Conversational Interfaces for Task-oriented Spoken Dialogues

Design Aspects Influencing Interaction Quality

ANDREEA NICULESCU

Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

Andreea Ioana Niculescu

PhD dissertation committee:

Chairman and Secretary:
Prof. dr. ir. Ton. J. Mouthaan, Universiteit Twente, NL
Promotor:
Prof. dr. ir. Anton Nijholt, Universiteit Twente, NL
Assistant-promotor:
Dr. Betsy van Dijk, Universiteit Twente, NL
Members:
Prof. Alan Dix, Computing Department, Lancaster University, UK
Dr. Li Haizhou, Institute for Infocomm Research (I^2R), Singapore
Prof. dr. Gerrit van der Veer, Open University, NL
Prof. dr. Vanessa Evers, Universiteit Twente, NL
Prof. dr. Dirk van Heylen, Universiteit Twente, NL
Paranymphs:
Riham Abdel Kader
Christian Mühl

CTIT Dissertation Series No. 11-212
Center for Telematics and Information Technology (CTIT)
P.O. Box 217 – 7500AE Enschede – The Netherlands
ISSN: 1381-3617

Taaluitgeverij Neslia Paniculata
Uitgeverij voor Lezers en Schrijvers van Talige Boeken
Nieuwe Schoolweg 28 – 7514 CG Enschede – The Netherlands.

Human Media Interaction
The research reported in this thesis has been carried out at the Human Media Interaction research group of the University of Twente.

SIKS Dissertation Series No. 2011-49
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

The cover images represent Danbo and his little brother Mini Danbo. The characters are small cardboard robots commissioned by Amazon Japan from Kiyohiko Azuma, the Japanese writer who designed Danbo to appear in his comedy manga serie *Yotsuba*. The name Danbo is a pun on the Japanese word *danboru* (ダンボ), meaning corrugated cardboard. The cute Danbo became a true Flickr sensation after attracting the interest of many photographers in the world.

ISBN: 978-90-752-9600-6
ISSN: 1381-3617, No. 11-212

CONVERSATIONAL INTERFACES FOR TASK-ORIENTED SPOKEN DIALOGUES: DESIGN ASPECTS INFLUENCING INTERACTION QUALITY

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee
to be publicly defended
on Tuesday, November 22, 2011 at 16.45

by

Andreea Ioana Niculescu

born in Bucharest, Romania

This thesis has been approved by:
Prof. dr. ir. Anton Nijholt, University of Twente, NL (Promotor)
Dr. Betsy van Dijk, University of Twente, NL (Assistant-promotor)

ISBN: 978-90-752-9600-6

Acknowledgements

I have finally managed to come to the most read part of a PhD thesis: the acknowledgments. The list is very long since I was very fortunate to meet lots of wonderful people during the time I was doing my PhD I hope I haven't forgotten anybody!

My biggest thank goes to my promoter Anton Nijholt and my daily supervisor Betsy van Dijk. I don't have enough words to express my gratitude to Anton who gave me the chance to do my Ph.D. at the HMI group and who, in all these years, has never failed to encourage me and to believe in me. I stopped counting the times I emailed Anton in the middle of the night with desperately urgent paper deadline issues, receiving an answer only a few minutes later! Betsy was the best daily supervisor I could ever imagine. No matter how busy her schedule, she always found time to give advice or to discuss an idea. I will never forget the day I submitted my thesis and Betsy waited for me until very late in the night just to make sure everything was just fine. Also, Anton's son Nelson and Betsy's son Per came to help me with annotations - a big thanks to both of them as well! True fact, Anton and Betsy never let me down. My work with them was one of the most positive professional and human experiences I've ever had. Without their constant support this thesis wouldn't have been possible!

Further, I would like to express my gratitude to all my committee members who agreed to participate in my defense: Li Haizhou, Alan Dix, Gerrit van der Veer, Vanessa Evers and Dirk Heylen. A special thanks to Alan Dix for his detailed comments and suggestions for improvement to my thesis, in addition to very interesting research discussions.

The HMI group was a very pleasant and stimulating work environment during my PhD research. I thoroughly enjoyed working with Dennis Hofs on the IMIX experiment and with Yujia on the Crisis Manager. Yujia was one of my best friends in Enschede. Thank you so much for being there for me, Yujia! I greatly enjoyed our trips together to San Diego and China, our gatherings together with David and our Chinese cooking sessions. It was a lovely time that I will never forget!

Frans helped me out with nasty statistical calculations. Frans and Hayro, best buddies, we had such a good time in Dublin, during HMI outings, dinners in the city center and parties. I am going to miss you guys a lot! Together with Christian, Randy, Alejandro and Maral I had the most entertaining lunches ever.

Olga showed me some awesome CorelDraw 'secrets' and Thijs Verschoor always had good advice whenever I was stuck on some Java or Perl code. Iwan gave me a lot of help with annotators' agreement calculations, Alessandro Valitutti offered me many helpful literature hints while Andrea Minuto was very helpful in forcing LaTeX to finally do what it was supposed to do. In addition, I had interesting discussions about my work with Egon van den Broek, Dennis Reidsma, Christian Mühl, Femke and Laurens van der Werff. Another big thank you to Dennis Reidsma from whom I received the LaTeX thesis templates. Dolf, Sergio Duarte, Luis Escalona and Claudia encouraged me and gave me a lot of support. Claudia, I really enjoyed our talks and I am grateful for your constant moral support even after you left HMI! Further, a special thanks to Mark and Boris for many tips regarding the thesis printing and other administrative issues.

Among the many social events and 'cultural experiences' organized at HMI, I greatly enjoyed the Halloween party organized by Danny Plass Oude Boss, the dance workshop organized by Randy Klaassen and our HMI Bollywood dance production - a really memorable experience! Also, the tango evenings with Alessandro Valitutti were very enjoyable.

I would also like to thank to my office mates Dhaval and Bram for putting up with a talkative person like me. I enjoyed sharing the office with you guys! Dhaval, also many thanks for the nice dinners you organized at your place.

Many thanks to my housemates Bart, Alexandre, Christian, Alejandro and Ann for making my life more enjoyable and more fun. Bart is a fantastic cook and we had unforgettable dinners. Alejandro is an incredible housemate with whom I really enjoy sharing a house. Alexandre, Christian and Lisa and I had a wonderful time having a BBQ in the garden, playing guitar, having dinner with friends or watching movies. Christian, my first paranymph is a great friend from whom I received so much help. I will dearly miss those lovely German breakfasts and the philosophical talks we used to have in the morning before going to the office!

A special thanks to Hendri Hondrop who helped countless times with all sorts of technical problems. Hendri spent hours helping me to convert video files, to create complicated LaTex tables, to format this thesis and even to write Japanese or Romanian characters in LaTeX! Dear Hendri will never forget your help!

Many thanks also to Lynn Packwood who stoically proofread my papers and the numerous improved versions of the thesis. Dear Lynn, many thanks for your patience!

I am also very grateful to Nick Hamm, Clare Shelley-Egan and Jon Dennis who agreed to proofread parts of my thesis as well. Many thanks dear Nick, Clare and Jon for putting so much effort and time to help!

Charlotte and Alice have been very helpful on many occasions concerning administrative issues. Thank you! Charlotte, I really enjoyed the dinner at your place and I will never forget your help when I was sick and you drove me to the doctor. Many thanks also to Lilian Spijker for her support in many occasions.

Another very good friend is Riham Abdel Kader, my second paranymph whom I would like to thank for the unforgettable lunches, dinners, Saturday shopping and long talks. Thank you for being there Riham! I hope we will always be in touch.

During my PhD, I had the chance to spend six months at the Institute for Infocomm Research (I^{2}R) in Singapore with the help of Li Haizhou to whom I am deeply indebted. I^{2}R is a wonderful and stimulating work environment and Singapore is one of the greatest places on earth to be. The time I spent in Singapore was the best of my life. A big thanks goes to Swee Lan for being my friend. Swee Lan, I enjoyed so much our work discussion, lunches and dinners with Jess. I will never forget the Singapore tour you organized on my birthday! Further, I enjoyed working with Dilip, Adrian, Alvin and Yeow Kee. Special thanks to Yeow Kee who spent a lot of time helping me with my last experiments. Also, I am very grateful to George M. White for his constant support over the years. The talks we had were very inspiring for my research. Thank you George for being such a wonderful friend and work colleague!

A special thank you goes to my friends who contributed to making my stay in Singapore an unforgettable experience: Mahani, Julia, Megumi, Miguel, Ryna and George, Sejal, Rachel, Joe, Paul, Suryani, Karthik, Lakshmi. Jon and Weisi, thank you so much for your friendship: I enjoyed so much the trips to MacRichie Reservoir, the dinners in the hawker center and Weisi's birthday party. Staying at your place during the ICSR conference was really awesome. Many thanks to Sophia who also hosted me for an entire week and organized an amazing party for me at her place. It was indeed a wonderful week and I am so grateful for that. Another special thank goes to my dear friend Lux who encouraged me to study computer science and to pursue a PhD He taught me my first line of C programming many years back. Thank you for being there for me, dear Lux!

During my stay in Singapore I worked in Annalakshmi - a vegetarian restaurant run by volunteers for charity purposes. Apart from serving the most delicious Indian food, Annalaksmi is a unique restaurant in the sense that the menu has no fixed price: it is an *"Eat what you Want and Give as you Feel"* concept. Working in this restaurant was a very meaningful activity for me and I am grateful to Suresh Krishnan who accepted me as a volunteer. I greatly enjoyed working with Vasudha who became a good friend. Further, I would like to thank Ganesh, Jothi, Radha, Ravi, Subash, Nazir and the entire Annalakshmi crew for an unforgettable time.

Along the way I met a lot of very nice people whom I would also like to thank: I had a great time during parties, dinners and birthday celebrations with my Romanian friends Raluca, Mihai, Eugen, Stefan, Ileana and Luminita.

Anindita, Supriyo and little Samhita, thank you so much for the lovely dinners at your place and for the parties we had together.

A very big thank you to my friend Michel Rosin for the incredible moral support he offered me in difficult moments. I will never forget that dear Michel!

Eric and Sandra are two of my best friends here in Enschede. I enjoyed so much going out with you guys, dancing salsa, having dinner at the Greek restaurant or at your place together with Clare, Des and Kodo. I will never forget that lovely present you guys sent me all the way down from Japan on my birthday: a book with drawings of a little girl traveling around the world.

Lorena, Marta, Sergio and Martin van Essen, Julian, Lisa and David are other good friends with whom I shared wonderful moments. Some of the best parties in Enschede were held at Julian's place. Sergio and Blas helped me when I was moving into my current house. Thanks a lot for your help!

Maurizio is one of the best photographers I know. Many thanks dear Maurizio for your good heart and your beautiful pictures. You bring so much joy to your friends!

Flavia and Arun are other good friends whom I would like to thank. Flavinha, your inborn enthusiasm is contagious! I enjoyed so much your birthday party, going out for dinner and to Molly's with you, Arun, Giovanne, Aliz, Turi and all the other friends.

Alejandra, Hanik, Nick and Arturo: the dinners and parties at your place were awesome! Thank you Arturo for cooking dinner for me and buying my favorite kip sate krocket from *de muur*! You spent hours building a statistical model for my work and showed me the art school from Losser. It was a lovely time that I will always remember.

A special thanks goes to Anje and Bjoern from the chair massage who helped me so many times to relax during the stresful time of the thesis writing.

Another good friend of mine is Ria, to whom I am very grateful for her constant support and care. Thank you Ria for being such a great friend!

Kardelen I owe the nicest part of this thesis: the cover. Dearest Kardo, I really don't know what I would have done without you! Many, many thanks!

Another special thanks to Oana Buhan Ionescu and Simona Grecu for being my longstanding friends and helping me out on so many occasions. Oana, you are one of the most efficient and helpful people I've ever met! Simona I don't know what I would have done without your help for translations!

I would like to address a special thanks to Rogelio Murillo Vallejo for his moral support and care during the past years. Many thanks also for the most interesting talks. Knowing you, dear Rogelio was in many ways personally very enriching.

I am also deeply indebted to Irina and Warren Treadgold for their amazing support during my undergraduate studies in Bochum. I will never forget your help, dearest Irina and Warren!

My family, in particular my aunt Iarina, my cousins Tudor, Oana, Peti, my sister Smaranda, little Irinuca and Paul, Andrei, Ioana and Uca deserve a special thanks for cheering me up and reminding me that there are more important things in life than getting a PhD.

Last but not least, I would like to thank my parents for the incredible support and love I received over the years. Without their constant care, I wouldn't have made it. This thesis is dedicated to them.

Andreea Ioana Niculescu
Enschede, November 2011

Pentru mama

Contents

Chapter 1

Introduction

The dream of creating humanlike machines - driven either by simple curiosity or by the need to achieve a certain functional purpose - has fascinated human minds since early times. One of the earliest descriptions of humanoid automata was found in a text written in the 3rd century BC by Lie Zi. The account relates the encounter between the King Mu of Zhou (1023-957 BC) and the mechanical engineer Yan Shi. The engineer presented the king with one of his latest inventions: a human shaped figure made of leather, wood and artificial organs that could move around and sing [1], [2]. Also, in 1495 Leonardo da Vinci built what has been considered the first humanlike robot. The robot, representing a warrior, had the ability to stand, sit, walk, open and close its mouth and raise its arms [1].

Attempts to develop machines that were able to mimic human speech appear to have started in the second half of the 18th century. Such machines could produce humanlike sounds by using resonance tubes connected to organ pipes [3] or by deploying manually controlled resonators made of leather [4]. The first machine able to recognize isolated digits was developed In 1952 at the Bell laboratories [5]. Since then speech recognition technology has progressed rapidly from a simple machine that responds to a reduced set of words to sophisticated systems, such as conversational interfaces able to communicate fluently in spoken natural language.

But is it wise to build machines that look and talk like humans? Researchers are still arguing whether following the human model is appropriate when building and interacting with machines. Mashiro Mori formulated the theory of uncanny valley in which he refers to the point when the human likeness of a robot can trigger repulsion effects in people who perceive the robot as very similar but not exactly like themselves [6]. As for spoken interactions, it has been argued that human dialogues often contain frequent interruptions, overlapping, unclear, incomplete or incoherent statements, repetitions, self-corrections and thus are offering poor modeling material to follow [7].

On the other side, Reeves and Nass [8] have demonstrated across a wide variety of experiments that an increase in behavioral similarly between people and computers produces an increase in the human emotional response towards the machine, as people are primarily social beings even when interacting with inanimate entities. In their experiments the authors have shown that people were polite to computers, treated machines with female voices differently than those with male voices and showed preference for computers displaying a personality matching their own. While the current technology is still far from being able to produce artificial entities with highly similar human traits - thus, uncanny valley remaining a remote threat - we believe that human-human interaction, despite its imperfections can provide valuable insights for modeling and evaluating human-machine dialogues [9], [10].

1.1 What are conversational interfaces?

Conversational interfaces software programs enabling users to interact with computer devices using voice input and spoken dialogues. The term was most likely coined by Edwin Hutchins [11] who described conversational interfaces as a metaphor of human-human conversation functioning as an intermediary between users and machines.

Conversational interfaces use speech or natural language as their main communication modality. However, some of these interfaces may use additional input/output modalities, such as type, pen, touch, manual gestures and so on, to enhance system robustness and to lower users' cognitive load. In some cases speech can be a poor modality choice: when the output contains graphical information, such as maps, images, or large tables it becomes difficult to convert it into verbal explanations. Similar to human communication which is inherently multimodal, conversational interfaces can also be complemented by visual and sensory motor channels, allowing users to gesture, point, write and type on the input side and presenting graphics or facial expression and gestures (e.g. more typical for anthropomorphic agents or social robots) on the output side.

Such interfaces can be very useful in situations where users cannot use other input modalities (e.g. while driving, accessing the interfaces over the phone, using pocket size devices or when impaired) or do not know how to interact with the interface (e.g. new type of interface). Users neither need to learn nor to adapt to the designer's interaction style, since speech is learned since childhood.

Among experts in the field there is no consensus on which criteria are sufficient for a voice (also called speech-based) user interface to be considered conversational. In our view, a voice user interfaces can be considered conversational as long as the interaction between user and interface involves verbal sequence pairs implemented as question-answer, request-acceptance, suggestion-rejection, and so on. As such, conversational interfaces can vary from interfaces with rudimentary dialogue structures, where the computer has the complete interaction control requiring the user to answer a set of prescribed questions (e.g. interactive voice responses) to interfaces with more complex dialogue structures allowing a mixed dialogue initiative (e.g. interactive information systems [10]).

Thus, conversational interfaces is a global term which can refer to an interactive voice user interface, a spoken dialogue system, a multimodal question answering system or a social robot using speech to communicate.

1.1.1 System architecture

Figure 1.1 presents the major components of a typical conversational interface. The input in the form of speech, text, pen or hand is recognized and passed to an understanding component. The understanding component produces a meaning representation for the input. If the input is performed in parallel, partial meaning representations are generated and fused in the multimodal integration unit. If the information gathered from the meaning representation is ambiguous, the system may ask for clarification. Discourse information is maintained during the process in order to understand an utterance in context.

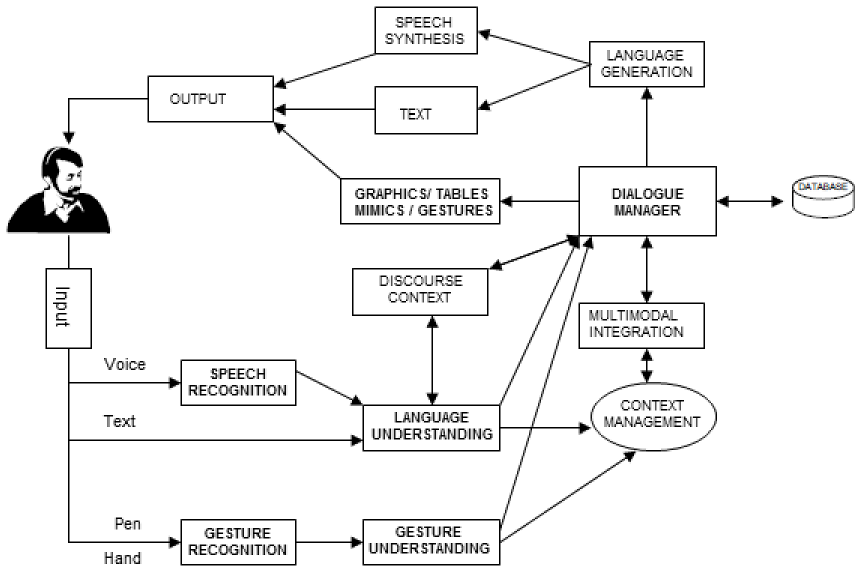

Figure 1.1: Typical architecture of a conversational interface (adapted from [10] and [12])

The meaning representation can be used by the dialogue manager to retrieve appropriate information in the form text, graphics, tables or in speech accompanied by mimics or gestures. Natural language generation and speech synthesis are used for the speech.

1.1.2 Conversational interfaces used in this work

Nowadays spoken conversational interfaces have multiple application domains such as interactive information systems, smart environments, automatic training and education, in-car applications, social robots, etc. In this work we used three types of conversational interfaces designed for different purposes. The interfaces had different degrees of anthropomorphisation and two of them had alternative input/output modalities.

- Our first interface was a prototype of a voice user interface application for mobile phone users. The application was meant to be used as a voice enabled user manual to help users become familiar with the phone functionalities. The dialogues were designed based on real user queries posted on the web and refined later through scenario-based human dialogues. The interaction modality with the interface was only through speech and the users had to perform the tasks using the phone's touch screen. Apart from speech the interface had no other anthropomorphic features (more details are presented in section 4.2).

- The second interface was a multimodal question answering system for medical queries. The interface consisted of a graphical and a voice user interface. Users could use speech, text or pen input to communicate with the system. The answers consisting of text and images were displayed on the screen and spoken by an anthropomorphic talking head. Further, users were allowed to refer to the pictures in the answer presentation when asking follow-up questions using verbal questions or encircling parts of pictures or words. The follow-up questions were designed with the help of human users while the entire dialogue structure was evaluated using human-human conversational protocols (more details are presented in section 5.2).
- The third conversational interface had the highest degree of anthropomorphisation and was a social robot acting as a receptionist. The robot used speech and gestures to communicate with users. Attached to the robot was a touch screen where additional information cues were displayed. Users could communicate with the robot using speech or the touch screen. The dialogue with the robot was designed after collecting and analyzing human natural dialogues between visitors and receptionists in scenario-based interactions (more details are presented in section 6.2).

1.2 Research focus

This PhD thesis focuses on the design and evaluation of conversational interfaces for task-oriented dialogues using speech as main interaction modality. Since spoken natural language is essential in interaction with this type of interfaces we chose to address two salient design aspects of this modality: voice characteristics and language features used by the system to communicate with the user. The reasons for choosing these two aspects are twofold: firstly, because of their proven impact on human social relationships [13], [14], [15], [16], [17] we expect similar effects to occur in human-computer relationship, conform CASA paradigm [18], i.e. computers are social actors. Secondly, both voice characteristics and language features are among all design variables the easiest ones to manipulate. As such, their manipulation can be beneficial for improving the users perception of the evaluated conversational interface at a very 'low' cost. Thus, the contributions of this thesis relate to the following research questions:

- *RQ 1:* What impact do voice characteristics, such as voice pitch, voice accent and voice consistency with physical look have on the evaluation of a conversational interface?
- *RQ 2:* What impact do social skills, empathy and humor (implemented as language features) have on the evaluation of a conversational interface?
- *RQ 3:* Which communicative interaction patterns are relevant for task-oriented human-human interaction with potential applicability in human-machine interaction?
- *RQ 4:* How can we use human communicative interaction patterns to test and enhance conversational interfaces?

1.3 Contributions of this thesis

This thesis makes four main contributions that can be relevant to the HCI community: theoretical, methodological, empirical and design related. These are in short:

Contribution 1 (theoretical) - a newly compiled set of guidelines for dialogue and interaction design for spoken conversational interfaces from the reviewed literature (chapter 2) and a taxonomy of conversational interaction quality focusing on hedonic and pragmatic quality aspects (chapter 3)

Contribution 2 (methodological) - a novel approach to evaluating the adequacy of conversational structures implemented in conversational interfaces using a new concept that we call 'verbal affordance' (chapter 5)

Contribution 3 (empirical) - results of experiments concerning the effects of voice characteristics (chapters 4, 5, 8) and language features (chapters 6, 8) on the evaluation of conversational interfaces, in particular on the overall interaction quality

Contribution 4 (design-related) - there are two design related contributions:
• design of a task-oriented human-robot conversational interface based on human dialogue interactions
• design of a novel application consisting of a voice enabled user manual for mobile phones

1.4 Thesis outline

This dissertation is divided into two parts: part I presents the theoretical background while part II is concerned with experimental studies. The outline of the dissertation is structured as follows:

- Chapter 2 presents various theoretical and practical modeling approaches of human-human communication with applicability in the design of conversational interfaces. The theories refer to the core structure of dialogues and help in understanding how verbal and non-verbal exchange occurs in natural circumstances.

- Chapter 3 is concerned with the evaluation of conversational interfaces from the perspective of interaction quality. Since the notion of quality is central in this work the chapter shows an overview of several definition approaches. Further, the chapter deals with evaluation methods and taxonomies of quality aspects. Elements presented in the taxonomies were later included in the evaluation questionnaires.

- Chapter 4 deals with two of our research questions: namely, how to design and improve a voice enabled user manual using written instructions and verbal human-human dialogues (RQ 4) and how the voice accent influences its evaluation (RQ 1).

- Chapter 5 focuses on how human verbal interaction patterns can be used to evaluate the adequacy of dialogue structures implemented in a multimodal question answering system (RQ 4) and how voice consistency with physical look influence the evaluation of the system (RQ 1).

- Chapter 6 addresses the evaluation of a social robot in an open uncontrolled environment. We used languages features (in combination with gestures and body movements) to design the social skills of a robot receptionist. The study was aiming to explore relationships between the robot's social skills (RQ 2) and the way users reacted and evaluated the robot.

- Chapter 7 focuses entirely on RQ 3, analyzing relevant verbal and non-verbal interaction patterns in task-oriented human-human interaction. The patterns were grouped in a set of recommendations which were further used to design the interaction with a social robot receptionist in chapter 8.

- The thesis ends with conclusions in chapter 9 where the thesis contributions, research questions and 'take away messages' from all our experiments are presented and discussed. This chapter also includes a discussion on future work directions.

Part I

Theoretical Background

Chapter 2

Discourse analysis and design approaches for conversational interfaces

Since natural language remains, despite several other means of communications, the most convenient form of interaction between humans there is a strong need for conversational interfaces to adequately adapt to this communication modality. In order to do so it is necessary to develop dialogue strategies able to overcome the vagueness and ambiguity of natural language, allowing a clear and intuitive way of interacting. In human face-to-face interactions communicative problems are often solved through context interpretation, repair strategies or through processing additional knowledge sources, such as mimics, gestures, body postures or gaze directions. Thus, studying the details of human-human dialogues and their modeling approaches can increase the chances of designing more appropriate human-computer spoken interfaces. With this purpose in mind in this chapter we will discuss theoretical and practical modeling approaches of human-human communication with applicability in the design of conversational interfaces. In section 2.1 we will present a general introduction to the process of communication. Section 2.2. will review approaches to spoken discourse including conversational analysis, pragmatics, speech act theory and dynamic interpretation theory. Section 2.3 will provide a short contrastive analysis between human-human and human-machine dialogues while section 2.4. will present an overview of the most important guidelines for dialogue and interaction design from the literature. The chapter will end with a summary in section 2.5.

2.1 About communication

The process of communication can be described as a two-way activity between two or more participants with the goal of transmitting information. The word *'communication'* is derived from the Latin *'communis'* which means *'common'*. Thus, communication refers to the process of establishing something *'in common'* with others.

During the communicative process participants share a common channel for the exchange of signals, a common language and a common discussion topic.

A general model of the communication process was described by Shannon and Weaver [19] (see figure 2.1).

