing technologies can be improved or new technologies can be implemented to support

cooperative work. The findings of this study provide a number of insights for CMC and CSCW systems and are outlined below.

Clark and Brennan's framework states that the affordances and constraints of a medium impose particular costs on the grounding process and on how grounding shapes the interaction through this medium (Clark and Brennan, 1991). As such, interaction through computers presents potential barriers to establishing mutual understanding, because it reduces the means through which the interlocutors can ground an utterance. These barriers have to do with constraints such as visibility, co-presence, co-temporality, audibility, simultaneity, sequentiality, reviewability and revisability. These predictions are validated by the results of the present study; when visibility was removed, collaboration with the system became more effortful and both users and 'robots' needed to rely on more sophisticated linguistic means to achieve mutual understanding and success. Therefore, it is important that these constraints are taken into consideration when designing systems.

Moreover, it is argued that the positive effect of shared visual information may vary based on task characteristics; that is, visual information could be essential for complex, dynamic and temporally-dependent tasks but may not make a substantial contribution to the performance in simpler tasks, or may even be detrimental if provided with temporal delays (Kraut, Gergle and Fussell, 2002). In addition to temporal delays, Whittaker (2003a) has argued that visually-enhanced interactions may present a close, but misleading approximation to face-to-face communication. As Schober (1993) pointed out, subtle differences between one's own visual perspective and his/her partner's are extremely difficult to recognise. These arguments relate to the empirical findings of this study. The experiment was designed to enable users to see what the 'robots' saw, but not the perspective of the 'robot' when it was stationary. As such, visual information led them to rely on their normal behaviour relaxing grounding criteria, an approach that failed because the perspective was not truly shared. This was evidenced by the higher frequency of incorrect instructions and non-understandings.

Many CSCW tools involve the collaboration between novices and experts. The results of this study showed that when visual information was not shared, the responsibility for task and understanding maintenance had to be equally distributed. Thus, in cases in which one of the collaborators is a novice, he/she will be unable to provide equal and accurate contributions, which will render the interaction inefficient and vulnerable to miscommunication. For such

applications, in which there is an expected prior asymmetry between the knowledge and proficiency states of the collaborators, the ability to effectively ground information should be reinforced. Developers should consider implementing visual functionality to support grounding that does not rely on language, such as remote pointers, highlighting tools, and other methods that ensure joint attention to objects. Thus, the expert can easily refer and draw the attention to landmarks and details of the context when working with the novice who may be unfamiliar with the terminology. The results of this study also showed that when users can view the partner's workspace, the partner's actions function as and replace verbal statements. As such, the novice may be able to use tools such as pointer trajectories (Gutwin and Penner, 2002; Fraser et al., 2007), which show the movement of the cursor, as a way to provide feedback without having to explain what he/she is doing verbally. Along the same lines, visually enriched interfaces can support non-native speakers of English, who may benefit by being able to rely on non-verbal information (as shown in a Map Task-based study by Veinott et al., 1999).

Clark and Brennan's framework and the present empirical study supported that different mediums impose different costs on how people ground information; this has implications for systems that offer real-time direct text-based communication. Instant messaging applications dominate people's interactions and are also often integrated in CSCW systems. Speech is ephemeral, so people engage in a frequent grounding process of small chunks of language. On the other hand, because typed communication is not ephemeral, it involves higher production costs, so interlocutors ground less frequently and longer utterances. Therefore, visual information is expected to present a larger benefit to text-based communication, by alleviating some of the higher cost of grounding.

Human-Robot Interaction

Building on the theories within the Collaborative Model, Klein et al. (2005) discussed the principles that people must follow in order to sustain common ground with the aim to effectively collaborate with their team members, and presented the argument that robotic and human agents have analogous responsibilities in team coordination. Also drawing on aspects of the Collaborative Model and the findings of the present study, this subsection discusses the

implications for robot and dialogue systems that are involved in situated interactions with their users.

The findings illuminated that the demands for spatial reasoning are higher for a robot that shares the same physical/visual space with the user than for a non-collocated robot. In particular, when sharing visual information, users relied on underspecified purely spatial instructions which often lacked boundary information. Since dialogue and execution were synchronous, the user was able to provide the next instruction with temporal precision, at the moment in which the 'robot' was observed to have completed the previous instruction. Thus, the 'robots' assumed that 'move forward' means 'move forward, until I tell you to stop or give you a new instruction', and the command 'stop' regularly appeared as an instruction (as shown in the dialogue example in Table 6.3 above). Such user input would be problematic for the majority of real-world robots that lack similar inferential capacity. It was additionally observed that although users in both conditions could not know the robot's orientation at all times, users that did not receive visual feedback were more inclined to find out before giving directions. Moreover, as noted above, users in the Monitor condition failed to specify that a building was the destination. These phenomena may relate to misplaced assumptions of common ground. Taken together, while visual co-presence is mostly beneficial in computer-mediated collaborations between people, it may be detrimental in interactions with artificial agents, as it may encourage users to employ strategies that entail that the robot possesses human-like perceptual abilities.

On the other hand, a robot which does not share its visual space with the user faces another challenge; when monitoring was not possible, the users continuously requested information about the current location of the robot. A 'human robot' was certainly able to provide rich descriptions of its surroundings. Indeed, providing effective feedback is crucial for task-oriented interactions, and especially in the dynamic setting of HRI, in which the user's instructions can be incomplete or outdated. However, this is a non-trivial task; feedback should be given at the right time and amount, or else, it compromises the interaction, as detailed below.

In particular, speech-enabled mobile robots are usually built on agent-based (distributed) architectures, which involve several components typically divided in two modules, one for interpreting and generating language and one for processing and executing the actions.

Situated dialogue entails instantaneous synchronisation and updating of these modules to include a continuous influx of information. Since the human agent can also send information at any time, distributed systems often face serious problems of synchronisation (Blaylock et al., 2002). So, in a typical scenario, the robot may receive a new instruction before having processed the previous one. The new instruction could be interpreted within the 'old' context leading to a wrong execution. Lack of synchronisation also occurs at the turn-taking level. Namely, the robot and user's utterances may overlap or the user may 'barge-in' while the robot is speaking. Thus, clarification requests and feedback need to be provided with high temporal accuracy, or else, they could impair the interaction and lead to confusion and errors (Tenbrink and Hui, 2007). The experimental setup of this study enabled us to observe the effect of synchronisation and spatiotemporal congruity problems on the interaction. The messages were formulated in a private window and were transmitted when the participants pressed 'enter', and, as a consequence, 5.2% of all turns overlapped. The partners typically ignored delayed input and proceeded with the dialogue. Occasionally, however, an ill-timed response by the 'robot' would cause the user to give erroneous instructions. In the example in Table 6.5 below, the first messages by the user and 'robot' were sent simultaneously. By the second user turn, the 'robot' had already executed, but the user still perceived the 'robot's' request as relevant and repeated the instruction. If instructions are 'stacked' for execution, this would lead the robot to execute the same instruction twice (as sometimes did the human 'robots' in this study). Moreover, providing redundant feedback compromises the 'naturalness' and efficiency of the interaction. The empirical results of the present study argued that the execution of the task is often sufficient feedback by itself. In brief, when and what kind of feedback to provide should be determined by a criterion that draws on several knowledge sources and is updated both within and between sessions (Brennan and Hulteen, 1995). These sources could be the dialogue history (e.g., how many times in the dialogue so far the robot and user have initiated repair?), model of the environment (e.g., is the robot at home, outdoors or at a crowded workplace?) and the task (e.g., is the route well-known, what are the consequences of errors?).

Table 6.5: Dialogue excerpt from the No Monitor condition [NMF1_P6-9].

Time-stamp	Utterance
14:31:15	*U:* turn right
14:31:16	*R:* I am at a junction
	[robot turned right]

14:31:25	*U:* turn right

The ability to accurately ground information is essential in the interaction with a robot. Technologies for non-invasive and computationally inexpensive eye gaze and gesture tracking are now available for many platforms. Thus, joint reference to objects and perspective may be resolved by multimodal input from these sources. As suggested in section 6.3.2 above, it might be necessary to draw a distinction between visual co-presence (only sharing a view of the work area) and full physical co-presence, given that grounding is determined by their respective affordances. In this experiment, the users did not generally employ simple, underspecified deixis, while it may be expected that in fully situated interactions with collocated robots, users will rely more on these expressions. Therefore, dialogue strategies to address such linguistic elements should be integrated in the dialogue manager of robots that are destined to interact with users within the same space. Behavioural indicators of attention such as eye gaze, head, posture and gesture may also be more prevalent and, thus, the aforementioned technologies may be of particular utility in such interactions. On the other hand, in remotely-controlled or (semi-) autonomous robots abilities to resolve underspecified deixis and reference could be less essential.

Because of the novelty of the field, the majority of work in HRI is conducted for data collection and modelling purposes within laboratory settings. As such, many aspects of the communication and coordination patterns in HRI remain speculative. Similar to the present study which suggested higher miscommunication in the Monitor condition, field work by Casper and Murphy (2002; 2003) and Burke et al. (2004) poses serious questions about the utility of sharing visual space with a robot. These studies tracked the behaviour of users tele-operating robots with on board cameras in urban search and rescue tasks. The operators were found to experience extreme difficulties building and maintaining situation awareness and consolidating data obtained from the robot's view with existing knowledge, and spent more time collecting information about the state of the robot and environment than actually navigating the robot. This can be diagnosed as a problem of establishing and sustaining common ground. The authors recommend that operators and the rest of the team should receive additional special training and argue that visual information from the 'robot' should be available to all team members, not only to the operator.

Finally, the study provided a detailed account of the range of linguistic options that users are likely to employ in Human-Robot navigation dialogues, in two realistic scenarios of supervised and unsupervised interaction. The fact that differences exist between how people provide instructions to humans compared to artificial agents in similar contexts is not counter-intuitive. However, the dimensions and extent of these differences merit additional in-depth research. Comparing the corpus collected in this study to similar corpora provides interesting insights into the subject. Studies that have used the same classification of instructions (action only, action + reference to environmental feature, etc.) across a variety of experiments and conditions report that simple action prescriptions do not exceed 19% of all instructions given (Denis 1997; Daniel and Denis, 2003, 1998). In Muller and Prévot (2009), the rate is even lower (5%). The common factor in all these studies, however, is that the 'follower' is a human. When the follower is a simulated robot, the proportion of action-only instructions rises – to, for example, 31% in the study by Tenbrink (2007) – suggesting that action-only instructions are less common when produced as part of navigation tasks for human participants. A likely reason for this is that people are generally naive about the linguistic and functional abilities of a robot, so they tend to employ a higher proportion of simple action-based descriptions that are not anchored on visually-recognised landmarks (see also studies by Moratz and Fischer (2000); Moratz et al. (2001)).

6.4 Gender and visual information

The previous section confirmed that visual information changed the structure and content of communication in ways that lead to the least amount of effort for the pair as a whole. Interestingly, while the pairs produced more efficient speech when sharing visual space, there was a trade-off with accuracy, which indicates that grounding involves satisficing. An incongruity with past research occurred when deictic and anaphoric expressions were found to be generally underused, suggesting that grounding processes may be weaker in a human-computer dialogue setting. After clarifying the mechanisms and characteristics of the effect of visual information, the investigation now focuses on how they relate to gender. This relation was targeted by Research Questions 6(a) and 6(b), which addressed whether the impact of visual information will be stronger on the performance and communication

strategies of females. The findings have direct relevance for computer-mediated collaboration and may also provide some suggestions for the domain of navigation in online environments.

6.4.1 Task Performance in visually-supported CMC

In their review in gender differences in navigation, Coluccia and Losue (2004) proposed that gender differences (favouring males) arise when the task is 'difficult' enough. Yet, the analysis in relation to Research Question 6(a) failed to confirm the ensuing prediction that females would perform better in the visually supported task. A possible explanation is that the potential benefit of visual information was cancelled out because in the visually supported condition, landmark references, on which women rely, were disused.

The lack of a significant effect in this study does not, in fact, challenge Coluccia and Losue's argument that gender differences arise depending on task complexity. Rather, it may lay on two phenomena. First, men have stronger visuospatial working memory and cognitive visuospatial abilities like mental rotation (see the discussion in Chapter 2, section 2.3.3.2). Hence, pronounced gender differences have been found to emerge for tasks that involved high memory load and to disappear when the memory load was lower; for instance, the presence of a map to guide navigation eliminated the performance gap between women and men (see, for example, Ward et al., 1986). As such, it may be argued that if the task employed in this study required that users memorised the map, or that 'robots' remembered parts of the route, or if the environment had fewer or more subtle landmarks (on which women rely), the differences between genders would have been greater. Second, females are less capable to perform tasks that involve 3D rotation (Hubona and Shirah, 2004). So, it could be argued that navigation in the 2D environment of this study prevented this limitation from surfacing, and that a 3D virtual environment could have brought out significant gender differences. However, it should be noted that navigation in 3D environments is generally more demanding for people irrespective of gender. Cockburn and Mackenzie (2002) showed that navigation performance of users deteriorated in a 3D environment compared to a 2D condition and users perceived the 3D environment as cluttered. In light of these, the aforementioned arguments have little relevance, since the primary aim of this study was not how to elicit differences because of the experimental setup, but how to keep the setup as

'neutral' as possible (that is, not disadvantage any gender by design) in order to observe true gender differences.

At the same time, there is empirical evidence that females and males have comparable navigation performance when females had the opportunity to complement a visual and map aid configuration (preferred by males) with written verbal instructions (Devlin and Bernstein, 1995). This finding supports the argument of this study that, all things being equal, suitable communicative means can diminish the performance gap between genders.

In conclusion, this section has discussed findings that indicate that the performance of males and females does not follow the predicted patterns, which may be, at least partly, attributed to communication processes that take place between collaborators during direct interaction. This claim is further developed in the next section.

6.4.2 Communication processes in visually-supported CMC

The absence of visual information did not result in significant variations to the performance of females, and, yet, it had a strong effect on how they used language (Research Question 3(b)). While female users employed the lowest number of location and destination references when visual information supported navigation, their use was prevalent when visual information was absent. Similarly, while female users assumed control of task coordination when they had visual feedback, they most eagerly solicited information from the 'robot' when they did not have it. When visual information was withheld, coordination became explicit with a sharp rise in the number of verbal acknowledgements issued by both participants. The three-way interactions further clarified these effects showing that the differences mostly occurred when both collaborators were female. At the same time, the behaviour of males was consistent across conditions.

The analysis of the interaction effects refined the picture of how 'robots' and users modify their strategies to deal with the lack of visual information, revealing that females are associated with the most dramatic adaptations. It also indicated that incorporating landmark references may not be the default strategy of females, but employed as dictated by the circumstances. Taken together, females appear to be more sensitive to interaction changes. It may also suggest that females orient themselves more strongly towards the principles of the

Collaborative Model putting more effort to compensate for impoverished interaction conditions. In light of the literature in task-oriented CMC, the patterns of behaviour of females in the condition of no visual information are not completely unanticipated (see past findings in Chapter 2, section 2.7.2). Female collaborators have been found to issue more questions, provide more elaborate descriptions and negotiate actions and plans. They also tended to be more attuned to the task and collaboration. On the other hand, male partners dominated the discussion in terms of more words or turns, and were less willing to agree with their partners (Prinsen et al., 2009; Ding et al., 2011; Prinsen et al., 2007; Richert et al., 2011). Yet, the question that naturally arises is why females exhibited the 'prototypical' female patterns of communication when visual information was not available, but assumed a communication style associated with males when they shared visual information. Social theories of CMC have predicted that text-based CMC would inhibit socio-emotional patterns of behaviour because of the lack of visual and auditory communication cues ('media richness', 'social presence' and 'cuelessness'). Rutter (1987, p.74) argues that 'cuelessness leads to psychological distance, psychological distance leads to task-oriented and depersonalized content, and task-oriented and depersonalized content leads to a deliberate, unspontaneous style and particular types of outcome'. While these theories remain contentious, they may offer the basis of an interpretative hypothesis for why females assume a non-normative communication style when the task is easy, but maintain their cooperative and interactive style when the task is more complex. In this study, people interacted in order to perform a well-defined task, as opposed to meeting social goals of conventional CMC, and believed that they interacted with a non-human agent. Therefore, since the focus was on the task and the information needed to complete it, there was no need for females to follow their normal supportive style when visual information made the task easy. When task conditions changed, females adapted their communication style in order to reduce uncertainty and attended to the partner. It should also be noted that in neither condition did females exhibit all the features commonly found even in task-oriented speech of females, such as personal language, signals of appreciation and frequent use of hedges (Sun, 2008).

Pair and group gender composition has been found to be a powerful predictor of coordination and communication styles of people in CMC (Savicki et al., 1996). In addition, the observation of the present study that all-female pairs exhibit different collaboration patterns compared to mixed pairs and all-male pairs has also been documented in past

research (Savicki and Keely, 2000; Choi et al., 2009). Denis et al. (1999) hypothesised that women would be more gravely affected from the CMC medium, because of its relative poverty of cues and lack of non-verbal cues on which women mostly rely. But Savicki et al. (2006) proposed that women (especially in all-female pairs; see Savicki and Keely (2000)), more than men, will be able to compensate for impoverished conditions by relying on linguistic resources. This study has provided empirical support to this claim, by providing evidence that when visual information was withheld females resorted to richer and more elaborate descriptions. The findings in relation to pair composition are revisited in the discussion in section 6.5.1.

6.5 Linguistic Alignment in HCI

As discussed in Chapter 2, section 2.8.4, no empirical data exists about whether the tendency to align one's vocabulary depends on gender. Research Questions 8(a) - 8(d) were formulated to address whether female speakers align more than male speakers and to male addressees, and whether same-gender pairs are more aligned than mixed-gender pairs. The results are discussed in the next section. Moreover, there is limited understanding with regards to how the alignment mechanism operates in the interaction between users and computer systems, and how it may be exploited to improve the efficiency of the interaction. Research Questions 7(a) – 7(e) were framed to help improve our understanding of alignment in the interactions with computer systems. Synthesising previous findings with the results of this study, section 6.5.2 empirically demonstrates the practical implications of alignment and provides design recommendations relevant to the development of computer systems with natural language interfaces. The section concludes by presenting a general model towards the integration of alignment in dialogue-based human-computer interaction.

6.5.1 Gender-related alignment

Research in sociolinguistics has suggested that people use gendered or, rather, gender-preferential language. For instance, female language is said to be more indirect, elaborate and affective, richer in intensive adverbs, questions, hedges, emotion words and changes in pitch, while male speech is more impersonal and succinct (Newman et al., 2008). These

characteristics are particularly prevalent in dialogue between same-gender partners. In mixed-gender interactions, people were found to moderate the use of gender-preferential features and the communication styles of males and females appeared less divergent. It was also argued that females tend to accommodate more strongly to their addressees, especially to male partners (Fitzpatrick et al., 1995). In effect, these findings hint at the presence of gender-related alignment. Yet, no experimental data exist with regards to task-oriented conversations and functional linguistic indicators.

The study reported in this thesis aimed to address this gap through Research Questions 8(a) – 8(d). The arguments that females align more strongly than male speakers and to male addressees were not confirmed by our analyses. Nevertheless, it was found that same-gender pairs aligned in the turn level ('input/output matching') during the interaction more strongly than mixed-gender pairs. F_uF_r and M_uM_r pairs also managed to conclude the interaction with a more concise vocabulary, which is also indicative of higher alignment. The effect of pair composition in interaction through computers is further reviewed in the next section.