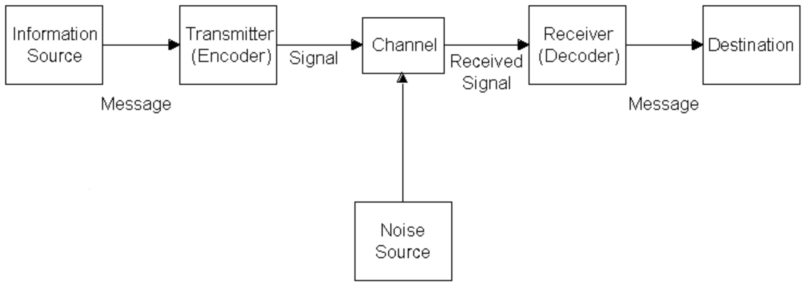

Figure 2.1: The Shannon-Waever model (1949)

The model explains how the flow of information begins when the message is encoded and sent by a sender. The message in the form of acoustic signals (words or non-verbal sounds) or visual signals (gestures, body movements, written words, images, etc.) is sent through a channel. Once arrived the message is decoded by a receiver, that is to say the receiver interprets the message in terms of meaning. A noise source can disturb the signal transmission which can reach its destination damaged, that is the message cannot be interpreted in the right context.

In conversational interactions participants take turns to talk with each other. Thus, by alternating sender and receiver roles the communicative process between participants becomes reciprocal [20] (see figure 2.2).

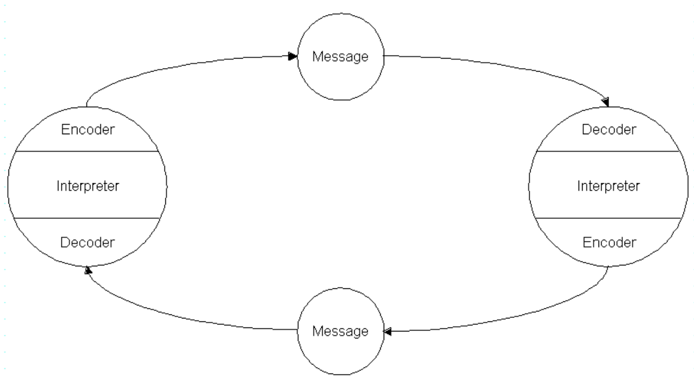

Figure 2.2: The Schramm's model of communication

Conversational interactions, also called dialogues, are a form of interpersonal communication in which specific thematic or situational, intentional controlled utterances are directed towards a partner. The interaction is influenced by the level of information, emotional charge and participants' interests. Most dialogues have a relatively short form and a simple syntactical structure and can be carried out between two or more participants. Their essential feature is the fact that each contribution is dynamically determined by the previous one. Cappella and Pelachaud [21] called this feature *responsiveness* and defined it mathematically as the contingent probability between two sets of behaviors: considering the conversation between two persons A and B, A has a behavioral repertoire set of $X=(X_1, X_2, ..., X_N)$ while B has a similar one defined as $Y=(Y_1, Y_2, ... , Y_K)$; the values of X and Y are the N and K discrete behaviors enacted at discrete intervals of time. Thus, the responsiveness can be modeled as:

eq. (1): $P[X_i(t+1) \mid Y_j(t)] > 0$
eq. (2): $P[X_i(t+1) \mid Y_j(t)] > P[X_i(t+1)]$

for at least some combination of the behaviors *i* and *j*. In words equation (1) states that B's behavior must influence the probability of A's behavior at some significant level while equation (2) specifies the fact that the size of the probability must be greater than the probability that A will emit the behavior in the absence of B's prior behavior.

Dialogues can be task-oriented or non-task-oriented. In non-task-oriented dialogues no task is provided, thus no boundaries are defined to mark the beginning or the termination of a dialogue with respect to a common goal. In contrast, task-oriented dialogues have well defined goals and interlocutors work together to achieve a task as quickly and efficiently as possible. Since our work deals with task-oriented dialogues our theoretical discussions will consider only this particular dialogue type.

How can we define a task-oriented dialogue? Assuming the dialogue has a single

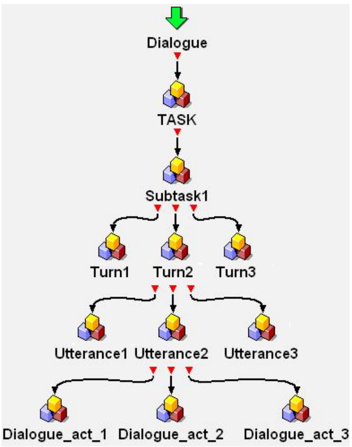

Figure 2.3: Hierchical dialogue representation in a tree

task goal we can describe it as a chain of transactions where each transaction fulfills a particular subtask (see figure 2.3). Further, each subtask can be divided into smaller segments consisting of one or more complementary units called turns. During a turn the dialogue control is temporarily assigned to one of the participants: it always starts when one participant begins to talk and ends when another participant takes the dialog control over. Each turn consists of one or several utterances and each utterance can have one or more communicative functions. Each of these functions is represented by a dialogue (or speech) act.

2.2 Discourse analysis

Natural language dialogues involve the exchange of multiple utterances between the participants. An attempt to model natural language dialogues would be mainly concerned with the coherence which 'glues' utterances together trying to explain how a new utterance can be understood given a certain context or how the context can be used to predict what will come next. Thus, such models require careful analysis of natural language dialogues.

Several theories subsumed under the general term of discourse analysis attempt to study natural language dialogues from different perspectives. In the context of our work we understand discourse as to be a spoken dialogical exchange unit (or turn) between two conversational partners in a social situated interaction. This understanding is equivalent with the most common definitions to discourse as language above the sentence (that is to say a unit which is larger than a sentence) and as language in use (language produced and interpreted in a real-world context) [22]. According to D. Schiffrin there are six major approaches concerning the study of discourse: conversational analysis, pragmatics, speech act theory, interactional sociolinguistics, ethnography of communication, and variation theory [23]. In the following we will focus only on the first three approaches which are relevant for this work.

2.2.1 Conversational Analysis

Conversational analysis (CA) is an ethnomethodological approach to spoken discourse which aims to understand from fine grained analysis how people manage ordinary spoken interactions in everyday situations. The approach originated from sociolinguistics and was developed by H. Sacks together with E. Schegloff and G. Jefferson in the late 1960s. CA is in particular interested on the sequence, structure and coherence of the verbal discourse examining several conversational elements such as turn-taking, adjacency pairs, feedback and repair, discourse markers and opening and closing procedures. Also, gestures and gaze can be included in the analysis as they may bring additional contextual information, emphasizing or even changing the meaning of the verbal exchange.

CA methods were used for research in two chapters of this thesis: in chapter 5 to test the adequacy of the conversational protocol implemented in multimodal question-answering system and in chapter 7 to analyze task-oriented conversations between two human test participants.

Turn-taking

One very important aspect in conversation which co-ordinates the changing roles between receiver and sender is the turn-taking. The basic rule in conversations is that one person speaks at a time and speech overlap is kept to a minimum. Turn allocation can be given explicitly by the current speaker or it can be taken by the interlocutor through self-selection during a 'transition' relevant place [24]. Such places are signalized through the completion of a syntactic unit or through the use of falling intonation followed by pausing. Additionally, the end of a turn can be signalized through gaze (eye contact) and body position movements. An 'aggressive' strategy for turn talking allocation by self-selection is to use repeated speech overlap and thus, to force the interlocutor to stop his discourse.

Adjacency pairs

In some cases turn-taking statements belong together, that is to say the first statement requires the second one. Schegeloff and Sacks [25] defined the concept of *adjacency pair* to refer to these statements. An example of an adjacency pair is a question followed by an answer, a request succeeded by a promise, and so on. Typically for adjacency pairs is that the speaker always allows the interlocutor to take over the turn. The notion of adjacency pairs has played an important role for practical approaches concerning the design of dialogue systems since the analysis of dialogues based on structural relationships facilitates interaction modeling [26].

Feedback and repair

One crucial condition for conversational interaction is that people understand each other. This assumes a common ground, that is to say both speakers maintain a mutual understanding about the issue under discussion during the conversation in order to collaborate, co-ordinate joint activities or share experience. For this purpose conversational partners try to establish mutual knowledge, beliefs and assumptions that are oriented towards a common goal [27]. One mechanism which enables interlocutors to control common ground is to provide feedback: the receiver sends regular messages about the state of the information processed (*auto-feedback*) while the sender may also check whether the information was correctly received (*auto-feedback elicitation*). During the conversational exchange communication problems can arise: when the decoding of a message goes wrong the receiver will signalize it by sending negative feedback; problems related with the production of an utterance are marked by hesitations (stalling segments, such as "eh", "mhm"). Clark [27] identified three different strategies to deal with communication problems before or after they occurred: *preventative*, which should avoid communication problems before they happen, *warning* signals, which will warn about an unavoidable communicative problem and *repair*, when miscommunication has occurred. These strategies correspond to the grounding acts categories defined by Dillenbourg et al. [28]: *monitoring*, *diagnosis* and *repair*.

Further, H. Clark and D. Wilkes-Gibbs explained the occurrence of presentation flaws, errors and consequently repair feedback statements through a principle called the least collaborative effort: *"In conversation, the participants try to minimize their collaborative effort - the work that both do from the initiation of each contribution to its mutual acceptance"* [29]. Since it takes more collaborative time to come up with well-structured utterances speakers tend to prefer improper formulations. In this way they shorten their effort enlisting their interlocutors' help, asking for confirmation or waiting for clarification questions. For example, speakers can present a difficult utterance in sequences and check for understanding after each sequence. Alternatively, they can ask interlocutors to complete an utterance they are having trouble with.

Discourse markers

Discourse markers are particles such as "oh", "well", "you know", used to increase the discourse coherence. However, they do not influence the utterance meaning expressing rather different types of relationship, such as between different utterance parts, between the speaker and the message or between the speaker and the hearer. Discourse markers are more common in informal conversations being syntactically independent, that is to say, if they are removed the sentence structure remains intact.

Opening and closing

Opening and closing procedures are integral parts of conversations showing the availability and willingness to start and respectively end a verbal exchange. Opening and closing occur mostly in adjacent pairs. Conversational openings include greetings or (self-) introduction statements while closing procedures contain valediction acts and pre-closing statements preventing the interlocutor for bringing a new topic into the discussion or proving a reason for the upcoming conversational end. In general, the tendency in conversation is to gradually close down the conversation because a simple ending of a conversation could be perceived as rude and even offensive [30].

Opening and closing are universal conversation features, however their practical realization may depend on the cultural context in which they are performed: for example in Arabic speaking countries initial greetings are followed by additional sequences of specific traditional greetings with predefined statements, in Latin cultures women's greeting and goodbyes are accompanied by a kiss on the right check while in Japan opening and closing procedures are often accompanied by bows.

Gestures and gaze

Speech is the primary means of conveying information, however human communication can also take place through gestures, such as hand and head movements, facial expressions and body postures. Additionally, gaze directions regulate the conversational flux giving feedback on the participants' attention focus. The multimodal nature of human interaction poses a challenge for the dialogue research since it involves information that is not easily described using a formal model [31].

Human gestures have been studied by a number of researchers with different purposes. A. Kendon [32] called hand gestures (as meaningful hand movements) *"visible actions as utterance"* and considered speech and gestures not only connected, but merely two surfaces of a single underlying utterance. In his view gestures do not originate from speech, but rather have the same origin as the speech. Kendon categorized gestures as used alone or co-produced with speech. The contribution gestures can add to an utterance can be in the form of *content* (mainly through emphasis), *deixis* (referring to objects) or in *conjuncts* with the speech, that is, with no lexical meaning.

D. McNeill extended on Kendon's [32] and D. Efron's [33] work and proposed a classification of four types of hand gestures [34]: *deictic* - gestures used to point to a person, object or a certain direction; *iconic* - gestures used to illustrate physical items; *emblematic* - gestures with a specific standard meaning, e.g. waving the hand to mark valediction; *beats* - rhythmic gestures with no particular meaning but performed to emphasize particular words or speech parts; *metaphoric* - gestures used to explain a concept.

Also head gestures, facial expressions or body postures exhibited alone or in combination with speech can be used for many purposes: for example, head nodding or shaking can be used for visual grounding, turn-taking or answering yes/no questions [35]; frowning can be used to express negative feedback [36] while eyebrow raising is often used to emphasis meaning [37] or to show surprise; smiles are used as polite markers to open and close interactions, to signalize mutual understanding [38], to cover embarrassment or accompanying excuses as signs of appeasement [39]. Body movements can be used with certain communicative functions, such as referencing, that is, 'pointing' the body in a certain direction, displaying a communicative attitude to indicate the willingness to engage in interaction or focusing on another (physical or abstract) spot by directing the body toward a new point of interest [40].

Research in the past [41], [42], [43], [44] has shown that gaze behavior seems to play a role in indicating addresses, displaying attentiveness, effecting turn transitions and in requests for back-channeling [45]. In conversation involving more than two participants the gaze behavior is a mechanism used to indicate the person to whom the current dialogue sequence is directed.

The direction of gaze gives important cues about the focus of attention during the dialogue. In fact, the gaze is the most basic way of showing positive evidence that the attendant is listening. Listeners are gazing at speakers showing they are listening while speakers are gazing back to check whether listeners are indeed paying attention [41]. This mutual gaze exchange covers about 60% of the conversation [46]. Gaze behavior may reflect the cognitive process of a dialogue participant: looking away is often used to avoid distraction, to concentrate, or to indicate one does not want to be interrupted [47]. It can also be used to reflect hesitation, embarrassment or to locate objects or persons in an abstract space (e.g. when pointing directions). Additionally, gaze is used to regulate turn management: for example speakers seek mutual gaze right at the beginning of a turn in order to allocate the interlocutor the next turn [42].

A good literature review on head gestures and gaze in context of face-to-face conversations can be found in [48].

The analysis of body posture, facial expressions, gestures and verbal behavior are also studied in behavioral analysis, a method commonly used in psychology to acquire knowledge about human social interactions with the goal of understanding and predicting behavior. For the sake of simplicity we listed them under the conversational analysis section. We will use gesture, gaze and verbal behavior analysis in chapter 6, 7 and 8.

2.2.2 Pragmatics

Pragmatics is concerned with the way the meaning of an utterance can changed according to the context in which the utterance is performed. A pragmatic approach to discourse is provided by the philosophical work of H. P. Grice. Grice's assumption was that when people communicate they perform an act of collaboration. Based on this assumption Grice formulated the following cooperative principle: *"Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged"* [49]. The cooperative principle consists of four specific maxims:

1. **Quantity**
   Make your contribution as informative as is required
   Do not make your contribution more informative than is required

2. **Quality**
   Do not say what you believe to be false
   Do not say that for which you lack adequate evidence

3. **Relevance**
   Make your contribution relevant

4. **Manner**
   Be clear:

   - Avoid obscurity of expression
   - Avoid ambiguity
   - Be brief (avoid unnecessary prolixity)
   - Be orderly

2.2.3 Speech act theory

The speech act theory originated by J. Austin [50] and extended by J.R. Searle [51] tackles the integration problem between semantics and pragmatics. The theory attempts to explain how speakers use language to accomplish certain actions and how hearers infer meaning from the context in which something is being said. According to speech act theory, utterances performed in a dialogue do not have a certain constant meaning attached, being rather affected by the situational context and by the speaker's and listener's intentions. Thus, an utterance, for example *"It is too cold in here"* can be analyzed from three different meaning perspectives:

1. *propositional/locutionary* – referring to the literal meaning of the utterance.

2. *illocutionary* – referring to the intended meaning of the utterance; this could be: an indirect request for someone to turn on the heating, an indirect refusal to open the window because someone is warm or a complaint expressed emphatically.

3. *prelocutionary* – referring to the effect of the utterance on others, that is, the utterance could result in someone turning on the heating [52].

Searle refined later the concept of illocutionary act by splitting it in two parts: indirect illocutionary speech act (which is not literally performed in the utterance but it is inferred from the context) and direct illocutionary speech act (which is literally performed in the utterance). The following example illustrates the concept of direct and indirect speech act:

> Speaker X: *"We should leave for the show or else we will be late."*
> Speaker Y: *"I am not ready yet."*

The indirect speech act performed in this dialogue sequence is the Y's rejection of X's suggestion to leave while the direct speech act is Y's statement that she is not ready yet [51].

A speech act can be considered to be the smallest functional unit in human communication. A. Cohen [53] extending Searles' work [54] classified speech acts in five categories based on the functions assigned to them. These were: *representatives* (assertions, claims, reports), *directives* (suggestions, requests, commands), *expressives* (apologies, complaints, thanks), *commissives* (promises, threats, offers), *declaratives* (declarations).

Other influential work on the speech act theory was done by B.J. Grosz and C.L. Sidner [55], M.E. Bratman [56], D.R. Traum and E.A. Hinkelman [57], D.G. Novick [58] and D.J. Litman and J.F. Allen [59]. Although speech act theory was not first developed as a means of analyzing spoken discourse, the fact that utterances are seen as context dependent relates the theory to discourse analysis [23].

2.2.4 Dynamic Interpretation Theory

The Dynamic Interpretation Theory (DIT) developed by H. Bunt [60] is a further development of the speech act theory. The theory emerged from the study of human-human task-oriented dialogues aiming to determine fundamental principles to design human-computer dialogue systems. From the perspective of DIT the dialogue can be seen as a sequence of dialogue acts which are defined as semantic units of communicative behavior produced by a sender and directed to an addressee [61]. The theory explains how the communicative behavior is changing the dialogue context and describes five context categories:

- *Linguistic context*: referring to previous and future planned contribution in terms of linguistic material
- *Semantic context*: referring to the current state of the underlying tasks and the properties of the task domain
- *Cognitive context*: referring to the participants' state of perceiving, interpreting, and evaluating their beliefs about the dialogue partner processing state.
- *Physical and perceptual context*: referring to the physical environment in the case of communication at distance
- *Social context*: referring to the communication rights, obligations and constraints of each partner

Further, these types of contexts are divided into two categories: *local* and *global*. The local context is the information that can be changed through the dialogue while the global context remains unchanged during the entire dialogue.

An important difference between the speech act theory and DIT is the fact that utterances are considered to be multifunctional [62], meaning that they can perform several dialogue acts at once. In contrast, the speech act theory assumes the utterance encodes a single speech act [36].

Dialogue acts have a *semantic content* and a *communicative function*. While the semantic content specifies the elements, objects, events, situations, relationships that the dialogue act is about the communicative function specifies how the semantic content updates the interlocutor's context.

DIT, similarly to the model used by Traum [63], distinguishes between *task-oriented acts*, that is to say acts which are directly motivated by a task and contributes to its achievement, and *dialogue control acts*, that is to say acts which are concerned with the interaction itself. A more detailed description of DIT dialogue acts will be provided in section 7.2.1 where they were used to analyze a task-oriented dialogue corpus consisting in dialogues exchanged between two human participants.

2.3 Human vs. human-machine spoken dialogues

Humans are experts in communicating being equipped with a large set of cognitive capabilities which enable them to deal efficiently with complex verbal and gestural interactions. Some of the most important skills both interlocutors are controlling in a spoken conversation are listed below:

1. Recognition of spontaneous speech utterances, including their intentional meaning, regardless of gender, age, dialect variations, background noise or signal intensity
2. Controlling a wide vocabulary on various topics
3. Ability to understand and interpret complex, prosodic, elliptic or anaphoric constructions within a certain context, such as interruptions ”um”, ”ehm”, word repetitions, error corrections or a certain type of background noise (coughing, sneezing)
4. Ability to establish semantic relationships between the actual content and other related topics
5. Ability to perceive the context dimensions in which the conversation takes place and adjust to stylistic, semantic and topic changes, as well as to the interlocutor’s mental model
6. Ability to easily make corrections and give explanations
7. Ability to continue the dialogue despite spontaneous interruptions
8. Ability to alternate intonation and pronunciation

In contrast, human-machine dialogues show a highly asymmetrical relationship between the interlocutors, since most of the human communicative skills are transferable only to a limited extent. Thus, the machine as conversational partner is confronted with:

1. Reduced recognition capabilities whose performance depend on vocabulary and topic limitations, background noise and pronunciation features, such as indistinct or dialectal phoneme articulation
2. Control of a thematically limited vocabulary
3. Limited abilities to handle elliptic utterances, word repetitions, hesitations or false starts
4. Ability to produce semantic relations only through cross-references
5. Limited abilities to detect context dimensions and to adjust to stylistic, semantic and topic changes

6. Limited abilities to perform meta-communicative strategies such as corrections, explanations and repetitions

7. Limited abilities to react appropriately to spontaneous interruptions

These are only a few typical characteristics in which human-machine dialogues differ from human-human spoken interactions. Thus, it becomes obvious that, even following the same sequential conversational steps machines have to overcome huge hurdles in order to compete with a human interlocutor.

2.4 Guidelines for dialogue and interaction design

Human language processing is a very complex task and building a machine with the full conversational abilities of a human being is not realistic. However, several design guidelines and practical advices were formulated in the past with the goal of modeling human conversational behavior in spoken dialogue systems in such a manner that they would be perceived as having humanlike communicative functions. In this section we present a set of guidelines compiled from the literature ([64], [53], [65], [66]) on dialogue and interaction design for task-oriented conversational interfaces. The list, far from being complete shows a variety of factors that should be taken into account during the design of the prompts. Most of the guidelines are based on the generic (**GG**) and specific (**SG**) guidelines for cooperative communication developed by Bernsen and colleagues [64]. The principles extend the Gricean maxims presented in section 2.2.2 aiming to make them usable for the interface design and evaluation. Some of these guidelines were used to design the dialogues with an interactive voice user manual for mobile phone users (section 4.2.1) and with a social robot receptionist (section 8.3.4).

2.4.1 Dialogue design

1. Take into account users' background knowledge and expectations
Before starting the prompt design of a conversational interface a designer should think of the target user group for whom the dialogue is intended. Users may have different speech behavior according to their background knowledge and expectations of the system [64].

a) Background knowledge (GG11)
The distinction between novice and expert users is important for tailoring the system output to the informational needs of the user group. Usually, more experienced users need less explanation, since they possess the information required to understand the system's functionality.

b) Expectations (SG6)
Differences in expectation towards the system caused by possible inferences by analogy from related task domains may invite users to ask clarifying, out-of domain

questions that the system cannot handle.

Example:
S: *"O.K. I booked a one-way ticket on Friday, at 9.30."*
S: *"Do you have more questions?"*
U: *"Hm... Can I get a discount?"* (this example was taken from [64])

The user wants a ticket discount, but does not know that such an option is unavailable on one-way journeys. Thus, the system should take into account the user's expectation by mentioning that one-way tickets have no discount options.

2. Distribute information load wisely
a) First prompt
The design of the first system prompt has an important role for the entire interaction and should contain **what** kind of information the system basically provides and **how** the user should interact with the system [53].

Example:
S: *"Good morning, welcome to BoRIS, the Bochumer Restaurant Information System. BoRIS permits you to search for a restaurant according to the following criteria: cuisine type, meal price, meal time, restaurant location or restaurant opening hours. Please formulate your inquiry"* (this example was taken from [67])

b) Summarization
If the dialogue flowchart has a complicated structure which requires several user inputs or the user himself has changed his input several times, the system should briefly repeat the commitments made earlier [64].

Example:
S: *"You reserved a train ticket for the 14th of July 2003, from Amsterdam to Munich, departure time 9.33 am. Is this correct?"*

c) Informativeness (GG1)
The system's answers should not contain more information than required for the subtask they are designed for. Normally one question should handle one particular piece of information (e.g. questions about departure time should contain only information related to departure and should not refer to ticket prices). Too many questions at the same time can confuse the user [64].

d) Feedback (SG2)
Immediate feedback provides users with an opportunity to detect misunderstandings quickly. The sooner the misunderstanding can be corrected, the better. There are three possibilities to provide feedback (examples were taken from [64]):

Examples :
•**Echo feedback:**
S: *"Where does the journey end?"*
U: *"In Copenhagen."*
S: *"In Copenhagen. Do you want a return ticket?"*

•**Implicit feedback** (including in the next prompt that recognizes user input):
S: *"At what time?"*
U: *"Afternoon."*
S: *"In the afternoon on Sunday January 29th there is a departure from Aalborg to Copenhagen at 17:00. Do you like this departure time?"*

• **Explicit feedback:**
S: *"How many persons are traveling?"*
U: *"One person."*
S: *"You said 'one person'. Is that correct?"*

The implicit and echo feedback are better choices when compared with explicit feedback which extends and unnecessarily complicates the dialogue [64].

3. Highlight partner asymmetry (GG10)
This guideline refers to differences that exist between the interlocutors. These differences are likely to influence the interaction course. When learning to speak, people implicitly learn what to expect from a 'standard' conversational partner. When interacting with a 'non-standard' interlocutor people adjust their manner of speaking according to the partner's abilities, such as when speaking to children, hearing impaired people or interlocutors who find themselves in noisy environments. The computer is in many respects a 'non-standard' partner thus, it is recommended to highlight this partner asymmetry to avoid miscommunication. This can be achieved by providing a clear indication about the system competence. Research in the past has demonstrated that people who tend to make more well-formed phrases use a reduced vocabulary when they assume they are talking with a machine, [68], [69].

4. Ensure an appropriate expression manner
The aspect of manner concerns the way in which the intended meaning is being expressed.

a) Avoid ambiguity (GG7)
Avoid too open or non-specific formulations. Such formulations, apart from inviting the user to take the initiative and ask out-of-domain questions, may lead users to hesitation, false starts or revisions. L. Karsenty suggested that using explicit requests helps users to structure their responses and to avoid long utterances [70].

b) Avoid obscurity of expressions (CG6)
Obscure formulations lead to doubts and therefore require clarifications for which

the system might not be prepared.

Example:
S1: *"For restaurant locations you have the following options: downtown, university center, other possibilities or no preference."*
U2: *"Other possibilities."*
S2: *"The other possibilities are: Grumme, Werne and Langandreer."*
U3: *"None of those."*
S3: *"Sorry, I cannot understand"* (this example was taken from [67])

The system utterance S2 is a typical example of an obscure formulation: the data base contains information about 12 city districts, from which the users can choose. In order to shorten the prompt nine city districts were omitted which obviously confused the user. As the user tries to make a 'blank' selection (*"none of those"*) the dialogue fails.

c) Be orderly (GG9)
The task-relevant topics during the interaction should be presented in the order expected by the user. If a certain topic is addressed earlier than expected it might cause a request for clarifications. Studying similar tasks in human-human conversation can support the design of an orderly interaction [64].

Example:
S: *"On which date will the return journey start?"*
U: *"Preferably Sunday."*
S: *"At which time do you want a departure?"*
U: *"I would like a departure late in the afternoon. Is there ... any kind of discount possibility?"* (this example was taken from [64])

d) Be short (GG8)
Prompts should be short, if possible. However, since in some particular dialogue contexts the system prompts cannot be short the prompt should contain a dialogue focus at the end 'pointing' to the next dialogue sequence [64].

Example:
U: *"I want information on discounts for children."*
S: *"Accompanied children between two and eleven years of age may obtain discount on return journeys: red discount at 323 kroner or green discount at 400 kroner. Children between 12 and 19 years of age may obtain green discount at 550 kroner[..]. Do you want information on other discount possibilities or do you want to return to the main menu?"* (this example was taken from [64])

e) Announce breaks during the dialogue interaction
If the system needs time to process the information it should inform that it will take a few seconds to give the requested information [53].

f) Ensure uniformity of expression (SG3)
In limited interaction domains uniform formulations (e.g. same statement in similar contexts) tend to remove ambiguities and to encourage the use of a reduced vocabulary, since users are expected to adjust their input after the system's utterances.

2.4.2 Interaction design

1. Provide a mixed dialogue initiative
Speakers have the ability to formulate a question in many different ways, a fact which makes any anticipation difficult. This problem can be solved if the system takes over the entire dialogue initiative - that is to say, the system asks questions and 'moderates' the dialogue. On the other side, an interaction relying only on the system initiative lacks flexibility and user friendliness. A mixed dialogue initiative would permit both users and system to interfere in the dialogue flow creating the conditions for a initiative balanced interaction. Thus, the system could initialize and lead the dialogue, allowing users to ask for clarifications, information repetitions, to interrupt, continue or stop the conversation and skip preplanned dialogue segments [53].

2. Meta-communication strategies (GG13)
Meta-communication strategies are applied in cases of communication failure. Such strategies requesting for clarifications generally tend to be difficult to handle as they may raise unplanned questions. The system needs to handle such situations by presenting alternative input options. If no input comes from the user, it is advisable to repeat the inquiry options not more than twice and to break off the conversation by saying goodbye and advising the user to a human information source. It is desirable that the system does not hang up abruptly [53].

3. Avoid overlapping
Human interlocutors often overlap each other. However, overlapping talk is problematic for speech recognition and if possible it should be prevented. A recommended strategy to avoid simultaneous talk is to turn the system silent immediately if interrupted by the user. Commonly, detection of silence sequences are used to establish the end of a turn; in these cases overlapping occurs only when long pauses are mistaken for turn-endings [71].

4. Give the user the choice of input modality
This guideline is particularly relevant for conversational interfaces with multimodal input options [65]. Users should have the possibility to choose and even switch between input options. Also, some modalities are more appropriate [72] for some tasks than others: for example, pointing is more appropriate to determine a certain location on an interactive touch screen while selecting a type of cuisine for a restaurant can be ideally done by speech. Essential for this guideline is that the modalities complement each other to compensate shortcomings [73].

The following three principles were formulated by Rosset and colleagues [66] as *ergonomic choices* that should be taken into account while designing spoken language interactions.

5. Freedom and flexibility
When designing the interaction with a conversational system imposing constraints on the order in which the information is provided should be avoided as long as the dialogue flows well. The users should be allowed to ask for a particular piece of information from any dialogue stage.

6. Negotiation
Users should have the possibility to accept or refuse system proposals by adding or changing constraints. This feature is particularly important in situations where no database entry satisfies the constraints specified by the user. However, the system should inform the users about this particular system option.

7. Navigation
This feature concerns the identification of changes in tasks. Also, it is meant to provide users with additional help: users should have the possibility to ask about the system functionality and about the type of information available.

2.5 Summary

The main focus of this chapter was to present an overview of modeling and design approaches for conversational interfaces.

Since natural language remains the most convenient form of interaction between humans there is a strong need for conversational interfaces to adapt to this requirement. In order to meet this requirement it is necessary to develop dialogue strategies that can overcome the vagueness and ambiguity of natural language and allow a clear and intuitive way of interacting.

Conversational interactions are based on dialogues which can be task-oriented or non-task oriented. In non-task-oriented dialogues no task is provided, thus no boundaries are defined to mark the beginning or the termination of a dialogue with respect to a common goal. In contrast, task-oriented dialogues have well defined goals and interlocutors work together to achieve a goal as quickly and efficiently as possible. Our work is focused only on task-oriented dialogues.

Several theories subsumed under the general term of discourse analysis attempt to study natural language dialogues from different perspectives. Three out of six major approaches to discourse analysis were presented in this chapter. These approaches concern the structure and the sequence of the discourse (conversational analysis approach), the discourse pragmatics (Gricean maxims) and the integration of semantics and pragmatics (speech act theory and DIT approaches).

Conversational analysis (CA) is a structural approach to spoken discourse which aims to understand from fine grained analysis how people manage ordinary spoken interactions in everyday situations. The approach examines several conversational elements such as turn-taking, adjacency pairs, feedback and repair, discourse markers and opening and closing procedures. Also, gestures and gaze can be included in the analysis as they may bring additional contextual information, emphasizing or even changing the meaning of the verbal exchange.

Pragmatics is concerned with the way the meaning of an utterance can be change according to the context in which the utterance is performed. A pragmatic approach to discourse is provided by the Grice which formulated cooperative principle consisting in four maxims.

The speech act attempts to explain how speakers use language to accomplish certain actions and how hearers infer meaning from the context in which something is being said. According to speech act theory, utterances performed in a dialogue do not have a certain constant meaning attached, being rather affected by the situational context and by the speaker's and listener's intentions. A further development of speech theory is represented by the dynamic interpretation theory (DIT). An important difference between the speech act theory and DIT is the fact that utterances are considered to be multifunctional, meaning that they can perform several dialogue acts at once. In contrast, the speech act theory assumes the utterance encodes a single speech act

Human-machine dialogues show a highly asymmetrical relationship between the conversational partners: compared with the human, the machine has reduced recognition capabilities, a limited vocabulary, reduced abilities to handle elliptic or anaphoric utterances, to adjust to stylistic, semantic and topic changes or to react appropriately to spontaneous interactions.

Several design guidelines and practical advices were formulated with the goal of modeling more humanlike conversational behavior in spoken dialogue systems. Such guidelines concern the design of the dialogue and the interaction with the conversational interface.

The dialogue design guidelines concern taking into account the user's background knowledge and expectations, distributing the information load, highlighting the partner asymmetry and ensuring an appropriate expression manner.

The interaction design guidelines concern the dialogue initiative, meta-communication strategies, speech overlapping, system flexibility, negotiation and navigation features.

Chapter 3

Evaluating interaction quality with conversational interfaces

The technological growth of the past decades brought conversational interfaces to a level of maturity which allows widespread application [74]. Examples include interactive information systems, in-car applications, smart environments, media guides, training and educational systems, social robots, and so on. Thus, the need for adequate evaluation methods for such interfaces has been steadily increasing. One of the central aspects in the evaluation of conversational interfaces is the assessment of the interaction quality. Quality is a multidimensional concept that cannot be easily defined nor measured. In this chapter we will focus on this concept aiming to provide a general understanding of its various dimensions. Section 3.1 will give an overview of several quality definition attempts. Section 3.2 will concentrate on approaches to product and service quality while section 3.3 will specifically targeting the quality of human-machine interaction and will present two taxonomies of quality aspects for multimodal conversational interfaces. Section 3.4 will be concerned with quality assessment methods concerning objective, subjective, expert and predictive evaluations. The chapter will end with a summary in section 3.5.

3.1 What is quality?

We all have experienced at least once a sense of disappointment when something we have bought did not meet our expectations: that could be an electronic device that was immediately broken after a first use, a book that turned out to be a flop, or a help desk department that did not return our call. At the heart of meeting such expectations is the notion of quality [75]. But what is quality exactly? As we will show in this section the answer to this question is more complex than it would seem at first glance. And obviously, if we want to measure or improve the quality of a product or service we need a clear picture of what quality means [76].

3.1.1 A definition attempt

The word *quality* is originates from the Latin *qualitatem* (nom. *qualitas*). It is said to have been translated by Cicero from the Greek *poiotes* [77], meaning *of-what-kind-ness*. This meaning of *kind of* or *property* can still be found in the semantic description area of the word *quality* as shown in the American Heritage Dictionary of the English Language [78]:

a. a specific characteristic or property of an object (*"the qualities of water"*)
b. character or nature as belonging to or distinguishing a thing (*"the quality of water", i.e. waterness*)
c. character with respect to fineness, or degree of excellence of an object (*"good quality water", i.e. not of inferior grade*)
d. the meaning of excellence itself (*"wood grain of quality"*)
e. a personality or character trait: (*"kindness is one of her many good qualities"*)

These definitions suggest the complex nature of the concept *quality* which seems to incorporate ambivalent meanings, such as constant characteristics or variable features.

We can find early philosophical debates on the definition of quality by Aristoteles who defined the term *quality* as a category with concrete and functional value, such as having the color white or being literate [79].

A further epistemological and metaphysical distinction between *primary* and *secondary* qualities was made by G. Galilei, [80], R. Descartes [81] and J. Locke [82]. The distinction was motivated by the fact that modern science has shown how unaided sensory perception gives false or incomplete information about the intrinsic qualities of physical objects [83]. From this perspective, primary qualities, such as shape, quantity or motion are objective properties of things independent on any observer. In contrast, secondary qualities, such as odor, taste, sound or color exist only through human perception, being therefore subjective properties.

W. Shewhart was the first quality expert of the modern era who adopted the distinction made between primary and secondary qualities, renaming them in objective and subjective quality aspects. According to Shewhart [84], most of the product evaluations measure objective aspects of quality, even though the subjective aspects of quality are the ones relevant for commercial interests.

Further definitions of quality were focused on the customer's subjective perspective. For example, J.M. Juran building on Shewart's work, defines quality as *"fitness for use"* [85]. For A.V. Feigenbaum quality is the *"best for certain customer conditions"* [86] while for A. Parasuraman and colleagues it means *"meeting or exceeding customer expectations"* [87]. One of the most progressive definitions of quality is given by the American Society of Quality and relies entirely on the customer's evaluation: *"quality is a subjective term for which each person has his or her own definition"* [88].

D.A. Garvin, investigated the notion of quality intensively from different perspectives and identified five major definition approaches [89]:

1. Transcendent approach: *"Quality is absolute and universally recognizable"*. The concept *quality* is understood in this approach as synonymous with *"innate excellence"* suggesting the concept is a construct with universal applicability.

2. Product-based approach: *"Quality is a precise and measurable variable"* which takes into account all components of a product that describe its degree of excellence. Differences in quality would reflect differences in quality of different product attributes.

3. User-based approach: *"Quality is fitness for intended use"*. In this approach the user perspective is highlighted implying that the customer's opinion should be considered in the design process of a product.

4. Manufacturing-based approach: Quality is *"conformance to (engineering and manufacturing) requirements"*. This approach is congruent with the ISO 9000 standard which defines quality as *"the degree to which a set of inherent characteristics fulfills requirements"* [90]. However, the approach was criticized by several experts, such as W.E. Deming [91] who argued that conformance to requirements does not necessarily fulfill customer's needs. Also, G. Taguchi pointed to the flaws of the approach by developing a quadratic quality loss function which shows that the loss increases exponentially as a parameter deviates from its target value even if it still meets requirements [92], [76]. Others, such as P. Crosby [93] argued that the approach is sustainable as long as the requirements are derived from customer requirements.

5. Value-based approach: Quality is defined in terms of costs and prices: *"A quality product is one that provides performance at an acceptable price or conformance at an acceptable cost"*. This definition extends the manufacturing approach from an economic perspective, that is to say quality is considered as the expense of non-conformance to requirements.

While the first approach (transcendental) can be seen as a philosophical debate around the meaning of quality the other approaches attempt to define quality from an engineering and economic perspective considering objective (product-based and manufactured-based) and subjective (user-based) aspects. However, the debate around the concept of quality still continues between experts and no agreement on a common definition approach has been found yet.

3.2 Product and service quality

Even though experts cannot agree on a common definition, there is a widespread consensus to consider quality a multidimensional construct. A number of researchers in the quality assurance field have developed lists of quality dimensions for product and service quality [76].

3.2.1 Product quality

D. A. Garvin [94] formulated 8 quality dimensions generally applicable to products (see table 3.1). The importance of each of these dimensions is variable depending of the main functional purpose of the product. Thus, the product design should not aim to simultaneously maximize each quality dimension, but merely to find suitable trade-offs by focusing on those with the highest relevance for the product functionality (e.g. the aesthetic of a computer is less relevant for the functionality of a computer when compared with the power of its CPU) [76].

Dimension	Description	Example
1. Performance	A product's primary operating characteristics	Clock speed; RAM Hard drive size
2. Features	Characteristics supplementing basic functioning	Wireless mouse, flat screen monitor; DVR-RW
3. Reliability	Probability of a product malfunctioning within a specific time period	Meantime between failures
4. Conformance	Degree to which a product's design and operating characteristics meet standards	Underwriter laboratories labeled; mouse, monitor keyboard included with CPU
5. Durability	Expected product life	Time to technical obsolescence; rated life of monitor
6. Serviceability	Speed, courtesy, competence and ease of repair	Warranty conditions; availability of customer service and replacement parts
7. Aesthetics	How a product looks, feels, sounds, tastes or smells	Computer housing color scheme; keyboard "touch"
8. Perceived quality	Reputation and others indirect measures of quality	Brand name; advertising

Table 3.1: *Garvin's [94] dimensions of product quality*

Garvin [89] and others [95], [96] suggested the applicability of the dimensions to both products and services. However, since services have different characteristics not all product quality dimensions would fit them - for example, the dimension *durability*, referring to technical obsolescence can hardly be seen as quality feature of a service.

3.2.2 Service quality

Services are defined as *"social act(s) which take place in direct contact between customers and representatives of the service company"* [97]. In general, the quality of service (**QoS**) is more difficult to measure objectively since service represents a rather volatile concept.

To describe quality of service J. Evans and W. Lindsay developed a list with eight dimensions of service quality [98] (see table 3.2 left side). The list represents an improvement over Garvin's product dimensions to services. However, no empirical basis has been provided to support the veracity of the dimensions. Also, A. Parasuraman and colleagues developed a shorter list with 5 quality dimensions [87]. The dimensions, embedded in an evaluation tool called SERVQUAL were tested in four types of service industries (see table 3.2, right side).

Evans and Lindsay [98]	Parasuraman et al. [87]
1. Time	1. Tangibles
2. Timeliness	2. Reliability
3. Completeness	3. Assurance
4. Courtesy	4. Empathy
5. Responsiveness	5. Responsiveness
6. Consistency	
7. Accessibility and Convenience	
8. Accuracy	

Table 3.2: *Service quality dimensions by Evans and Lindsay [98] and Parasuraman et al. [87]*

These lists are good starting points for evaluations, but current research indicates that at least in terms of service quality, the dimensions may differ between the industry types, that is to say not all dimensions have general applicability [76]. Therefore, it is recommendable to develop lists with quality dimensions for those specific service industries to be evaluated by determining what is important to their customers.

3.3 Interaction quality

From a commercial point of view task-oriented conversational interfaces represent a rather dual concept: on one side they are (interactive) products, on the other side they offer (informational) services. In this context the interaction can be seen as synonymous to service. Since product quality can influence service quality the identification of quality relevant dimensions for the interaction with a conversational system should also take product characteristics, that is to say system features, into account.

Summarizing the definitions presented in the previous section we believe that the quality of the interaction with a conversational interface can be mainly addressed from two separate points of view: one concerns objectively measurable criteria related to the system performance, reliability, durability and conformance to standards, i.e. whether the system is able to provide the functions it has been designed for (these criteria are exemplified by product- and manufacturing-based approaches); the other point of view refers to subjective criteria, i.e. secondary qualities which entirely depend on the user's perceptions of the system features and performance (these criteria are reflected by the user-based approach). Both points of view need to be considered when designing and implementing conversational interfaces.

However, while objective criteria (e.g. conformance to requirements) can be helpful in the first steps of the design stage, subjective criteria give the ultimate evaluation vote. Moreover, research in the past has found that objective measures, such as performance (response time, error rate) do not always have as direct a link to better evaluation scores in terms of interaction quality [99], [100]. This, of course, does not mean the system's effectiveness is unimportant for the evaluation, but it emphasizes the fact that user's perceptions of quality are not always influenced by a good performance (measured in objective terms). In the context of subjective criteria quality was defined by U. Jekosch [101] as:

> *"[...] the result of appraisal of the perceived composition of a unit with respect to its desired composition [...]. The perceived constitution contains the totality of the features of an entity. For the perceiving person, it is a characteristic of identity of the entity"* (cited in [74], page 7).

This definition indicates that quality assessment is a perceptual event which results from the comparison between the user's desires and expectations on one side and the perceived characteristics of the system, on the other side. Thus, the quality judgment is highly dependent on the situation in which the perception takes place [102], i.e. user profile, (social) interaction context. Also, the interactivity - an intrinsic characteristic of conversational interfaces - makes the subjective quality judgment process more complex since users actively participate in the interaction, being a part of it.

3.3.1 Taxonomy of quality of service and quality of experience

Taxonomies are good ways to structure empirical results, to provide definitions and to identify relationships between the components offering a common ground on which comparable evaluations can be performed. A taxonomy of the most relevant quality dimensions of service and quality of experience for multi-modal human machine interaction was developed by S. Moeller and colleagues [74] (see figure 3.1). The taxonomy is based on earlier versions developed by Moeller for spoken dialogue systems [67] and contains additional quality aspects, as well as a broader range of input and output modalities. In the following we will briefly describe the components of the taxonomy as presented in [74] .

Quality of Service (QoS)

QoS refers to performance criteria and influencing factors of the system. It contains two factor layers, one considering influencing factors on the interaction behavior, the other one referring to interaction parameters. Most of the **QoS** factors - except for user's perceptual and physical response effort - can be quantified by a person external to the interaction process, such as the system developer or an expert evaluator. The influencing factors on the interaction refer to:

- *User factors*: static (e.g. age, gender, background knowledge, expectation) and dynamic (e.g. goals, emotional state) user characteristics
- *System factors*: agent (e.g. technical, such as speech and/or gesture recognition, multimodal fusion, and non-technical characteristics, such as aesthetic appearance) and functional factors (e.g. task type, number, complexity, etc.)
- *Context factors*: physical environment characteristics (e.g. home, office, mobile, transmission channel) and service factors (e.g. system accessibility and availability)

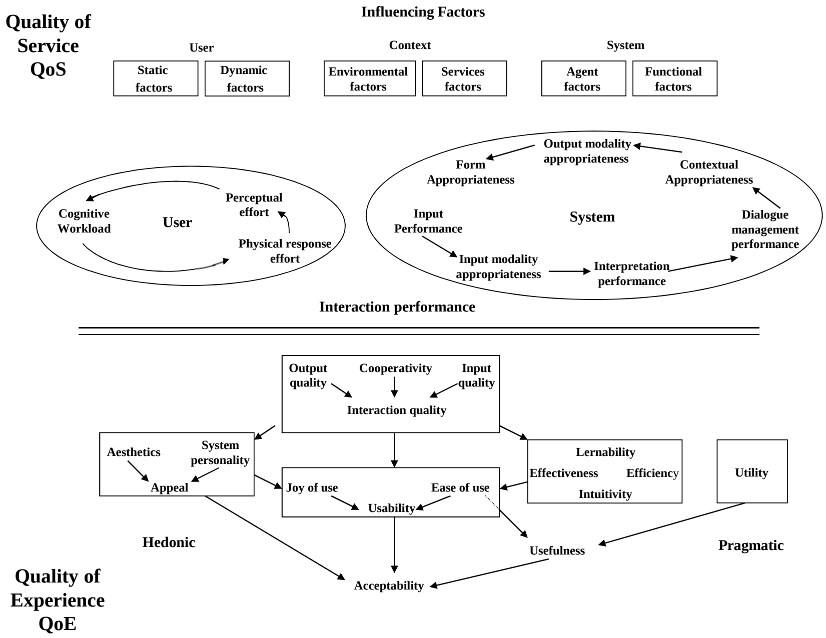

Figure 3.1: Taxonomy of **QoS** and **QoE** aspects developed by Moeller et al.[74]

The interaction performance includes aspects concerning both system and user. On the system side the following aspects are described briefly as:

- *Input performance*: refers to the recognition accuracy, response time and user's input coverage
- *Interpretation performance*: concerns the system's capacity to extract meaning accurately from the user input
- *Dialogue management performance*: concerns dialogue success, dialogue initiative distribution and ability to correct misunderstandings
- *Input/output modality appropriateness*: most appropriate input/output modality selection depending on the interaction context [103]
- *Contextual appropriateness*: refers to the appropriateness of an utterance within the dialogue context
- *Form appropriateness*: concerns the intelligibility and comprehensibility of the system's output, as well as the capability to convey specific information, including emotions, turn-taking, back channels

On the user side aspects include:

•*Perceptual effort*: listening effort required to understand the system messages

•*Cognitive load*: degree of concentration required to perform the tasks

•*Physical response effort*: physical effort required to communicate with the system

Quality of Experience(QoE)

QoE addresses the subjective criteria of users' perception of the interaction and refers to M. Hassenzahl's concepts of hedonic and pragmatic quality which influence the user perception of an interactive system [104]. Pragmatic quality refers to functional aspects, such as the utility and usefulness of a product while hedonic quality refers to non-functional aspects, that is to say aspects with no obvious relation to the task the user wants to accomplish with the system, such as the product's aesthetics or personality. The **QoE** includes the following aspects:

•*Interaction quality*: perceived input/output quality of the system, as well as distribution of the dialogue initiative (cooperativity)

•*Effectiveness*: accuracy and completeness with which users can complete a certain task

•*Efficiency*: effort and resources required for the system to achieve accuracy and completeness

•*Learnability*: easiness with which users feel they can handle the system

•*Intuitiveness*: extent to which users are able to interact with the system without a priori knowledge

•*Usability*: defined in ISO 9241 [105] as the *"extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*; as such, usability is one sub-aspect of the quality of the experience; the taxonomy considers two aspects of usability: the ease of use and the joy of use

•*Aesthetics, system personality and appeal:* hedonic quality aspects related to the overall appeal of the system consisting of users' sensory experience elicited be the system (*aesthetics*) and users' perceptions of system characteristics (*personality*) originated from the agent factors

•*Utility* and *usefulness*: service utility and usefulness in context of the user's needs referring to whether a specific user can resolve his task with the help of the system, frequency of use etc.

•*Acceptability*: refers to whether a user is prepared to use the system

3.3.2 Taxonomy of quality of conversational interactions

Inspired by the above presented taxonomy of **QoE/QoS** [74] and the work of Hassenzahl and colleagues [104] we developed a new taxonomy of influence factors relevant for the quality of conversational interactions from a user perspective (see figure 3.2).

The novelty of our taxonomy consists of mapping the **QoE/QoS** taxonomy on the framework of hedonic and pragmatic quality proposed by Hassenzahl, extending the categories with new elements and relationships from other relevant literature studies on product quality [98], service quality, [87], [106], [72] and guidelines for cooperative communication [64]. Additionally, the taxonomy indicates the impact of several categories on the users' mood and feelings towards the system during the interaction (marked in figure 3.2 in orange). The taxonomy represents a structural model with no exhaustive claims, meaning that it can be further extended with new factors and categories.

As shown in figure 3.2 the quality of conversational interaction (**QoCI**) is influenced, similarly to **QoE/QoS**, by external factors, such as *user, system* and *context factors*. Further, the **QoCI** depends on several internal factors grouped into two main categories: hedonic and pragmatic aspects. The hedonic aspects the question *"Did I enjoy the way of getting the information?"* while pragmatic aspects relates to the interaction *utility*, responding to *"Did I get the information I was seeking?"* In the following we will describe the components of each of these two main aspects in more detail.

Hedonic aspects

The attribute *hedonic* is derived from the Greek *hedone* which means *pleasure* [107]. Engaging in hedonic pursuits, or *hedonia* means seeking personal pleasure, enjoyment, and comfort. *Hedonia* can be achieved whether through physical means, such as sensory experience including visual, acoustical, haptic, or through emotional-cognitive means, such as enjoyment of social interaction, art, and so on [108]. As such, we argue that hedonic aspects in face-to-face conversational interaction include three main components: *attractiveness*, *fun*, i.e. entertainment degree and *comfort* experienced during the interaction with the system.