The analysis of the data suggested that male users tended to accommodate their gender-preferential strategies to suit the needs of their partner (as discussed in section 6.2.2 above) as often as female users. In particular, the frequency of landmark references correlated with the gender of the addressee and not the speaker, such that female 'robots' received a larger number of instructions with landmark references as a navigation aid, despite the fact that males normally prefer omitting landmark references. Overall, these findings may hint that females do not align more than males in *task-oriented interactions*. However, if this interpretation is accurate, it does not necessarily negate previous observations that females are stronger aligners in social interactions; in such interactions, social identities, and, possibly, stereotypical expectations are more foregrounded. Given that past research has focused on non-functional linguistic elements, a possible direction of future work could be to determine whether women also align their vocabulary choices in social interactions.

The effect of pair composition in CMC

The basic premise of this thesis and point of departure from past differential research is that gender-related performance and language use arise as functions of inter-individual processes.

Consequently, interaction effects were anticipated between the genders of the collaborators. These effects are discussed in light of previous literature in the effect of pair composition on CMC.

Studies in CMC settings provide evidence that pair composition has an effect on several dimensions of communication, collaboration efficiency and experience. For instance, in a study in asynchronous (email) CMC, Savicki et al. (1996) found that female-only pairs used more words and reported higher satisfaction than mixed-gender and male-only pairs. In another study, in which pairs of young pupils used a collaborative learning system, mixed-gender pairs showed lower levels of engagement and co-operation and were less interactive than same-gender pairs (Underwood et al., 2000). Finally, Ding et al. (2011) confirmed that partners in mixed-gender pairs used divergent strategies to analyse problems, while same-gender pairs were more coordinated in their strategies. While the performance of males in mixed-gender or same-gender pairs was similar, females in female-only pairs scored higher in a post-test than females in mixed-gender pairs. The authors argued that in synchronous text-based learning systems, females may benefit from same-gender collaborations compared to mixed-gender collaborations. However, it is unclear to what degree the reported effect was confounded by the influence of stereotype threat. Indeed, females undertaking mathematics, physics and spatial tasks face social prejudice and pressure that impair their performance (see Chapter 2, final part of section 2.3.3 and section 2.7.2). However, it has been previously exemplified that reducing stereotype threat with special experimental manipulations can successfully narrow the performance gap in even the 'hardest' spatial tasks (Brownlow et al., 2011). In the study reported in this thesis, the stereotype threat is circumvented by masking the human partner altogether. The findings of this study clearly showed a strong effect of pair composition in three respects. First, as discussed in a previous section (section 6.4.2), female-only pairs exhibited strong collaborative and adaptive behaviour and upheld the principles of route communication (as listed in the CORK framework, see section 2.5.1 in Chapter 2), when the interaction conditions were impoverished. The pairs exchanged information-rich messages of high granularity, explicitly acknowledging correct execution and understanding, female users increased the number of questions, and interaction responsibilities for task maintenance were equally distributed. These results empirically confirm the hypothesis by Savicki et al. (2006) that female-only groups will overcome the paucity of CMC settings by putting to use their superior linguistic skills timely and as necessary. Moreover, it is argued

that the experience and output of female collaborators may benefit from environments that make social cues less salient and relevant. In fact, research in computer science education indicated that programming in pairs has narrowed the gender gap in performance between male and female novice programmers and reduced failure rates for students of both genders (Berenson et al., 2004; McDowell et al., 2003). Second, as discussed in section 6.2.1, there was a pair composition effect in terms of miscommunication, with users in mixed-gender pairs failing to provide accurate instructions. Users in mixed-gender pairs were also more likely to omit action boundary information in their instructions, which could be disruptive to the performance of their partners. Therefore, this study adds to past research that identified inter-gender miscommunication in social contexts and offered the interpretation that it arises as result of the divergent communication and coordination styles of females and males. Third, as discussed in section 6.5.1 above, mixed-gender pairs were more weakly aligned compared to same-gender pairs and converged on a wider vocabulary.

6.5.2 Alignment in Human-Computer communication

There are at least three important reasons for seeking to better understand and characterise alignment in human-computer dialogues. First, clear insight in processes that play a part in the interaction between users and computer systems may help inform more naturalistic system designs. Second, if alignment is indeed a precondition for communicative success, systems that do not support this mechanism are destined to fail. Third, alignment may help 'prime' desirable user input and inhibit out-of-vocabulary words. As such, five research questions were framed to clarify whether: alignment occurs in HCI (Research Question 7(a)); it is mutual (Research Question 7(b)); it is influenced by changes in the affordances of the communicative situation (Research Question 7(c)); it is locally disrupted by miscommunication (Research Question 7(d)); it impacts user perceptions of interaction success (Research Question 7(e)). Extending Research Question 7(d), the analysis also sought to discover whether lexical innovation tendency, especially after miscommunication, depends on gender (Research Question 8(e)).

The results indicate that alignment is present, resulting in the gradual reduction and stabilisation of the vocabulary-in-use, and that it is also proportional (users that strongly aligned to robots, who also strongly aligned to them) The results also indicate that alignment

in human-computer interaction may involve strategic component, being used as a resource to compensate for less optimal interaction conditions. Further, the results suggest that when system and user errors occur, the development of alignment is temporarily disrupted and users tend to introduce novel words to the dialogue. This effect is, however, highly dependent on gender. Moreover, lower alignment (particularly, in system-generated input) is associated with less successful interaction, as measured by user perceptions. These issues are elaborated in the following sub-sections where the findings from this study are translated into design recommendations which are subsequently used to inform the development of a framework of dialogue management that incorporates linguistic alignment.

Alignment in human-computer communication develops early and reciprocally

Section 5.8.2 in Chapter 5 reported a one-to-one coupling of user and 'robot' inputs at the adjacency pair level. The analysis demonstrated a trend, according to which the more aligned one participant is, the more aligned their partner will be. Hence, it is likely that a computer dialogue system which consistently matches the input of the user will trigger similar user tactics. In turn, as these expressions become grounded, the use of different lexical items by the user may well be more inhibited. In addition to local priming, the analysis in section 5.8.1 demonstrated its operation over the course of the dialogue: the interlocutors, although presented with different landmarks and environment configurations during the session, began to rely more and more on previously-used expressions. This led to a small-sized working vocabulary that peaked and stabilised after only 70 dialogue turns. As such, speakers simply drew from the preceding dialogue to formulate future utterances. Taken together, these observations provide strong evidence that alignment operates in human-computer dialogues through both local priming and longer-lasting alignment of vocabulary.

In summary, there is symmetry in the linguistic input and output of system and user which gains stability over time. That is, the user aligns with the system and the system aligns with the human at the utterance pair level, which eventually results in a relatively stable set of expressions that are being re-used. As such, alignment appears instrumental in addressing the 'Vocabulary Problem', allowing prediction and constraint of the linguistic input of the user. These observations suggest that, through their output, dialogue managers should seek to prime users such that they are more likely to input in-grammar terms and structures.

Production and interpretation are coupled processes, so system prompts should contain no syntactic or lexical items that the system itself cannot interpret. In addition to this, specific design issues arise with regards to how the system's dialogue manager supports lexical alignment in order to restrict the vocabulary in use, and these will be considered in the final subsection of section 6.5.2 as part of a proposed dialogue model.

Lack of alignment is linked to lower user perceptions of task success.

Previous work in human communication emphasises that linguistic alignment is the basis of stable, successful communication (Pickering and Garrod, 2004; 2006). Indeed, a study by Reitter and Moore (2007a) in which people collaborated in a spatial task offers empirical support , reporting a strong correlation of task success and long-term alignment of syntactic structures, though no effect was found for local priming. The authors concluded that lexical and syntactic alignment is a reliable predictor of task success, and that 'successful dialogue requires syntactic alignment' between human interlocutors in a spatial task (Reitter and Moore, 2007b, p. 1).

The question that naturally follows from the analysis of relevant work in human communication, and which motivated this study, is whether alignment is also a precondition for successful communication with computer systems. The results presented in Chapter 5, section 5.8.5 suggest that it is; demonstrating a link between lower perceived task success and lower lexical alignment achieved by the end of the dialogue. While there is literature that reports that systems that aligned to their users in terms of prosodic or other paralinguistic elements are rated more positively (e.g., Nass and Lee, 2001; Bailenson and Yee, 2005), to the author's knowledge, no other study has presented evidence that more effective interactions are possible when systems align to users. Taken together, these findings present a potential effect of alignment on perceived communication success. In effect, they reverse the priorities, bringing the role of system-generated responses into the foreground, and suggest that alignment by the computer system is of key importance to the success of the interaction. As such, though important, system prompts designed to prime the user to provide desirable input (as recommended in the first part of section 6.5.2above) may not suffice to yield effective interactions. Rather, it is suggested that alignment can be instrumental in interaction success if the system is also primed to repeat user output. This suggests that,

through their output, dialogue managers should seek to repeat user outputs to promote alignment. This recommendation will be revisited in the final part of section 6.5.2to explore its place in the development of a dialogue management model.

While interesting for the purposes of this exploratory study, these results remain preliminary, given that they were produced by correlational analyses. On the basis of the results, it is possible to argue for an association, but it remains unknown whether high miscommunication and low success perception is because of low alignment. To give evidence of causation, it would be necessary to replicate this study using appropriate experimental manipulations in order to test the directional hypothesis that *'aligned robot responses reduce miscommunication and increase user satisfaction'*. As mentioned in the in section 4.3.3, this could be achieved by the replication of the study involving two groups of trained 'robots' instructed to either systematically repeat the same lexical items as the user or use different forms, and measuring the effect in terms of user perceptions and frequency of miscommunication.

Such approach essentially follows the software development methodology for dialogue systems proposed by Levin and Passonneau (2006), which applies the AI concepts of *ablation* and *comparison* in WOz studies. That is, a number of studies are performed that involve incrementally restricting the capabilities of a human wizard in the direction of a dialogue system, starting from natural, unconstrained WOz setup to a fully automated one. Thus, by removing one dimension of the wizard's communication resources and replacing with an automatic component at a time, it is possible to develop focused hypotheses that aim to identify which of these features and strategies have the most severe impact on specific aspects of efficiency, effectiveness and user satisfaction.

The effect of visual information on alignment in HCI

Studies by Brennan, and Branigan and her colleagues (discussed in Chapter 2, section 2.8.3) have demonstrated strong presence of linguistic alignment in HCI which suggests that it is an automatic mechanism that invariably manifests in communication. Later research has added that it is also a strategy that is consciously-employed based on the speaker's beliefs about the linguistic competence of the interlocutor (for example, in the case where users aligned more

to ‘basic’ computers than to ‘advanced’ ones and more to computers than to human partners (see Pearson et al. 2006)). As one explanation, Branigan et al. (2010) have suggested that since computers are perceived as less competent interlocutors, alignment is more prevalent in HCI than human-human interaction, and has a stronger strategic component. Unifying this body of results, Branigan et al. (2011) concluded that lexical alignment is mediated by beliefs about interlocutors, and that speakers align more strongly when they believe that this will facilitate interaction success.

It is difficult to interpret the findings of the present study to contribute to the debate around the nature of alignment. Yet, from a different standpoint, they reiterate the conclusions offered by Branigan and her colleagues. The analysis presented in section 5.8.3 showed that the extent of alignment in HCI was determined by the interaction condition; in particular, alignment was prevalent when visual feedback was absent, and yet comparatively scarce in the condition of visual co-presence. When users could not readily establish joint reference, monitor task status or have instantaneous evidence of the system’s understanding and execution, speakers aligned more strongly. Therefore, the results of this study add weight to those previous findings that argue that alignment in HCI is used when communication success appears to be at risk and as a ‘safeguard’ against a perceived elevated likelihood of miscommunication.

From a wider practical perspective, awareness of how visual information affects collaboration and communication patterns is important for the design of CMC, CSCW systems and agents in situated interactions. Previous studies in CMC have discussed how visual information (particularly of the work area) increases awareness of the current state of the task and facilitates conversation and grounding, such that interlocutors can use linguistic shortcuts and simpler language (see, for instance, Gergle et al., 2013). It was found that it profoundly changes the structure and content of dialogue, since utterances may be substituted or complemented by actions and gestures[24]. Inspection of the dialogue corpus of the present study reiterates these observations and extends them to the domain of human-computer

[24] As expected, in the Monitor condition, many responses by the robot were carried out through a physical action rather than verbal means. It should be clarified that this was not annotated as a mismatch.

dialogues; when visual feedback is withheld, interlocutors tend to use more explicit and longer utterances (i.e., more words). Adding to this, the results of the analysis showed that users and 'robots' also systematically repeat each other's words. This is exemplified in the dialogue excerpt in Table 6.6 below.

Table 6.6: Dialogue excerpt from the No Monitor condition [NMF4_T82-93].

Speaker	Utterance
U:	Now turn to your right, walk straight ahead until you reach the ***road junction***
R:	Yes, I am at the ***road junction*** now
U:	Once you are at this ***road junction***, please turn to your left and walk straight ahead until you reach a right turning in the road
R:	Ok, I can see a ***right turn***
U:	Good, please turn right at this ***right turn*** and follow the road until you reach a ***roundabout***
R:	Ok
R:	I am in the ***roundabout*** now
U:	Good, you are at the ***roundabout.***
U:	Is there a ***car park on your left***?
R:	Yes
R:	The ***car park*** is ***on my left*** now
U:	Please turn to your left and take the first left exit off the roundabout, keeping ***the car park on your left***

Qualitative examination of the dialogues also revealed an interesting phenomenon; in a few exchanges in the No Monitor condition, users were inclined to repeat even the erroneous or idiosyncratic vocabulary of the 'robots' (as illustrated in the example in Table 6.7). This observation appears to echo the findings from the aforementioned studies by Branigan and colleagues, which indicated that human users tended to repeat the term that the computer used, even if it was less accurate or normal. As noted in the previous section, it may be worthwhile to explore the validity and extent of this phenomenon and its specific impact on aspects of interaction success using a controlled experimental setup, in which 'robots' systematically produce such terms.

Table 6.7: Dialogue excerpt from the No Monitor condition [NMF_T69-73].

U:	turn left
R:	Go pass the bridge?
U:	go forward
R:	**Go pass** the town hall?

U:	**go pass** Tesco

This study, then, has illustrated the impact of visual feedback on the patterns of communication when other parameters are kept the same. The results suggest that users align more strongly to systems when visual feedback is not possible, increasing the necessity to implement linguistic alignment capabilities in the dialogue manager of systems that are not physically or visually co-present with their users.

The effect of miscommunication

Miscommunication is a natural and ubiquitous phenomenon within communication, both between humans and, perhaps even more so, in computer-based dialogue systems. In interaction with such systems, miscommunication manifests as user errors, system errors and non-understandings. The ability to predict what users will do in terms of linguistic choices after the occurrence of errors is a matter of enormous practical significance. Addressing Research Question 7(d), section 5.8.4 in Chapter 5 explored how users reacted when they detected miscommunication.

It was found that, after miscommunication, users were more likely to use new words, whereas successful utterances were typically followed by responses that exclusively reused lexical items from the dialogue history. A simple explanation of this phenomenon is that, as the dialogue progresses, interlocutors build up a body of aligned expressions that seems to be mutually intelligible and that functions successfully. When miscommunication occurs, interlocutors lose confidence in the efficiency of these expressions and the interaction as a whole and introduce new expressions. This user behaviour was more pronounced when visual feedback was absent. This is likely to be because visual evidence offers a more effective and economic way of grounding compared to verbal-only evidence (Brennan, 2005). Thus, it can be argued that the status of lexical items that are grounded under a visual co-presence condition is less susceptible to the impact of miscommunication.

Two specific recommendations can be drawn from these findings. First, as suggested in the third part of section 6.5.2, dialogue managers should account for different interaction conditions of visual and verbal-only feedback. In particular, when miscommunication is detected in visual co-presence conditions the system should adhere to the vocabulary

established in the course of the dialogue. In verbal-only conditions, the system should anticipate novel words in the user input, and 'expect' departure from those previously recorded in the dialogue history.

The second recommendation concerns the miscommunication (or error) handling functionalities of the dialogue manager. The efficiency of dialogue systems is often compromised by their inability to detect speech recognition and language understanding errors. In turn, it has been found that humans do not typically provide explicit cues that a misunderstanding has occurred, but prefer implicit strategies such as reformulating their statements or even moving on (Skantze, 2005; Koulouri and Lauria, 2009; Bohus and Rudnicky, 2005). Therefore, the detection of out-of-vocabulary words may be used by the dialogue system as an indicator that an error has occurred. These recommendations will be incorporated in the dialogue model discussed in the final part of section 6.5.2.

Gender and miscommunication

As explained above, users tended to introduce new words in the dialogue when execution errors and non-understandings were detected. Research question 8(e) sought to clarify whether this effect is mediated by gender. It was found that while male users had stronger tendencies to try novel expressions after miscommunication, female users preferred to adhere to the old vocabulary. This suggests that females are more conservative and males more explorative when handling communication breakdowns. It is interesting to note that even under problem-free communication, females were more likely to re-use previous words than men. While there may be many possible explanations, these findings are argued to relate to gender differences in risk and cost perceptions. In the HCI domain, this translates to a user being less willing to try a useful but unfamiliar feature. Previous research argues that females perceive higher risks when they are involved in decisions or situations (for example, Finucane et al., 2000; Blais and Weber, 2001), especially in tasks that involve mathematical or spatial reasoning (Byrnes et al., 1999). As such, several researchers have hypothesised that females will be less likely to explore and experiment with unfamiliar features compared to males. Studies in various application domains, from programming IDEs (Beckwith et al., 2006a; Burnett et al., 2010; Cao et al., 2010) and spreadsheet software tools (Burnett et al., 2011) to web-based databases (Rosson et al., 2007) have confirmed that females are less

confident to use novel software features while men typically engage in exploratory behaviour. Taken together, it is argued that females' tendency to reuse vocabulary and not attempt a new strategy even when these messages ostensibly failed forms part of their general fear of 'tinkering'. The fear of trying new features has also been traced back to females' perceptions of low confidence and self-efficacy, which were also reported in this study (see section 6.2.3) (Beckwith et al., 2006b). Such findings should be considered by interface developers so that when unfamiliar or new features and strategies have to be adopted in the interaction with a system, techniques such as tutorial snippets, examples of what to say/do and short strategy explanations may help some users to feel more comfortable to utilise them. In a gender-neutral interface, such features should be customisable in order to avoid compromising the experience of a gender.

Towards an alignment-driven approach to dialogue management

Dialogue systems are typically based on modular pipeline architectures. Depending on the application domain, a basic architecture consists of modules for natural language understanding (including components such as speech recognition and language parsing), natural language generation (including speech synthesis) and dialogue management. In the case of spoken dialogue systems, the speech signal is captured and the speech recogniser produces a hypothesis which is passed to the natural language understanding (NLU) component. Speech recognition and NLU typically use language modelling to predict the next word given the identities of the previous words. The NLU component parses this input and submits a semantic representation to the dialogue manager, which determines the next system action, based on the dialogue history and other knowledge sources. This action is forwarded to the natural language generation (NLG) component which creates a system response. The speech synthesiser outputs the response. Text-based dialogue systems omit speech recognisers and synthesisers but use the rest of the core architecture. The NLU and NLG components typically use static data from application-specific grammars and lexicons – the set of allowed structures and words (sometimes collectively referred to as grammar). The dialogue manager also makes use of the same linguistic resources. Figure 6.2 summarises the interactions between the modules in such an architecture.