The *attractiveness* refers to the form[1] in which the interface presents itself to the outside world: how attractive it looks (*visual attractiveness*), how pleasant its voice sounds (*acoustical attractiveness* and how nice its *presentation style* is designed. Both visual and voice attractiveness are sensory experiences referred to under the global term of *aesthetics* in **QoS/QoE** [74]. Both *visual* and *acoustical attractiveness* share a

[1] *'Form'* should be understood in this context as opposed to *'content'*

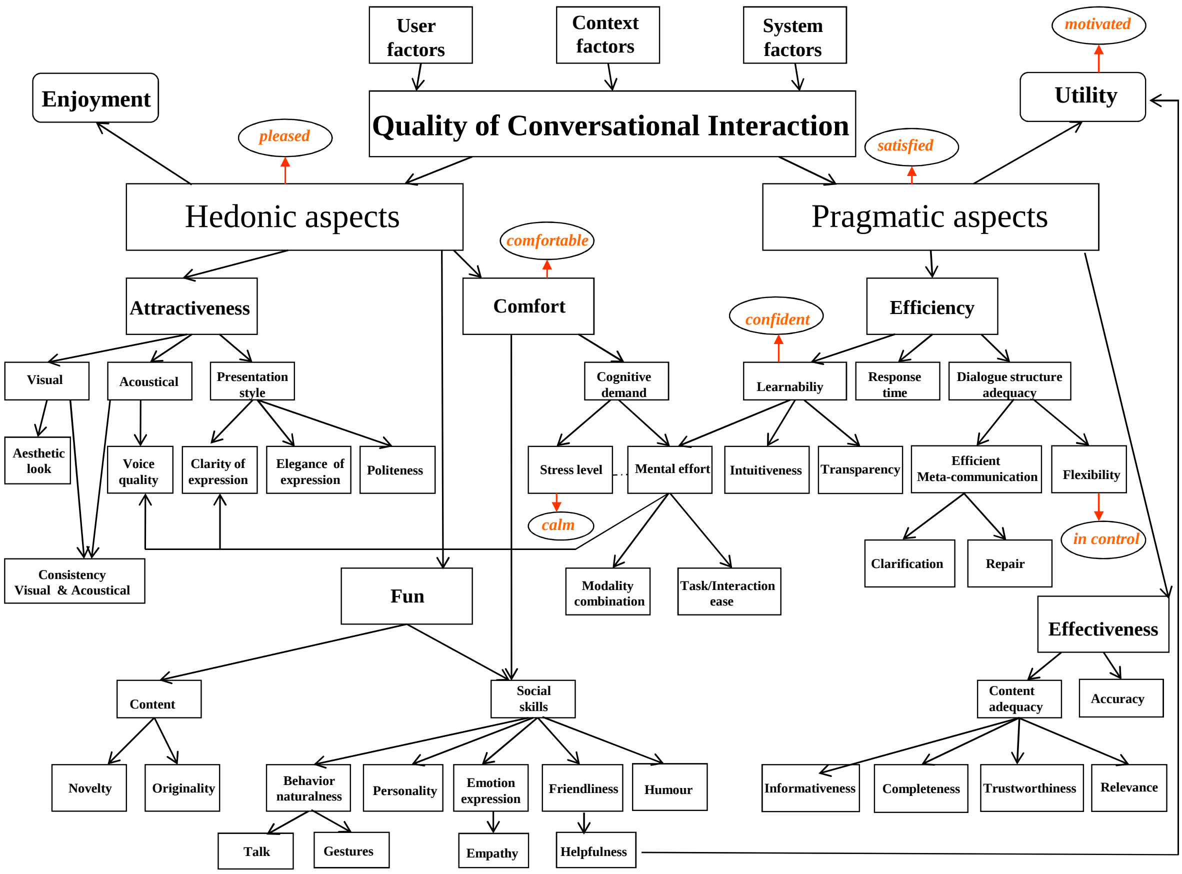

Figure 3.2: Taxonomy of quality of conversational interaction

common factor which refers to the *consistency* between visual appearance and voice. We added this factor since research has shown that people tend to like others whose appearance can be labeled consistently [15], [109].

Presentation style is a new addition to the *attractiveness* category and refers to *clarity of expression* (related with the factor *form appropriateness*, **QoE/QoS**,[74]), *politeness* (inspired by the dimension *courtesy* in the **QoS** developed by Evans and Lindsay [98]) and *elegance of expression* (inspired by the hedonic attribute pair *stylish/tacky*, Hassenzahl et al. [104]).

The hedonic aspect of *fun* refers to the entertainment degree of the interaction and is determined by the *content* presented and/or by the system's *social skills*. The *content* has two sub-aspects referring to *novelty* and *originality* of the information (inspired by Hassenzahl et al. [104]). The social skills are related with the system's behavioral expressivity and include five components: *naturalness, personality, emotion, friendliness* and *humor*.

Humor and *friendliness* are factors inspired by the hedonic attribute pairs *funny/ boring* and *sympathetic/unsym-pathetic*, (Hassenzahl et al. [104]). One common friendliness aspect is the *helpfulness* which is the system's perceived readiness to help users to perform a tasks. The system's *helpfulness* is a new addition to the taxonomy, being intuitively related with the overall pragmatic aspect of *utility*.

The factor *emotion expression* is also a new addition and refers to the system's capability to convey adequately and at the appropriate moment emotions that match the social interaction context. One example of emotion expression with positive effects in social interactions is the expression of *empathy*. The factor was inspired by the **QoS** taxonomy, (Parasuraman et al. [87]).

The system's *personality* (inspired by **QoE/QoS**, Moeller at al. [74]) refers to verbal and non-verbal cues which lead the categorization of the system as belonging to a certain personality type such as introvert, extrovert and so on.

The last factor in the social skills category is the *behavior naturalness* referring to the talk and gestures naturalness. The *behavior naturalness* is a new addition inspired by a questionnaire item mentioned in the ITU-T P.851 framework [106] (*"The system reacted in the same way as humans do"*).

The system's *social skills* can influence the level of *comfort* experienced by users during the interaction with the system. The second aspect influencing the *comfort* refers to the *cognitive demand* (**QoE/QoS**,[74]) imposed by the conversational interaction. The *cognitive demand* encloses the perceived level of effort needed to perform the task with the system (*mental effort*) and user feelings arising from this effort, i.e. (*stress*)[106]; the stress level determines intuitively the users' degree of feeling *calm*. Stress and mental effort are related, but the relationship is however, not compulsory: mental effort does not always accompany stress while stress can also be caused by other situational or context-specific factors, such as time limitation, noises, frequent occurrence of errors, and so on. The mental effort is further determined by the *task/interaction ease* [110], but also by the way the information is presented in terms of *voice quality, clarity of expression* and (optimal) *modality combination* [72].

The aspect of *enjoyment* (inspired by the *joy of use*, **QoS/QoE** taxonomy [74]) can be used to measure the overall hedonic quality experienced by users who would feel *pleased* or *annoyed* depending on how the system meets their 'hedonic' expectations. The attribute pair *pleased/annoyed* is inspired by Hassenzahl et al. [104]).

Pragmatic aspects

Under pragmatic aspects the taxonomy includes communication *efficiency* and *effectiveness* (both included in the **QoE/QoS**,[74]).

Communication *effectiveness* refers to the quality of the solution offered [111] and has two subcategories: one refers to the system's performance in terms of recognition *accuracy*, the other one refers to *content adequacy* and includes the *informativeness*, *relevance* and *completeness* of the statements, as well as their *trustworthiness* value. The factors *informativeness* and *relevance* are inspired by the guidelines developed by Bernsen and colleagues [64] based on the Gricean maxims (see section sections 2.2.2 and 2.4.1). The factors *completeness* and *trustworthiness* are inspired by the ITU-T P.51 framework questionnaire (*"The provided information was com-*

plete"/"You would rate the information as wrong/true").

Communication *effectiveness* refers to the resources expended in achieving an effective communication; these are mostly related to the task completion time which depends on the perceived system's *response time*, *dialogue structure adequacy* and *learnability* [111].

The adequacy of the dialogue structures (addressed in the **QoS/QoE** [74] in the *input quality* category) refers to the system's *flexibility* in allowing a mixed dialogue initiative and barge-ins; a flexible system would induce the feeling of being in control of the interaction. Further, dialogue adequacy refers to *meta-communication strategies* in terms of *repair* and *clarification*.

The last factor in the effectiveness category is the *learnability* which relates to the system's *intuitiveness* and *transparency*. The *intuitiveness* is presented in the **QoS/QoE** [74] as a separate item. However, we believe that the definition given to the intuitiveness subordinates this concept under the more global factor *learnability*. The *transparency* refers to the ability of the system to highlight the partner asymmetry and to communicate what it expects from the user. The system *learnability* can induce the feeling of being *confident* during the interaction. The factor is also related to user's *mental effort* which will increase or decrease depending on how intuitive and transparent the interaction is perceived.

The aspect of *utility* (similar as in **QoS/QoE** taxonomy) can be used to measure the overall pragmatic quality experienced by users who feel satisfied or frustrated depending on the way the system fulfills their 'pragmatic' expectations. The attribute pair *satisfied/frustrated* was inspired by Hassenzahl et al. [104]).

We used the **QoCI** taxonomy to guide the design of the evaluations performed in chapter 4, 5, 6, 7 and 8.

3.4 Evaluation approaches

In order to evaluate the quality of an interactive product two approaches can be followed: one deals with the measurement of interaction parameters concerning the performance of the system components and the interaction flow; the second deals with evaluations performed with test users or expert evaluators: quantitative questionnaires and qualitative interviews can be performed with test users while usability inspection methods (e.g. heuristics, cognitive walkthroughs, etc.) are usually applied in expert evaluations. In addition, prediction algorithms can be used to provide a quality estimation by mapping interaction parameters to subjective questionnaire results [102].

3.4.1 Interaction parameters

In the **QoS** taxonomy developed by Moeller et al. [74] several components include interaction parameters that can be quantified by a person external to the interaction process, that is to say without relying on user judgment. These interaction parameters can be divided, as proposed by Moeller and colleagues in [102], into the following subcategories:

Input-related parameters: such as *input performance*[2], *interpretation performance* and *input modality appropriateness*. Input performance can be measured in terms of speech, object, gesture recognizer accuracy at word/sentence/item/movement level counting the error rate, number of errors per unit, and so on. Interpretation performance can be quantified in terms of accuracy of filling in correct attribute-value pairs based on expert-driven estimation of 'correct' interpretation. Input modality appropriateness can be assessed with the help of modality properties as proposed by Bernsen and colleagues [103]. Input performance measurements were used in the analysis reported in chapter 6.

Dialogue and communication-related parameters: refer to the *dialogue management performance*, total dialogue duration, system/user response delay, initiative distribution (system/user turn duration), number of words/utterances exchanged, number of system/user questions, and parameters describing how fast new information can be processed by the system (query density, concept efficiency) [102]. We measured several dialogue and communication-related parameters in chapter 6, 7 and 8.

Meta-communication-related parameters: refer to 'dialogue problems' and are also related to the *dialogue management performance*; such problems can be help requests, time-out prompts, ASR rejections, error messages, clarification requests initiated by users or cancel attempts. Also, parameters quantifying the system recovery such as the system/user correction rate, and the implicit recovery can be quantified [102]. Such measurements were used in the research described in chapter 6.

Cooperativity-related parameters: refer to *contextual appropriateness* and can be measured in terms of violating the Gricean cooperativity principle [49] (see section 2.2.2) using the contextual appropriateness parameter [112].

Task-related parameters: refer to the task success for the whole dialogue or for only a sub-dialogue, that is to say whether the dialogue achieved its goal to give the users the requested information; it is related to the *dialogue management performance* and can be measured as a binary value (success/failure) or as calculated using as the K coefficient to prove the correctness of the answer given by the system (see below in section 3.4.4 the PARADISE model [110]).

Behavior-related parameters[3]: refer to the user behavior in terms of interactive response during the interaction. Several categories such as gestures (head nods, body posture, facial expression, hand gestures), gaze behavior and speech patterns (key-

[2]The parameters addressed in both **Qos/QoE** and **QoCI** taxonomies are displayed in cursive font

[3]This category is not listed in [102]. However, we added it because it relates to the interaction flow and it can be measured objectively.

words vs. sentences, dialogue initiative) can be observed and logged. They can be interpreted in terms of degree of involvement in the interaction, positive/negative responses to the system's personality, level of attention, etc. Further, correlations can be measured between these parameters, other interaction parameters and subjective user judgments. We collected behavior related parameters in the experiments described in chapters 6 and 8.

3.4.2 User evaluation

Quantitative questionnaires can be used to collect information about the user experience with the system in a measurable form. Here four types of questions can be distinguished which relate to: (1) user personal data, (2) functional and (3) non-functional interaction aspects and (4) overall impression of the system. Questions belonging to all four of these questions types are usually mixed together in a single questionnaire. For the evaluation of interactive systems and of their components there are several standard questionnaires that can be used.

SASSI is one of the most famous questionnaires used to evaluate the usability of uni-modal speech-based interfaces and was developed by K.S Hone and R. Graham [113]. The questionnaire addresses five different dimensions: system response accuracy (e.g. *performance*, likeability, *cognitive demand*, *annoyance*, habitability (e.g. *intuitiveness, transparency* and speed (*response time*) measured on a 7-point Likert scale. Items of SASSI questionnaire were used in the experiments described chapter 5, 6 and 8.

ITU-T Recommendation P.851 contains questionnaire items for evaluating the subjective quality evaluation of telephone services based on spoken dialogue systems [106]. The questionnaire uses a 5-point Likert scale and includes questions related to the user background, individual interaction with the system and user's overall impression of the system. Items from this questionnaire were used in the evaluations described in chapters 4 and 5.

• **User background**
The questions concerning the user background (e.g. *user factors*) refer to personal information (e.g. age, gender, profession, etc.), task-related information (e.g. frequency of the task, motivation, etc.) and system-related information (e.g. experience with similar systems).

• **Individual interaction**
The questions concerning individual interactions with the system include questions related with the quality of the information obtained and the speech input/output *clarity of expression, performance, listening effort*, system's interaction behavior (e.g. *transparency*, perceived *system personality* (e.g. *politeness*, user's feeling towards the system and perceived task fulfillment (e.g. *task/dialogue success*).

• **Overall impressions**
The questions concerning the overall impression of the system include questions addressing the overall impression about the system/service but also questions focus-

ing on the system's *manner of expression*, *meta-communication* strategies, perceived interaction control (*flexibility*, *dialogue initiative*), *comfort*, perceived *usability*, enjoyment, likeability, appropriateness of fulfilling the tasks, *utility*, *acceptability* and future use of the service.

AttrakDiff is a questionnaire addressing the attractiveness of interactive products, e.g. the product's ability to satisfy users' practical and emotional needs. The questionnaire developed by Hassenzahl and colleagues [114] uses semantic differential scales and measures the users' perception of a given product concerning its pragmatic qualities, hedonic qualities and overall appeal.
• **Pragmatic qualities (PQ)**
PQ include the following variables: comprehensible/incomprehensible, supporting/ obstructing, simple/complex, predictable/unpredictable clear/confusing, trustworthy/ shady, controllable/uncontrollable, familiar/strange.
• **Hedonic qualities (HQ)**
HQ contain the following variables: interesting/boring, costly/cheap, exciting/dull, exclusive/standard, impressive/nondescript, original/ordinary, innovative/conservative.
•**Overall appeal**
The overall appeal refer to the following variables: pleasant/unpleasant, good/bad, aesthetic/ unaesthetic, inviting/rejecting, attractive/unattractive, sympathetic/ unsympathetic, motivating /discouraging, desirable/undesirable.
The word-pairs are rated on an expanded 7-step Likert scale (from -3 to +3, including a null value).
Hedonic aspects concerning the *fun*, and *visual* and *content* attractiveness can be quantified using AttrakDiff. The questionnaire was used in the experiments presented in chapter 8.

To evaluate the *voice attractiveness* the **ITU MOS** questionnaires [115] can be used which measures the voice quality on a 5-point Likert scale along 8 dimensions: pronunciation, articulation, speaking rate, pleasantness, listening effort, comprehension problems and overall impression. This questionnaire was used in the study described in chapter 4.

The overall workload, concerning *mental, perceptual* and *physical effort* can be assessed via the **NASA TXL** questionnaire. The questionnaire includes additional dimensions concerning temporal demands, own performance, effort and frustration and uses a 20 level scale for ratings. The *mental effort* and the *stress* associated with can also be measured using physiological measurements, such as the heart rate variability (HRV) and the galvanic skin response (GSR).

Additionally, qualitative interviews can be used for the system evaluation. Such interviews are particularly useful for getting the story behind the participants' experience and the interviewer can pursue in-depth information around the topic investigated. The evaluation can be performed as a standardized open-ended interview, that means the same questions are posed to all users; another option is a closed, fixed-response interview - where all interviewees are asked the same questions to be responded to from a list of possible answers. Standardized open-ended interviews were used in the experiments presented in chapters 5 and 7.

3.4.3 Expert evaluation

Expert evaluations is another assessment option available for testing interactive products. For example the evaluation of *agent* and *functional* factors can be done by comparing them against the specifications written in the system documentation. Most of the agent factors are specified in the documentation by system developers or design experts while functional factors can be best specified by domain experts. Also, *contextual* and *input/output appropriateness* (i.e. modality combination) can be assessed using a domain expert [74]. In general, expert evaluation involves a usability expert inspecting a system to identify any usability problems. According to Nielsen and Mack [116] there are eight types of usability inspection methods:

1. Heuristic evaluation: an informal method involving usability specialists judging whether the dialogue elements follow established usability principles

2. Heuristic estimation: a method in which inspectors are asked to estimate how usable a system will be on a variety of quantitative criteria.

3. Cognitive walkthrough: developed by Wharton [117], this method is meant to help designer teams to quickly evaluate interaction systems from the early stages of development, that is to say, it does not require a fully functioning prototype. The method emphasizes cognitive aspects, such as learnability by analyzing users' mental processes required for each step. The experts perform the test taking into account the potential user perspective with the purpose of identifying problems that might arise during the interaction. After each scenario the experts are asked to answer specific questions stated in a questionnaire. This method will be described in chapter 5 to detect design flaws in the dialogue structure implemented in a question-answering system.

4. Pluralistic walkthrough: a method which involves developers and usability experts going through a scenario and discussing usability issues associated with each scenario step. It requires the development of a series of tasks presented to a panel of users who are asked to write down the actions they would take to complete the task.

5. Consistency inspections: a method conducted to determine whether multiple products from the same development team are consistent in their design and operation.

6. Standard inspections: involves an expert on an interface standard inspecting the interface for compliance.

7. Formal usability inspections: combines individual and group inspections in a six-step procedure with strictly defined roles to combine heuristic evaluation and a simplified form of cognitive walkthrough.

8. Feature inspections: a method in which sequences of features used to accomplish typical tasks are analyzed; the method checks for long sequences or cumbersome steps, such as steps that would not be natural for users to try or that would require extensive knowledge/experience.

3.4.4 PARADISE prediction model

The PARADISE (PARAdigm for DIalogue System Evaluation) model is an approach to predict user satisfaction based on interaction parameters proposed by Walker et al. [110]. The model combines a set of interaction parameters K (task success) and c_i (cost function of the dialogue, including dialogue efficiency, recognition score, repair and help) in an estimation of user satisfaction (US) in the following way:

$$US = \alpha \; N \; (\kappa) - \sum_{i=0}^{n} (w_i) \; N(c_j)$$

where user satisfaction (US) is calculated from the arithmetic mean over several user quality judgments and N is the Z-score normalization function.

The interaction parameters and the user quality judgments have to be determined in controlled experiments. The weighting coefficients α and w_i are calculated using linear regression models. Once the parameters are determined the function can be used to predict the user judgments for unseen dialogues, that is to say without directly asking the users [102]. The percentage of variance covered by the model is measured with R^2 and is based on the comparison between the predicted and measured values for each dialog. For the prediction of the training data, R^2 is usually below 0.6 [118] while for unseen data the results have typically even lower R^2 [119] making PARADISE a rather unsuitable evaluation method for user satisfaction.

Beringer et al. proposed to transfer the PARADISE approach to multimodal interactions by defining a methodology for end-to-end evaluations (PROMISE)[120]. However, the method has, as PARADISE, limited predictive power.

However, even though PARADISE and PROMISE were not particularly successful, the idea of predicting subjective user judgments is nevertheless, very attractive, especially because evaluations with test users are costly and time consuming. In chapter 6 we will propose to use information gathered from behavioral analysis to find correlations and possible parameters with potential predictive power for evaluation results. Alternatively, the information can be used for user modeling.

3.5 Summary

In this chapter we have discussed several definitions of the notion quality and its underlying aspects, as well as different evaluation and assessment methods.

The concept quality represents a complex term that can be defined from different perspectives. However, there is a general tendency to define *quality* as a multidimensional construct in which objective and subjective aspects of a product or a service are combined.

A number of researchers in the quality assurance field have developed lists of quality dimensions for product and service quality, such as Garvin [89], Evans and Lindsay [98], Parasuraman [87]. These lists are good starting points, but current research indicates that at least in terms of service quality, the dimensions can be different depending on the industry type and may not apply in all cases.

Task-oriented conversational interfaces can be seen as interactive products which offer informational services. However, the intrinsic characteristics of conversational interfaces - the interactivity - makes the judging process of their quality far more complex as compared with other non-interactive products and services since users actively participate in the interaction being a part of it.

A taxonomy of the most relevant quality dimensions for multi-modal human machine interaction was developed by Moeller and colleagues [74]. The taxonomy contains two main aspects concerning the quality of service (**QoS**) and the quality of experience (**QoE**). **QoS** refers to performance criteria that can be mostly objectively measured by external judges. It contains two factor layers, one considering influencing factors on the interaction behavior, the other one referring to interaction parameters. **QoE** refers to the subjective criteria addressing the users' perception of the interaction. Inspired from the **QoE/QoS** and the work by Hassenzahl and colleagues [104] we developed a new taxonomy which focuses in particular on influence factors relevant for the quality of conversational interactions (**QoCI**) from a user perspective.

In order to evaluate the quality of interactive products two approaches can be followed: one deals with the measurement of interaction parameters concerning the performance of the system components and the interaction flow; the second deals with evaluations performed with test users or expert evaluators; quantitative questionnaires, qualitative interviews (for test users) or usability inspection methods (for expert evaluators) can be used.

Additionally, prediction algorithms such as PARADISE and PROMISE have been proposed which provide a quality estimation by mapping interaction parameters to subjective questionnaire results. Although these models are helpful in the system design process, their predictive power is limited, covering usually 40-60% in the variance of the user judgment used for training. On the other side, the idea of predicting subjective user judgments is very attractive, especially because evaluation with test users are costly and time consuming.

Part II

Experimental studies

Chapter 4

When the voice has the wrong accent

In this chapter we will discuss the importance of voice characteristics in the context of voice user interfaces (VUI) design. In particular, we will address the issue of choosing the right voice accent for a VUI application designed for Singaporean mobile phone customers. The application was meant to serve as a voice enabled user manual for a Motorola A388. The dialogue script implemented in the VUI prototype was designed using written instructions. Further, it was tested and enhanced with human-human dialogues. Thus, the chapter aims to make a contribution to our forth research question (RQ 4) concerning how human communicative interaction patterns can be used to test and enhance conversational interfaces. We used the VUI prototype to perform an experiment aiming to determine whether the system's evaluation would improve if the interface were to have a voice accent matching the user's accent. The experiment represents the main topic of the chapter and tries to respond to our first research question (RQ 1) concerning the impact of voice accents on the user evaluation. The content of this chapter was published in [121].
The chapter is structured as follows: a short introduction about use, design challenges and design opportunities of voice user interfaces will be presented in section 4.1. In section 4.2 we will present the design steps of the voice user manual concerning the structure of the dialogue and characteristics of the VUI persona focusing on the choice of the accent. In section 4.3 we will present relevant literature concerning voice accents while section 4.4 will focus on the experiment design. In section 4.5 we will show the results gathered from our study. The results will be discussed in section 4.6. The chapter will end with a summary in section 4.7.

4.1 Voice user interfaces - a brief overview

Currently, we live in a world surrounded by computer controlled devices which enable us to access huge amounts of information available online using PDAs, tablet PCs, mobile phones, desktops, etc. Such devices provide different access modalities to information requiring the interfaces to be permanently adapted to a dynamically changing context. However, as applications become more pervasive and devices much smaller such adaptations become more difficult due to minimal visual displays, tiny input buttons and lack of an alphanumeric keyboard or mouse [122].

Voice user interfaces (VUI) offer a solution to such input-output limitations as they are able to recognize the voice of their user, to understand voice commands, and to provide responses to them, usually, in real time. The state-of-the-art in speech technology already enables the development of automatic systems designed to work in real conditions. In fact, VUIs have become popular in the last decade as interfaces for communicating information, controlling devices or accessing data.

Typical examples of VUIs commonly deployed in business areas since the late '90s are interactive voice responses (IVR) applications. These applications allow users to retrieve information, such as bank balances, weather forecasts, flight schedules or movie show times from any telephone line. Users interact with an IVR application combining voice input and touch-tone keypad.

As built-in functionalities VUIs are deployed in cars and enable drivers to use their voice for placing and accepting calls while driving, entering an address, selecting a radio station or controlling the navigation system settings. In smart home environments VUIs are used to control audio-visual equipment, phones, household appliances, doors or curtains.

VUIs offer a wonderful hands free and eye-free solution for circumstances in which typing is problematic: for mobile devices, where the size of the buttons can be critical for accurate input, speech presents an excellent way to enter and retrieve information; for people with temporary (e.g. repetitive strain injuries) or permanent (e.g. visual or locomotor) disabilities VUIs have potentially enormous benefit since speech recognition can serve as a replacement for both keyboard typing and mouse control while speech synthesis can be used to read information presented on screen to substitute reading.

VUI design is an integral part of developing any speech based application and is

> *"perhaps the most critical factor in the success of any automated speech recognition ASR) system, determining whether the user experience will be satisfying or frustrating, or even whether the customer will remain one" [53].*

Even though many design principles used for speech application are the same as for graphical user interfaces (GUIs), VUIs have two main characteristics that pose unique design challenges but also opportunities as compared with GUIs: the modality is auditory and the interaction is based on speech.

The auditory modality enables communication only through non-persistent messages. This means, the users hear a message and subsequently the message is gone. In contrast, GUIs display persistent messages, i.e. messages that remain visible on the screen, so the users are allowed to interact with the system at their own pace. With VUIs the interaction pace is at the output rate, such as in human conversations.

The ephemeral nature of the auditory modality increases the users' cognitive load during the interaction with a VUI. This shortcoming can be overcome by a proper interface design, in which the users are not overloaded with too much information at once and have the possibility to control the interaction pace.

The use of speech, as communication modality imposes additional challenges for the VUI designer: compared with GUIs, where accomplishing a task depends on specific learned actions created by designers, i.e. drag and drop icons, communicating through speech is learned intuitively at early age. So, designers must anticipate the users' convention and interaction patterns rather than creating underlying conversational elements.

On the other hand, using the auditory communication modality offers unique design opportunities: since people extract meaning not only from the content but also from the voice's quality, intonation, accent and rhythm, designers can easily create a consistent system personality by carefully choosing a voice talent for recording the VUI's prompts. In this way the application can be tailored for target groups creating an appropriate user experience [53].

4.2 A voice enabled user manual for mobile phones

Due to the great potential of VUIs a project aiming to create voice enabled applications for mobile phone users was initiated at the Institute of Infocomm Research (I^2R), Singapore. The project was part of the larger program to design and implement practical systems for automating help desks, Yellow Pages, and various other telephone-based information services. In the current project the focus was on creating a voice enabled user manual for mobile phones.

Manuals usually assist users to become familiar with a new purchase providing the support needed for performing several tasks. However, when using a manual on paper to search for information, users are limited by their own ability to find the desired information and by the clarity of manual's structure. Some manuals are available online, however many questions arise when the user is operating the phone. Thus, searching user manuals available online or through hundreds of paper pages can be frustrating, inefficient and time consuming. Instead, a voice user interface could offer an alternative solution to support users in finding phone related information in real time while performing tasks on the same phone.

For our study we choose a Motorola A388 (see figure 4.1), a smart phone with many functionality such as e-mail, handwriting recognition, IM, EMS, fax, Bluetooth, Internet access, PDA, and so on. Users need to navigate through a list of options using interactive touch-screen menus on the phone. The menus are efficiently designed, but nevertheless it is still hard to find information about the more subtle features, configuration settings, sending options, etc.

4.2.1 Dialogue design

The first step in developing our prototype was to determine what requirements (i.e. what informational need) our interface should satisfy. For this purpose we conducted a first study with eight novice phone users. Firstly, each user was asked to get familiar with the phone for five minutes. None of the users had access to the user

Figure 4.1: Motorola A388

manual. Next, the users were required to write down a list of unclear topics related with the phone functionality and to each of these topics to formulate questions.

Almost half of the questions asked by our participants were related with SMS messaging. Therefore, we choose for our first prototype to focus only on this particular topic. A list with all questions related to SMS mentioned by our test subjects was prepared. These questions could be split into four thematic groups concerning problems encountered by sending and/or receiving text messages, using or installing new T9 language packages, embedding/saving pictures in/from messages and sorting/searching for messages (see table 4.1) .

Nr.	Question
1.	How do I configure my Motorola to send an SMS?
2.	Why can I not receive an SMS?
3.	How can I switch off T9?
4.	How can I download and install T9 languages?
5.	How can I attach a picture to my SMS?
6.	How to save a picture received by SMS in another folder?
7.	How can I sort my messages by name?
8.	How can to search for a keyword in a message?

Table 4.1: Example questions formulated by test participants

To answer the questions we consulted the Motorola user manual and found that more than one third of the questions were not answered. This finding reinforced our initial thoughts about the potential usefulness of our application. Further, we searched on the Motorola web page and other user forums in order to find the requested answers.

The problem that we encountered by answering questions using the Motorola manual and the Internet websites was that explanatory information was often presented in visual form, as shown in figure 4.2.
Thus, in order to design the prompts we had to analyze the screens in detail and to translate them into a verbal representation that closely mirrored the original. The descriptions were kept simple and as short as possible in order to keep the user's cognitive load to a minimum. An initial dialogue script was designed in accordance

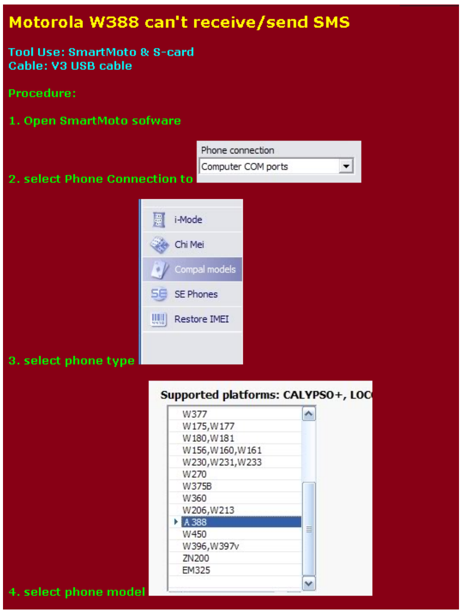

Figure 4.2: *Explanatory information from an online user forum*

with the dialogue design guidelines presented in section 2.4.

Since the interface was meant to serve the needs of both novice and expert users two levels of prompts were designed: the prompts designed for the first level were short, but contained enough information to enable expert users to proceed to the next step; the second level prompts were longer, as they included more detailed information. The users could chose at the dialogue begin which explanation level they preferred.

Example:
S0: *"Hello, this is the Motorola voice enabled user manual. You can ask questions related to SMS sending, SMS receiving or other options. Do you know this system?"*
U0: *"No".*
S1: *"The system can support you to find information also about sorting, searching for keywords in messages, download, install or switch off T9 languages. You can use it by answering the system's questions. In addition you may use two special commands,* ***repeat*** *and* ***change*** *to have the most recent information repeated or changed. The system will understand you only when you answer its questions briefly and one at a time."*
S2: *"Please ask your question."*[1]

Due to the novelty of our application we did not fear any inferences with other similar task domains. Thus, no particular user expectations were taken into account.

The first system prompt included indications on what kind of information the system was providing and how users could interact with the system (see in the example above the design of the prompt S0). The system's prompts were formulated

[1]For an expert user the system will skip statement S1 and will go directly to statement S2.

as closed-ended questions or short answers handling single pieces of information.

The dialogue initiative was mixed: the system was leading the interactions, but the users were allowed to ask for repetitions and clarifications, to skip dialogue segments, to accept or to refuse system proposals. However, due to a relatively rigid wizard interface we chose to implement an explicit feedback strategy and no navigation features.

Once the dialogue script was designed a small pilot study was conducted with five test participants and one tester using a person-to-person conversation. The test participants received a list with the questions related to message sending topics (as mentioned in table 4.1) while the tester got a list with the designed prompts. Both users and tester talked over a simulated phone line. They were instructed to imagine themselves in a help desk scenario where people call to ask questions and are attended by a help-desk employee. The conversations were recorded and the dialogues were manually transcribed. The transcriptions were then compared with the original dialogue script. The comparison enabled us to identify possible problematic dialogue units and to enhance the script with additional questions asked by the test users. Eventually, the dialogue was refined to a final version.

4.2.2 VUI persona

Amongst all the variables VUI designers can manipulate to improve the user experience (e.g. voice quality, ability to accommodate interruptions, flexibility of the dialogue flow, length of prompts, repair strategies) the easiest to control are the voice characteristics and the language features used for prompting. Such characteristics refer to the VUI 'persona', i.e. the personality of the system reflected by the choice of the voice actor, including his/her gender, accent or speech style. These choices depend on the company hosting the application, but also on the target user group. For our interface we chose a female voice as recommended in [123] and a formal speech style as being more appropriate for interfaces giving instructions, i.e. from which users have to learn [124].

Concerning the accent our first thought was to use Standard English, since we were targeting Singaporean customers. Singapore (a former British colony) is a complex network of multi-ethnic, multilingual speech communities who use English (one of the four official languages) to understand each other. Singaporeans learn Standard English from childhood along with their mother tongue, either Mandarin, Malay or Tamil. However, the English spoken in Singapore has a very distinct, local accent which explains the official name given, Singaporean Standard English (SSE). Apart from a few words reflecting local realities SSE differs from British Standard English (BSE) mainly on the phonological level [125], [126] - a list summarizing the main pronunciation differences between SSE and BSE can be found in appendix A.

Thus, the question that has to be answered here is: which voice accent is more suitable for a VUI targeting customers with almost native English proficiency, but speaking the language with a certain local pronunciation?

4.3 Related work: which accent is better?

The impact of voice accents was the subject of many socio-psychological research studies in the past [127]. Their results showed that listeners react subjectively to accents by assigning personality traits on the basis of ethnic identity [13].

In general, foreign non-native accents are judged less favorably on social attractiveness than local native accents - locally accented speakers appearing to be more trustworthy, dependable and friendly [128] [129]. Moreover, speakers converging most to the language of the listeners are perceived more positively [130]. The reason is that accent similarity is unconsciously interpreted by listeners as a sign of equivalent status, values, interests, needs [131], [132], and consequently, is perceived as being socially more attractive.

Experiments at Stanford University by Nass and Brave [124] transposed the effects of (human) spoken language accents to VUIs. Nass and Brave indicated that gender, place of origin or personality simulation of a VUI strongly affects users' impressions of that interface. They showed that, as in human-human interactions users seem to be more trusting, more forgiving of errors and report greater satisfaction with VUI personalities and ethnicities that match their own.

In one of their experiments Nass and Brave tested the effect of regionally marked speech with 96 subjects, half Caucasian Americans and half first generation Korean Americans. They had to rank the overall quality of two voice agents of an e-commerce website, one speaking with a Korean accent and language style and the other one using an Australian accent and language style. Results showed that agents were rated much more positively when they spoke with an accent that matched the user ethnicity. For the Korean Americans the shared accent was a strong mark of similar socio-cultural background evoking familiarity in geographical context where their parents were foreigners. On the other hand, the Caucasian Americans clearly preferred the Australian accent as being culturally much closer to their own.

Nass and Brave's conclusion is that at a fundamental psychological level, users tend to make no distinction between voices coming from a human being and those coming from a computer. Therefore, users will apply the same social categorization to real and virtual speakers showing preference on those perceived as being on the same "team" as themselves [8].

4.4 Experiment design

To test the user preferences for a particular VUI accent we replicated Nass and Brave's study trying to verify if this preference stereotype can be confirmed in a different geographical location and sociological context. Following the Nass and Brave's findings our hypothesis was that users in their native country environment would tend to prefer speaking with individuals of their own socio-cultural educational background when seeking help to find information. In particular, we speculated that Singaporeans would prefer to speak with a VUI assistant with a Singaporean accent as opposed to a British accent.

In order to carry out the experiment two sets of prompts were prerecorded in two content-identical versions: one corresponding to the Singaporean Standard English (SSE) accent and one corresponding to the British accent. Both prompts versions were read by a professional voice talent, a native born and raised Singaporean female who was professionally trained to speak BSE.

Further, a Wizard-of-Oz (WOZ) simulation of a spoken dialogue system (SDS) was created using the CSLU toolkit [2]. The speech recognizer module of the system was replaced by the wizard to ensure a frictionless interaction. The spreadsheet with all dialogue prompts was saved into a text file. A Perl script was written to convert the prompts into Tcl/Tk scripts for the CSLU interface. The SDS's role was to act as a voice enabled user manual, since our prototype was not implemented yet.

The experiment was conducted with 59 subjects. The subjects were informed they were going to interact with two different systems using automatic speech recognition software. Each test subject was given two simple problems to solve concerning SMS messaging: 1) how to set up a proper connection with the server to send/receive SMS and 2) how to attach pictures to a message. Each problem required interacting with the simulated voice enabled manual represented by a virtual voice assistant. Subjects were told they were talking to fully automated SDS systems. In fact, they were interacting with the wizard. To ensure the same conditions for all test participants we included only second level dialogue prompts, i.e. dialogue prompts meant for novice users. Accordingly, we selected only non-experienced users for our experiment.

To remove any potential bias from being exposed to one system before the other the test subjects were organized into two groups. One group was asked to evaluate the BSE system first, then the SSE version. The second group did the opposite: they evaluated the SSE system first and afterwards the BSE version.

After completing each test with a system the participants were asked in a first questionnaire to rank on a five-point Likert scale two hedonic aspects - voice quality and politeness - and two pragmatic aspects - dialogue easiness and trustworthiness. For our evaluation we used the ITU MOS questionnaire [115] to assess the voice quality and items from ITU-T Recommendations P.851 questionnaire [106] for the other variables - both questionnaires were discussed in section 3.4.2.

At the end of the experiment the users were also asked in a second questionnaire to estimate the usefulness of the application and intention of use (on a three-level scale, *yes, maybe, no*). Additionally, they were asked to indicate how much they enjoyed interacting with the VUI (on a five-point Likert scale), as well as to make comments about their general impression of the system and improvement suggestions[3].

The experiment was recorded to allow the detection of potential dialogue shortcomings and other interaction flaws.

---

[2]http://www.cslu.ogi.edu/toolkit/

[3]The questionnaire used in this study can be found in appendix B.

4.5 Results

Thirty-six male and 23 female native Singaporeans participated in the experiment (see figure 4.2). All subjects were educated and familiar with the use of the English language and there was no communication barrier for them to understand the instructions to undertake the experiment and do the evaluation. All subjects were novice users of the mobile phone A388 and many of them (89%) were first-timers in using the services of an automated SDS system.

Age	
Range	%
<25	11.86%
25-34	54.23%
35-44	27.11%
45-54	6.77%

Mother Tongue	
Language	%
English	16.94%
Chinese	72.88%
Malay	6.77%
Tamil	3.38%

Educational Background	
Education	%
Technical	71.18%
Non-technical	28.81%

Table 4.2: Distribution of age, mother tongue and educational background among subjects

All participants, except for one managed to successfully complete the task. Having completed the WOZ experiment the test subjects were first asked to rank the voice quality of the system. They were not told it was the same voice playing the SDS assistant in both cases. The results showed that 65% of the subjects voted for the BSE while 45% did the same for the SSE (the percentages refer to the cumulated positive values - good and very good).

The subjects were then asked to state their opinion regarding the systems' politeness: around 88% of the subjects rated the BSE system as being very polite or polite while only 50% felt so with the SSE system, despite the fact the spoken prompts of both systems were identical.

Concerning the dialogue structure the cumulated positive values indicate that 82% of the test subjects considered that the BSE system was easy and very easy to interact with while 71% chose the same categories for the SSE system.

Next, the subjects were questioned about the trustworthiness of the systems. The BSE system appeared trustworthy and very trustworthy for 51% of the participants while the SSE system did the same for only 43%.

The repeated measurements ANOVA performed on our data showed that except for trustworthiness there is a statistically significant difference in the way the subjects perceived the accents (see table 4.3). The mean rating for each tested factor

Nr.	Factors	F(1.58)	p-value
1.	Politeness	15.79	$p<.001$
2.	Voice quality	4.65	$p<.050$
3.	Dialogue easiness	4.10	$p<.050$
4.	Trustworthiness	1.38	p=n.s.

Table 4.3: Results of the significance test with ANOVA

was higher with the BSE system than with the SSE system (see figure 4.3). Results within groups showed no significant differences.

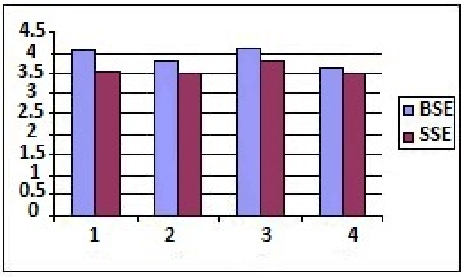

Figure 4.3: Mean rating for measured factors for BSE and SSE 1-politeness, 2-voice quality, 3-dialogue easiness, 4-trustworthiness

From the numerous comments and improvement suggestions gathered from the second questionnaire, as well from the analysis of the experiment recordings, we detected several flaws in our dialogue design.

One of the problems detected was the system's inability to produce prompts uttered in earlier stages of the dialogue without going back to the main menu. Some of the user's comments criticized this issue specifying the wish to have a less rigid system which would allow them to indicate which prompts they would like to have repeated. A solution to this problem would be the implementation of a global navigation feature that could be added to the system to enable access commands from any point of the dialogue. This finding confirms the importance of design features concerning flexibility features and navigation capabilities (see section 2.4).

Another problem was the explicit feedback strategy implemented which incremented the dialogue length unnecessary. A better option could be to use an implicit feedback strategy, i.e. to incorporate the feedback statement into the explanation, as shown in section 2.4. Also the ability of the system to recover easily from errors and to react quickly to barge-ins was addressed in the user's comments.

However, 71% of the subjects indicated the system responses were fast and the interface *"well done"*, *"efficient"* having a *"nice design"*, and so on. 82% of the subjects mentioned they found the application useful. Regarding the future use of the voice enabled user manual 49% of all subjects said they would like to use it and 38% said they would maybe use it. 75% of the participants indicated they enjoyed the interactions. These findings are encouraging for future design of intelligent mobile phone interfaces that incorporate voice enabled services.

4.6 Discussion

Contrary to our expectation users seemed to prefer the British accent over the Singaporean regardless of their mother tongue, age, educational background or gender. This outcome was especially visible in the category *politeness* and might have several possible explanations.

One influencing factor might have been the fact that the SSE accent resembles the Singaporean Colloquial English (SCE) accent, an indigenized variety of English (also called Singlish) spoken by Singaporean locals [133]. At governmental level the use of SCE is strongly discouraged as it is believed to hinder the proper learning of Standard English [134]. The fact is that linguistic features of SCE, such as a mixed vocabulary of English, Chinese, Malay, Tamil words and a syntax similar to Southern varieties of Chinese give the impression of 'broken English'. Even though still popular in casual situations there is a general tendency to avoid Singlish in the workplace or in other formal settings. Therefore, the contamination through association with a frowned upon idiom in a formal context of public information seeking might explain the users preference for the other foreign accent considered to be an absolute standard. This conclusion is also reinforced by numerous studies in linguistics and social psychology that confirmed the preference of listeners within a certain language area for standard accents. For example, the studies performed in the United Kingdom by H. Giles [128] [135], revealed that speakers using the BSE pronunciation were rated as more competent, self-confident and educated than speakers using a non-standard accent (Somerset Accented Speech) or regional accent (South Welsh).

On the other hand, the foreign accent used for comparison belongs to a culture which is highly valued in Singapore. It is a well-known fact that the island was borne into modernity under British administration [136]. Thus, the elevated opinion of the British culture might have dominated the expected effect of common ethnic background, the standard accent being considered a marker of power [137]. A similar conclusion was drawn in socio-linguistic studies from New Zealand (NZ), where listeners surprisingly devalued NZ English on all characteristics when it was compared with other English accents such as North American English and BSE [138], [139].

Eventually, the geographical location of our investigations might have also been of influence: Singaporeans were confronted with their own accent in their native environment. Therefore, the psychological impact of 'home accents' on test subjects (such as in Nass and Brave's experiment) did not take place.

4.7 Summary

The main focus of this chapter was to analyze the importance of choosing the right voice accent in context VUI development. A first prototype of a VUI meant to serve as a voice enabled user manual for a mobile phone (Motorola A388) was presented. The prototype was developed within a project belonging to a larger program aiming to design and implement practical systems for automated help desk services.

The input of eight novice test participants was used to formulate the initial informational requirements for the voice enabled user manual. The participants were asked to write down questions related to unclear functionalities of the mobile phone. To answer the questions we used the Motorola user manual as well as information posted online.

An initial dialogue script was designed, tested with another five test participants and further improved to a final version.

Since our interface primarily targeted Singaporean customers we thought to include in the design of our VUI persona user specific language features, i.e. to use prompts spoken in SSE. Following findings from the literature we speculated that Singaporeans would prefer to speak with a VUI assistant with a Singaporean accent as opposed to a British accent. SSE is widely spoken in Singapore and differs from BSE mainly in terms of pronunciation.

The hypothesis was tested in a WOZ experiment with 59 native Singaporeans subjects who ranked two hedonic quality aspects (voice quality and politeness), and two pragmatic quality aspects (dialogue easiness and trustworthiness) for two virtual voice assistants: one speaking with a Singaporean accent, the other one speaking with a British accent.

To design our questionnaire we used items from the ITU MOS [115] questionnaire focusing on the voice quality and the ITU-T Recommendation P.851 questionnaire [106] for the remaining variables.

Contrary to our expectation and despite the identical content of the information presented the British accented assistant was ranked higher in all categories than its Singaporean counterpart. The ranking was significant for politeness, voice quality and dialogue easiness.

Our experiment could only partly confirm the theories presented in [124], namely that voice characteristics, in particular the voice accent is a critical design issue for VUIs, as it strongly affects users perceptions of other system features, such as dialogue structure or sound quality. On the other hand, the results demonstrate that widely recognized stereotypes, such as ‘similarity attracts‘ illustrated in the same book might be inconsistent in situations where cultural and psychological biases interfere. Therefore, mindlessly designing interfaces conform to such stereotypes might be unjustified and even detrimental to the everlasting goal of improving user satisfaction. Further research in this area should consider not only the distinction between standard/non-standard or local/foreign accent, but also the geographical location, cultural identity and attitudes of the speaker group towards the selected idioms chosen for an experiment.

The experiment also gave an important feedback on the flaws of the currently implemented dialogue strategy: improvements concerning global navigation features, implicit feedback strategies, quick error recovery and reaction to barge-ins should be considered for the future design.

Chapter 5

Experimenting with IMIX and its embodied conversational agent Ruth

In this chapter we will present two studies performed with the multimodal question-answering system IMIX. The system consists of a user interface with multimodal input/output modalities and a virtual agent. The first study will focus on the user interface and aims to analyze the adequacy of the conversational structure implemented from the perspective of the affordance concept. The analysis is performed by comparing IMIX conversational structures with those occurring in natural human conversations. The study aims to respond to the research question RQ4, namely how human communicative interaction patterns can be used to enhance conversational interfaces. In the second study we focus on the virtual agent investigating how its gender ambiguous appearance - female voice combined with a rather masculine look - influences the users' perceptions of the system. The study responds to our first research question RQ1 regarding how voice consistency with physical appearance influences the evaluation of a conversational interface. The contents of this chapter were published previously in [140], [141], [142].
The chapter is structured as follows: section 5.1 presents a general overview about question answering systems, section 5.2 introduces the IMIX system, section 5.3 and 5.4 respectively present the two studies, including the background work, preliminary studies, methods, results and discussions. Finally, section 5.4 concludes the chapter.

5.1 Question answering systems - a brief overview

Question answering (QA) systems are information retrieval systems that aim to respond to questions formulated in natural language by giving concise answers back rather than lists of documents. They can be especially useful in situations where users are looking for very specific information and/or do not have time to read all of the available documents [143].

Research into QA systems addresses with a wide range of question types including factoid, lists, definition, *'how-'*, *'why-'*, hypothetical, semantically constrained, complex temporal, cross-lingual questions, and so on (see examples in table 5.1).

Question type	Example
Factoid	Who is the president of USA?
Lists	Which are the countries adhering to NATO?
Definition	What is aspirin?
How	How many cups are in a gallon?
Why	Why is the sky blue?
Hypothetical	What soft drink would provide me with the biggest intake of caffeine?
Semantically constrained	What color is the top stripe on the U.S. flag?
Complex temporal	Who was spokesman of the Soviet Embassy in Baghdad during the invasion of Kuwait?
Cross-lingual questions	*Answer is given in a different language as the question*

Table 5.1: Examples of question types

To answer questions a QA system is highly dependent on good search corpora which can vary in size, from small local document collections, to internal organization archives, a newswire or even the World Wide Web.

The first QA systems were developed in the '60s as early artificial intelligence systems. Two of the best known systems developed during that time were BASEBALL [144] and LUNAR[145]. BASEBALL could answer relatively complex questions about the US baseball games, such as *'How many games did team X play in June?'* or *'On how many days in July did eight teams play?'* and was used over a period of one year. LUNAR, in turn was designed to respond to questions related to geological analysis of rocks returned by the Apollo moon missions. The system was successfully demonstrated at a lunar science convention in 1971 where it was able to answer 90% of the questions posed by geologists, without prior instructions on the system [146]. Both systems were very effective in their chosen domains. Their success relied not only on their advanced IR and NLP techniques, but also on the amount of information processed: both systems were **closed-domain** systems, i.e. they were expert systems dealing with questions based on a specific, limited domain.

In contrast, **open-domain** question answering systems allow users to put questions on any topic. These systems retrieve relevant passages from very large text collections (mostly the World Wide Web) to find answers to the specific question. Some famous examples of open domain QA systems available online are: Start[1] (one of the first online QA system operating since 1993), AnswerBus[2], NSIR[3]. The first large-scale evaluation of open-domain QA systems was held by TREC in 1999 and it continues until present [147].

The systems mentioned above are classical QA systems, 'behaving' rather passively and providing users with only one single input modality. Apart from them, there are other QA systems with more interactive capabilities allowing the exchange of much more complex conversational structures, such as clarifications, feedback or query refinement. Additionally, the users have the possibility to lead the system or to allow the system to guide them through the whole dialogue sequence. Here, two different interaction approaches can be distinguished. The first one focuses purely

[1] http://start.csail.mit.edu
[2] http://www.answerbus.com/index.shtml
[3] http://clair.si.umich.edu/clair/NSIR/html/nsir.cgil

on text-based dialogues, resembling a chat program. An example of such systems is the YourQA system [148]. The second is graphically oriented. Examples of such systems in which the dialogues are embedded in a graphical environment, are the IMIX system [149] and the HITIQA system [150].

Another type of interactive QA systems is the chatterbot. Chatterbots are computer programs meant to simulate an intelligent conversation with a human interlocutor. Some famous examples of chatterbots implemented in the '1960s are ELIZA, a system simulating a conversation with a psychologist and SHRDLU, a program simulating the operation of a robot moving objects - cubes, cones, blocks - in a virtual world. Both systems were simple having a relatively rudimentary way of answering questions: while SHRLDU was domain specific, ELIZA's conversational capabilities were unlimited due to straightforward pattern matching rules that detected important words in the interlocutor's input and used them to post 'her' question.
More recent chatterbots available online are UltraHal[4], Elbot[5] and Suzzete[6] - these chatterbots were awarded the Loebner Prize[7] in 2007, 2008 and 2010 for holding the most humanlike conversation.

Currently, there is a growing interest in integrating question answering into Web searches. Companies such as Google, Microsoft and Ask.com have already made the first attempts. Additional aspects, such as question understanding, knowledge representation and reasoning have been explored as useful methods to enhance the QA systems' quality.

5.2 The IMIX system

IMIX (Interactive Multimodal Information eXtraction) is a multimodal interactive QA system that helps users to find information in the medical domain [149]. The system's language is Dutch. IMIX can answer closed-domain, non-factoid [8], "encyclopedic" questions, that is to say, general medical questions that can be answered by selecting suitable text passages from an encyclopedia. Thus, these questions do not require expert knowledge. More difficult questions, such as diagnostic questions or complex medical analysis are not supported by IMIX. The information is intended to cover the needs of users with non-professional medical knowledge, who would probably make use of IMIX only occasionally [149]. No special training is required to interact with the system [151]. A typical IMIX answer is 1-3 sentences long [152]. The length and depth of the details depend on the users' informational needs. The answers are presented in the form of speech, text and pictures.