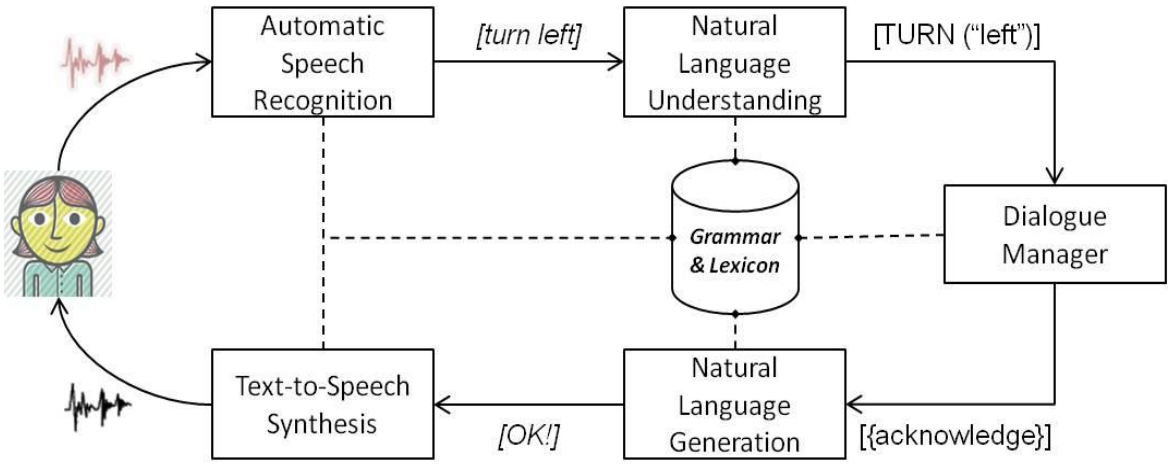


Figure 6.2: A typical dialogue system architecture. Text-based dialogue systems omit the speech recognition and synthesis modules.

The dialogue manager determines the next system action, based on the current input, dialogue state, dialogue history, task status and some dialogue policy. It then updates the dialogue and task states accordingly. Communication with an external database is also involved for various applications. In effect, the dialogue manager is responsible for planning and maintaining the flow and development of the dialogue (Bohus and Rudnicky, 2009). These functions are non-trivial and, as such, have led to a proliferation of different approaches to dialogue management (see McTear, 2002, for a review).

The dialogue manager supports some basic natural conversation phenomena. For instance, systems allow for turn-taking, where the system and user may make successive contributions, and often support user interruptions and back-channels. Another function of the dialogue manager is miscommunication (error) handling. This includes: error prevention, by appropriate design of system prompts (e.g., Cohen et al., 2004); detection, monitoring the dialogue for cues that might indicate that an error has occurred (Skantze, 2007); prediction, monitoring features in the dialogue so far to predict problems (e.g., Walker et al, 2000b); and recovery, system strategies that include asking the user to repeat or rephrase, or issuing relevant questions (e.g., Bohus and Rudnicky, 2005). Many systems may also use 'grammar switching' in order to improve the accuracy of speech recognition and natural language processing (Lemon, 2004). In particular, form-based, finite-state and Information State Update dialogue managers activate different language models for different slots, states and contexts in the dialogue, respectively. For example, if the latest dialogue turn is a yes/no

question, the language model activated at that point will be defined by a context-free grammar that only covers utterances like ‘yes’, ‘no’, ‘that’s right’, etc.

After this brief review of the architecture and technologies of a typical dialogue system, this section concludes by incorporating the recommendations detailed in the previous sub-sections into a high-level dialogue model for task-oriented interactions with a computer-based system. In particular, the model focuses on the dialogue manager’s interaction with the grammar for determining the content of the next system action and adapting the lexicon, as a result of the processes of linguistic alignment on which the study reported in this thesis focused. Thus, all other dialogue management functions are treated as black boxes. Finally, it should be noted that although this study and framework focus on lexical alignment, alignment is expected to operate in identical ways across all other linguistic levels. Therefore, although the following account deals explicitly with lexical items, it could be extended to apply to, for example, syntactic structures.

Let us assume a basic dialogue manager that considers three sources of information: task state (information required to complete a task); user commands; and system responses. Its operation based on the proposed model will be illustrated though a simplified dialogue example from a human-robot supervised navigation scenario. The dialogue example corresponds to a task completed within one transaction: a user utterance instructing the robot to turn left at a junction, and the robot executing the instruction. Based on empirically collected data, the environmental feature, *junction*, was more or less accurately referred to as ‘v-shaped junction’, ‘three-way junction’, ‘y-junction’, ‘intersection’, ‘crossroad’, ‘cross junction’, ‘fork’ and ‘t-junction’ by different users (as observed in this study). At the beginning of the dialogue, the grammar contains all possible synonymous lexical items. A weighting feature is assigned to each lexical item, indicating its frequency of use in the dialogue. Thus, all lexical items begin by having equivalent weightings.

The user initiates the interaction using the instruction ‘*turn left at the fork*’. At this point, there are three communication outcomes: correct understanding; non-understanding; or misunderstanding. In the cases of correct understanding and non-understanding, the system gives positive or negative evidence of understanding, respectively.

First, in the case of correct understanding, the dialogue manager triggers a verbal acknowledgement followed by the physical action of the system. The execution is based on particular expressions that referred to actions and objects in the interaction situation. If the understanding was indeed correct, as evidenced by the user acknowledging successful execution, the expression is taken to be conceptually-equivalent for both user and system to refer to the relevant actions and objects. As such, the dialogue manager should perform two *grammar updates*, which reinforce the use of this lexical item in subsequent similar situations: (i) *the expression should be mapped to a particular situation (object or action);* and *(ii) the expression's weighting should be increased, meaning that it will subsequently be favoured over synonymous expressions in the grammar.*

Then, following the basic 'input/output alignment' principle in the Interactive Alignment Model and the recommendations in the first two parts of section 6.5.2, the system should immediately repeat the expression by generating a verbal acknowledgement which reinforces the expression used (i.e., *'I have turned left at the fork'*). This system output, in turn, should further prime the user to re-use the expression to refer to this object, inhibiting the use of any alternative term. This will eventually lead to the particular expression becoming 'fixed', and routine for this dialogue (Pickering and Garrod, 2004). As described in Chapter 2 (section 2.8.1), 'routines' (following the Interactive Alignment Model) or 'conceptual pacts' (following the Collaborative Model) are linguistic constructs that are agreed between the interlocutors to refer to an entity in the situation model.

Following the process described so far, as the dialogue progresses the working grammar will be gradually reduced in variation and size, with some expressions being dispreferred and others being favoured until, ideally, the grammar becomes stabilised and only consists of dialogue routines. It is proposed that the downsized working grammar should feed back to the speech recognition component, which can incorporate the positive weightings of the frequently-used lexical terms to re-score the recognition hypotheses lattice or list of the language model. This approach could complement existing grammar switching techniques for tuning language models (as outlined above). Following these techniques, if recognition fails based on the re-scored language model, user input is re-processed using the original language model and grammar (as described in Lemon, 2004). As demonstrated in Hockey et al.'s (2003) system, it is also possible for recognition to run simultaneously using both approaches.

Second, in the case in which the instruction is not understood, the dialogue manager will implement the strategy specified in the error recovery module of the dialogue manager. As mentioned above, these strategies include asking the user to repeat or rephrase the problematic utterance, or, if the system has advanced inferential capabilities, asking task-level reformulations, such as 'turn left after the bridge?' (see Bohus and Rudnicky, 2005; Gabsdil, 2003). The results in this study suggest that when miscommunication occurs, users lose confidence in the efficiency of established dialogue routines and introduce new expressions (see fourth part of section 6.5.2). Therefore, in case of non-understanding, the initial system response should be not to increase the weighting of any expressions used. Similarly, no grammar update is performed in cases of misunderstanding (execution errors in the user/robot scenario from the study in this thesis). The recommended dialogue system actions for the three communication cases discussed are summarised in Figure 6.3.

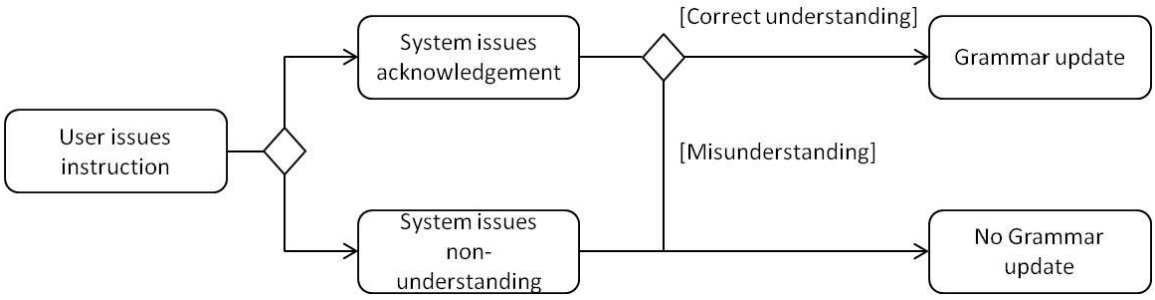


Figure 6.3: Model of a dialogue sequence showing the three communication outcomes.

The process of routinisation, discussed in the case of correct understandings, not only simplifies language understanding and generation, but also allows the system and user to take advantage of the *local principle of contrast* (Clark, 1993; Pickering and Garrod, 2004; 2006). This means that if the user decides to use a competing expression for the referent, for instance, 'crossroad', instead of the established term 'fork', the system should deduce that the user is not referring to the same type of intersection but trying to introduce a new concept. In effect, the dialogue routine that consisted of a linguistic expression and a referent is questioned, triggering a clarification sub-dialogue by the system (i.e., the system might ask: 'When you said "crossroad", did you mean "fork"?'). In such cases, Brennan (1996) found that when a system attempts to clarify user input, users naturally adopt the system's clarified term in their subsequent utterances (with users aligning in 94% of cases).

In terms of dialogue management, this clarification sub-dialogue has two possible outcomes: the user accepts the expression or rejects it. In particular, if the user replies 'yes' to the example question given above, updates for both expressions should be performed and, consequently, 'fork' will remain more highly-weighted than the recently used 'crossroad' (assuming that was the position in the dialogue before this point). This instance should then lead to the generation of the verbal acknowledgement by the system, *'I have turned left at the fork'*, reinforcing the more highly-weighted word.

In the case of the user replying 'no', and thus accepting 'crossroad', this lexical term should be associated with the situation and its weighting increased as part of the two-stage grammar update (as previously described). Following the classification of non-understandings by Hirst et al. (1994) and Gabsdil (2003) (see section 4.4.2in Chapter 4), this process of clarification and update corresponds to when the system has obtained uncertain or multiple interpretations. As explained above, the third case of non-understanding (that is, the system obtained no interpretation) naturally leads to no grammar update by the system. Taken together, the model of the dialogue presented in Figure 6.3 can be refined to incorporate the clarification sub-dialogue (see Figure 6.4).

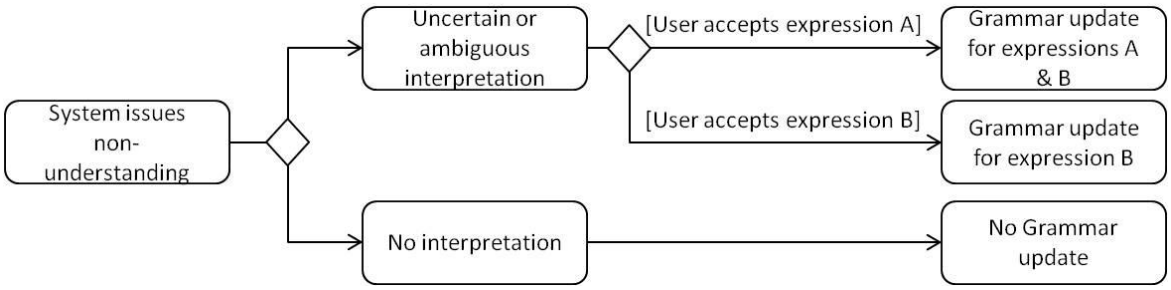


Figure 6.4: Refined version of the 'system issues non-understanding' model component showing clarification sub-dialogue options. Expression A denotes the previously used term for a referent while expression B denotes a novel alternative term.

The observation that users tend to introduce new lexical items when they perceive an error may also be translated into a guideline for late error detection. Error detection involves monitoring the dialogue for cues that an error has occurred, and is classified into early and late error detection. Early error detection techniques usually rely on the speech recognition confidence score. Additional knowledge sources may be used, including features from the

NLU and dialogue manager components. For example, in Walker et al. (2000b), errors were detected by online monitoring of features like degree of context shift, parsing confidence, grammar coverage and preceding system prompt, while Litman et al. (2000) employed prosodic cues, and Skantze (2005) used dialogue and referent history. In late error detection, the system error is detected after execution or after several turns. The typical technique for late error detection is to look at what the user actually said. Krahmer et al. (2001) refer to these cues as 'go on' versus 'go back' user instructions. Their approach is to monitor the content of user turns for a combination of cues such as longer turns, particular word order, repetitions, rejections and instances of no new information being provided. Hence, drawing on the recommendation framed in the fourth part of section 6.5.2, it is suggested that late error detection approaches should include testing for the presence of alternative lexical items in user turns (that is, words that currently hold lower weightings in the grammar compared to the most frequently-used expression for a situation) as a valid negative cue to detect errors.

In summary, this subsection has outlined a high-level model to illustrate how linguistic alignment can be supported by the dialogue manager. The dialogue manager performs two types of updates as a function of the usage of an expression over the course of a single dialogue: it creates an association between the lexical item and a referent; and changes its weighting within the lexicon. The model has also integrated miscommunication-handling. Possible benefits of the suggested approach include: enhanced recognition accuracy, owing to rescoring of word probabilities based on their weightings; improved intelligibility of system generated output, owing to it consisting of recurring words; and user interaction with the system that is more natural and cognitively easy.

It should be noted that the model is generic in order to serve as a springboard for researchers to incorporate linguistic alignment in detailed dialogue models. In effect, using this procedure of semantic updates, the model can be extended to enable the development of idiosyncratic word uses between system and user. It can also be used as a theoretical account for developers of dialogue systems, who can make decisions with regards to how these updates should be performed based on the specific system architecture and computational algorithms implemented (for instance, techniques involving neural networks or latent semantics analysis).

Following the Interactive Alignment Model, the proposed dialogue model assumes all linguistic choices are driven by purely mechanistic processes of priming and repetition. As such, it excludes more complex alignment phenomena resulting from social considerations (as described by the Communication Accommodation Theory), such as persuasiveness, authority and politeness, and from interlocutors modelling each other's knowledge states (as argued by the Collaborative Model). Future research efforts should focus on the integration of all three dimensions of alignment; that is, supporting mechanistic priming, social considerations, and dynamic inferential grounding processes.

6.6 Chapter summary

This chapter has provided a reflection on the findings of the empirical study using two viewpoints; it discussed gender differences in performance and language use from a human communication perspective and an HCI perspective. It exemplified how interactive communication modulates gender differences. It illustrated how visual information affects the interaction through situation awareness and grounding, and how these effects depend on gender. Implications for communication theory development, collaborative and dialogue system design and navigation in online environments were presented. It clarified the nature of alignment, which was found to depend on gender and pair composition, and its practical relevance for natural language interfaces was demonstrated. The following chapter draws the preceding chapters together, presents overall findings, discusses the contributions that are made by the research, describes the limitations that have been identified and details potential avenues for future work.

7 Conclusions and Future Directions

7.1 Introduction

Although gender is among the most influential of the factors underlying differences in spatial abilities, human communication and interactions with and through computers, our understanding of its influence on these areas remains largely incomplete. This thesis has argued about the importance of gaining better awareness of gender differences. To this end, it has reported a multidisciplinary investigation into gender differences as they pertain to spatial navigation and communication with humans and computer systems.

Past research has offered important insights into gender differences in navigation and language, but, having been derived from non-interactive or artificial experimental settings, its generalisability is argued to be limited. Moreover, little is known about differences in gender strategies and preferences in various domains of HCI, including collaborative systems and systems with natural language interfaces. Targeting these gaps, the thesis has aimed to address the main question of how gender differences emerge in navigation dialogues, through a process of teasing apart the elements of navigation and communication and formulating specific research questions. The thesis revolved around the qualitative and quantitative analysis of performance and dialogue data collected using a custom system that supported synchronous navigation and communication between a user and a simulated robot. It used experimental evidence to describe the key role of direct interaction in gender differences in performance and communication and to illuminate the phenomena of linguistic alignment, miscommunication and visual co-presence (originally, an experimental manipulation) in human communication and HCI and how they are mediated by gender. In particular, the thesis has produced three sets of contributions: methodological; theoretical; and practical. The methodological contributions resulted from the use of dialogue as a research paradigm.

The theoretical contributions articulated novel findings in gender differences in navigation and communication, the role of visual information in collaborative interactions, and the dynamics of communication. The practical contributions include design guidelines for natural language interfaces and implications for the development of gender-neutral interfaces in specific domains of HCI.

This final chapter begins by a review of the central findings within each of the three areas of contribution (sections 7.2 and 7.3). In section 7.4, reflections are offered with regards to the limitations of the research, which motivate further research. The chapter concludes the thesis by discussing potential avenues for future work.

7.2 Research questions and central findings

This section briefly revisits the high-level findings in response to the thesis' research questions before enumerating their theoretical, practical and methodological implications.

Research Questions 1(a) – 2(b) explored the predictions that females are less accurate and efficient than males in navigation tasks and rely on landmarks for instruction giving and following. The results indicated that such observations may be too simplistic to occur in dyadic interactions. It was found that users in single-gender pairs are more accurate and that users of *both* genders gave landmark-rich instructions to female 'robots'. Research question 3 sought to verify the recurring phenomenon of females rating their performance lower than men in spatial and computer-based tasks; it was confirmed that, regardless of actual performance, female users' perception of success was lower than males'. At the same time, females gave higher scores to the system than did males.

Research Questions 4(a) – 5(b) investigated past findings from task-oriented CMC and human communication studies with regards to the effect of visual information on performance and discourse patterns in the interaction with a computer system. The findings substantiated a robust effect of visual information, but did not fully support the argument that visual information leads to more successful communication. Several findings resonated with past research in that visual evidence allowed users and 'robots' to collaborate using underspecified and less detailed utterances, lacking important elements like landmark references, boundary information and frame of reference. Responsibilities for maintaining the

task and mutual understanding were assumed by users, such that the discourse resembled a cycle of user instruction followed by 'robot' execution, reserving additional contributions only when miscommunication was at hand. Yet, other findings contradicted the expectation that visual information leads to faster and more accurate interactions; miscommunication was reported to be more frequent in this condition and no significant difference was found in terms of time. The use of deictic expressions was low across conditions, which implied that sharing visual workspace with a robot does not offer the same affordances for grounding as in human-human interaction.

Research Questions 6(a) and 6(b) extended the investigation of the effect of visual information to the gender factor in an effort to determine whether gender accounts for the magnitude of the effect; in particular, whether females are more susceptible or responsive to the absence of visual information (poorer interaction conditions). The expectation was that the performance of females would drop when visual information was withheld, but no significant differences were found across measures (such as miscommunication, time, number of turns, etc.). Most importantly, it was found that significant differences due to the effect of visual information on communication processes were exacerbated between female users and all-female pairs in the two conditions. In the condition deprived of visual cues, female users and all-female pairs adopted a highly interactive and collaborative style, providing more information, acknowledgements and queries, compared to females in the Monitor condition. Male behaviour remained consistent across conditions. These findings provided strong evidence that females are more attuned to task conditions and suggested that the anticipated performance detriment due to lack of visual information was avoided by employing richer means of communication and collaborative strategies. Moreover, the findings further hint at a single-gender pair composition advantage.

Research Questions 7(a) – 7(e) aimed to provide a better understanding of the occurrence and characteristics of lexical alignment in HCI. The results indicated that alignment occurs reciprocally in HCI, and leads to the gradual downsizing and stabilisation of the vocabulary-in-use. Further, it was found that the development of alignment is locally disrupted and users tend to introduce new vocabulary to the dialogue when miscommunication is detected. The results also indicated that alignment may be used to maintain understanding under impoverished interaction conditions (lack of visual feedback). Finally, lower alignment was found to correlate with poorer ratings of task success.

Research Questions 8(a) – 8(e) explored the uncharted area of whether the strength and development of lexical alignment depends on gender and pair composition. Observations from studies in phonetics and sociolinguistics suggested that females align more strongly than male speakers and to male addressees. The findings in relation to research questions 8(a) – 8(c) did not find significant effects attributed to the genders of the user or 'robot', but of the interaction of genders. In particular, single-gender pairs appeared to be more aligned at the adjacency pair level and were able to 'agree upon' a more concise vocabulary. While no focused attempt was made to clarify the direction of alignment, research question 8(d) investigated whether users would align their gender-related communication strategy to the strategy of their interlocutor. It was reported that male users employed landmark-based instructions to navigate female 'robots'. This finding indicated that while people of different genders may have preferential strategies, they switch depending on the interaction situation. Finally, research question 8(e) offered insight into the behaviour of genders when communication breakdowns occur. It was found that while male users had stronger tendencies to try new words after miscommunication, female users preferred to fall back to the old vocabulary. In fact, males were more likely to introduce new expressions at any point in the dialogue. This finding suggests that females exhibit more conservative behaviour and males are more explorative when interacting with a computer system.

While past research has provided consistent answers about how gender differences arise, these patterns are highly skewed when gender is considered in dialogue. In fact, it appears that interactive mechanisms have the capacity to moderate some gender-related disadvantages. Taken together, the findings uphold the central hypothesis of the thesis that gender differences will be modulated within interactive communication.

7.3 Contributions

The contributions of this work are situated at the intersection of the fields of HCI and human communication and relate to gender, alignment and visual information. They are divided into three main categories: theoretical; practical; and methodological. Section 7.4.1 discusses how these findings add to theoretical frameworks of human communication and differential research in gender. Section 7.4.2 reports the practical relevance of these findings for a range

of HCI applications. Section 7.4.3 presents the contributions that can be drawn from the experimental approach employed in this study.

7.3.1 Theoretical contributions

The theoretical implications of this work mainly address gender differences. In addition, important insight was gained with regards to the effect of visual information in coordination. Finally, the findings advance our knowledge with regards to route communication protocols and practices. In particular, the thesis made the following theoretical contributions:

i. The study empirically demonstrated that the interaction of genders has a greater impact on performance and communication processes and strategies than individual gender, and provided a novel account of gender differences in performance and language use in dialogue. It also presented unique evidence that the strength of linguistic alignment in task-oriented interaction depends on gender, showing that same-gender pairs are more strongly aligned.

ii. It clarified elements within the Collaborative Model for a new interaction domain, and refined the understanding of coordination mechanisms, through the investigation of the effects of visual co-presence.

iii. It produced a corpus of route instructions generated and interpreted by females and males in real-time dialogic interaction. The results of the analysis are valuable for existing monologue-based frameworks of route communication protocols.

Contributions in relation to gender differences in spatial navigation and communication

The interaction of gender has a greater impact on performance and communication processes and strategies than individual gender.

Many studies have identified robust gender differences in communication related to spatial navigation in real and virtual worlds. The majority of this research has focused on individual communication and performance, that is, how people either give or follow route instructions. This study identified a conspicuous gap with regards to research evidence from navigation dialogues and whether gender differences arise in the way they are conducted. This study

validated a strong gender effect in the dialogue domain. However, the findings illustrated that it is the interaction of genders– the combination of genders and role (i.e., instructor or follower) – that has the most significant impact. To the author's knowledge, it is the only study that has empirically and by design validated the theoretical claim that gender differences will be moderated in dyadic interactions.

Previous non-interactive studies showed male superiority in a variety of spatial tasks. This study suggested that males (as instructors, followers or both) are neither more accurate nor faster than females. Previous research has also argued that females and males have distinct and default strategies and communication styles, such that females employ landmark-based strategies to navigate others or themselves and have an overall 'cooperative' communication style, irrespective of interaction situation and addressee. This study demonstrated that the interaction of genders and context is found to be a stronger predictor of the choice of communication strategies and coordination patterns. Both females and males use landmark references only when the addressee is female, which implies that people are tuned to, and use, the strategies that suit the addressee's needs. Most importantly, female-only pairs appear to adopt a cooperative behaviour and assist each other by providing information-rich descriptions, balancing interaction responsibilities and explicitly acknowledging understanding and execution, but only when necessitated by interaction conditions. At the same time, in a visually-supported collaboration, females resorted to strategies and styles which would have been typically categorised as 'male'.

Therefore, the results of this thesis refine previous findings from studies in monologic settings or social conversations in two respects. First, they illustrate that a landmark-based strategy to give route instructions is not used exclusively and by default by females. Second, they show that a cooperative and interactive communication style is not automatically adopted by females, but instead they assume it when the task is more complex and the conditions impoverished. Taken together, the results of this thesis do not directly challenge past research, but complement it by demonstrating that any female disadvantage in accurate and efficient way-finding and direction-giving can be mitigated through dialogic interaction and its natural mechanisms.

Stronger linguistic alignment in same-gender pairs in task-oriented interactions

As indicated above, this study provided unique findings with regards to whether the tendency and strength of alignment in task-oriented interactions depends on the genders of the interlocutors. The study illustrated that same-gender pairs are more lexically aligned to each other than mixed-gender pairs. Females were not found to align more than males or to males; this finding does not necessarily contradict past research, but it may imply that alignment, either as an automatic mechanism or strategy, is used by women in overt, social interaction settings, but less so in goal-oriented collaboration with a computer system.

Contributions in relation to the effect of visual information on task-oriented interaction

Detailed empirical support to the Collaborative Model of human communication and further insight into coordination mechanisms.

Visual information was used as an experimental manipulation because there was sufficient evidence that it facilitates collaboration. The significant results produced by this manipulation provided support to specific concepts within the Collaborative Model of communication and additional insight to the mechanisms of human communication. First, the study added to the empirical evidence in favour of the contribution model (Clark and Schaefer, 1987) and against unilateral accounts of communication, such as Searle's (1992), that treat language production and understanding as two autonomous processes. The experimental conditions involved the presence or absence of a monitoring window on the user's interface, while the 'robot's' interface was the same. Large differences in language use by both participants were observed across the conditions. If language production was indeed an independent process, 'robots' would not modify their language in response to what their partners could/could not see. In harmony, although what the 'robots' knew did not change across conditions, users also provided more explicit and detailed route information. Second, the findings were consistent with the principles of mutual responsibility and least collaborative effort (Clark and Wilkes-Gibbs, 1986). In particular, when visual information was withheld, responsibilities for grounding were equally distributed, but when users could monitor the 'robot's' actions, the responsibility to decide that understanding had failed/succeeded shifted to the users, which reduced the communication effort for both (since 'robots' minimised queries and other verbal contributions). Third, the results validated Clark and Brennan's

(1991) framework on the costs and tradeoffs of grounding in different communication media; when visibility was removed, collaboration with the system became more effortful and both users and 'robots' needed to rely on complex linguistic means to maintain understanding. Fourth, the findings added to our understanding of the mechanism of grounding. Brennan (2005) suggested that grounding criteria are only as precise as they need to be. The rise of miscommunication in the visually-supported condition revealed that more or stronger evidence does not necessarily lead to more successful interactions, but successful interactions depend on how well people are able to tune their grounding criteria. So, when visual information was not shared, users navigated 'robots' closer to the destination before declaring successful completion. These speakers stipulated a stricter criterion, which minimised possibilities for miscommunication, while for interlocutors sharing visual information, the grounding criteria were set lower so that 'close enough was good enough', which at times did not suffice to ensure successful understanding. Fifth, it was observed that when visual information was shared, 'robots' communicated and grounded information through their movements, instead of vocal signals. This finding helps make the case that discourse models should also account for how physical actions ground meaning and how they are combined and integrated with language in task-oriented interaction.

Contributions in relation to route communication protocols

A large language and task corpus of route instructions spontaneously produced by females and males in real-time, situated interaction.

The experiments were designed to elicit natural dialogues which contained spontaneously-generated route instructions within a controlled spatial network. Two realistic scenarios, of supervised and unsupervised navigation, were used. As previously discussed, the large majority of studies explore spatial language and route instruction in monologue settings, with subjects following or formulating instructions independently. Yet, there is growing understanding that spatial language and dialogue need to be seen together (see Coventry et al., 2009). This corpus consists of spatial descriptions produced in dialogue and are, thus, dynamically formulated and revised to meet the current needs of the addressee and task state, and are synchronised with their execution. The corpus linguistics approach to the data analysis followed existing schemes and advanced our knowledge with regards to the choice

and frequency of components in route instructions. Moreover, it provides additional support to the principles and 'best practices' of route communication, as presented in the CORK framework (see section 2.5.1), extending their applicability in a dialogue domain. However, it demonstrated that the reliance on these principles is dramatically reduced when the interaction task becomes less demanding (for instance, through the provision of visual feedback). As discussed above, the results further indicated that people's reliance on them also depends on gender and pair composition; for example, all-female pairs relied heavily on these principles in deprived interaction conditions and mixed pairs consistently omitted boundary information.

7.3.2 Practical contributions

Owing to the rich effects of dyadic interaction and visual information, this work has produced results of immediate practical relevance for a wide range of applications. The following subsection outlines the main contributions in relation to the influence of gender in HCI. Moreover, this study provided insight into the mechanisms that visual information modifies performance and communication patterns, which has important implications for collaborative systems. Finally, the investigation of alignment led to unique insight into its operation in HCI, producing concrete design recommendations for dialogue systems. In particular, the practical contributions of this work were the following:

i. This study demonstrated that gender is a powerful factor in the novel domain of dialogue systems and human-robot collaboration, in task-oriented CMC, CSCW and virtual world navigation. It proposed that any female 'disadvantage' is mitigated through interactive mechanisms, which were identified. It also confirmed that female users employ conservative strategies (particularly after system errors), while males have more explorative behaviour, and that female users rate their performance lower than males, while rating the system more favourably.

ii. It described the ways in which visual information can enhance or disrupt collaborative work and communication, offering recommendations on how it can be used to best support interactions with situated agents, CMC and CSCW systems.

iii. It provided a complete account of the operation of alignment in naturalistic human-computer communication. It distilled the findings into design recommendations for

systems with natural language interfaces and proposed a simple dialogue model that leverages alignment.

Contributions in relation to gender differences in HCI

There is a documented gender divide in how people use and experience various technologies, such as in searching and navigating the web, using web applications (Large et al., 2002; Roy and Chi, 2003; Jackson et al., 2001; Chen and Macredie, 2010), and collaborative systems (Richert et al., 2011) and exploring and adopting software (Burnett et al., 2011) and hardware features (Czerwinski et al., 2002). Despite this body of evidence, in system design, 'the user' remains genderless (Bardzell, 2010). By making implicit decisions, developers run the risk of unintentionally creating technology that favours the needs or preferences of one gender while marginalising the other.

This study can be situated within the 'Gender HCI' subfield of HCI (Beckwith et al., 2006a), which consists of studies that focus on the differences in how males and females interact with systems and, by taking gender issues into account, how systems can be designed to be equally effective for both men and women (Fern et al., 2010). The position held in this body of research is that software design determines how well female problem solvers can make use of the software. Understanding how gender influences strategies, behaviours and success is the first step towards design that promotes successful behaviours and strategies by users of both genders. Following this paradigm, researchers have identified gender differences in VR navigation (Czerwinski et al., 2002), debugging (Fern et al., 2010), and spreadsheet software use (Burnett et al., 2011), and have implemented solutions to mitigate the impact of these differences. Along the same lines, this thesis aimed to delineate gender differences in terms of performance, strategies and perceptions in a new domain of HCI and articulate ways to moderate them.

Robust gender differences in the domains of dialogue systems and human-robot interaction, in task-oriented CMC and CSCW and virtual world navigation.

The comprehensive analysis of performance and dialogue data revealed robust gender differences emphasising the necessity of being aware of the influence of gender on the interactions with and through technological artifacts. First, the study reported in this thesis contributes to 'Gender HCI' by detecting gender differences in the novel domains of Human-Robot Interaction and spoken dialogue systems, which are prime examples of collaborative/goal-oriented interaction between humans and computer systems. Second, this work extended findings from studies in social computer-mediated interactions, and provided an account of communication processes and performance differences in task-oriented CMC and CSCW. Third, the results of this study may be useful for studies in virtual environment navigation. While previous research has shown that females' performance deteriorates in virtual world navigation, female 'robots' in this study had similar performance outcomes. It was attributed to collaborative aspect of the task – the fact that female 'robots' received suitable verbal directions – or even to the experimental manipulations, such as simple interface controls, salient landmarks and 2D environment. This resonates with a previous study that demonstrated comparable performances when females were given the opportunity to complement the visual and map aid configuration (preferred by males) with verbal instructions (Devlin and Bernstein, 1995). Therefore, it is argued that through neutral interfaces that offer the possibility of customisable settings (for instance, the provision of verbal aids) the needs and preferences of both genders can be met.

Interactive mechanisms reduce gender performance gap.

The theoretical insight that in dyadic interactions, females are able to do as well or better than males is of immense practical value. The results suggested that any female 'disadvantage' is offset by interactive mechanisms and pointed to dialogue elements that may equally benefit genders. Related empirical evidence can be found in other domains: studies in computer science education reported that pair programming reduced the gender performance gap between male and female programmers and failure rates for students of both genders dropped (Berenson et al., 2004; McDowell et al., 2003). However, there has been no focused attempt to pinpoint which elements of the collaboration underlie this effect. The results of this thesis help outline these elements. In this study, female 'robots' overcame the navigation barrier through receiving landmark-rich instructions. All-female pairs managed to compensate for

the ‘cueless’ interaction condition, by applying their verbal skills in a timely way and as necessary, and collaborated through elaborate, detailed messages. Female partners working in pairs exhibited strong collaborative and adaptive behaviour, putting more communicative effort and successfully reducing uncertainty and attending to the partner when the interaction conditions were poorer. The study also illustrated that same-gender pairs enjoyed efficient interactions, an effect which correlated with stronger linguistic alignment. Taken together, it is argued that the interactive mechanisms of partner and context adaptation and mutual lexical alignment are of key importance towards the development of gender-neutral natural language interfaces. Moreover, the observation that females perform better when using collaborative systems has important implications for the adoption of these systems in education and the workplace. This benefit may also relate to the fact that ‘social genders’ in task-oriented CMC become less relevant.

Female users employ conservative strategies (particularly after system errors), while males have more explorative behaviour.

The study yielded novel findings in relation to what females and males do when they are faced with communication breakdowns. Females fall back to the old vocabulary/strategy, while males try novel words. Even in smooth communication, females are less willing to ‘experiment’ with new expressions compared to male users. This observation is consistent with gender differences in problem-solving strategies, such that females avoid using unfamiliar software features, while males tend to engage in ‘tinkering’ when using software (Burnett et al., 2011; Cao et al., 2010). Such findings should be taken on board in system design so that when unfamiliar features and strategies have to be adopted when interacting with an application, techniques, such as tutorial snippets, examples of what to say/do and short explanations, are available to guide users.

Certainly, inclusive design does not mean that the user experience of males should be compromised. Therefore, these functionalities should be made optional. Since gender is a stable user profile characteristic, such options can be easily implemented in an adaptive system. It should be clarified, however, that females or males are not a homogeneous group of users exhibiting all the characteristics and preferences that are statistically associated with

their gender. It is highly likely that many males are affected by the same interface complexities as females, and many females may enjoy the same software features as males, as argued in Beckwith et al. (2006a). This underscores the importance of gender-neutral software that supports all users, rather than developing gender-specific systems. An interesting idea has been proposed by Ljungblad and Holmquist (2007), according to which, designs that are informed by the needs of a 'marginal' population may also benefit the wider user population; so, for instance, the textual aids originally meant for encouraging female users to adopt a software feature may be also enjoyed by another group of users, such as users with field-dependent cognitive style (Magoulas et al., 2004), or provide essential support to subpopulations, such as users with visual impairments. Finally, for many application domains, explorative and innovative user behaviour is not desirable. A good example is certainly the domain of this study, of systems implementing natural language interfaces, for which innovative and unpredictable user input is the main source of system failures.

Female users rate their performance lower than male users.

Resonating with previous research, the findings of this study show that females rated their own performance more poorly than male users. On the other hand, males appear to assess that the system less favourably than females. This finding empirically supports notions that females are more likely to blame themselves and their lack of skill if they face problems when performing a task using a computer system, while men will attribute the problems the system.

Contributions in relation to the effect of visual information in collaborative systems

The thesis described ways in which visual information can enhance or disrupt collaborative work and communication processes, offering recommendations on how it can be used to best support interactions with situated agents, CMC and CSCW systems.

CMC and CSCW systems typically integrate visual information. Awareness of how visual information affects collaboration and communication patterns can enable interface

designers to take full advantage of its benefits and avoid the pitfalls. The findings reiterated theoretical and empirical observations that communication media entail different affordances that make the communication more or less effortful. The study generally confirmed the advantages of visually-supported collaborations by showing that sharing the workspace facilitates situation awareness. Visual copresence enables users of collaborative systems to complete their tasks more efficiently, with shorter interactions and simpler language. It was also observed that when visual information is shared, much of the communication was carried out through physical actions and not verbal means. Visual information is expected to be more useful in typed communication, since it alleviates the resource-intensive task of grounding textually. Moreover, these findings have implications for enhancing novice-expert collaborations: the expert can point to an object instead of using a possibly unfamiliar term (by utilizing features like highlighters or pointers); the novice can show an action without having to verbally explain it (by using features like pointer traces).

However, the results of this study also describe the potential pitfalls: visually-enhanced interactions may present a close, but misleading approximation to face-to-face communication, leading people to assume continuous joint perspective and to relax their grounding criteria, thus causing miscommunication. Visual copresence may also give rise to inflated assumptions of common ground and of the addressee's (human or computer) perceptual abilities. Users in visually-supported conditions may provide less information verbally than needed, which will impair the performance of the partner. The analysis also revealed that visual copresence entail distinct different requirements for mobile robots, such that visually copresent robots may need to disambiguate underspecified deixis, while non-collocated robots may need to be able to provide elaborate feedback. This observation emphasises the importance of system development being informed by corpus collection studies in realistic deployment conditions.

Contributions in relation to alignment in HCI

The study provided a complete account of the operation of alignment in naturalistic human-computer communication. It distilled the findings into design recommendations for systems with natural language interfaces and proposed a simple dialogue model that leverages alignment.

Speakers tend to repeat their own and each other's linguistic choices in dialogue, a phenomenon which arguably underlies communication success. The study reported in this paper has drawn on the Interactive Alignment Model and existing work in HCI in order to investigate alignment in task-oriented dialogues with computer systems. The experimental data, obtained from naturalistic human-robot navigation dialogues, have helped to address important questions about the operation and role of alignment in the effectiveness and success of the interaction. In addition, the analysis has led to design guidelines which were subsequently used in the development of a simple alignment-driven approach for dialogue management. It is hoped that the model presented in this thesis will serve as a starting point for exploring the potential of alignment within computer-based dialogue models and system implementations.