One particularity of the IMIX system is the ability of its dialogue manager to handle follow-up questions enabling the users to continue their information search by extending their previous questions or the system's answers. Additionally, the users

[4]http://www.ultrahal.com
[5]http://www.elbot.com
[6]http://ai.bluemars.com/chat
[7]http://www.loebner.net/Prizef/loebner-prize.html
[8]Non-factoid questions can include *'why'*, definition and *'how'* question types

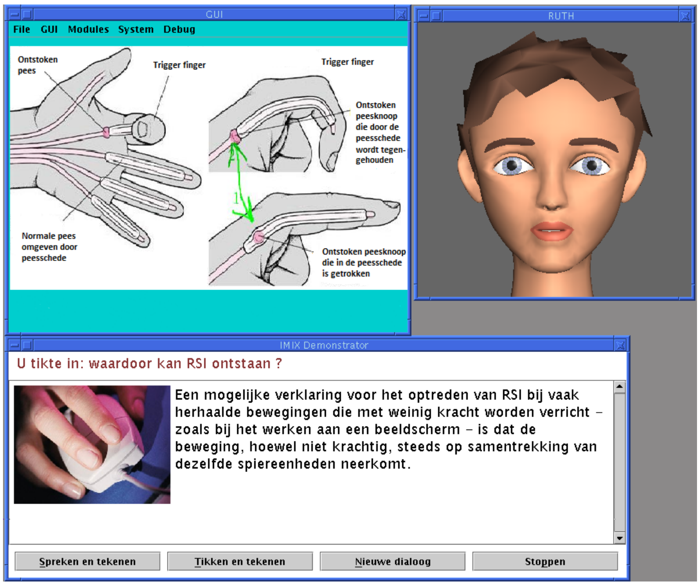

Figure 5.1: Screenshot of the IMIX system. The left-upper part shows an explicative picture. In the right corner the talking head Ruth can be seen. The bottom part shows a window where textual information is presented. The window contains several buttons which allow the user to select the input modality, speech or text, and to signal the start of a new dialogue.

can ask the system for clarification or make corrections on the system's interpretations. As a result the users are integrated into the information searching process in an interactive and cooperative way. On the other side, the system may also ask verification questions and adopt error handling strategies, if mistakes are discovered. Further, the dialogue with IMIX includes greetings, thanks and feedback to the users' questions.

To interact with IMIX users can ask questions by typing or speaking. These both input modes can be combined with pen input to point or to draw [152]. This input option is particularly useful for follow-up questions where the users can encircle unknown details of a picture they want to know more about.

The IMIX system is personified by a virtual agent attached to the interface (see figure 5.1) The agent is a Dutch-speaking version of the talking head Ruth, developed at Rutgers University [9]. The agent's function is symbolic: it represents an information searching assistant who, like a librarian, provides help by retrieving relevant documents from the 'library' [149]. Apart from its symbolism, adding an embodied character to an interface has been shown to have positive effects on the users' perceptions and users' interaction behavior: regardless of how primitive the agent looks users have an increased feeling of presence [153] and more often tend to formulate full sentences instead of keywords. More details on the system architecture and dialogue manager components can be found in [152].

[9] www.cs.rutgers.edu/village/ruth/

5.3 Affordances in conversational interaction with IMIX

Since IMIX 'intentions' are to communicate with users simulating a dialogue between a librarian and a common information seeker it is important that adequate conversational structures are implemented in the system. A way of testing the adequacy of the structures is to prove the correct orientation of each communicative action towards a preceding action. We call this orientation leading to a certain response the affordance of the communicative action. The affordance is a concept used in interaction design which focuses mostly on visual elements, regarding verbal units as simple cognitive support for the graphics. However, multimodal QA systems, such as IMIX are a special case of GUIs in which graphical elements are combined with large text units. The rapport between verbal and graphical elements here is reversed: the text units are supported by graphical elements being the quintessence of the QA interaction. Consequently, we believe that the affordance concept can be used not only to evaluate the design of graphical interfaces, but also the design of interfaces using natural language. To prove whether our assumption is correct we performed a study to detect and evaluate verbal affordances implemented in IMIX by comparing them against those occurring in human conversation.

The concept of affordance has been subject of intense controversy in HCI debates. It was introduced by D. A. Norman [154] but many researchers, such as W.W. Gaver, R.H. Hartson and even Norman himself struggled with several definitions and categorizations whilst trying to make it operational. In the following section we will discuss these theoretical considerations in more detail.

5.3.1 What are affordances?

The concept of affordance was developed by the American psychologist J.J. Gibson [155],[156]. It is a significant part of his ecological theory of direct perception. Gibson defined the term *affordance* as latent action possibilities existing in an environment independent of the individual's ability to perceive them. These action possibilities are in relation with the actor's capabilities of action, being independent of his culture, prior knowledge or expectations. Any substance, any surface, any layout has some affordances with respect to a certain actor. According to his theory, the action possibilities indicated by affordances are perceived visually in a direct way that does not require mental information-processing activity. Hence, the immediate perception of the environment will inevitably lead to a certain action.

Gibson's theory of non-conscious information pick-up was criticized because it fails to explain how actors assign meaning to what they see and decide whether to perform an action. Even if the existence of affordances is independent of the actor's experience and culture, the ability to perceive such affordances may be dependent on them. Therefore, the actor may need to learn first to discriminate the information he or she gets from the environment, in order to perceive it directly [157]. Although Gibson rejected the involvement of mental activities in the process of direct perception, he conceded that even for the most 'basic' affordance perception might need to develop somehow, thus, suggesting that learning could be involved in it.

Other criticisms about Gibson's theory refer to the fact that affordances are defined as being relational, but the nature of the relation between actors and environment are not further discussed. Even though the theory presents some illustrative examples of affordances, it does not provide an analytic way of identifying affordances [158].

Despite all the criticism, Gibson's theory of affordance brought radical changes in the field of perceptual psychology and was successfully adopted as a key concept in other fields, such as cognitive science, robotics, artificial intelligence, design and so on.

In the HCI field the term *affordance* was introduced by Norman in his book *The Psychology of Everyday Things* [154]. Norman adopted Gibson's concept and used it to address design aspects of artifacts, considering technology as part of the environment. He defined affordances as:

> *"perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used [...]. Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning [...]. When affordances are taken of, the user knows what to do just by looking: no picture, label, or instruction needed" [154]*, page 9.

However, even though Norman borrowed the term from Gibson he disagreed with his theory about whether the mind processed or simply 'picked up' information (see [154], [159] and [160]). Consequently, Norman departs from Gibson's theory and considers affordances as perceived properties of objects that may or may not exist. The perception is determined not by the action capabilities of the actor, as Gibson says, but by his mental and perceptual capabilities. Moreover, past knowledge and experience are neatly coupled with affordances in Norman's view [157].

For Norman, the notion of affordance becomes a mixture between actual (shape, material, color) and perceived properties of the object, where the perceived properties are in fact the suggestion of how the object might be used.

The lack of separation between physical properties and perceptual information about their use created a rather ambiguous definition of affordance and generated lively discussions about the meaning of the term in the HCI community.

A substantial contribution to clarify the concept bringing the Gibsonian thinking in the HCI world was made by the interaction designer W. W. Gaver. Similar to Gibson, Gaver argued that affordances do exist independent of their perception, being ontologically but not epistemologically relevant. He defined the affordance concept as the property " *[...] of the environment relevant for action systems*" [161] and proposed a taxonomy where the affordance concept is separated from the perceptual information available about it.

In Gaver's framework, affordances (aff.) and their perceptual information (per. inf.) are defined as entities taking binary values, such as *yes* and *no*. Their combinations result in four types of affordances called: perceptible, hidden, false and correct rejection.

- **Perceptible affordances**: (aff.: *yes*, per. inf.: *yes*) offer a link between perception and action by signalizing a possible action in a visible way

- **Hidden affordances**: (aff.: *yes*, per. inf.: *no*) offer no link between perception and action; actions are possible but there is no signal acknowledging their existence (e.g. a hidden door)

- **False affordances**: (aff: *no*, per. inf.: *yes*) offer a link between perception and a non-existing action possibility; actions are mistakenly signalized as being possible (e.g. a door might appear to afford opening, but it will not afford if it is locked)

- **Correct rejection**: (aff: *no*, per. inf: *yes*) refers to the situation where no action is afforded or signalized (i.e. no affordance)

While conveying the categorization to interaction design, field designers should avoid false and hidden affordances because they are a sign of weak design. False affordances put users on a wrong path while hidden affordances waste resources, because users will probably encounter difficulties on detecting their existence. Instead, designers should concentrate on making affordances perceptible or creating situations where the lack of affordance is correctly rejected.

Another attempt to extend and refine Norman's concept of real and perceived affordance came from R.H. Hartson, [162]. Hartson proposed four complementary types of affordance in the context of interaction design and evaluation: cognitive, physical, sensory and functional affordance.

Cognitive affordance, corresponding to Norman's perceived affordance, is associated with the semantics of the interfaces and refers to design features that help users to know something (e.g. the label of a button indicating what will happen if a user clicks on it).

Physical affordance, corresponding to Norman's real affordance, is associated with characteristics concerning the 'operability' of the interface, and refers to design features that help users to accomplish accurately a physical action in the interface (e.g. the size of a button that is large enough to allow users to click on it).

Sensory affordance is related to 'sense-ability' characteristics of the interface and targets design features that help users to perceive (e.g. see, hear, feel) something (e.g. the font size of a label). Sensory affordance plays a critical supporting role to cognitive and physical affordances.

The last category, functional affordance, addresses design features that help users to accomplish their work (e.g. the internal system ability to sort numbers invoked by a user who clicked the 'sort' button, [162]).

There are several other interesting interpretations and formulations of the affordance concept, but we have discussed only a few of them in details, those we considered as being the most relevant to our analysis.

5.3.2 Practical dimensions of affordances

The study of affordance goes beyond theoretical speculation. Authors like L. Vainio and colleagues [163] validated the affordance concept as part of an interesting empirical study. They showed that during several tests participants could identify objects faster if they were congruent with an observed action prime (e.g. power grasp - power grasp compatible object) rather than if they were incongruent (e.g. power grasp - precision grasp). Their conclusion was that motor knowledge plays an important role in object identification and, consequently, action-related information associated with an (graspable) object is an inseparable element of that object's representation.

In the HCI field the affordance concept has found its practical setting as design model and analysis tool for physical and graphical user interfaces.

J.G. Sheridan and G. Kortuem [164] proposed an affordance-based design model of physical interfaces for ubiquitous environments. They proposed an experimental method to study object affordances, showing how the method can be applied to the design of concrete physical interface artifacts.

L. Ping et al. [165] used the affordance design model as theoretical basis and methodological underpinning to evaluate an e-learning program on mammogram reading.

Hartson [162] explored the relationship between the affordance types associated with usability problems and provided examples and a methodology scheme for practitioners on how to identify affordance issues involved in flawed design cases.

However, the theory of affordance and its practical applications concern mainly the visual perception of environment objects analyzing verbal element only marginally, e.g. only when they are meant to support visual elements (see Hartson's cognitive affordance). Since new media technologies such as interactive information systems (such as QA or dialog systems) are design artifacts that use elaborated conversational structures with preponderate verbal text elements there is a strong need to consider such elements as integral parts of the interaction design. Therefore, we propose a design evaluation that uses the affordance concept to analyze text units and graphical elements.

5.3.3 Affordances in conversational interactions

An important characteristic of conversational interactions in general is the fact that they are deeply anchored in the cultural context of use; that means they are based on conventions and constraints of the socio-cultural environment following a rigorous protocol course.

Contrary to Norman, who argued that the socio-cultural world is placed outside the domain of affordance [154], Gaver emphasized the role of the culture together with other factors such as experience and learning involved in the process of perception of affordances [161]. We also consider these factors are not to be considered affordances, but have an important function in making affordances visible.

A way of detecting these factors and implicitly affordances in conversational interactions is to apply conversational analysis (CA). In CA, conversations can be considered 'environments' where certain types of actions such as gestures, mimics, verbal statements are afforded in certain circumstances - in terms of non-violating coherence principles and cultural constraints. Interlocutors can express their communicative intentions both verbally (through speech) and non-verbally (through gesture and mimic). Analyzing the organization of each conversational sequence we can determine what kind of action possibilities (affordances) have the participants in a certain moment of the conversation. We consider these affordances from a pragmatic perspective, i.e. action possibilities oriented to achieve a certain goal.

Conversations are by nature interactive and follow a relatively strict turn-based protocol. In this study we address only the case of closed-domain question-answer interactions, since the system that we will analyze deals with this type of interaction.

In general, question-answer conversations are fully structured into adjacent pairs, meaning that all exchanged turns are functionally related to each other in such a manner that the first turn requires a certain type (or range of types) of the second turn [25]. The adjacent pairs are grouped in three separate categories, corresponding to a conversation initialization, termination (both containing greeting-greeting pairs) and body sequence (containing question-answer pairs). These three categories together form what we call a conversation protocol.

Conversation initialization and termination

Starting and ending a conversation are levels of phatic communication with social functions. They are responsible for establishing rapport or quitting the interaction 'circle' in a polite way.

A conversation usually starts with a signal showing the readiness to engage in a conversation. Such signals are salutation forms, self-presentation (if the interlocutors have not met before), non-verbal gestures (handshake, hugging, kissing, hand waving, gazing), mimicry (smiling) and changing the corporal position towards the interlocutor.

A similar protocol for ending the conversation includes farewells and thanks exchanging, waving, hugs, kisses, glancing away, re-orienting body posture away from interlocutor and so on.

Each performed action affords in principle a similar one in return: a greeting, a smile, a self-introduction affords symmetrical responses from the interlocutor. However, the realization of conversation initialization might differ across culture taking into account participants' gender, age, social position and degree of acquaintance. For example, in Western cultures a stretched out hand will afford handshaking; in Muslim countries such a gesture will afford handshaking usually if both interlocutors are men. Women instead will press their hand upon the chest to signalize salutation response at the same time avoiding direct contact with the opposite gender.

Conversation body

The conversation body contains the essential part of the interaction, namely the information exchange (also called informational communication).

The information exchange may start with a short explanation of the intended nature of the conversation-to-be preceding (the first question-answer exchange). At this point a common ground (a set of propositions that make up the contextual background for the utterances to follow) can be established.

From a pragmatic point of view a question may afford the following responses: a matching answer, acknowledgment of ignorance, a suggestion for asking someone else (re-routing), intermediary questions to clarify a previous question, postponement, refusal to provide an answer, feedback showing that the question was understood or a request for time to process the question, and so on.

In the case of miscommunication, repair strategies occur in the form of explanatory adjacent question-answer pairs.

The turns can be accompanied by non-verbal cues, such as gestures and mimicry used to emphasize the content. For example, gaze signalizes attention and readiness for interaction, raised eyebrows shows surprise and smile acknowledges agreement.

Question and answer pairs must respect the coherence principle by being semantically and meaningfully related to each other. In order to achieve coherent information exchange, syntactical features such as anaphoric, cataphoric and deictic elements may be used. Additionally, logical tense structure, as well as presuppositions and implications connected to general world knowledge are deployed to connect coherently answers to questions [166].

In common practice, interlocutors do not perform their utterances at the same time. Speakers usually take turns to talk. Overlapping and simultaneous talk is generally seen in Western cultures as unpleasant. The turn-taking usually occurs at the utterance ends, often signalized by silence.

Interruption might be allowed if one of the interlocutors has signalized verbally or by gesture the wish to take the turn [167].

5.3.4 Methods

Even though conversational interactions with multimodal systems differ in many aspects from their human counterpart, they generally follow the same conversation protocol consisting of initialization, body sequence and termination. This similarity is intentionally simulated by designers in order to increase the system's ease-of-use and to make users' answers predictable.

As theoretical framework we adopted Gaver's taxonomy to identify affordance values and Hartson's scheme to establish affordance types. The analysis followed the conversational protocol steps described above.

The test was carried out using the cognitive walkthrough (CW) method (presented in section 3.4.3).

Prior issues to the test

Before starting the test the evaluators were given sheets of paper with preliminary information about the test goal, a short description of the term *affordance*, a detailed explanation of the human conversation protocol and a general description of the IMIX system.

Afterwards, the evaluators received the scenarios containing specific tasks to accomplish, a list of correct actions required to complete each of these tasks and a separate questionnaire for each scenario. Three scenarios covering the conversational structures implemented in the IMIX system were developed. We tried to design the scenarios to be as pleasant and humorous as possible in order to achieve an enjoyable interaction. Each scenario focuses on a specific task.

In the first scenario, evaluators were asked to put a single question and analyze the corresponding conversation protocol. The scenario identifies the situation of a naive user with little medical expertise who uses IMIX to find out what repetitive strain injury (RSI) means.

In the second scenario the evaluators had to concentrate on the special case of follow-up questions using drawing and/or typing options. The user profile of this scenario corresponds to a subject with search engine expertise and interests in the medical domain. He or she uses IMIX to find information related to liver functions.

The third scenario addresses repair and meta-communication strategies when the answer to a question is not found. The user profile addresses an expert user who uses IMIX for entertainment purposes. He or she is seeking for information about the SARS virus.

The evaluators were asked to keep in mind the aim of the test: to look at the way the user is invited to interact with the system and NOT at the answer quality he/she might get back. They also had the possibility to repeat a scenario several times if they wished to.

Questionnaire design

For each scenario a separate questionnaire was developed. The questionnaires were designed in accordance with the cognitive walkthrough (CW) method [117]. However, the questions were adapted to fit the special case of verbal interactions and were grouped in three units, each one corresponding to a separate conversation protocol category (see table 5.2). The purpose of the questionnaires was to detect affordances of the conversational sequences implemented in the protocol. The structure of the questions was similar for each unit. Evaluators had first to detect elements signalizing the current protocol. Then, they had to anticipate users' reaction given a certain conversational sequence. Furthermore, the evaluators had to determine whether the users' responses were acknowledged by the system and how they would perceive this feedback. Eventually each question unit ended with a question about potential violations of the conversation protocol. This last question was meant to catch issues that might have 'escaped' the evaluator's observation.

Classical CW	Adapted CW
1. Will the users be trying to produce whatever effect the action has?	1.0 What elements are used to signalize [*a certain action*]? 1.1 Does the word [..] suggest [*a certain action*]? 1.2 What kind of statements affords the question [..]?
2. Will users be able to notice that the correct action is available?	2.0 Will the users see there is [*a certain*] option ? 2.1 Will the users understand this signal as an invitation to [*do a certain action*]?
3. Once users find the correct action will they know that it is the right one?	3.0 Will the users know how to use [*a certain*] option?
4. After the action is taken, will users understand the feedback they get?	4.0 Will there be a feedback to acknowledge the action performed by the users? 4.1 Will the users understand these kind of feed-back?
	5.0 Does the conversational protocol get in any way violated?

Table 5.2: *Classical CW versus Adapted CW*

Pilot study

Before starting the experimental run, a first pilot study with one expert evaluator was carried out. From the pilot study three main observations were gathered:

1) The relatively high degree of difficulty of the question demanded the presence of experienced evaluators having some affinity with the CW method.

2) Typical CW questions like *"Will the user notice the conversational starting as the correct action available?"* are too general. Precise formulations similar to *"Will the user understand this signal as an invitation to start a conversation?"* seemed to be more appropriated even if the questionnaire size increased, because each signal requires a separate question.

3) The answers to the CW questions are often not straightforward, because they require some deliberation time. Hence, it seemed wise to record the testing session. In this way a considerable amount of time could be saved and no observations were lost.

5.3.5 Results

The test was completed by five evaluators with design expertise. All were recruited from our department. All evaluators, except one, were novices at using the system. The results of their evaluations are summarized below, following the conversational protocol categorization.

Conversation initialization

The conversation initialization implemented in IMIX does not afford symmetrical response. The conversation's start is signalized by a textual welcome message, a short system presentation, a 'start' button and a talking head emerging from the background gazing and rising eyebrows. At this point the only afforded action is the pressing of the 'start' button to begin the 'conversation'. No other actions, such as greetings or salutation gestures, are afforded, even though according to the conversation protocol a greeting affords another greeting. The occurrence of a signal (greeting message) combined with the lack of an adequate response (no greeting in return) indicates the presence of a false cognitive affordance.

The talking head appearance is not a very convincing invitation to talk, even though it its blinking eyes indicate a waiting behavior. One of the evaluators argued that the presence of speech, e.g. a welcome message read by the talking head would increase the users' feeling of being involved in a conversation.

An adjustment in the head mimic would be beneficial as well because gazing behavior combined with smiling is a more appropriate way to start a conversation.

The short system presentation was criticized as being too technical, especially since less experienced users would have difficulties in understanding the meaning of having a "multimodal dialog" with the system.
The text color should be uniform. One criticism addressed the presence of colored words in the text message, a fact that could mislead Internet experienced users to click on it, because it is commonly associated with websites with embedded links. This would be a false sensory affordance.

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press 'start' button
Sensory & Physical	False	Click on highlighted words
Cognitive	False	Give symmetrical response

Table 5.3: *Affordances in conversation initialization*

Table 5.3 shows the affordance types and values encountered in the conversation initialization.

Conversation body

The conversation body includes single question-answer sequences, follow-up questions, meta-communication and repair strategies.

Single question-answer sequences

The conversation body begins once the 'start' button is pressed. The users arrive on a new screen where they receive some brief instructions on how to interact with the system.

At this stage of the conversation the users have to choose between two input options: speech or typing. All evaluators agreed that input selection modalities seem to be afforded in a proper manner: the buttons are intuitively labeled and it was estimated that none of the user categories would have difficulties in selecting an input option. A suggestion was made to use additionally explanatory icons like a pen for the typing option and a microphone for the speech option.

There was criticism for the presence at the same level of two other buttons: one for the stop option and the other one for the new dialog. The 'stop' button should be placed in a corner -in order to be congruent with the typical design of closing buttons while the 'new dialog' button should be removed because it is functionless at that point and indicates a false physical affordance.

A further remark was to adapt the head position towards the typing input field while users are typing a question in order to increase the interactive feeling and to give a certain feedback. Due to the lack of adequate mimic reactions and synchronization with the current conversation stage, the talking head gives the impression that it does not belong to the system.

Selecting the option 'typing' causes an input field to appear on the same window. The input field affords sentence-like questions as well as keywords. All, except one, evaluators considered that the full sentence capability would not be easily perceived by more experienced users. This is because they would probably associate the system's functionality with that of a typical search engine and consequently would use keywords. The presence of a relatively extended input field is not a clear indication of the expected input and, if sentence-like input is desired, a short how-to-ask example should be provided. It can be concluded that the input field has hidden physical affordances.

The input field is introduced by the question *"What would you like to ask or say?"*. The designer's intention was to let users know the system is able to handle different types of statements, such as full-sentence questions, greetings or even transition formulations (*"OK"* or *"thank you"*). However, being rather too open, the question suggests it can deal with any kind of statement, which is a false cognitive affordance.

On the other hand, all evaluators, except one, concluded that, probably experienced users would not be aware of what exactly they can utter: greetings and transition statements affordances remain hidden because nothing indicates their possible usages.

Affordance type	Affordance value	Action
Cognitive	Perceptible	Chose input options
Physical	Perceptible	Type in the input field
Physical & Cognitive	False	Click 'new dialog' button (first dialog screen)
Cognitive	False	Perform whatever question
Physical	Hidden	Put sentence-like question
Cognitive	Hidden	Perform greeting
Cognitive	Hidden	Perform transition statement
Physical & Cognitive	Hidden	Press 'new dialog' button to type
Physical & Cognitive	Hidden	Press 'follow-up question' button to type

Table 5.4: *Affordances in single question-answer sequences*

We continue the analysis considering the case where a naive user will enter a greeting. The system will logically respond repeating the same question *("Hello! What would you like to ask or say?")*, but will not indicate how to continue the dialog as the input field disappears. So far a direct answer is not afforded. The users need to press the button for either new-dialog or follow-up question in order to get to the input field to type in, a fact that complicates the conversational flow.

None of the labeled buttons suits the actual conversational situation semantically. The 'new dialog' button should be used in situations where a dialog session re-initialization is wanted, while the 'follow-up' button refers to situations where users require more detailed information in the medical answer. Therefore, it can be concluded that both buttons afford a hidden action in this conversational sequence, namely to allow users to get back to the typing field.

Table 5.4 presents a summary of the affordance types and values found in the question-answer sequences.

Follow-up questions
After receiving the first answer the users have the option to continue the information exchange on the same topic by selecting a 'follow-up question' button. The button is labeled with text, indicating that users can type or point at something in the answer. However, the pointing option is not intuitive and has not a specific usage indication. Furthermore, not only pointing but also drawing is supported, a fact that the label does not specify. All evaluators agreed on the fact that none of the user categories would know what the option does and how to use it. Moreover, it is not clear which advantages it has compared to the typing option. Therefore, we identify here a hidden physical affordance. It is also not very clear the way a 'drawn' follow-up question is entered for further processing. This is because the 'OK' button, located in the proximity of the input field, can also be used for this purpose. The evaluators concluded: the button has hidden physical and cognitive affordances. There is a feedback statement to acknowledge the waiting pause and the users' query, but the statement does not specifically address follow-up questions.

Affordance type	Affordance value	Action
Physical	Perceptible	Type in the input field
Physical	Hidden	Use the mouse to "drawn" a question
Physical & Cognitive	Hidden	Use the 'ok' button to enter a "drawn" question

Table 5.5: *Affordances in follow-up questions*

In table 5.5 the affordance types and values found in follow-up questions are shown.

Meta-communication and repair strategies
When the answer to a question is not found, the system displays a message requesting rephrasing. The function of the rephrasing request should help users to become more successful in finding the desired information.

Affordance type	Affordance value	Action
Cognitive	Perceptible	Rephrase question
Physical & Cognitive	Hidden	Press 'new dialog' button to rephrase
Physical & Cognitive	Hidden	'Follow-up question' button to rephrase

Table 5.6: *Affordances in meta-communication*

All evaluators found the request not supportive at all. According to their estimation, even expert users would experience problems rephrasing their questions.

After the rephrasing request, users can choose between the follow-up question (in the form of typing or pointing) or the new dialog option in order to get cumbersomely back to the typing field. Both options were considered inadequate for this particular stage of the conversation. Just as in the follow-up paragraph, these two buttons indicate the presence of hidden physical and cognitive affordances.

Table 5.6 shows a summary of affordance types and values in meta-communication.