This study offered primary knowledge in alignment in HCI owing to the experimental approach it adopted. Unlike past research that investigated alignment in HCI, this study did not use an artificial experimental task such as object-naming, but a corpus of spontaneously produced utterances; it measured alignment both as it operated at the adjacency pair level and as it developed over the full course of the dialogue, under conditions of smooth and problematic communication; finally, it looked at 'priming effects' for both user and system.

The study provided unique insight into linguistic alignment in HCI. In particular, the results indicated that alignment is present in HCI, resulting in the gradual reduction and stabilisation of the vocabulary-in-use, and that it is also reciprocal. Further, the results suggested that when system and user errors occur, the development of alignment is temporarily disrupted and users tended to introduce novel words to the dialogue. The results also indicated that alignment in human-computer interaction may involve a strategic component, being used as a resource to compensate for less optimal interaction conditions. Moreover, lower alignment (particularly, in system-generated input) is associated with less successful interaction, as measured by user perceptions. It should be noted that this observation is also of prime theoretical importance for models of human communication, being the first empirical validation of the link between lexical alignment and interaction success, as predicted by the Interactive Alignment Model.

To the author's knowledge, this is also the first account of alignment in HCI that integrates miscommunication. The study showed that after system/user errors, users lose

confidence in the efficiency of previous expressions and introduce novel words. This has two practical implications; first, the ability to predict what users will do after the occurrence of errors is a matter of enormous significance for interactive systems (and their natural language understanding components). Second, detecting the presence of novel lexical items in user turns may function as a valid negative cue for an error detection algorithm.

The study concluded by incorporating these recommendations into a high-level dialogue model, which aimed to illustrate how linguistic alignment can be supported by the dialogue manager of a system. Following the dialogue model, the dialogue manager performs two types of update as a function of the usage of an expression over the course of a single dialogue: it creates an association between the lexical item and a referent; and changes its weighting within the lexicon. The model also integrated system error-handling. Possible benefits of the suggested approach include: enhanced recognition accuracy, owing to rescoring of word probabilities based on their weightings; improved intelligibility of system generated output, owing to it consisting of recurring words; and user interaction with the system that is more natural and cognitively easy. The model is by no means complete, but should primarily serve as a springboard for researchers to incorporate linguistic alignment in detailed dialogue models.

7.3.3 Methodological contributions

Dialogue is a valid research paradigm to study gender and spatial language.

The major methodological contribution of this work concerns the successful application of dialogue as a naturalistic, and yet experimentally sound, research paradigm to study gender and spatial language. Additional contributions include: (i) the experimental manipulations that helped avoid gendered perceptions of the interaction and experimental biases, (ii) a realistic task and language corpus of human-robot navigation dialogues, (iii) a multidisciplinary, detailed approach for evaluation of performance, communication processes and strategies and (iv) a fine-grained linguistic analysis of the corpus, and (v) a custom system that allows for controlled and fine experimental manipulations to investigate interactive phenomena.

This study proved the validity of dialogue as a research/data collection paradigm. First, by empirically confirming that gender differences assume a different form in dyadic interactions, it illustrated that accounts of gender differences that are only limited to monologic data are incomplete and, for many purposes, incorrect. Moreover, a merit of the approach of the experimental setup was that the genders of the interlocutors were effectively masked, thus confounding social perceptions that influence behaviour and performance were minimised. The naturalness of the data was ensured by modifying the Wizard of Oz paradigm to involve naïve participants.

Second, scientists in applied and theoretical fields have only recently begun to jointly investigate spatial language and dialogue and, as such, our knowledge remains incomplete. By using dialogue in a naturalistic but carefully controlled spatial setting, the study provided a realistic account of the range of linguistic options that users are likely to employ in spatial Human-Robot Interaction, in two different interaction scenarios. A large language and task corpus was collected, which is hoped to be a useful resource for the community. From a wider perspective, this analysis may also be useful for researchers and designers to better understand how spatial information should be displayed or communicated by systems and how the availability and presentation of such information may change the behaviour and experience of users of different gender. For example, a major research area in the field of geographic information/wayfinding systems is how route instructions should be configured to be appropriate for different individual groups of users in terms of landmarks, their frequency and salience (Montello and Sas, 2006).

Third, the study illustrated the value and feasibility of employing a multidisciplinary approach to perform data analysis. The data analysis largely followed established classification schemes from different research fields and tied together a novel corpus of spatial language in real-time direct interaction with a simulated robot, performance data, miscommunication analysis and detailed sequential analysis (turn-level alignment analysis). Therefore, a significant contribution of this thesis concerns the development and use of an evaluation methodology for spatial collaboration appropriate for interactive systems, which combined subjective and objective metrics; it is hoped that this evaluation framework can be employed in future studies in this domain.

Fourth, previous research in the effect of visual information employed coarse coding schemes based on dialogue acts. In this study, a detailed component-based scheme complemented the dialogue act analysis and illustrated a refined picture of how visual information influences the choice and distribution of utterance constituents.

Finally, the results obtained from the study's computer-based system that enabled real-time dialogue and collaboration were highly consistent with past research in human communication that had used 'traditional' face-to-face, physical settings. This validates the suitability of the experimental approach for thorough investigations of human communication processes. This approach enabled monitoring of the unfolding language actions and how they integrated with visual actions and contextual entities. Moreover, unlike real-world experimental settings, the system offers an excellent platform for fine-grained exploration of interactive phenomena, as it allows for controlled, systematic and subtle experimental manipulations.

7.4 Limitations and future research

The contributions of this thesis have revealed further research questions and elucidated interesting directions for future experimental investigation. Two areas of future research are proposed. The first of these arises as a result of the limitations of this study. These limitations are discussed in the next section. The second area of future research describes the research steps forward incited by the knowledge gained from the current thesis.

7.4.1 Limitations of the study

Reflection on the methodology and results of this research has led to the identification of a number or limitations. Some of these have already been noted in the previous chapter in the discussion of individual findings. This section presents some of the most important limitations of this study, most of which arise from the experimental approach adopted in this thesis.

The first limitation of the study lies on the sample size. To minimise the risks associated with small sample sizes (that is, failing to reject a false null hypothesis), both the

experimental design and analysis involved additional steps. Firstly, the design aimed to reduce the variance in the measurements by employing some common techniques (Sauro and Lewis, 2012, p. 121): using relatively homogeneous participant groups (see the ensuing limitation discussed below); using a simple rather than a complex experimental task; ensuring that participants understand what they need to do; enabling participants to get familiar with the experimental set-up and environment by giving them some practice time. Secondly, the statistical analysis involved particular caution: first, for the parametric tests, the shape of the distributions was examined to ensure that the assumptions of normality are not grossly violated; post-hoc tests and error bar graphs were used to assess significant effects; and effect size measurements were reported for all statistically significant results. Taken together, it is acknowledged that the sample of this study is problematic, and possibly inappropriate in the context of experimental psychology. Yet, HCI theory and practice have demonstrated that following careful procedures of experimental design and analysis enable researchers to reach valid statistical conclusions with sample sizes as small as 2-5 users (Sauro and Lewis, 2012, p.10). Indeed, small sample sizes appear to be more acceptable in the HCI field; a review sampling ninety-seven high-quality HCI publications found that the median number of users per study was 10 (64% of tests had between 8 and 12 users and 80% had fewer than 20 users) (Sauro and Lewis, 2009), while another study analysed over seventy papers in top HCI journals and reported that the number of users ranged from 6 to 181 (Hornbaek and Law, 2007).

The second limitation has to do with the sampling population. That is, the participants of this study exclusively consisted of university students. As discussed above, this approach helps reduce variance but, at the same time, it also adversely affects the external validity and generalisability of the study. It is highly likely that education level, previous experiences and age influence the magnitude of gender differences, verbal and spatial abilities (see, for example, Golding et al., 1996) and patterns of using computers and computer-mediated communication tools.

Furthermore, the study did not take into account the cognitive profile of participants, which may be considered another limitation. That is, the spatial and verbal abilities of the individuals were not measured and not included in the experimental design. Yet, low/high spatial and verbal abilities are expected to play an important role in navigation performance

and the way that people communicate route information (Vanetti and Allen, 1988). Section 6.2.3 discussed this limitation.

The research reported in this thesis does not make a distinction between gender and sex, while acknowledging that sex and gender may not be equivalent. Following related literature, it employs the term 'gender differences' as an inclusive term to signify differences in outcomes between females and males, with no attempt to categorise these differences as biologically- or socially- based. This appears to harmonise with previous research (for example, Halpern et al., 2007) which advocated the 'biopsychosocial model', according to which, since cognitive abilities are a product of the interactions of biological and environmental factors, it is neither possible nor useful to make a distinction between gender and sex. However, it is necessary that all research in the field clarifies how these terms are methodologically used (Institute of Medicine, 2001, p. 6). For the experimental part of the study, participants were classified by the dichotomous variable of sex. A classification based on gender would involve testing and placing participants on the continuum between masculinity and femininity. In addition, as pointed out in medical research (Greenspan et al., 2007), group differences in variables that involve perception (for example, pain or user experience) may be attributed to both sex and gender. So, it is recommended that, when possible, both gender and sex are explored in order to determine their relative contributions to the effect. Moreover, it is argued that women generally have 'female brains' and men have 'male brains', but this does not hold in all cases. For example, a study found that 17% of participants had traits of empathy and systemising typical of the opposite sex (Goldenfeld et al., 2005). Although it is expected that the large majority of participants of the present study have the cognitive profile associated with their biological sex, it is recognised that the classification by sex without exploring the relationship with gender may be a shortcoming of the experimental design.

The fifth limitation relates to the use of typed communication. The use of text-based dialogue enabled important parameters of the experiment such as masking the 'wizard' and his/her gender and avoiding communication processes taking place through paralinguistic means. Existing literature gave confidence to this approach by showing that spoken and typed task-oriented HCI have few differences (Hauptmann and Rudnicky, 1988), the modality (speech or text) did not affect how route instructions are formed (Moratz and Tenbrink, 2003), and lexical alignment emerged in similar ways for speech- and text-based interaction

with computer systems (Brennan, 1996). Yet, there are known differences that may have had an effect on the communication patterns observed in this study. In particular, typed communication is 'quasi-synchronous' (Garcia and Jacobs, 1999); that is, the recipient sees the message in its entirety the moment his/her partner presses 'enter', whereas in spoken dialogue, interlocutors start formulating their response whilst listening to their partners' utterance. This 'quasi-synchrony' may have also disrupted the sequential cohesion of dialogue, such that the second of two successive turns may not actually be the response to the first one. Moreover, spoken dialogue involves more frequent grounding of shorter utterances. Finally, grounding is performed via auditory and gestural cues, while in text-based communication, mutual understanding is established through more explicit means. As such, it needs to be further validated whether the findings of this study can be extended to communication through other modalities.

The sixth limitation relates to the characteristics of the navigation environment used in the experiment, and, in particular, its dimensionality and configuration in terms of environmental features. While the production of route instructions may not be significantly different across settings (Tenbrink, 2007), the possibility that three-dimensional virtual worlds and real-world outdoor environments exacerbate gender differences cannot be eliminated. Moreover, the urban environment used in the study contained a number of salient features which may have benefited females more than males. At the same time, it should be pointed out that real-world urban environments are typically rich in landmarks and that the study aimed to set up a 'gender-neutral' experimental environment and not to disadvantage any gender by design, by, for instance, excluding buildings (following Hubona and Shirah's (2004) argument, see Chapter 2, section 2.7.5).

Finally, a number of limitations can be identified in the investigation of alignment and the design recommendations presented in this thesis. First, this study only dealt with the operation of lexical alignment in human-computer dialogue, with no attempt to broaden the scope to alignment at other linguistic levels such as syntax and pragmatics. Second, following the Interactive Alignment Model, this account viewed alignment as an automatic input/output matching, discounting alignment modulated by audience design, social factors such as politeness, affect and relationship and other conscious decisions. Third, the proposed guidelines and model are not suitable for interactions in which the user is less expert than the

computer system and tends to use linguistic terms incorrectly, as in case of interactions with educational and tutoring systems.

7.4.2 Future work

The limitations identified above illustrate the routes through which this work may be advanced. Addressing the first limitation, this study could be complemented by running another round of experiments. The second limitation had to do with the fact that the sample consisted of female and male university students. As such, the study could be replicated using participants drawn from populations of different demographic profile; the characteristics that are likely to be relevant for computer-based spatial tasks are age, education and computer expertise. This work should provide more power and generalisability to the results of the study. Moreover, it could illuminate the individual simple effects and interaction effects of these factors – gender, age, education, computer expertise – on navigation, collaboration and communication processes with computer systems, and contribute to the debate whether gender differences decline or increase as a function of age. A third limitation was that the study did not account for the cognitive abilities of participants. As such, the study could also involve participants taking psychometric tests to assess their verbal and spatial abilities. Novel findings are expected to emerge and help define the relation between cognitive spatial and verbal abilities and navigation performance in dialogue. Furthermore, it should be noted that the present study also collected, but not included in the analysis, data on field dependent/field independent cognitive styles. Thus, the work will be extended to the analysis of this data, which could reveal interactions between cognitive styles and the factors already considered in this study. Similarly, it may be significant to distinguish between sex and gender and understand their relative contributions to the differences observed in this study. To this end, specialised tests that are used to determine feminine and masculine traits in individuals should be administered before the experiments. These tests range from spatial ability and object location memory tests to tests on empathy and emotional responses. Comparisons can be, then, made between pairs that are configured in terms of gender, rather than biological sex. Taken together, future research efforts in response to these issues could produce a complete framework of individual differences, enabling designers to make informed decisions in the development of collaborative systems and systems with natural language interfaces.The fifth and sixth limitations stemmed from concerns with regards to

whether the findings can be applied to spoken dialogue and other navigation settings. These limitations could be addressed by further experiments using speech or multi-modal interaction (for example, combining language-based communication with pointing or drawing), adding to theoretical accounts of how grounding is performed across different communication modalities. Future studies could employ and explore the effect of different maps that vary in terms of their configuration and frequency of landmarks. The custom system developed for the purposes of this study is capable of supporting new experimental conditions of this kind with minimal modification and tuning. It would also be interesting to extend the investigation of gender differences in navigation performance and route communication in immersive virtual worlds.

Finally, this study was limited to lexical alignment as input/output matching in HCI. This account may serve to instigate research interest in the other levels of linguistic alignment in HCI. In particular, it is of immense practical relevance to identify and describe syntactic alignment, preferably under conditions of naturalistic task-oriented dialogue with a system. Future research should also be directed towards producing complete accounts and models of linguistic alignment that integrate mechanistic priming as well as alignment for affective and social purposes and audience design. Building on such thorough investigations, alignment processes can be well understood and incorporated into computational dialogue models which could lead to innovative system implementations that leverage and capture the potential of this natural mechanism. Moreover, studies in human communication have illustrated that alignment between interlocutors extends beyond a single dialogue (Brennan and Clark, 1996), such that the set of ‘preferred’ referring expressions is pair-specific and persists for subsequent interactions. Interactions with dialogue systems embedded on personal devices (such as assistive robots, computers and mobile phones) are destined to have long-term interactions with users. As such, the mechanism of alignment could prove to hold great value for more efficient and effective performance of these applications. Therefore, it is important to investigate this claim with appropriate follow-up or longitudinal studies. Finally, this work has indicated that alignment is instrumental in interaction success. Thus, as previously suggested by Reitter and Moore (2007), alignment rates may be used as a reliable predictor of interaction success in task-oriented dialogues with systems and call centre representatives. Future work should test this hypothesis and include the application of

regression analytical techniques on the data of this corpus to determine the predictive power of alignment.

In addition to addressing the observed limitations, this work has revealed areas that merit further exploration outside the confines of the thesis.

While the investigation of the effect of visual information was not a primary target of the thesis, it presented rich opportunities of experimental research. The ensuing robust effects argued for the necessity to gain awareness of how visual elements modify behaviours, interact with language and integrate with it. Therefore, further systematic research should focus on the role of visual information as a resource in collaborative work in computer-mediated communication, virtual environments and interaction with software and robotic agents. Such work could add to our theoretical understanding of communication processes as well as enable better design of collaborative systems.

The present study focused on dyadic interaction. An interesting continuation of this work is to investigate the effects of gender composition in group interactions. Better understanding of the role of gender in group dynamics has important ramifications for the success of teamwork in organisational settings and videoconferencing (Molyneaux et al., 2008). There are a few empirical studies in teamwork that suggest that mixed-gender groups including at least one female have more felicitous interactions, reporting higher levels of satisfaction and social presence and performance (Wong et al. 2004; Houldsworth and Mathews, 2000; Hamlyn-Harris et al., 2006). Given that the present study and past research have demonstrated that same-gender pairs had more successful collaborations than mixed-gender pairs, it may be the case that findings from research in pair dynamics cannot be extrapolated to group work scenarios and vice versa. In light of this, future work should be directed towards understanding how the gender composition of the group influences computer-mediated collaboration, communication and learning, and which gender composition leads to optimal outcomes.

This study produced a wealth of findings with regards to how people produce and follow route instructions in dialogue. As such, future work could exploit these findings for the development of software systems that are capable of interpreting and following route instructions, a development methodology successfully employed for robotic assistants (Lauria

et al., 2004; Shi and Tenbrink, 2009; MacMahon, 2004). Similarly, these findings could be applied to the domain of route instruction generation systems such that geographic information systems and maps in kiosks, the web, smartphones and cars are capable of producing route instructions that are natural and easy to follow for people of both genders.

7.5 Closing remarks

The influence of gender manifests in numerous expressions of human cognition and behaviour, including the ways females and males interact with artefacts. The effect of gender and the ways it interacts with the characteristics of the technology are not trivial. By making implicit decisions about user preferences and needs, designers are bound to build applications that disadvantage many users. Yet, our knowledge of how gender differences arise in the communication with and through computers is rudimentary. At the same time, existing differential research from behavioural sciences being based on non-interactive or artificial settings provides insight of limited value for HCI. This thesis synthesised insights from research in spatial cognition, sociolinguistics and task-oriented human communication and assumed an HCI perspective in order to examine the gender factor in computer-mediated communication and in language-based communication with systems. The investigation revolved around the HCI themes of interaction efficiency, effectiveness and user perceptions and yielded findings that underscored the importance of taking gender differences into account in the development of collaborative and interactive software.

The thesis detailed the effects of linguistic alignment, miscommunication and shared visual information in HCI and human communication and how they vary as a result of gender. The thesis presented the argument that interactive mechanisms are instrumental in how males and females perform and coordinate in HCI; and if these mechanisms are translated into design criteria, they have the potential to promote successful behaviours and strategies that lead to enhanced performance and experience for users of both genders. Most importantly, the thesis should serve to stimulate further research in gender and the other human factors in various domains of HCI. The fruits of these endeavours are environments and applications that support users equally, without marginalising a particular group.

References

Abdi, H 2010. Greenhouse-geisser correction. In NJ Salkind (ed.), *Encyclopedia of research design*, SAGE Publications, Inc., Thousand Oaks, CA, 545-549.

Agresti, A. (2007). *An introduction to categorical data analysis* (Vol. 423). Wiley-Interscience. 2nd edition.

Allen, G. L. (2000a). Men and women, maps and minds: Cognitive bases of sex-related differences in reading and interpreting maps. In O'Nuallain, S. (ed.), *Spatial Cognition: Foundations and Applications*. Amsterdam: John Benjamins, 3-18.

Allen, G. L. (2000b). Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, *14*(4), 333–359.

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*(3), 327-355.

Allwood, J. (1995). An activity based approach to pragmatics. In *Abduction, Belief and Context in Dialogue, Studies in Computational Pragmatics*, Amsterdam. John Benjamins. 47-80.

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human–computer speech interaction. *International Journal of Man–Machine Studies*, *38*(4), 547–566.

American Psychological Association. Publication Manual of the American Psychological Association (5th ed.). Washington, DC: American Psychological Association, 2001, p. 63.

Anacta, V., & Schwering, A. (2010). Men to the East and Women to the Right: Wayfinding with Verbal Route Instructions. *Spatial Cognition VII*, 70-84.

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2011). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of learning and memory*.

Annett, M. (1992). Spatial ability in subgroups of left-and right-handers. *British Journal of Psychology*, *83*(4), 493-515.

Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioural brain research*, *151*(1), 103-115.

Bae, S., & Lee, T. (2011). Gender differences in consumers' perception of online consumer reviews. *Electronic Commerce Research*, *11*(2), 201-214.

Bailenson, J. N., & Yee, N. (2005). Digital chameleons—automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*(10), 814–819.

Bardzell, S. (2010). Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, USA, 1301-1310.

Bargh, J. A. (1989). Conditional automaticity: varieties of automatic influence on social perception and cognition. In J. S. Uleman, J. S. & J.A Bargh, (Eds.), *Unintended thoughts*, (3–51). Guilford Press.

Barkowsky T., Knauff M., Ligozat G. & Montello, D. R. (eds.) (2007). *Spatial Cognition V: Reasoning, Action, Interaction*, Lecture Notes in Computer Science, Berlin: Springer.

Barnett, M. A., Vitaglione, G. D., Harper, K. K., Quackenbush, S. W., Steadman, L. A., & Valdez, B. S. (1997). Late Adolescents' Experiences With and Attitudes Toward Videogames1. *Journal of Applied Social Psychology*, *27*(15), 1316-1334.

Beckwith, L. (2007). Gender HCI Issues in End-User Programming, (Ph.D. Thesis, Oregon State University). Retrieved from https://ir.library.oregonstate.edu/xmlui/handle/1957/4954

Beckwith, L., & Burnett, M. (2004, September). Gender: An important factor in end-user programming environments?. In *Visual Languages and Human Centric Computing, 2004 IEEE Symposium on* (pp. 107-114). IEEE.

Beckwith, L., Burnett, M., Grigoreanu, V. & WiedenBeck, S. (2006a). HCI: What about the software? *Computer*, pp. 83–87.

Beckwith, L., Kissinger, C., Burnett, M., Wiedenbeck, S., Lawrance, J., Blackwell, A., & Cook, C. (2006b, April). Tinkering and gender in end-user programmers' debugging. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 231-240). ACM.

Beckwith, L., Burnett, M., & Wiedenbeck, S. (2007). Gender HCI Issues in End-User Software Engineering Environments. In M. H. Burnett, G. Engels, B. A. Myers, & G. Rothermel (Eds.), *End-User Software Engineering*.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological review*, *88*(4), 354.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1956). The differential aptitude tests: An overview. *The Personnel and Guidance Journal*, *35*(2), 81-91.

Benyon, D. and Höök, K. (1997). Navigation in information spaces: Supporting the individual. In Human Computer Interaction: INTERACT '97, 39-45.

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human–computer interaction. *Proceedings of the International Congress of Phonetic Sciences*. 2453–2456.

Berenson, S. B., Slaten, K. M., Williams, L., & Ho, C. W. (2004). Voices of women in a software engineering course: reflections on collaboration. *Journal on Educational Resources in Computing (JERIC), 4*(1), 3.

Bernard, M., Mills, M., & Friend, C. (2000). Male and female attitudes towards computer-mediated group interactions. *Internet Publication for Usability News, 2*(2).

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language & Communication*. *8*(3), 183-194.

Bimber, B. (2000). Measuring the gender gap on the Internet. *Social Science Quarterly, 81*(3), 868-876.

Blais, A. R., & Weber, E. U. (2001). Domain-specificity and gender differences in decision making. *Risk Decision and Policy*, *6*(1), 47-69.

Blaylock, N., Allen, J., & Ferguson, G. (2002, July). Synchronization in an asynchronous agent-based architecture for dialogue systems. In *Proceedings of the 3rd SIGdial workshop on Discourse and dialogue-*. *2*, 1-10. Association for Computational Linguistics.

Bohus, D. & Rudnicky, A. I. (2005). Sorry, I didn't catch that! – an investigation of non-understanding errors and recovery strategies. *Proceedings of SIGdial2005*. Lisbon, Portugal.

Bohus, D. & Rudnicky, A. I. (2009). The RavenClaw dialog management framework: Architecture and systems. *Comput. Speech Lang.*, *23*(3), 332-361.

Boiano, S., Borda, A., Bowen, J., Faulkner, X., Gaia, G., & McDaid, S. (2006). Gender issues in HCI design for web access. *Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities, Section III, Gender Issues*, 116-153.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, *23*(2), 119-147.

Bosco, A., Longoni, A. M., & Vecchi, T. (2004). Gender effects in spatial orientation: Cognitive profiles and mental strategies. *Applied cognitive psychology*, *18*(5), 519-532.

Branigan, H.P & Pearson J. (2006). Alignment in Human-Computer Interaction. In K. Fischer (Ed.) *Proceedings of the Workshop on How People Talk to Computers, Robots, and Other Artificial Communication Partners*. pp. 140-156. Delmenhorst, Germany: HWK.

Branigan, H. P., Pickering, M., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(B), 13–25.

Branigan, H.P., Pickering, M.J., Pearson, J.M., McLean, J.F., & Nass, C. (2003) Syntactic alignment between computers and people: the role of belief about mental states. *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*. pp. 186-191. Mahwah: Erlbaum.

Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F., Nass, C.I., & Hu, J. (2004). Beliefs about mental states in lexical and syntactic alignment: Evidence from human-computer dialogs. *Proceedings of the CUNY Conference on Human Sentence Processing.*

Branigan, H.P., Pickering, M.J, & McLean, J.F., & Stewart, A. (2006). The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and Cognitive Processes, 21*, 974-1010

Branigan, H.P., Pickering, M.J., Pearson, J. & McLean, J.F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*, 2355–2368.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41-57.

Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of the International Symposium on Spoken Dialogue*, 41–44.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions.* (95-129). Cambridge, MA: MIT Press.

Brennan, S. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1482–1493.

Brennan, S. E., & Hulteen, E. A. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems, 8*(2), 143-151.

Brockmann, C., Isard, A., Oberlander, J., & White, M. (2005). Modelling alignment for affective dialogue. In *Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling* (UM-05).

Brown, L. N., Lahar, C. J., & Mosley, J. L. (1998). Age and Gender-Related Differences in Strategy Use for Route Information A" Map-Present" Direction-Giving Paradigm. *Environment and Behavior*, *30*(2), 123-143.

Brownlow, S., Janas, A. J., Blake, K. A., Rebadow, K. T., & Mellon, L. M. (2011). Getting by with a little help from my friends: Mental rotation ability after tacit peer encouragement. *Psychology*, *2*(4), 363-370.

Burke, J. L., Murphy, R. R., Coovert, M. D., & Riddle, D. L. (2004). Moonlight in Miami: Field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human–Computer Interaction*, *19*(1-2), 85-116.

Burnett, M., Fleming, S. D., Iqbal, S., Venolia, G., Rajaram, V., Farooq, U., ... & Czerwinski, M. (2010, September). Gender differences and programming environments: across programming populations. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM.

Burnett, M. M., Beckwith, L., Wiedenbeck, S., Fleming, S. D., Cao, J., Park, T. H., ... & Rector, K. (2011). Gender pluralism in problem-solving software. *Interacting with Computers*, *23*(5), 450-460.

Burns, P. C. (1998). Wayfinding errors while driving. *Journal of Environmental Psychology*, *18*(2), 209-217.

Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing Research*, *12*(2), 147-158.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological bulletin*, *125*(3), 367.

Cánovas, R., Espínola, M., Iribarne, L., & Cimadevilla, J. M. (2008). A new virtual task to evaluate human place learning. *Behavioural brain research*, *190*(1), 112-118.

Cao, J., Rector, K., Park, T. H., Fleming, S. D., Burnett, M., & Wiedenbeck, S. (2010, September). A debugging perspective on end-user mashup programming. In *IEEE Symp. Visual Languages and Human-Centric Computing*.

Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist; American Psychologist*, *40*(7), 786.

Carletta, J., & Mellish, C. S. (1996). Risk-taking and recovery in task-oriented dialogue. *Journal of Pragmatics*, *26*(1), 71-107.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., & Anderson, A. (1996). HCRC dialogue structure coding manual. *Universities of Edinburgh and Glasgow*.

Carr, T., Cox, L., Eden, N., & Hanslo, M. (2004). From peripheral to full participation in a blended trade bargaining simulation. *British Journal of Educational Technology*, *35*(2), 197-211.

Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science*, *17*(1), 21-25.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in math college entrance test scores across diverse samples. *Developmental Psychology, 31,* 697-705.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology, 35*, 1237-1247.

Casper, J. L., & Murphy, R. R. (2002). Workflow study on human-robot interaction in USAR. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*. *2,* 1997-2003. IEEE.

Casper, J., & Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *33*(3), 367-385.

Cassell, J. (1998). Chess for girls?: Feminism and computer games. In J. Cassell & H. Jenkins (eds.). *From Barbie to Mortal Kombat: Gender and Computer Games*. Cambridge, MA: MIT Press

Castelli, L., Latini Corazzini, L., & Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human behavior*, *24*(4), 1643-1667.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1972). Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *14*(6), 487-509.

Chen, C. H., Chang, W. C., & Chang, W. T. (2009). Gender differences in relation to wayfinding strategies, navigational support design, and wayfinding task difficulty. *Journal of environmental psychology*, *29*(2), 220-226.

Chen, S. Y., & Macredie, R. (2010). Web-based interaction: A review of three important human factors. *International Journal of Information Management*, *30*(5), 379-387.

Chipman, S. F. (1994). Research on the women and mathematics issue. In A. M. Gallagher & J. C. Kaufman (eds.) *Gender Differences in Mathematics: An Integrative PsychologicalApproach*. Cambridge: Cambridge University Press.

Choi, K. S., Deek, F. P., & Im, I. (2009). Pair dynamics in team collaboration. *Computers in Human* Behavior, *25*(4), 844-852.

Chrisler, J. C., & McCreary, D. R. (2010). *Handbook of gender research in psychology* (Vol. 1). Springer.

Clark, E. V. (1993). *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127-149.

Clark, H.H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Memory & Language J.*, *50*, 62-81.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. *Psycholinguistics: Critical Concepts in Psychology*, *3*, 414.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of memory and Language*, *26*(2), 209-225.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, *13*(2), 259-294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.Coates, J. (1986). Women, men and languages: Studies in language and linguistics. *Longman. London*.

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics*, *10*(4), 417-451.Cockburn, A., & McKenzie, B. (2002, April). Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves* (pp. 203-210). ACM.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Lawrence Erlbaum.

Cohen, M.H. Giangola, J.P. & Balough, J. (2004). *Voice User Interface Design*. Addison Wesley.

Colom, R., Contreras, M. J., Botella, J., & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences*, *32*(5), 903-912.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*(3), 329-340.

Coluccia, E., Bosco, A., & Brandimonte, M. A. (2007). The role of visuospatial working memory in map learning: New findings from a map drawing paradigm. *Psychological research*, *71*(3), 359-372.

Contreras, M. J., Rubio, V. J., Peña, D., Colom, R., & Santacreu, J. (2007). Sex differences in dynamic spatial ability: The unsolved question of performance factors. *Memory & cognition*, *35*(2), 297-303.

Coupland, N., Coupland, J., Giles, H., & Henwood, K. (1988). Accommodating the elderly: Invoking and extending a theory. *Language in Society*, *17*(1), 1-41.

Coupland, N., Giles, H., & Wiemann, J. M. (Eds.). (1991). *" Miscommunication" and problematic talk* (pp. 1-18). Newbury Park, CA: Sage publications.

Coventry, K., Tenbrink, T., & Bateman, J. (2009). Spatial language and dialogue: Navigating the domain. In Coventry, K. R., Tenbrink, T., & Bateman, J. (Eds.). (2009). *Spatial language and dialogue* (Vol. 3). Oxford University Press., 1-8.

Cowan, B. R., Beale, R., & Branigan, H. P. (2011, May). Investigating syntactic alignment in spoken natural language human-computer communication. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (pp. 2113-2118). ACM.

Crowston, K., & Kammerer, E. (1998). Communicative style and gender differences in computer-mediated communications. In B.Ebo *(Ed.), Cyberghetto or cybertopia? Race, class, and gender on the Internet* (pp. *185–203*). Westport , Conn : Praeger.

Czerwinski, M., Tan, D. S., & Robertson, G. G. (2002, April). Women take a wider view. In Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves (pp. 195-202). ACM.

Dabbs, J. M., Chang, E. L., Strong, R. A., & Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior*, *19*(2), 89-98.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, *6*(4), 258-266.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments?. *British Journal of Educational Technology*, *41*(1), 10-32.

Daniel, M. P., & Denis, M. (1998). Spatial descriptions as navigational aids: A cognitive analysis of route directions. *Kognitionswissenschaft*, *7*(1), 45-52.

Daniel, M. P., & Denis, M. (2003). The production of route directions: Investigating conditions that favour conciseness in spatial discourse. *Applied cognitive psychology*, *18*(1), 57-75.

De Goede, M., & Postma, A. (2008). Gender differences in memory for objects and their locations: A study on automatic versus controlled encoding and retrieval contexts. *Brain and cognition*, *66*(3), 232-242.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de psychologie cognitive*, *16*(4), 409-458.

Denis, M., Pazzaglia, F., Cornoldi, C., & Bertolo, L. (1999). Spatial discourse and navigation: An analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, *13*(2), 145-174.

Devlin, A. S., & Bernstein, J. (1995). Interactive wayfinding: Use of cues by men and women. *Journal of environmental psychology*, *15*(1), 23-38.

Dickhauser, O., & Meyer, W. (2006). Gender differences in young children's math ability attributions. *Psychology Science*, *48*(1), 3.

Dillon, A., & Watson, C. (1996). User analysis in HCI: the historical lesson from individual differences research. *International Journal of Human-Computer Studies*, *45*(6), 619-637.

Ding, N., Bosker, R. J., & Harskamp, E. G. (2011). Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Computers & Education*, *56*(2), 325-336.

Djamasbi, S., & Loiacono, E. T. (2008). Do men and women use feedback provided by their Decision Support Systems (DSS) differently?. *Decision Support Systems*, *44*(4), 854-869.

Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., & Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology*, *55*(4), 580.

Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, *15*(2), 95-105.

Egan, D. E. (1988). Individual differences in human-computer interaction. *Handbook of human-computer interaction*, 543-568.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests* (pp. 109-113). Princeton, NJ: Educational Testing Service.

Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: some issues and experiences. *Communications of the ACM*, *34*(1), 39-58.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850-855.

Fern, X., Komireddy, C., Grigoreanu, V., & Burnett, M. (2010). Mining problem-solving strategies from HCI data. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *17*(1), 3.

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block Talk: Spatial Language During Block Play. *Mind, Brain, and Education*, *5*(3), 143-151.

Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: the'white male'effect. *Health, risk & society*, *2*(2), 159-172.

Fischer, K. (2006). The role of users' preconceptions in talking to computers and robots. *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, 112.

Fitzpatrick, M. A. (1988). *Between husbands & wives: Communication in marriage*. Beverly Hills, CA: Sage Publications.

Fitzpatrick, M. A., Mulac, A., & Dindia, K. (1995). Gender-preferential language use in spouse and stranger interaction. *Journal of Language and Social Psychology*, *14*(1-2), 18-39.

Fox, A. B., Bukatko, D., Hallahan, M., & Crawford, M. (2007). The Medium Makes a Difference Gender Similarities and Differences in Instant Messaging. *Journal of Language and Social Psychology*, *26*(4), 389-397.

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, *5*(1), 81-99.

Fraser, M., McCarthy, M. R., Shaukat, M., & Smith, P. (2007). Seconds matter: improving distributed coordination by tracking and visualizing display trajectories. In *CHI2007*(Vol. 2, p. 1303). ACM.

Gabsdil, M. (2003). Clarification in spoken dialogue systems. In *Proceedings of the 2003 AAAI Spring Symposium. Workshop on Natural Language Generation in Spoken and Written Dialogue* (pp. 28-35).

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and individual differences*, *14*(1), 53-65.

Garden, S., Cornoldi, C., & Logie, R. H. (2001). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology*, *16*(1), 35-50.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy?. *Trends in cognitive sciences*, *8*(1), 8-11.

Gergle, D. R. (2006). The value of shared visual information for task-oriented collaboration (PhD Thesis, Carnegie Melon University).

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Language efficiency and visual technology minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, *23*(4), 491-517.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction*.*28* (1).

Giles, H., Coupland, N., & Coupland, I. (1991). 1. Accommodation theory: Communication, context, and. *Contexts of accommodation: Developments in applied sociolinguistics*, 1.

Gill, A., Harrison, A., & Oberlander, J. (2004). Interpersonality: Individual differences and interpersonal priming. In *Proceedings of the 26th annual conference of the cognitive science society* (pp. 464-469).

Gluck, J., & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing*, *3*(3), 293-308.

Goldenfeld, N., Baron-Cohen, S., & Wheelwright, S. (2005). Empathizing and systemizing in males, females and autism. Clinical Neuropsychiatry, *2*(6), 338-345.

Golding, J. M., Graesser, A. C., & Hauselt, J. (1996). The process of answering direction-giving questions when someone is lost on a university campus: The role of pragmatics. *Applied Cognitive Psychology*, *10*(1), 23-39.

Golledge, R. G., Dougherty, V., & Bell, S. (1995). Acquiring Spatial Knowledge: Survey Versus Route-Based Knowledge in Unfamiliar Environments. *Annals of the Association of American Geographers*, *85*(1), 134-158.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*(5), 659-670.

Gravetter, F. J., & Forzano, L. A. B. (2011). *Research methods for the behavioral sciences*. Wadsworth Publishing Company.

Green, A., Eklundh, K. S., Wrede, B., & Li, S. (2006, October). Integrating miscommunication analysis in natural language interface design for a service robot. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 4678-4683). IEEE.

Greenspan, J. D., Craft, R. M., LeResche, L., Arendt-Nielsen, L., Berkley, K. J., Fillingim, R. B., ... & Traub, R. J. (2007). Studying sex and gender differences in pain and analgesia: a consensus report. *Pain*, *132*(Suppl 1), S26-45.

Grice, H. P. (1975). Logic and conversation. *1975*, 41-58.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, *68*(1), 155.

Guilford, J. P., & Zimmerman, W. S. (1948). The Guilford-Zimmerman Aptitude Survey. *Journal of applied Psychology*, *32*(1), 24.

Guindon, R., Shuldberg, K., & Conner, J. (1987, July). Grammatical and ungrammatical structures in user-adviser dialogues: evidence for sufficiency of restricted languages in natural language interfaces to advisory systems. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics* (pp. 41-44). Association for Computational Linguistics.

Guttman, L. (1954). A new approach to factor analysis: the Radex. In P. F. Lazarsfeld, (ed.),*Mathematical Thinking in the Social Sciences*. Glencoe: The Free Press, 1954, pp. 258–348.

Gutwin, C., & Penner, R. (2002, November). Improving interpretation of remote gestures with telepointer traces. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 49-57). ACM.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Lawrence Erlbaum.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1-51.