Conversation termination

The interaction can be interrupted by clicking the 'stop' button. A real conversational termination is not afforded. Users do not have the possibility to express verbally an intention to leave the conversation, since no typing field was designed for this stage of the conversation. They could click on the 'new dialog' button and type a farewell greeting because the system affords such statements. However, this option seems rather counterintuitive since it is unlikely that anybody will probably think to start a new dialog when in fact he/she wishes to stop it. Furthermore, even if the system replies logically to the farewell greeting it does not allow a verbal termination of the conversation.

There is no feedback to acknowledge the end of the conversation and users get the general impression of a system crash by clicking the 'stop' button. We certainly face a conversation protocol violation.

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press the 'stop' button
Cognitive	False	No symmetrical response

Table 5.7: *Affordances in conversation termination*

Table 5.7 shows a summary of affordance types and values encountered in conversation termination.

5.3.6 Discussion

Extrapolating the affordance definition given by Gaver, in this study we considered interactive information systems as artificial environments where verbal and graphical elements are artifacts leading users to perform certain actions. Therefore, we proposed a design evaluation in which not only graphical, but also verbal elements, can be analyzed under the framework of the affordance concept. The analysis was performed by comparing IMIX conversational structures with those occurring in natural human conversations. The conversational protocol was indicative of which verbal statements are afforded in a particular situation.

Our results revealed several inefficient structures that were identified by analyzing affordances of conversational structures. The conversation initialization and termination implemented in IMIX do not afford a symmetrical response from the user perturbing the natural flow of the dialog. The system's question formulations are too open, a fact that might generate false expectations or disorientation. The labeling of buttons should reflect the actions induced by the buttons. A question should automatically generate a response environment avoiding unnecessary pressing of additional buttons.

The study of affordances also showed unnecessary functionalities that might be removed or adapted in order to become useful. For example, the presence of buttons leading to certain actions should be in accordance with the conversational sequence they are designed for: it makes no sense to start a 'new dialog' when no other dialog had been started before. The affordance of certain conversational structures, such greetings in the middle of an interaction shows a cooperative behavior. However, it is unlikely that someone would use greetings at that particular conversational stage.

Special features like pointing or drawing on a virtual surface should be introduced to users. It is rather unlikely that someone would use an unfamiliar option to ask questions when he/she has more natural choices, like typing or speaking.

The affordance analysis also provided important observations about the system's ease-of-use. Users may not understand the system's description, as it seems to be too technical, may not be aware of its full sentence capabilities, may not know whether other transition statements are allowed, may experience difficulties using the pointing/drawing option or rephrasing their questions and may probably feel annoyed when they expect to be able to type a question and no input field is provided.

Finally, the affordance analysis of verbal elements and its associated human conversational protocol proved to be beneficial for the evaluation and future enhancement of the interface, confirming the fact that many problems associated with natural language-based interaction originate from the lack of deeper understanding of communicative structure: most of the false and hidden affordances identified were cognitive in nature (6 pure cognitive and 6 physical-cognitive out of a total of 14).

We conclude that understanding affordances of verbal and graphical elements and being aware of their roles in conversational interaction design can help practitioners in diagnosing usability problems from an early stage of development. This is because the affordance analysis using CW methods provides a useful and informational rich perspective for qualitative evaluations of prototypes, without requiring costly user studies.

5.4 The gender-ambiguous agent Ruth

While we were performing the interviews from the first study, one of our participants made a very interesting remark concerning the virtual agent Ruth: *"Ah .. what is this Ruth actually? Is it a guy or girl? I am bit confused ..."*. His confusion came from the fact that Ruth's look could be classified as either a young male or a female with slight masculine traits (see figure 5.2). Additionally, the agent's low pitched feminine voice and the name 'Ruth' which designates a female person, increased the ambiguity, letting the participant clueless about the agent's gender. This remark let us wondering, whether the gender-ambiguous appearance of an agent is important in conversations, whether people feel disturbed conversing with gender ambiguous characters and how much this would eventually affect the system evaluation.

It is well-known that physical characteristics, such as age, gender and ethnicity are important cues in human social perception, cognition and behavior[168]. They represent a basic 'business card' that tells people how to approach a potential

Figure 5.2: A 'portrait' of Ruth

conversational partner.

Research shows that humans prefer to engage in conversation with those whose physical appearance can be labeled consistently [15]. The reason is the human tendency to simplify the interlocutor's representation by framing her into pre-defined categories (e.g. young, female, Asian) [169]. This framing lightens the cognitive load and gives a secure feeling of dealing with predictable situations [170].

Among all salient visual cues related to physical appearance, gender seems to be of fundamental importance, being one of the first visual information people exchange in an encounter. The explanation goes beyond the cognitive load lightening and relates to our evolutionary history where gender related information assured the correct orientation toward a potential mating candidate. Since the decoding of such information has powerful impact in social interactions we believe that its lack would be perceived as unpleasant. In other words, we assume humans would prefer to interact with those whose gender they can label consistently and they would maintain this preference even when they interact with artificial entities such as embodied conversational agents (ECAs).

5.4.1 Virtual gender issues in HCI: am I a guy or a girl?

The virtual gender issue has become an important topic to the HCI community, since many computer media systems have started to use representative human avatars. Previous research has demonstrated that humans treat computers as if they were social actors, even though they do not exhibit anthropomorphic traits [8]. By adding a face and embodiment to an interface the social relationship between user and computer becomes even more explicit. Clothing, facial expression, hairstyle, gender and age cues displayed by an agent bring the rich and complex world of human social interactions into the interface [153].

Several researchers have studied the effects of 'virtual' gender on the way people perceive conversational agents and build relationships with them.

Zimmerman et al. [171] concluded that people prefer agents displaying gender stereotypes conforming to specific roles - female agents were preferred for tasks traditionally undertaken by women (librarian, matchmaker), male agents for tasks undertaken by men (fitness trainer). They also found that men prefer embodied agents more than women do, and that female agents were preferred over male agents, by both male and female users.

Baylor et al. [172] investigated how the attitudes of female undergraduates towards engineering were influenced by agents' age, gender, and 'coolness'. They found that, after interacting with a female agent, test subjects reported more positive stereotypes of engineers. But after interacting with a male agent test subjects regarded engineering as being more useful.

Catrambone et al. [173] suggested that male and female test subjects might have different ways to personify agents. Their study showed that 54% of the female participants used a personal pronoun (he/she) to refer to an agent, while only 13% of the male participants did the same.

De Angeli and Brahnam [169] found the gender of the virtual embodiment impacts the incidence of sex talk: agents that clearly signalized their genders (female, male characters) were more prone to verbal abuse than those that did not do so (robot character).

With the exception of [169] there are no other studies known to us investigating human perceptions of gender-ambiguous characters. This is surprising, because computer applications often display avatars or agents whose physical appearance does not point to any particular gender. This look is intentionally created by designers with the purpose that both male and female users could relate to the character. Therefore, in this study we propose an experiment investigating the impact of agents' gender-ambiguous versus gender-marked appearance (voice and look) on the perceived interaction quality of a multimodal question answering (QA) system.

5.4.2 General experiment design

First, based on the agent Ruth we developed additional agents by adding simple but very explicit feminine or masculine characteristics. The modifications were kept to a minimum in order to make the comparison between the agents sustainable, that means factors such as beauty, facial symmetry or hair color should not influence subjects' preferences for a particular agent.

The female versions wore earrings, have narrow eyebrows, a lighter skin color and a feminine hair style. Each head had a different hair color while the masculine heads have a much darker skin color, mustache and/or beard and short hair. One of the heads is blond, the other two have brown hair (see figure 5.3).

5.4.3 Physical look and gender - preliminary study

To determine whether Ruth's look is indeed perceived as gender ambiguous we conducted a short preliminary study in which the agents' images were printed on paper and were shown to 48 test subjects (24 males and 24 females). The participants were aged between 16 and 73 years, originating from 10 different countries across Europe, Latin America, Asia and Middle East. They were asked to rate the agents' gender identity, as well as their degree of feminity or masculinity on a 5 point scale. No direct question addressed the gender ambiguity, in order to avoid priming effects.

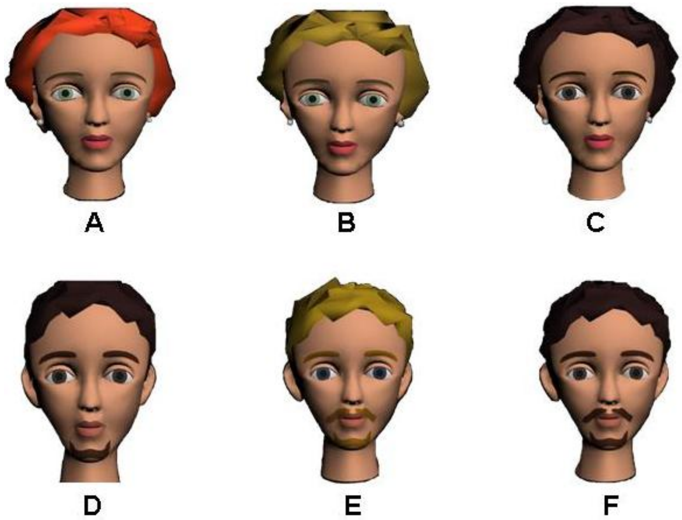

Figure 5.3: Female and male versions of Ruth

We used letter identifiers (A, B, C, etc.) in order to exclude any semantic association test subjects might unconsciously make between name and look.

To remove any potential bias from being exposed to one image category (gender ambiguity) before the other (gender explicitness) test subjects were organized into two equal groups: one half had to rate the gender marked images first and then the image representing Ruth (test order 1: TO1). The other group did the opposite (test order 2: TO2).

Apart from determining people's perception of Ruth's gender look, this preliminary study was also aimed to determine which of the gender marked agents best represent the feminity and masculinity concept as opposed to gender ambiguity.

5.4.4 Results

The results confirmed Ruth's gender ambiguity: even thought there was a slight trend to consider the agent a 'male' the answers of both test groups were given randomly, i.e. the test was not significant (χ^2(1, N=48)=3.00, p>.050). However, the trend was not particularly strong: measuring the degree of masculinity Ruth was considered 'less masculine' (M=3.43) than 'so-so' (M=3.00), t(46)=3.42, p<.001. For the feminity degree there was no statistically significant difference between the middle value 'so-so' (M=3.00) and the mean value of the answers chosen by the test subjects (M=3.07), t(44)=3.50, p>.05. This suggests again that there was no trend toward a particular gender direction.

No statistically significant difference could be found on how male and female test subjects decided on Ruth's gender identity. However, when the test order in which the participants rated the agents was taken into account the subjects' gender became statistically significant: females found Ruth more feminine in TO1 (M=3.00) compared to TO2 (M=3.83) and more masculine in TO2 (M=2.86) compared to TO1

(M=3.60); males found Ruth more feminine in TO2 (M=2.69) compared to TO1 (M=3.33), and more masculine TO1 (M=3.00) compared to TO2 (M=3.75). Thus, male and female subjects reacted oppositely when confronted with a different test order. The interaction effect between order and gender was confirmed to be statistically significant by a MANOVA test (Ruth's perception of feminity, $F(1,41)=7.71$, $p<.010$ and masculinity, $F(1,43)=9.11$, $p<.010$).

Regarding the most feminine agent a slight trend towards the blond (B) and brunette (C) agents could be observed. However, the trend was statistically not significant. No negative values ('less' or 'no' feminine) were found. Concerning the most masculine agent a significant difference ($\chi^2(2, N=48)=41.63$, $p<.001$) was found in the favor of the brown hair agent displaying beard and mustache (F).

5.4.5 Discussion

Our results showed that female and male test subjects had opposite perceptions on Ruth's gender, depending on the order they had to rate it (before or after seeing the gender marked agents). However, test order and gender, as influence parameters taken separately, were not statistically significant. Only when combined together in a single significance test were they statistically meaningful concerning their perception of Ruth's gender. The strong interaction effect between test order and subjects' gender supports the idea that females and males have quite different perceptions on masculine and feminine traits.

5.4.6 Voice, physical look and gender - main study

In our main study only three heads were used: Anna -the brunette agent 'C', Bart -the most masculine agent 'F' and Ruth. There was no significant difference between the feminity degrees of feminine heads, so we chose the brunette one in order to avoid too strong hair color contrasts between the agents (see figure 5.4).

Figure 5.4: Female and male versions of Ruth

5.4.7 Experimental set-up

In the current experimental setting the subjects used only the speech modality to interact with the system, they did not use the keyboard or the mouse. We chose the speech input option because it increases the naturalness of the interaction.

Pitch and frequencies of the synthesized TTS voices were adapted to fit the agents' gender. For Ruth, a female voice with grave pitches was chosen in order to increase the ambiguity of the gender perception.

To ensure homogeneity between the trials we applied the Wizard-of-Oz technique and replaced the speech recognition module by a wizard. However, we introduced one simulated speech recognition error in each evaluation session to avoid the impression that the system was controlled by a human operator. A list with all questions was prepared in advance for the wizard. During the interaction the wizard could quickly copy-paste the questions in the QA interface, minimizing the risk of delays or input mistakes that could have been caused by manual typing. The test subjects were informed that they were interacting with three differently configured systems (i.e. using different search algorithms), whereas they were interacting with only one system controlled by the wizard. Each system was represented by a different gender-marked agent.

The test subjects received a set of three scenarios per evaluation session and accomplished a total of 9 (3 x 3) trials. The scenario sets (A, B, C) were constructed in a similar manner to provide equal conditions in terms of answer quality and time spent to complete the tasks (see table 5.8).

Scenario Nr.	**Set** A	**Set B**	**Set C**
1.	1.What is the heart? 2.What does the picture represent?	1.What is the lung? 2.What does the picture represent?	1.What is the eye? 2.What does the picture represent?
2.	1.What is hay fever? 2.What are the symptoms? 3.What causes hay fever? 4.How can hay fever be healed?	1.What is RSI? 2.What has RSI with stress to do? 3.What are the symptoms? 4.What represents the picture?	1.What is asthma? 2.What are the symptoms of asthma? 3.How can asthma be cured? 4.What does the picture represent?
3.	1.What is the DNA? 2.What does the picture represent? 3.What is a chromosome? 4.What does the picture represent?	1.What is malaria? 2.What causes malaria? 3.What are the symptoms? 4.How can malaria be cured?	1.What is the sleeping sickness? 2.What are the symptoms? 3.How can the sleeping sickness be healed? 4.What does the picture represent?

Table 5.8: *Scenario sets*

Nevertheless, to overcome possible scenario weakness leading to a less positive system assessment we rotated the agents assigning them to a different scenario set each time.

We also randomized the order in which the participants interacted with the agents to exclude any potential bias that might have arisen, due to being exposed to one particular agent before the others.

Evaluation design

To determine whether test subjects perceived the interaction quality with the agents differently, we deployed two complementary evaluation methods: one quantitative short questionnaire to be filled in after each evaluation session and one in-depth qualitative interview conducted after the entire experiment.

A. Quantitative questionnaire

The purpose of the quantitative questionnaire was to give an idea of the preferences trend of the participants, i.e. no significant statistical results were meant to be achieved. Our short survey was inspired by the SASSI questionnaire [113] and the ITU [106] questionnaire (both questionnaires are discussed in section 3.4.2). For each SASSI dimension, except for the speed and the habitability, we chose one to three variables. We excluded the speed dimension because, according to our experiment settings, we were not expecting perceptible differences between the systems. Nevertheless, we included these two dimensions in the qualitative interview. From the ITU questionnaire we took two questions which were not mentioned in the SASSI questionnaire, but were relevant for our study. These were questions regarding the overall interaction quality and the response clarity. Additionally, we added a question regarding the users' overall feeling of comfort derived from our taxonomy presented in section 3.3.2.

Since the experiment was carried out within the limitation of a pilot study we used only 10 questions for the questionnaire and re-arranged the variables in two factor subscales. The first subscale measured interaction-related features and contained four variables: interaction ease, response clarity, system flexibility and system accuracy. The second scale referred to the interaction effects on users' mood. We called this subscale 'user feelings'. The subscale contains five variables: metal effort, tenseness, degree of confidence comfort and enjoyment. The survey ended with a question regarding the overall interaction quality (see table 5.9). All variables were rated on a 20-point scale to assure fine grained results. The questionnaire can be found in appendix C.

Interaction features	User feelings	Interaction quality
1. interaction ease 2. response clarity 3. system flexibility 4. system accuracy	1. mental effort 2. tenseness 3. degree of confidence 4. comfort 5. enjoyment	1. overall interaction quality

Table 5.9: *Structure of the quantitative questionnaire*

B. Qualitative interview

In the qualitative interview subjects were asked about their system preferences on several functional (pragmatic) and non-functional (hedonic) interaction aspects involved in the experiment (see table 5.10). Under functional aspects we included questions about system transparency (i.e. referred as habitability in SASSI questionnaire), response accuracy, response speed, response quality and feedback strategies. In the non-functional category we asked questions related to agent and interface aesthetics, voice quality, content formulation and trustworthiness. The interview ended with a question about the overall system preference. The interview is attached in appendix D.

Functional aspects	Non-functional aspects	System preference
1. system transparency 2. response accuracy 3. response speed 4. response quality 5. feedback strategies	1. agent aesthetics 2. interface aesthetics 3. voice quality 4. content formulation 5. trustworthiness	.overall interaction quality

Table 5.10: *Structure of the qualitative interview*

5.4.8 Results

The experiments and evaluation sessions lasted in total one hour. Additionally, the subjects were interviewed for another 25 minutes.

Eight test persons participated in the study. Half of them were male, half were female. We chose a small sample of participants in purpose, since we wanted to test first whether our study was worth pursuing. Most of the subjects belonged to the age group 26-30, except for two participants, whose ages were between 45-54 years. All participants, except for one, had a technical background and were knowledgeable about QA systems. Half of the subjects had even used a QA system in the past.

A. Quantitative questionnaire

The reliability analysis performed on our subscale shows acceptable internal consistencies (Cronbach's alpha values are between 0.6-0.8, as shown in table 5.11).

Factor category	Anna	Bart	Ruth
Interaction features	.816	.662	.756
User feelings	.761	.630	.841

Table 5.11: *Cronbach's alpha values*

The system represented by Ruth had the lowest mean average ratings for all factor categories (see table 5.12). On the other hand, the system represented by Anna was systematically rated better.

To check the significance level of the mean differences, we performed repeated ANOVA measurements followed by a paired t-test with Bonferroni correction adjustments for multiple comparisons. The ANOVA measurements indicated statistically significant differences between the systems $F(73.29,1.0)=4.35$, $p<.050$ (the degrees of freedom were corrected using Huynh-Feldt estimates). The paired t-test revealed significant differences between the Anna system and the systems represented by Bart (t=2.38, p<.025) and Ruth (t=2.46, p<.025) on the factor user feelings. For the factors interaction features, overall enjoyment and interaction quality no statistically significant differences were found. Furthermore, no significant differences were found between the evaluation scores given by female subjects and those given by male subjects. Interestingly, the agent Ruth achieved constantly lower mean average scores as compared with the other two gender marked agents.

Factor category	Anna	Bart	Ruth
Interaction features	15.48	14.85	14.23
User feelings	16.90	15.71	15.50
Overall enjoyment	15.12	14.62	12.87
Interaction quality	14.75	14.00	13.75
TOTAL	**15.71**	**14.85**	**14.49**

Table 5.12: *Mean averages for Anna, Bart and Ruth on all factor categories*

B. Qualitative interview

During the qualitative interview test subjects had the opportunity to talk openly about their experience with the IMIX system and its agents.

Regarding the system transparency, feedback strategies and general interface aesthetics no differences were found between the systems. Most of the test subjects (six persons) considered the systems relatively transparent. The interaction style appeared to be intuitive and subjects knew right from the beginning how to handle the systems. The feedback was considered to be sufficient by the majority (seven persons). Only one person complained about the reduced feedback visibility, because some feedback statements were placed on the interface most top corner and could be easily overlooked. The interface, which was identical during all evaluation sessions, except for the agent's look, was considered to be acceptable (six persons), but relatively simple and containing only basic features.

Regarding all remaining aspects the test subjects did find differences between the systems. When asked about their system preference more than half of the test subjects (5 persons) chose the Anna system. Among the reasons for preferring Anna were the response accuracy (three persons), response quality (four persons) and, surprisingly, response speed (two persons). Anna's answers appeared to be more nicely formulated, more informative and more relevant to subjects' queries (four persons). The agent Anna was also considered as to have the pleasantest look (seven persons) and a much nicer voice (5 persons). Her look appears to be ”more professional”, like a ”nurse” or a ”teacher”. Anna left the impression she was more knowledgeable, more trustworthy (seven persons) and more appropriate as a ”medical expert” (seven persons), as compared with the other two agents.

Bart appeared less trustworthy because of his beard, while Ruth appeared to be too young and quite ”dull”. The Bart system was preferred by two test subjects in terms of answer quality (two persons), response speed (two persons), and agent look (one person).

Only one single subject showed a general preference for the Ruth system but immediately added that Anna had a nicer face. Interestingly, even in situations where the content delivered by the Ruth system was, according to the participant's own statements, better - the participant still declared he would prefer the Anna system blaming the scenario setting for making Anna unable to give the desired answer (!).

5.4.9 Discussion

Despite the small number of test subjects our results are astonishing: even if test subjects interacted with the same system they felt significantly more comfortable, more confident and less tense with the Anna system, enjoying the interaction much more, as compared with the other systems. The Anna system also appeared to perform better than Bart or Ruth, yet, this result did not prove statistically significance. During the qualitative interviews most of the test subjects confirmed their preference for the Anna system.

In general, test subjects seemed to prefer the gender-marked (i.e. voice and look consistent) agents Anna and Bart over the gender-ambiguous (i.e. voice and look inconsistent) agent Ruth, while Anna, got most of the preference 'votes'. Thus, our study shows encouraging results for our hypothesis concerning the human preference for interacting with consistently gender labeled entities. On the other side, we are aware that creating three different agents we created not only three different gender representation but also three different characters. Hence, the particular details of the three agents evaluated may influence subtly many dimensions of user attitude towards the agent beyond the gender. Since gender effects cannot be studied in isolation from the face or voice of a character these influences are inherent. We believe by keeping the facial modification to a minimum, to have reduced the unavoidable impact of these influences on our manipulations.

5.5 Summary

In this chapter we presented two experiments conducted with the QA system IMIX and its virtual agent Ruth.

In the first study we evaluated the design of conversational structures implemented in the IMIX system using the concept of affordance and human communicative interaction patterns grouped under the concept of conversational protocol.

The term affordance was developed by the American psychologist J.J. Gibson refers to the latent action possibilities existing in an environment independent of the individual's ability to perceive them.

Extrapolating the affordance definition given by Gaver we considered interactive information systems as artificial environments where verbal and graphical elements are artifacts leading users to perform certain actions. Therefore, we proposed a design evaluation in which not only graphical but also verbal elements can be analyzed under the framework of the affordance concept. The analysis was performed by comparing IMIX conversational structures with those occurring in natural human conversations. The conversational protocol was indicative of which verbal statements are afforded in a particular situation.

Results showed that studying affordances in conversational interaction conversational protocol helps to detect inefficient constructions leading to disruptions in the dialog flow, to spot unnecessary functions and provides important insights on a system's ease-of-use. As such the analysis proved to be beneficial for the evaluation and future enhancement of the interface.

In the second study we investigated the effects of the agent's gender appearance (voice in combination with look) on the way users' perceive and judge the interaction quality with the IMIX system. Even if test subjects interacted with the same system they felt significantly more comfortable, more confident and less tense with the Anna system as compared with the other systems.

In general, test subjects seemed to prefer the gender-marked (i.e. voice and look consistent) agents Anna and Bart over the gender-ambiguous (i.e. voice and look inconsistent) agent Ruth which achieved constantly lower mean average scores as compared with the other two gender marked agents. The agent Anna got most of the preference 'votes'. Thus, our study shows encouraging results for our hypothesis concerning the human preference for interacting with consistently gender labeled entities.

Both studies complement each other by evaluating the system from two different perspectives: one concerning the IMIX interface, the other one concerning its virtual agent.

Future research should consider studies with a larger number of participants and an additional set of agents displaying similar gender-ambiguous versus gender-marked characteristics in order to gain statistical evidence for our hypothesis.

Chapter 6

Meet Olivia - the cute social receptionist robot

In this chapter we will present our first experiment with the social robot Olivia. The experiment was performed in an open environment, during the two days of a technological exhibition. The main focus of the study is to explore relationships between the robot's social skills, user behavior and the overall interaction quality. The study aims to respond to our research question RQ2 concerning the impact of social skills on the interface evaluation. Furthermore, the study aims to determine additional important interaction quality features with potential general validity. The content of this chapter was previously published in [100] and [174].
The chapter is structured in 5 sections. Section 6.1 will give a short overview on experiments with a social robot in open environments. Section 6.2 will introduce Olivia, the social robot receptionist. The section will present the robot's social skills design and technical features. Section 6.3 will focus on the experimental set-up while section 6.4 will discuss the results of the study. The chapter will end with a summary of the chapter in section 6.5.

6.1 Social robots outside the lab - a brief overview

Since the technology advances in engineering and computer science of the last decade brought the usage of robots outside their traditional industrial "playground" there is a growing interest in designing socially competent robots for entertainment, educational purposes, health care assistance, as museum tour-guides, or receptionists.

Many human-robot evaluations presented in the social robotic literature were carried out under controlled lab conditions, where the human social "landscape" was artificially re-constructed. These studies are especially useful for experiments aiming to determine the effects achieved through different variable manipulations. On the other hand, such experiments do not provide insights in how people would interact with the robot in spontaneous real-life situations, nor are the testing conditions comparable, i.e. systems that work well in the lab are often less successful in noisier field environments. Thus, as more and more social robots become available to the general public there is an increased trend to perform such studies in real-life settings, outside the laboratory where the robots are meant to function. These settings enable the gathering of useful observations that can be used to design truly social responsive robots for the human needs.

Some of the first field experiments in the social robotic area were performed in Japan by Asoh et al.[175] with the conversant mobile robot Jijo; the robot interacted with office employees and learned from them how to orient in the office space.

Burgard and colleagues [176] developed a museum tour guide robot and evaluated it over a period of six days in the environment it was created for: the German museum of the city Bonn. Another example of an interactive museum tour guide is the robot Minerva [177]. The robot could express four basic facial emotions and perform simple speech utterances. Studies with the robot Minerva explored short-term spontaneous interactions with visitors coming to the Smithsonian museum.