Hargittai, E., & Shafer, S. (2006). Differences in Actual and Perceived Online Skills: The Role of Gender. *Social Science Quarterly*, *87*(2), 432-448.

Harris, L. J. (1978). Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), Asymmetrical function of the brain (pp. 405-521). London: Cambridge University.

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication*, *11*(4), 910-931.

Hauptmann, A. G., & Rudnicky, A. I. (1988). Talking to computers: an empirical investigation. *International Journal of Man-Machine Studies*, *28*(6), 583-604.

Hausmann, M., & Schober, B. (2012). Sex and gender differences: New perspectives and new findings within a psychobiosocial approach. Zeitschrift für Psychologie/Journal of Psychology, *220*(2), 57-60.

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

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

Henley, N., & Kramarae, C. (1991). Miscommunication, power, and gender. Coupland, Nicolas; Wiemann, John. M.; Giles, Howard.(Eds.),*"Miscommunication" and Problematic Talk*, 18-43.

Herlitz, A., Airaksinen, E., & Nordström, E. (1999). Sex differences in episodic memory: the impact of verbal and visuospatial ability. *Neuropsychology; Neuropsychology*, *13*(4), 590.

Herring, S. (1996). Bringing familiar baggage to the new frontier: Gender differences in computer-mediated communication. *CyberReader*, 144-154.

Herring, S. C. (2000). Gender differences in CMC: Findings and implications. *Computer Professionals for Social Responsibility Journal*, *18*(1).

Hert, P. (1997). Social dynamics of an on-line scholarly debate. *The Information Society*, *13*(4), 329-360.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*. *13*(2), 141-154.

Hirst, G., McRoy, S., & Heeman, P. E. P. & Horton, S:(1994) Repairing Conversational Misunderstanding and Non-Understandings. *Speech Communication*, *12*, 213-229.

Hockey, B. A., Lemon, O., Campana, E., Hiatt, L., Hieronymus, J., Gruenstein, A., & Dowding, J. (2003, April). Targeted Help for Spoken Dialogue Systems: intelligent feedback improves naive users' performance. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1* (pp. 147-154). Association for Computational Linguistics.

Honda, A., & Nihei, Y. (2009). Sex differences in object location memory: The female advantage of immediate detection of changes. *Learning and Individual Differences*, *19*(2), 234-237.

Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, *6*(3&4), 287-303.

Houldsworth, C., & Mathews, B. P. (2000). Group composition, performance and educational attainment. *Education + Training*, *42*(1), 40-53.

Howell, D. C. (2009). *Statistical methods for psychology*. Wadsworth Publishing Company.

Howell, D. C. (2010). *Fundamental statistics for the behavioral sciences*. Wadsworth Publishing Company.

Hubona, G. S., & Shirah, G.W. (2004, January). The gender factor performing visualization tasks on computer media. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (pp. 9-pp). IEEE.

Hui, S. & Tenbrink, T. (2009). Telling Rolland where to go: HRI dialogues on route navigation. *Spatial Language and Dialogue*, Coventry, KR, Tenbrink, T., and Bateman, J.(Eds.), Oxford University Press, Oxford.

Hund, A. M., & Padgitt, A. J. (2010). Direction giving and following in the service of wayfinding in a complex indoor environment. *Journal of Environmental Psychology*, *30*(4), 553-564.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, *60*(6), 581.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26.

Iachini, T., Sergi, I., Ruggiero, G., & Gnisci, A. (2005). Gender differences in object location memory in a real three-dimensional environment. *Brain and Cognition*. *59*(1), 52-59.

Iaria, G., Lanyon, L. J., Fox, C. J., Giaschi, D., & Barton, J. J. (2008). Navigational skills correlate with hippocampal fractional anisotropy in humans. *Hippocampus*, *18*(4), 335-339.

Institute of Medicine, Committee on Understanding the Biology of Sex and Gender Differences, Board on Health Sciences Policy. Exploring the Biological Contributions to Human Health: Does Sex Matter? Washington, DC: National Academy Press,

2001, p. 1 - 6. Retrieved from: http://www.nap.edu/openbook.php?record_id=10028&page=1

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*(5), 365-371.

Ishikawa, T., & Kiyomoto, M. (2008). Turn to the left or to the west: verbal navigational directions in relative and absolute frames of reference. *Geographic information science*, 119-132.

Jackson, L. A., Ervin, K. S., Gardner, P. D., & Schmitt, N. (2001). Gender and the Internet: Women communicating and men searching. *Sex roles*, *44*(5), 363-379.

James, T. W., & Kimura, D. (1997). Sex differences in remembering the locations of objects in an array: Location-shifts versus location-exchanges. *Evolution and Human Behavior*, *18*(3), 155-163.

Jokinen, K., & McTear, M. (2009). Spoken Dialogue Systems. *Synthesis Lectures on Human Language Technologies*, *2*(1), 1-151.

Joseph, R. (2000). The evolution of sex differences in language, sexuality, and visual–spatial skills. *Archives of Sexual Behavior*, *29*(1), 35-66.

Karsenty, L. (1999). Cooperative work and shared visual context: An empirical study of comprehension problems in side-by-side and remote help dialogues. *Human-Computer Interaction*, *14*(3), 283-315.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity?. *Intelligence*, *35*(3), 211-223.

Kaushanskaya, M., Marian, V., & Yoo, J. (2011). Gender differences in adult word learning. *Acta psychologica*, *137*(1), 24-35.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47*(3), 193-198.

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless?. *Computers in Human Behavior*, *28*(1), 241-250.

Kimura, D. (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current opinion in neurobiology*, *6*(2), 259-263.

Kimura, D. (2000). *Sex and cognition*. MIT press.

Kinsey, B. L., Towle, E., O'Brien, E. J., & Bauer, C. F. (2008). Analysis of self-efficacy and ability related to spatial tasks and the effect on retention for students in engineering. *International Journal of Engineering Education*, *24*(3), 488-494.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, 139-184.

Klippel, A., Richter, K. F., & Hansen, S. (2009). Cognitively ergonomic route directions. *Handbook of Research on Geoinformatics*, 230-238.

Klippel, A., Tappe, H., & Habel, C. (2003). Pictorial representations of routes: Chunking route segments during comprehension. *Spatial cognition III*, 1034-1034.

Koulouri, T., & Lauria, S. (2009, September). Exploring miscommunication and collaborative behaviour in human-robot interaction. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 111-119). Association for Computational Linguistics.

Koulouri, T., Lauria, S., Macredie, R. D., & Chen, S. (2012). Are we there yet?: The role of gender on the effectiveness and efficiency of user-robot communication in navigational tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *19*(1), 4.

Krahmer, E., Swerts, M., Theune, M., & Weegels, M. (2001). Error detection in spoken human-machine interaction. *International journal of speech technology*, *4*(1), 19-30.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction*, *18*(1), 13-49.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002, November). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 31-40). ACM.

Kucian, K., Loenneker, T., Dietrich, T., Martin, E., & Von Aster, M. (2005). Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychology Science*, *47*(1), 112.

Lakoff, R. T. (1975). *Language and woman's place* (Vol. 56). M. Bucholtz (Ed.). New York: Harper & Row.

Landau, B., & Jackendoff, R. (1993). " What" and" where" in spatial language and spatial cognition. *Behavioral and brain sciences*, *16*, 217-217.

Landau, B. (1998). Spatial cognition and spatial language: What do we need to know to talk about space? *AAAI Technical Report WS-98-06.*

Landow, G. P. (1992). *HyperText 2.0: The Convergence of Contemporary Critical Theory and Technology (Parallax: Re-visions of Culture and Society Series)*. Johns Hopkins University Press.

Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative web searching behavior: an elementary school study. *Information Processing & Management*, *38*(3), 427-443.

Larsson, S., & Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, *6*(3 & 4), 323-340.

Lauria, S., Bugmann, G., Kyriacou, T., Bos, J., & Klein, A. (2001). Training personal robots using natural language instruction. *Intelligent Systems, IEEE*, *16*(5), 38-45.

Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex roles*, *30*(11), 765-779.

Lawton, C. A. (1996). Strategies for indoor wayfinding: The role of orientation. *Journal of Environmental Psychology*, *16*, 137-145.

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles*, *47*, 389-401.

Lawton, C. A., Charleston, S. I., & Sieles, A. S. (1996). Individual and gender-related differences in indoor wayfinding. *Environment and Behavior*, *28*, 204-219.

Lawton, C. A., & Morrin, K. A. (1999). Gender differences in pointing accuracy in computer-simulated 3D mazes. *Sex Roles*, *40*, 73-92.

Lawton, C. A. (2010). Gender, spatial abilities, and wayfinding. In J. Chrisler & D. McCreary (Eds.), *Handbook of gender research in psychology*, New York: Springer.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*. Wiley.

Lea, M., & Spears, R. (1992). Paralanguage and social perception in computer-mediated communication. *Journal of Organizational Computing and Electronic Commerce*, *2*(3-4), 321-341.

Leaper, C., & Robnett, R. D. (2011). Women Are More Likely Than Men to Use Tentative Language, Aren't They? A Meta-Analysis Testing for Gender Differences and Moderators. *Psychology of Women Quarterly*, *35*(1), 129-142.

Lemon, O., Gruenstein, A., Battle, A., & Peters, S. (2002, July). Multi-tasking and collaborative activities in dialogue systems. In *Proceedings of the 3rd SIGdial workshop on Discourse and dialogue-Volume 2* (pp. 113-124). Association for Computational Linguistics.

Lemon, O. (2004). Context-sensitive speech recognition in ISU dialogue systems: results for the grammar switching approach. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue, CATALOG'04*.

Levin, D. T., Killingsworth, S. S., & Saylor, M. M. (2008). Concepts about the capabilities of computers and robots: A test of the scope of adults' theory of mind. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 57-64). ACM.

Levin, D. T., Killingsworth, S. S., Saylor, M. M., Gordon, S. M., & Kawamura, K. (2013). Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human–Computer Interaction*, 28, 161-191.

Levin, E. & Passonneau, R. (2006). A WOz Variant with Contrastive Conditions, In *Proceedings of the Dialog-on-Dialog Workshop, INTERSPEECH 2006 ICSLP*. Pittsburg, PA.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, *28*(4), 612-625.

Levinson, S. C. (1983). *Pragmatics.* Cambridge, England: Cambridge University.

Levinson, S. C. (2003). Spatial language. *Encyclopedia of cognitive science*. pp. 131-137. London: Nature Publishing Group.

Lewis, J. R. (1993). Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction*, *5*(4), 383-392.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57-78.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development*, 1479-1498.

Litman, D. J., & Pan, S. (2002). Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction*, *12*(2), 111-137.

Litman, D. J., Hirschberg, J. B., & Swerts, M. (2000, April). Predicting automatic speech recognition performance using prosodic cues. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 218-225). Morgan Kaufmann Publishers Inc.

Ljungblad, S., & Holmquist, L. E. (2007, April). Transfer scenarios: grounding innovation with marginal practices. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 737-746). ACM.

Lohman, D. F. (1979). *Spatial Ability: A Review and Reanalysis of the Correlational Literature* (No. TR-8). STANFORD UNIV CALIF SCHOOL OF EDUCATION.

Lövdén, M., Herlitz, A., Schellenbach, M., Grossman-Hunter, B. Krüger, A., & Lindenberger, U. (2007). Quantitative and qualitative sex differences in spatial navigation. *Scandinavian journal of psychology*, *48*(5), 353-358.

Lovelace, K., Hegarty, M., & Montello, D. (1999). Elements of good route directions in familiar and unfamiliar environments. *Spatial information theory. Cognitive and computational foundations of geographic information science*, 751-751.

MacMahon, M. (2004). A framework for understanding verbal route instructions. In *Proceedings of AAAI Fall Symposium on the Intersection of Cognitive Science and Robotics: From Interfaces to Intelligence.* (pp. 97-102).

Magoulas, G., Chen, S., & Dimakopoulos, D. (2004). A personalised interface for web directories based on cognitive styles. *User-Centered Interaction Paradigms for Universal Access in the Information Society*, 159-166.

Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2004). Selective sex differences in declarative memory. *Memory & cognition*, *32*(7), 1160-1169.

Malinowski, J. C., & Gillespie, W. T. (2001). Individual differences in performance on a large-scale, real-world wayfinding task. *Journal of Environmental Psychology*, *21*(1), 73-82.

Maltz, D. N., & Borker, R. A. (1982). A cultural approach to male-female miscommunication. In *Language and Social Identity*. pp. 196- 216. Cambridge University Press.

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior*. *28*(3), 809-819.

McCoy, L. P., Heafner, T. L., Burdick, M. G., & Nagle, L. M. (2001). Gender Differences in Computer Use and Attitudes on a Ubiquitous Computing Campus. In *Annual Meeting of the American Educational Research Association* (Vol. 2001, No. 1).

McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2003, May). The impact of pair programming on student performance, perception and persistence. In *Proceedings of the 25th international conference on Software engineering* (pp. 602-607). IEEE Computer Society.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: method, theory and practice*. Cambridge University Press.

McGlone, J. (1980). Sex differences in human brain asymmetry: A critical survey. *Behavioral and Brain Sciences*, *3*(2), 215-227.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological methods*, *11*(4), 386.

McGuiness, D., & Sparks, J. (1983). Cognitive style and cognitive maps: Sex differences in representations of a familiar terrain. *Journal of Mental Imagery*. 7 (1983), pp. 91–100.

McRoy, S. W. (1998). Achieving robust human-computer communication. *International Journal of Human-Computer Studies*, *48*(5), 681-704.

McTear, M. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surverys*, *34*(1), 90–169.

McTear, M. (2008). Handling miscommunication in spoken dialogue systems: why bother? In L. Dybkjaer , L & W. Minker (Eds.), *Recent Trends in Discourse and Dialogue.* (101 – 122). Springer.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, *84*(4), 857.

Michon, P. E., & Denis, M. (2001). When and why are visual landmarks used in giving directions?. *Spatial information theory*, 292-305.

Mills, G. J., & Healey, P. G. (2006, September). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of Brandial06 The 10th Workshop on the Semantics and Pragmatics of Dialogue. University of Potsdam, Germany* (pp. 122-129).

Mills, G.J. (2007). *Semantic co-ordination in dialogue: the role of direct interaction.* (PhD thesis, Queen Mary University, London, UK). Retrieved from http://www.dcs.qmul.ac.uk/tech_reports/RR-07-07.pdf.

Mills, M.E. (2011). Sex Difference vs. Gender Difference? Oh, I'm So Confused! Psychology Today blog post. Retrieved from: http://www.psychologytoday.com/blog/the-how-and-why-sex-differences/201110/sex-difference-vs-gender-difference-oh-im-so-confused

Moè, A., & Pazzaglia, F. (2006). Following the instructions!: Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*(4), 369-377.

Moffat, S. D., & Hampson, E. (1996). A curvilinear relationship between testosterone and spatial cognition in humans: possible influence of hand preference. *Psychoneuroendocrinology*, *21*(3), 323-337.

Moffat, S. D., Hampson, E., & Hatzipantelis, M. (1998). Navigation in a "virtual" maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior*, *19*(2), 73-87.

Mohler, J. L. (2009). A review of spatial ability research. *Engineering Design Graphics Journal*, *72*(2).

Montello, D. R., Lovelace, K. L., Golledge, R. G. & Self, C. M. (1999). Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American Geographers*, 89(3), p.515–534.

Montello, D. R. (2009). Cognitive geography. In R. Kitchin & N. Thrift (Eds.), *International encyclopedia of human geography, Vol. 2*. pp. 160-166. Oxford Elsevier Science.

Montello, D. R., & Pick Jr, H. L. (1993). Integrating knowledge of vertically aligned large-scale spaces. *Environment and Behavior*, *25*(3), 457-484.

Montello, D. R., & Sas, C. (2006). Human factors of wayfinding in navigation.

Money, J. (1987). Propaedeutics of diecious GI/R: Theoretical foundations for understanding dimorphic gender-identity/role. Masculinity/femininity: Basic perspectives, 22-43.

Moon, Y., & Nass, C. I. (1996). How ‘‘real’’ are computer personalities? Psychological responses to personality types in human–computer interaction. Communication Research, 23(6), 651–674.

Moore, D. S., & Johnson, S. P. (2008). Mental Rotation in Human Infants A Sex Difference. *Psychological Science*, *19*(11), 1063-1066.

Moratz, R. & Fischer, K. (2000). Cognitively adequate modelling of spatial reference in human-robot interaction. *In 12th IEEE International Conference on Tools with Artificial Intelligence.* Vancouver, British Columbia, Canada, 13-15 November.

Moratz, R., Fischer, K & Tenbrink, T. (2001). Cognitive modeling of spatial reference for human-robot interaction. *International Journal on Artificial Intelligence Tools*, 10 (4), p.589-611.

Moratz, R., & Tenbrink, T. (2003, October). Instruction modes for joint spatial reference between naive users and a mobile robot. In *Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on* (Vol. 1, pp. 43-48). IEEE.

Moshell, J. M., & Hughes, C. E. (2002). . Virtual Reality as a Tool For Academic Learning. In *Handbook of Virtual Environments: Design, Implementation, and Applications.* K. M. Stanney (Ed.) Lawrence Erlbaum Associates.

Mulac, A., & Bradac, J. J. (1995). Women's style in problem solving interaction: Powerless, or simply feminine?. In *Gender, power and communication. P. J. Kalbfleish & M. J. Cody Eds*., pp. 83-104. Hillsdale, NJ: Lawrence Erlbaum.

Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language and Communication*, *14*, 299-299.

Mulac, A., Erlandson, K. T., Farrar, W. J., Hallett, J. S., Molloy, J. L., & Prescott, M. E. (1998). “Uh-huh. What's That All About?” Differing Interpretations of Conversational Backchannels and Questions as Sources of Miscommunication Across Gender Boundaries. *Communication Research*, *25*(6), 641-668.

Mulac, A., Seibold, D. R., & Farris, J. L. (2000). Female and male managers’ and professionals’ criticism giving differences in language use and effects. *Journal of Language and Social Psychology*, *19*(4), 389-415.

Muller, P., & Prévot, L. (2009). Grounding information in route explanation dialogues. Spatial Language and Dialogue, Coventry, KR, Tenbrink, T., and Bateman, J.(Eds.), Oxford University Press, Oxford.

Mutlu, B., Osman, S., Forlizzi, J., Hodgins, J., & Kiesler, S. (2006, September). Task structure and user attributes as elements of human-robot interaction design. In *Robot*

and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on (pp. 74-79). IEEE.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422-432.

Nass, C. I., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied, 7(3), 171–181.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, *29*(5), 1093-1109.

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology*, *27*(10), 864-876.

Nomura, T., & Takagi, S. (2011, November). Exploring effects of educational backgrounds and gender in human-robot interaction. In *User Science and Engineering (i-USEr), 2011 International Conference on* (pp. 24-29). IEEE.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of Human Behavior in Human--Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots. *Robotics, IEEE Transactions on*, *24*(2), 442-451.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, *15*(5), 625-632.

O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences*, *24*(5), 595-601.

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

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

Oulasvirta, A., Engelbrecht, K. P., Jameson, A., & Moller, S. (2006). The relationship between user errors and perceived usability of a spoken dialogue system. *ISCA/DEGA*.

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *11*(3), 300-328.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*, 2382.

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006, April). Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1177-1180). ACM.

Piccardi, L., Risetti, M., Nori, R., Tanzilli, A., Bernardi, L., & Guariglia, C. (2011). Perspective changing in primary and secondary learning: A gender difference study. *Learning and Individual Differences*, *21*(1), 114-118.

Pick, H.L., Montello, D.R., & Somerville, S.C. (1988). Landmarks and the coordination and integration of spatial information. *British Journal of Developmental Psychology*, 6, 372-375.

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(02), 212-225.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language & Computation*, *4*(2), 203-228.

Picucci, L., Caffò, A. O., & Bosco, A. (2011). Besides navigation accuracy: Gender differences in strategy selection and level of spatial confidence. *Journal of Environmental Psychology*, *31*(4), 430-438.

Porzel, R. (2006). How people (should) talk to computers. In K. Fischer (Ed.) *Proceedings of the Workshop on How People Talk to Computers, Robots, and Other Artificial Communication Partners* (7-37). Delmenhorst, Germany: HWK.

Postma, A., Izendoorn, R., & De Haan, E. H. (1998). Sex differences in object location memory. *Brain and Cognition*, *36*(3), 334-345.

Presson, C.C. & Montello, D.R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology*, 6, 378-381.

Prinsen, F. R., Volman, M. L. L., & Terwel, J. (2007). Gender-related differences in computer-mediated communication and computer-supported collaborative learning. *Journal of Computer Assisted Learning*, *23*(5), 393-409.

Prinsen, F. R., Volman, M. L. L., Terwel, J., & Van den Eeden, P. (2009). Effects on participation of an experimental CSCL-programme to support elaboration: Do all students benefit?. *Computers & Education*, *52*(1), 113-125.

Purver, M. R. J. (2004). *The theory and use of clarification requests in dialogue* (PhD thesis, University of London).

Rahman, Q., Bakare, M., & Serinsu, C. (2011). No sex differences in spatial location memory for abstract designs. *Brain and cognition*, *76*(1), 15-19.

Raux, A., & Eskenazi, M. (2004). Using task-oriented spoken dialogue systems for language learning: potential, practical applications and challenges. In *InSTIL/ICALL Symposium 2004*.

Reitter, D., & Moore, J. D. (2007a, June). Predicting success in dialogue. In *Annual Meeting-Association for Computational Linguistics* (Vol. 45, No. 1, p. 808).

Reitter, D. & Moore, J.D. (2007b). Successful dialogue requires syntactic alignment. *20th Annual CUNY Conference on Human Sentence Processing.*

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & cognition*, *27*(4), 741-750.

Richert, D., Halabi, A., Eaglin, A., Edwards, M., & Bardzell, S. (2011, May). Arrange-A-Space: tabletop interfaces and gender collaboration. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (pp. 1495-1500). ACM.

Rodríguez, K. J., & Schlangen, D. (2004). Form, Intonation and Function of Clarification Requests in German task-oriented spoken dialogues. *on the Semantics and Pragmatics of Dialogue*, *19*, 101.

Rogelberg, S. G., & Rumery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small Group Research*, *27*(1), 79-90.

Ross, S. P., Skelton, R. W., & Mueller, S. C. (2006). Gender differences in spatial navigation in virtual space: implications when using virtual environments in instruction and assessment. *Virtual Reality*, *10*(3), 175-184.

Rosson, M. B., Sinha, H., Bhattacharya, M., & Zhao, D. (2008). Design planning by end-user web developers. *Journal of Visual Languages & Computing*, *19*(4), 468-484.

Roy, M., & Chi, M. T. (2003). Gender differences in patterns of searching the web. *Journal of Educational Computing Research*, *29*(3), 335-348.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in" desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, *3*(2), 143.

Ruggiero, G., Sergi, I., & Iachini, T. (2008). Gender differences in remembering and inferring spatial distances. *Memory*, *16*(8), 821-835.

Rutter, D. R. (1987). *Communicating by telephone*. Pergamon Press.

Saccuzzo, D. P., Craig, A. S., Johnson, N. E., & Larson, G. E. (1996). Gender differences in dynamic spatial abilities. *Personality and Individual Differences*, *21*(4), 599-607.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696-735.

Sandstrom, N. J., Kaufman, J., & A Huettel, S. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research*, *6*(4), 351-360.

Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies?. *Behavioral neuroscience*, *116*(3), 403.

Saucier, D., Bowman, M., & Elias, L. (2003). Sex differences in the effect of articulatory or spatial dual-task interference during navigation. *Brain and Cognition*, *53*(2), 346-350.

Saucier, D., Lisoway, A., Green, S., Elias, L., Beschin, N., Robertson, I. H., ... & Eals, M. (2007). Female advantage for object location memory in peripersonal but not extrapersonal space. *Journal of the International Neuropsychological Society*, *13*(4), 683-686.

Sauro, J., & Lewis, J. R. (2012). Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

Savicki, V., Foster, D. A., & Kelley, M. (2006). Gender, Group Composition and Task Type in Virtual Groups. *Gender and communication at work*, 270.

Savicki, V., & Kelley, M. (2000). Computer mediated communication: Gender and group composition. *CyberPsychology & Behavior*, *3*(5), 817-826.

Savicki, V., Kelley, M., & Lingenfelter, D. (1996). Gender and group composition in small task groups using computer-mediated communication. *Computers in Human Behavior*, *12*(2), 209-224.

Schechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation. In *Proc. CHI 2003* (pp. 281-288).

Schermerhorn, P., Scheutz, M., & Crowell, C. R. (2008, March). Robot social presence and gender: Do females view robots differently than males?. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 263-270). ACM.

Schlangen, D. (2004, April). Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue* (pp. 136-143).

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of personality and social psychology*, *85*(3), 440.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review*, *115*(2), 336.

Schmitz, S. (1997). Gender-related strategies in environmental development: Effects of anxiety on wayfinding in and representation of a three-dimensional maze. *Journal of Environmental Psychology*, *17*(3), 215-228.

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

Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. Spatial Language and Dialogue, Coventry, KR, Tenbrink, T., and Bateman, J.(Eds.), Oxford University Press, Oxford., p.23-39.

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

Searle, J. R. (1992). Conversation. In J. R. Searle, H. Parrett, & J. Verschueren (Eds.), *(On) Searle on conversation* (pp. 7–29). Amsterdam, Philadelphia: J. Benjamins Pub. Co.

Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-computer interaction*, *10*(4), 401-444.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: the impact of self-relevance and the manner of stereotype activation. *Journal of personality and social psychology*, *83*(3), 638.

Siegel, M., Breazeal, C., & Norton, M. I. (2009, October). Persuasive robotics: the influence of robot gender on human behavior. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on* (pp. 2563-2568). IEEE.

Silverman, I., & Eals, M. (1992). Spatial sex differences: Evolutionary theory and data. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. N.Y.: Oxford University Press, 487-503.

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of sexual behavior*, *36*(2), 261-268.

Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior*, *21*(3), 201-213.

Simon, H. A. (1981). *The sciences of the artificial*. 2nd ed. Cambridge MIT press.

Skantze G. (2005). Exploring human error recovery strategies: implications for spoken dialogue systems. *Speech Communication*, *45*(3), 207-359.

Skantze, G. (2007). *Error handling in spoken dialogue systems: managing uncertainty, grounding and miscommunication.* (PhD Thesis, KTH, Stockholm, Sweden). Retrieved from http://www.speech.kth.se/prod/publications/files/3101.pdf

Skantze, G. (2008). Galatea: a discourse modeller supporting concept-level error handling in spoken dialogue systems. *Recent Trends in Discourse and Dialogue*, 155-189.

Smith, S. P., & Marsh, T. (2004). Evaluating design guidelines for reducing user disorientation in a desktop virtual environment. *Virtual Reality*, *8*(1), 55-62.

Sorby, S. A. (2007). Developing 3D spatial skills for engineering students. *Australasian Journal of Engineering Education*, *13*(1), 1-11.

Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In *Spatial information theory. Cognitive and computational foundations of geographic information science* (pp. 37-50). Springer Berlin Heidelberg.

Spiers, M. V., Sakamoto, M., Elliott, R. J., & Baumann, S. (2008). Sex differences in spatial object-location memory in a virtual grocery store. *CyberPsychology & Behavior*, *11*(4), 471-473.

Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Communications of the ACM*, *30*(1), 32-47.

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology*, *83*(3), 361.

Stoia, L., Shockley, D. M., Byron, D. K., & Fosler-Lussier, E. (2006, July). Noun phrase generation for situated dialogs. In *Proceedings of the fourth international natural language generation conference* (pp. 81-88). Association for Computational Linguistics.

Strauch, B. (2004). *Investigating human error: Incidents, accidents, and complex systems*. Ashgate Pub Limited.

Stubbe, M. (2010). " Was That My Misunderstanding?": Managing Miscommunication and Problematic Talk at Work. (PhD Thesis. Victoria University, Wellington.) Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/1462

Stupka, R. (2011). Communication Accommodation in Mixed Gender Dyads. *Oshkosh Scholar*. Vol. 6.

Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of applied developmental psychology*, *15*(1), 13-32.

Sun, X. (2008). Why gender matters in CMC: gender differences in remote trust and performance with initial social activities (PhD Thesis, Drexel University).Retrieved from http://idea.library.drexel.edu/handle/1860/2930.

Suzuki, N., & Katagiri, Y. (2007). Prosodic alignment in human–computer interaction. *Connection Science*, *19*(2), 131-141.

Tan, D. S., Czerwinski, M. P., & Robertson, G. G. (2006). Large displays enhance optical flow cues and narrow the gender gap in 3-D virtual navigation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 318-333.

Tan, D. S., Czerwinski, M., & Robertson, G. (2003, April). Women go with the (optical) flow. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 209-215). ACM.

Tannen, D. (1989). Talking voices: Repetition, dialogue and imagery in conversational discourse (Vol. 6). Cambridge University Press.

Tannen, D. (1990). You just don't understand: Men and women in conversation. *New York: Morrow*.

Tannen, D. (1994). *Gender and discourse*. Oxford University Press, USA.

Tenbrink, T., & Hui, S. (2007). Negotiating spatial goals with a wheelchair. *Proceedings of the 8th SIGdial*.

Tenbrink, T. (2007). Methods for analyzing natural discourse: Investigating spatial language in HRI vs. in a no-feedback web study. In A. G. Cohn, C. Freksa, & B. Nebel (Eds.), *Proc. Dagstuhl Seminar on Spatial Cognition: Specialisation and Integration*.

Tenbrink, T., Ross, R. J., Thomas, K. E., Dethlefs, N., & Andonova, E. (2010). Route instructions in map-based human–human and human–computer dialogue: A comparative analysis. *Journal of Visual Languages & Computing*, *21*(5), 292-309.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, *53*(5), 433-441.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2007). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, *22*(7), 996-1013.

Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language?. *Psychological Science*, *12*(2), 171-175.

Thurstone, L. L., & Thurstone, T. G. (1958). *Manual for the SRA Primary Mental Abilities: Ages 11 to 17*. Science Research Associates.

Tlauka, M., Brolese, A., Pomeroy, D., & Hobbs, W. (2005). Gender differences in spatial knowledge acquired through simulated exploration of a virtual shopping centre. *Journal of Environmental Psychology*, *25*(1), 111-118.

Tversky, B., & Lee, P. (1999). Pictorial and verbal tools for conveying routes. Spatial information theory. Cognitive and computational foundations of geographic information science, 752-752.

Tversky, B., Heiser, J., Lee, P., & Daniel, M. P. (2009). Explanations in gesture, diagram, and word. Spatial Language and Dialogue, Coventry, KR, Tenbrink, T., and Bateman, J.(Eds.), Oxford University Press, Oxford.

Ullman, M. T., Miranda, R. A., & Travers, M. L. (2008). Sex differences in the neurocognition of language. *Sex differences in the brain: From genes to behavior*, 291-310.

Underwood, J., Underwood, G., & Wood, D. (2000). When does gender matter?: Interactions during computer-based problem solving. *Learning and Instruction*, *10*(5), 447-462.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2012). The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. Psychological Bulletin.

Vandenberg, SG, & Kuse, AR (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.

Vanetti, E. J., & Allen, G. L. (1988). Communicating Environmental Knowledge The Impact of Verbal and Spatial Abilities on the Production and Comprehension of Route Directions. *Environment and Behavior*, *20*(6), 667-682.

Veinott, E. S., Olson, J., Olson, G. M., & Fu, X. (1999, May). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* (pp. 302-309). ACM.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological bulletin*, *117*(2), 250.

Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic bulletin & review*, *14*(1), 23-38.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817.

Walker, J. T., Krasnoff, A. G., & Peaco, D. (1981). Visual spatial perception in adolescents and their parents: The X-linked recessive hypothesis. *Behavior Genetics*, *11*(4), 403-413.

Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997, July). PARADISE: A framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics* (pp. 271-280). Association for Computational Linguistics.

Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1998). Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer speech and language*, *12*(4), 317-348.

Walker, M., Kamm, C., & Litman, D. (2000a). Towards developing general models of usability with PARADISE. *Natural Language Engineering*, *6*(3 & 4), 363-377.

Walker, M., Wright, J., & Langkilde, I. (2000b). Using natural language processing and discourse features to identify understanding errors in a spoken dialogue system. In *Proceedings of ICML*.

Waller, D. (2000). Individual differences in spatial learning from computer-simulated environments. *Journal of Experimental Psychology: Applied*, *6*(4), 307.

Ward, N., & Nakagawa, S. (2004). Automatic user-adaptive speaking rate selection. *International Journal of Speech Technology*, *7*(4), 259-268.

Ward, S. L., Newcombe, N., & Overton, W. F. (1986). Turn Left at the Church, Or Three Miles North A Study of Direction Giving and Sex Differences. *Environment and Behavior*, *18*(2), 192-213.

Webley, P. (1981). Sex differences in home range and cognitive maps in eight-year old children. *Journal of environmental Psychology*, *1*(4), 293-302.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: how do they work and when are they useful?. *International Journal of Man-Machine Studies*, *39*(5), 813-842.

Whittaker, S. (2003a). Things to talk about when talking about things. *Human–Computer Interaction*, *18*(1-2), 149-170.

Whittaker, S. (2003b). Theories and methods in mediated communication. *The handbook of discourse processes*, 243-286.

Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, *9*(2), 152-165.

Williams, C. L., Barnett, A. M., & Meck, W. H. (1990). Organizational effects of early gonadal secretions on sexual differentiation in spatial memory. *Behavioral neuroscience*, *104*(1), 84.

Williams, J. D., & Young, S. (2004, October). Characterizing task-oriented dialog using a simulated ASR channel. In *Proceedings of the ICSLP*.

Wong, Y.K., Shi, Y., & Wilson, D. (2004). Experience, gender composition, social presence, decision process satisfaction and group performance, *ACM International Conference Proceeding Series, 58***,** 1-10.

Woods, S., Dautenhahn, K., Kaouri, C., Boekhorst, R., & Koay, K. L. (2005, December). Is this robot like me? Links between human and robot personality traits. In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on* (pp. 375-380). IEEE.

World Health Organization. (2007). What do we mean by "sex" and "gender". *Gender, Women and Health*. Retrieved from: http://www.who.int/gender/whatisgender/en/

Zysno, P. V. (1997). The modification of the phi-coefficient reducing its dependence on the marginal distributions. *Methods of Psychological Research Online*, *2*(1).

Appendix

I. Instructions to participants in the 'robot' role

The experiment investigates how people converse with robots.

The aim of the project is to achieve natural communication with a robot drawing upon the principles of human conversation. To this end, we are collecting data from interactions between people.

You will interact and collaborate with another person to complete a task. This person will talk to you under the impression that he/she is talking to a robot. However, it is important that you talk naturally to him/her and not modify your language so that you would "sound like a robot".

The experiment involves reaching six particular locations by navigating in a little town. Each task is completed, when the robot (you!) arrives to the destination.

The user has access to the full map of the town and he/she will give you the instructions on how to go to a particular location.

The user is instructed to start each interaction with **"Hello"**, respond with **"Hello"**, too. Remember to end the task with **"Goodbye"** (important!).

The user can terminate a task at any point and start a new one. In this case, you will receive an "abort task" message. Respond with **"Goodbye"** and move on to next task.

You move the robot with the arrows. You interact with the user by typing in messages and you can also read his/hers. You can also view the user's old messages.

Finally, the robot has the ability to **learn and remember** a route. For instance, if it has previously gone from the PUB to the LIBRARY, when requested, it could execute this route again without asking for instructions. To simulate this ability, when a route is completed, in the next task, a new button appears on your screen for this route (e.g., *PUB to LIBRARY*). Thus, if the user explicitly requests to reuse a previous route, you can click on the button and the robot magically appears outside that location (e.g., the library).

The buttons that represent the routes you remember should be used only when the starting point and your current location are the same. For instance, if the user asks you to take again the route PUB to LIBRARY, you can click on the corresponding button if and only if you are currently outside the PUB. Therefore, if you are not outside the PUB, you should not follow this instruction.

In a nutshell, the focus of the experiment is the verbal interaction and collaboration with the user not to test your ability to follow directions.

II. Instructions to participants in the user role

The experiment investigates how people converse with robots.

Your task is to guide the robot to several destinations through a little town. The interface that you will use to communicate with the robot consists of a map and an application to exchange messages.

You have access to the full map of the town. On the other hand, the robot can only see the area that surrounds it, so it has to rely on your instructions. [In the upper right corner of your map, you see what it sees.][25]

You should NOT use UP, DOWN, NORTH, SOUTH etc.

The experiment involves guiding the robot to six different locations. When the robot reaches the location, the task is completed.

Always start the interaction with the robot by typing in "**hello**" and when the task is completed type in **"goodbye".** The robot will give the same responses.

The robot is fluent in understanding spatial language and can produce appropriate verbal responses.

The robot is able to **learn and remember** routes. For instance, if the robot has gone from the PUB to the LIBRARY, when you ask it to follow that route again, it will execute and arrive at the LIBRARY without requesting instructions. Thus, any known routes can be used within your instructions for new destinations. For instance, if you want it to go from the PUB to the BANK and the LIBRARY is on the way, you could directly ask it to go to the LIBRARY and then give further directions to the BANK.

[25] The text within the square brackets was removed in the instructions for users in the No Monitor condition.

Note: you should make sure that the starting point is the same as the current location of the robot. If the robot is not outside the PUB you should not instruct it to take the known route from the PUB to the LIBRARY.

You can terminate a task at any point. At the end of each task, there are a few brief questions for you to answer about the interaction. Even if you abort a task, you should complete the questionnaire and continue to the next task.

If you wish to terminate, type in **"abort task"** and proceed to the next task. Do not try the same task again.

The focus of the experiment is the verbal interaction with the robot not to test your ability to guide it to a location.

III. Post-task user questionnaires

SYSTEM SETUP:	
Task No:	

Tick your level of agreement for each of the following statements:

	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
I did well in completing the task							
The system was easy to use							
The system was accurate							
The system was helpful							
I am generally satisfied with this interaction							

IV. Consent Form

Data Collection and Analysis

Experimenter: Theodora Koulouri

School of Information Systems, Computing and Mathematics, Brunel University

The experiment investigates how people converse with robots.

If you agree to participate, you will be given six navigation tasks to complete by typing messages on a desktop messaging application. The interaction will be recorded anonymously for subsequent analysis. You may also be asked to complete anonymous questionnaires after each task and the experiment about your experience.

You are free to terminate the experiment at any time and request the log to be deleted and you would still receive payment in full. For your participation in this study, you will be paid £10.

There is no risk to you in this study. The purpose is to collect dialogue data, not to assess other skills.

By signing this form, you agree to allow the experimenters to collect and analyse the recorded interaction.

_____

Your name (printed)

_____ _____

Signature Date

_____ _____

Experimenter Date