Hayashi et al. [178] introduced their humanoid robot at the train station from Osaka. They used the robot as a communication medium to present travel information in an 8-days experiment. Similarly, the robot ACE [179] was used on the pedestrian area in Munich in a one-day experiment to provide by-passers with information about the surroundings, weather forecast, latest news, and about itself.

In the medical domain studies performed with the robot Pearl [180] investigated how the robot's social skills helped elderly people to improve their task performances. Pearl's tasks were to remind them to eat, to drink or to take medicine; additionally, the robot guided the elderly people from room to room while chatting about the weather or TV programs.

Other social robots were deployed as English tutors in schools; for example, the study of Kanda et al. [181] explored friendship relationships between two social robots and Japanese school children in an 18-days experiment.

Also, robotic applications meant to be commercialized, such as AIBO or PARO were tested with real users on long-term field experiments [182], [183].

6.2 Olivia - the social robot receptionist

Our study was performed with the social robot receptionist Olivia during the two-days annual exhibition TechFest, organized in October 2009 at I^2R, Fusionopolis (Singapore). Olivia is the 4th service robot model[1] developed by the A*STAR robotic team from I^2R. The robot's tasks were to inform and entertain the Techfest visitors by presenting information about building amenities, daily horoscopes and by playing a simple game consisting of recognizing and tracking different objects. Apart from being a receptionist Olivia's role during the exhibition was to represent the institution as a kind of mascot.

[1]A new version 2.1 of the social robot Olivia was released at Robocup 2010, http://www.asoro.a-star.edu.sg/robots_olivia.html

6.2.1 Designing a robot with social skills

Engaging socially in verbal interactions, as simple as it might appear for humans, is in fact a highly complex process, requiring a synchronized interplay of affective, conversational and behavioral related cues. Fong et al. [184] translated these cues into a list of design characteristics that robots aiming to exhibit human social behavior should possess. These characteristics are:

1. express and/or perceive emotions
2. communicate with high-level dialogues
3. learn/recognize models of other agents
4. use natural cues (gaze, gesture, etc.)
5. exhibit a distinctive personality or character

All these characteristics, except for the 3rd, were implemented in Olivia's behavior design. One of Olivia's most distinctive features is her role: she represents a

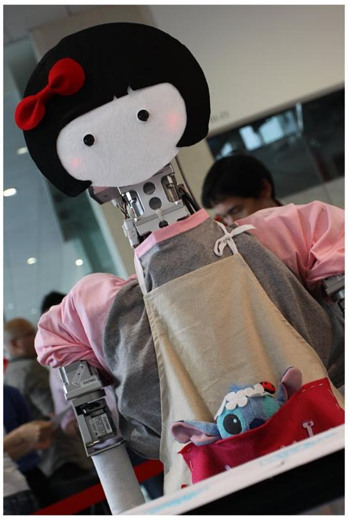

Figure 6.1: A portrait of Olivia

robot mascot that looks and talks like a child; dressed up in a cute pink skirt and wearing a red 'hair' ribbon Olivia speaks to visitors with the typical charm of a very young person (see figure 6.1). Olivia uses her childish charm to draw adults' attention so that they will interact with her. Since it has been proven that humans often treat artificial entities as though they were real [8] we hoped that Olivia's cute behavior would induce the sympathy people usually feel for young children and consequently, her overall abilities would be more positively assessed.

Considering the fact that her 'job' as receptionist and entertainer requires interacting with many people Olivia's personality was designed to be extrovert - as shown in the literature [185], extroverted individuals have enhanced social skills that allow them to communicate easily with others.

Olivia's personality profile was derived from Eysenck's [186] personality extrovert model. The model contains 7 traits such as being outgoing, talkative, lively, carefree (e.g. cheerful), responsive, easygoing (e.g. cooperative), leadership (e.g. dominant). These personality traits were implemented as follows:

1) **Outgoing**: Olivia's outgoingness manifests in a very friendly way of approaching people: always ready to engage in a conversation the robot usually makes the first 'move', greeting people passing by and asking them to spend time with her (see table 6.1).

2) **Talkative**: Olivia loves to talk and often adds a very personal touch to her discourse: visitors are informed not only about building amenities or horoscopes but also about Olivia's family members living in the building, about her preference for kaya toast or her passion for swimming. Because talkative people often use gestures to communicate, Olivia's statements are accompanied by head, arm and body movements meant to emphasize the intended message: for example, the robot uses her arm to point at relevant information cues on the screen or to show direction, shakes her head to express dizziness, waves her hand to greet people or rotates her arms to demonstrate how she swims.

3) **Lively & cheerful**: Olivia's speech and gestures unveil highly emotional features that leave the impression of a cheerful character with a highly animated personality: using a colorful intonation and many interjections Olivia shows surprise ("*wow*"), when a visitor's horoscope sign matches the one of her 'mommy', fear ("*oh*") when she meets a dangerous Scorpio person or joy ("*hey*") when she comes across a Cancer man as she likes the "*yummy*" taste of "*chili crabs*"; she yells for joy when she finds an object during the visual recognition game, 'yawns' to show boredom or complains of getting 'dizzy' when the tracking game lasts too long; in the end she gives visitors a cute onomatopoeic good-bye kiss.

4) **Responsive & cooperative**: Olivia's responsiveness and cooperativeness is expressed at three different levels: at the dialogue structure level, semantic level and gestural level. Through an implicit feed-back strategy ("*Oh I see, you are a Leo!*") the visitors are directly addressed by the robot and confronted with the internal processing state of their inquiry. On the semantic level the robot shows her interests in people's horoscopes revealing often positive characteristics: Taureans are flattered for being big and strong and Virgos for being intelligent people. Also, at the gestural level Olivia shows her readiness to help by leaning her upper body and head towards the touch screen in an attempt to look for corresponding answer images that she can point at.

5) **Dominant**: since our experiment was carried out in uncontrolled environment settings there is a need to guide the visitors in order to maintain a smooth interaction. Thus, this personality trait derived from the leadership characteristics was added to Olivia's personality model. Olivia's dominance is expressed on the dialogue structure level - the conversations is initiated, lead and ended by the robot - and on the semantic level - the robot uses the first person to refer to herself often

displaying an assertive verbal behavior (e.g. *"I like kaya toast!"*).

Amenities dialogue	**Game dialogue**
Olivia: Hi *(waving hand)*! I am Olivia! Nice to meet you! **User**: Nice to meet you too! **Olivia**: Would you spend some time with me? **User**: Sure! **Olivia**: Hmhm (*clearing her 'throat')* I know a lot about amenities here *(makes a round movement with the arm showing the amenities depicted on the screen)*. Tell me which one you like to know more? **User**: I would like to know more about "Fitness First". **Olivia**: Hm ... my daddy works out at "Fitness First", located at level 23 *(points with the hand on the screen where the fitness center details are displayed)*	**Olivia**: Now, let's play a game! **User**: Ok, what game? **Olivia**: Hmhm (*clearing her 'throat')*... pick up my toy *(points to her book left on the screen and looks up at the visitor)* and move it slowly in front of me, as I follow the motion. **User:** *(moves the book too fast)* **Olivia:** Hey! It's too fast! I can't catch up with you! **User**: *(moves the book slowly)* **Olivia**: It is fun! *(moves her head following the book; after sometime starts 'yawing' and pushes her upper body closer to the book; after sometime starts shaking her head and brings back her upper body)*. I am getting dizzy! Let's stop here! **User:** *(puts the book down)* **Olivia:** Muac *(kiss sound)*! Thank you for playing with me and have a nice day (*waving hand)*!

Table 6.1: *Excerpt from two conversations with Olivia*

6.2.2 Technical features

Olivia is approximately 152 kg and 1.6 m tall. The robot has 13 degrees of freedom in total: head (3 degrees), body (2 degrees) and hands (2x4 degrees). It is built on a PowerBot base mobile platform and equipped with several hardware/mechanical components, including actuators (servomotor, harmonic gear system, drive unit and harmonic drive servo actuators) a laser (Hokoyu URG-04LX), cameras (Bumblebee2 and DVN1501 mono camera), microphones and speakers. The robot has several independent software modules for controlling and executing several functions: a motion control (MC), a dialog management system (DMS) and a vision understanding (VU). The MC module employs advance motion control algorithms, such as nonlinear task space control and joint space control to control the robot's movements. The DMS module utilizes the Loquendo 7.52 text-to-speech (TTS) software to generate a female, childlike voice with an American English accent (timbre and pitch=70, speech rate=30, volume=50). The TTS enables the use of several emotion cues, such as hesitation sounds, coughs, yawning, etc. To increase the speech recognition accuracy the DMS' acoustic model was trained with 13.5 hours of read speech data, collected from 40 English non-native speaker subjects (mostly male). For the data collection a 200 word vocabulary was used; additional word entries related to the two main conversation topics (building amenities and horoscopes) were included (50 words per topic). The VU module deploys a multi-model fusion maximum likelihood method by integrating four different approaches: stereo-based human detection, HOG- based human detection, color-based tracking, and motion estimation for human detection and tracking. All software modules run on two PC boards: one Intel Corei7 (2.8 GHz) and one Atom processor (1.2GHz).

6.3 Experimental set-up

The robot was placed in the entrance hall of the building and was one of the inventions presented at the exhibitions. A total of 120 visitors spontaneously interacted with Olivia. Attached to the robot was a touch screen where additional information cues were displayed (see figure 6.2). Visitors could communicate with Olivia using speech or the touch screen. The topics and the games were randomly initiated by the robot: being equipped with visual-recognition capabilities Olivia was able to detect a person standing in front of her and accordingly, could initiate the conversation naturally. A conversation with Olivia typically lasted around 3-4 minutes. Olivia was accompanied by a human assistant standing at 2-3 meters distance. Visitors were free to talk with the assistant and ask questions about the robot. After interacting with Olivia visitors were kindly asked to fill in an evaluation questionnaire.

Figure 6.2: Visitor interacting with the robot

6.3.1 Questionnaire design

Since we are interested in the relationship between the robot's social skills and the perceived overall interaction quality, it is important to find adequate ways to measure them. Additionally, we are interested in finding the most relevant conversational aspects contributing to a better interaction assessment. Thus, our questionnaire has two parts: the first parts addresses the interaction and robot's evaluation while the second part refers to the ranking of the most important conversational aspects.

Robot's evaluation

A tool widely used in behavioral research for social skills evaluation was developed by Gresham and Elliott [187]. The tool is meant to assess human social skills along five categories: cooperation, assertion, empathy, self-control and responsibility. These categories were found to match social abilities aspects involved in human-robot interaction [188] being related to the design characteristics presented by Fong et. al [184] (see section 6.2.1).

Translated to Olivia, these abilities, partly overlap with her extrovert personality characteristics and are expressed in the following way: cooperation manifests in her readiness to help others by sharing information in a highly sociable manner, referring to her ’own’ experiences and using gestures to enhance explanations. Assertiveness relates to Olivia’s extrovert personality, as she initiates the conversation, introduces herself and shows openly her preferences and dislikes. Olivia expresses empathy through emotional, verbal interjection. For self-control and responsibility no direct related aspects were found. Since many authors [189] suggested that humor has an important role in interpersonal relationships, being a social skill in itself, we included it in our investigations. Olivia’s humor is, however, expressed only through a personalization effect: the robot often refers to itself as if it would be human, creating a hilarious impression (see table 6.1).

Consequently, we built up a social skills subscale with 5 items: the ability to socialize (i.e. ability to be friendly), to use natural gestures, to express emotion, personality and humor. These items are hedonic (i.e. non-functional) quality aspects and can contribute directly to the overall interaction enjoyment - for reference, see the **QoCI** taxonomy in section 3.3.2.

To evaluate the interaction quality we used the SASSI [113] questionnaire as inspiration. The questionnaire was developed to evaluate the usability of uni-modal speech-based interfaces and it addresses six different dimensions, as presented in 3.4.2. Because evaluating the interaction quality with a multimodal interface differs somewhat from assessing the usability (fit-for-use) of a uni-modal system we needed to modify the questionnaire to suit our purpose. Accordingly, we retained only items corresponding to the interaction features and their effects on users’ mood. Additionally, we replaced the accuracy dimension with a more precise category referring to the robot’s multimodal performance. We re-grouped the items semantically in two factor subscales: interaction features and user feelings.

The interaction features subscale contained 8 items: interaction easiness, level of concentration, response speed, flexibility, speech/object recognition and object tracking and usefulness.

The user feelings subscale includes only 3 items: calm, comfort and enjoyment. The comfort was not listed in the SASSI questionnaire, but it is often mentioned in the literature as contributing to the overall interaction quality perception [67].

The first part of thee questionnaire ended with a general question about the perceived overall interaction quality.

The visitors scored the subscale items using a 5-point Likert scale with ’strongly agree/disagree’ as endpoints.

Ranking conversational aspects

For the second part of our questionnaire, we selected from the SASSI, ITU [106], AttrakDiff [114] questionnaires a total of 16 items, applicable to social robots. 7 items were related to pragmatic, functional aspects, such as interaction speed, content relevance, clarity of answers, speech/visual recognition accuracy, system transparency and easy recovery from errors. The other 9 items were concerned with hedonic, non-functional aspects, such as voice and appearance pleasantness, friendliness, politeness, humor, emotion display, gestures and mimic, display of human-like physical characteristic (gender and age).

The visitors were asked to rank the functional and non-functional aspects according to their importance for the interaction quality. A 7-point scale with 'not important at all/extremely important' as endpoints was used for the ranking in order to ensure more differentiated results[2].

6.3.2 User behavior

To study the user reactions in interaction with the robot we used behavioral analysis. The method is commonly used in psychology to acquire knowledge about human social interactions by analyzing body posture, facial expressions, gestures and verbal behavior (see section 2.2.1).

Also, in human robot interaction the method was successfully applied in several studies with different purposes. For example, Sabanovic et al. [179] applied observational analysis with the purpose of improving the interactive capabilities of two social robots. Watanabe et al. [190] deployed behavioral analysis of human nonverbal interactions to develop a speech driven embodied interaction robot, while Breazeal [191] used the method to prove the salience of non-verbal cues in cooperative task-oriented interactions between humans and the robot Kismet.

In this study we applied behavioral analysis to determine certain behavior categories with the purpose of investigating whether these categories have specific evaluation patterns associated with them.

The interaction was recorded with three hidden cameras placed around the robot. We used the recording to annotate visitors' behavior concerning gaze, reactions to robot's humor, speech patterns, degree of participation and body posture.

6.4 Results and discussion

From 121 visitors who interacted with Olivia 88 filled in the questionnaire. 67.8% were male and 32.2% were female. 73.3% were of Chinese origin, 14.4% Indian, 12.3% other nationalities. The majority (71.1%) had an IT & engineering background, the rest sharing a background in business (13.3%), arts & humanities (5.6%) and other areas (9.9%). 66.7% were aged between 26-40 years, 20% between 18-25 years and 13.30% were above 41 years. More than half of the visitors

[2]The questionnaire can be found in appendix E.

(54.5%) were Master or Ph.D. holders; 34.4% had a Bachelor degree and 11.1% held other diploma degrees. Probably, due to a technical educational background a relatively high percentage (47.2%) had seen or read about robots and some visitors (24.5%) had even interacted with them; also other few (6.7%) had expertise in robot design & development. 21.6% had no knowledge of robots. A lower percentage of visitors (38%) had used speech recognition devices - mostly as input modality for mobile phones, video games, cameras, dictation systems; a very small number of visitors (2 persons) used the Microsoft SDK tool to build speech recognition applications; 62% of the visitors had no knowledge about speech recognition devices.

Next, we checked the internal consistency of the proposed subscales, as well as the cumulated negative ('disagree' + 'strongly disagree'), positive ('agree' + 'strongly agree') and neutral scores achieved by each subscale item. The reason behind listing the cumulated values lies in understanding the general item evaluation tendency. Subscales with α>.600, item total correlations *r*>.300 and a reduced number of items (<10) are generally considered as acceptable [192].

Robot's social skills	Items	Item-total correlation r	C_{neg}	Neutral	C_{pos}
	Socialize	.524	5.70%	42.00%	52.30%
	Nat.gesture	.536	15.90%	43.20%	40.90%
No. of items: 5	Personality	.536	15.90%	43.20%	40.90%
Cronbach α =.789	Emotions	.621	26.10%	39.80%	34.10%
	Humor	.499	15.90%	52.30%	31.80%

Table 6.2: *Robot social abilities subscale and item cumulated score values*

The analysis of the robot's social skills subscale (see table 6.2) revealed a high internal consistency (α=.789). Next, we performed a Friedman test and found significant differences between the ratings of the items($\chi^2(4)$=26.671, *p*<.001).

Thus, we conducted a post-hoc analysis with a Wilcoxon Signed-Rank test (W_+) applying a Bonferroni correction (BC) for multiple comparisons; a new p-value was set at *p*<.012. The test showed that the ability to socialize was significantly higher scored than all others subscale items (*p*<.001), except for the ability to express natural gestures (*p*=.016). The lowest rated item seems to be the ability to express emotion, however no significant difference with respect to the other items was found. All subscale items, except for the ability to socialize show high frequency distributions in the neutral category. This means that Olivia's social skills are acceptable, as she can 'socialize' but most of the features need improvements. Especially the ability to express emotions - a key item with the second highest subscale correlation (*r*=.621), but also highest negative ratings (26.10%) - should be given special attention in the future. The lack of mimicry on Olivia's face, most probably might have lowered the rating, as humans typically expect emotion expression to appear synchronized at both voice and face level.

The reliability analysis performed on the user feelings subscale proved an internal consistency of α =.696 (see table 6.3). Since all items showed relatively good scores we assume the majority of the visitors felt comfortable and calm while interacting with the robot, enjoying the conversation. The interaction features subscale showed an internal consistency of α=.645 (see table 6.4). Two items - attention

User feelings	Items	Item-total correlation r	C_{neg}	Neutral	C_{pos}
	Comfort	.433	10.20%	36.40%	53.40%
No. of items: 3	Enjoyment	.551	6.80%	29.50%	63.60%
Cronbach α =.696	Calm	.574	2.30%	21.60%	76.10%

Table 6.3: *User feelings subscale and item cumulated score values*

level required and interaction flexibility - were removed because of low correlations with all scale items (r=.091, and respectively r=.198). Correlations with the overall interaction were significant only for the flexibility (negative correlation, r=.356, p=.023).

Interaction features	Items	Item-total correlation r	C_{neg}	Neutral	C_{pos}
	Easiness	.438	13.60%	40.90%	45.50%
	Interaction speed	.331	27.20%	44.30%	28.40%
No. of items: 6	Usefulness	.361	1.10%	34.10%	64.80%
Cronbach α =.645	Speech recognition	.390	1.10%	22.70%	76.20%
	Object recognition	.385	4.50%	23.90%	71.60%
	Object tracking	.329	5.70%	37.50%	56.80%
Removed parameters	**Items**	**Item-total correlation r**	C_{neg}	**Neutral**	C_{pos}
No. of items: 1					
Cronbach α	Attention level required	.091	4.50%	15.90%	79.50%
(if included) =.435	Flexibility	.198	31.20%	36.42%	32.34%

Table 6.4: *Interaction feature subscale and item cumulated score value)*

A post-hoc analysis (W_+, BC, new p<.010) performed after the Friedman test ($\chi^2(5)$= 86.336, p<.001) showed significant differences between the interaction speed and easiness on one side and all the other items, on the other side (p<.001). Also, significant differences were found between speech recognition and object tracking capabilities (p<.001).

The response slowness was mostly caused by speech[3] and visual recognition difficulties in respectively, 54.40% and 28% of the cases. Since no feed-back or error recovery strategies were implemented, i.e. no reaction came when the recognition score was below a certain threshold, the visitors were left with the impression the robot's response was slow.

Table 6.5 presents an overview of response latencies[4] for speech and object recognition, as well as for visitors' tolerance level[5]to speech response latencies.

The robot's response latency in speech recognition error-free cases was on average 2.51 sec, a value still far behind that found in human face-to-face conversation - 0.97 sec, [193]. In cases with speech recognition errors response latencies were higher, lasting on average 3.18 sec. In such cases, total response time could achieve extreme (but luckily infrequent) values of even 45 sec.

Compared with human face-to-face conversations, where a delay of more than 2-3 seconds in providing a response was found to cause discomfort [194] it becomes clear that such response latencies are unacceptable. Interestingly, the delay value

[3]The speech recognition problems were caused by a noisy environment (92.80%), wrong pronunciations (4.8%) and other technical issues (2.4%).

[4]Response latency refers to the time elapsed between last user input and robot's response.

[5]The tolerance level refers to the time elapsed until a user re-prompts her input when no response is given.

Response latency	Mean	Median	Modus	Min.	Maxi.
ASR error free	2.51	2	2	0.75	7
ASR with errors	3.18	3	2	1	15.50
Total time until response-ASR errors	14.96	11.25	6	4	45
Visual recog. error free	1.77	1	1	0.25	11
Visual recog. with errors	13.52	12	11	4	27
Tolerance level	3.74	3.31	3	2	9.33

Table 6.5: *Robot's speech and visual average response latencies in seconds*

mentioned above corresponds roughly to the tolerance level of 3.74 sec measured in our study. Thus, we would expect high correlations between speech recognition errors and the speed scores.

But surprisingly, this was not the case: the speech recognition and the interaction speed have a correlation coefficient of only r=.279. A further detailed analysis revealed the speech recognition performance correlates with the speed scores in only 30% of the cases: many visitors (44.30%) scored the interaction speed as being neutral - neither fast, nor slow, even by high response latencies (23 up to 45 sec.) or by relatively low latencies (1.5-2 sec.).

This leads to the following three remarks. Firstly, an average response latency of 2.51 sec is too high. Secondly, the question referring to speech recognition abilities ("The robot was able to recognize my speech") should have been formulated more accurately (e.g. "The robot's ability to recognize my speech was very good"); since Olivia always provided a response, even after long response delay, it means she was 'able' to recognize speech; therefore, the question might have generated misleading responses. Thirdly, many visitors tended to avoid negative scores choosing instead neutral ratings; this tendency of scoring more positively in order to please the interviewer or to be helpful was also observed by other studies [195].

Items	C_{neg}	Neutral	C_{pos}
Overall interaction quality	2.00%	30.7%	63.6%

Table 6.6: *Overall interaction quality scale with cumulated score values*

Similar rating behaviors were observed between the speech recognition performance and the scores obtained in overall interaction quality (see table 6.6): the presence or absence of speech recognition errors corresponds to a negative/positive overall quality assessment in only 34.66% of the cases; visitors gave more neutral scores, even if the robot obviously failed to recognize their speech and her response had long delays. Past research studies also found that the perceived system understanding and the objective word accuracy is only weak correlated [99].

We also compared the ratings for enjoyment in cases with speech recognition error and found that in 60% of the cases nevertheless, the visitors gave high ratings. Looking at the enjoyment correlations we found the highest correlations with interaction easiness (r=.442), ability to socialize (r=.436) and overall quality (r=.418). Thus, the visitors' tendency towards more positive ratings as observed in [195] might have an additional, complementary explanation: people might have rated the interaction features and overall quality better because they experienced an enjoyable (and not particularly difficult) interaction with a sociable robot. Finally, we an-

Items	Min.	Maxi.	Mean	Std. deviation	Correlation with overall quality r
Robot social skills	2.00	5.00	3.2409	.55183	.444**
User feelings	2.00	5.00	3.6629	.55936	.435**
Interaction features	2.67	4.83	3.5568	.43468	.600**
Overall quality	**2.33**	**5.00**	**3.5083**	**.58880**	**1.001**

** p<0.01

Table 6.7: *Correlations between overall interaction quality and the subscales*

alyzed the correlation between the subscales and the perceived overall quality (see table 6.7). All subscales correlate significantly with the overall interaction quality, whereas the interaction features have the highest correlation coefficient (r=.600). The robot's social skills have a lower correlation coefficient (r=.444) as compared with interaction features.

Additionally, we checked the correlations between the overall interaction quality and each subscale item to detect the highest correlations; we found that the interaction easiness (r=.490), the ability to socialize (r=.435) and usefulness (r=.409) had the highest correlations (>.400) with the overall interaction quality. On the other hand, the interaction features were significantly better evaluated than the robot's social skills (W_+, $p<.001$).

Next, we analyzed the priorities ranks visitors assigned to different functional and non-functional aspects that might be involved in the face-to-face conversation with a robot (see table 6.8).

Despite a non-normal data distribution we chose the mean as sole option to build a differentiation order. However, the rank order can be validated by only applying a non-parametric significance test.

The mean scores show a demarcation line between the functional (1-7) and non-functional (8-16) aspects. The functional aspects were on average significantly higher ranked than the non-functional (W_+, $p<.001$). This finding is not surprising, since functional aspects are, from a pragmatic point of view, more important than non-functional aspects, e.g. the robot's nice appearance would not replace its poorly working speech recognition. Nevertheless, this does not mean non-functional aspects are unimportant. In fact, many studies proved the benefits of non-functional aspects such as emotion displaying, gesture and mimicry for the robot's social acceptance or human-like skills [10].

The result of post-hoc analysis (W_+, BC, new $p<.008$) performed after the Friedman test ($\chi^2(6)=18.38$, $p=.005$) revealed that the interaction speed was statistically higher ranked than the error-free speech/object recognition (p=.006/.007); this means that users could be more tolerant to errors, but less understanding if they have to wait too. No significant rank differences were found between the other aspects, except for the system transparency whose mean was significantly lower than those of the interaction speed (p=.005).

Among the non-functional aspects statistically significant differences could be found ($\chi^2(8)=175.70$, $p=.005$). The post-hoc test (W_+, BC, new $p<.006$) showed that the pleasant voice, friendly behavior and politeness were significantly higher ranked than the humor and the gender/age displaying. In fact, both gender/age displaying were included on the aspects list because of their relative importance in verbal addressing in Asian cultures; however, they achieved the lowest statistically significant ranking of all aspects (p<.001).

Interestingly, a pleasant voice achieved a statistically significant higher mean than a nice physical appearance (W_+, p=.005). This result could be explained as follows: even if the visual impression of the robot would impact visitors in the first place, its voice might play a more important role in the interaction, since it conveys the required information.

No.	Category	Mean	Significance level relative to the other item rank[a]		
1	Interaction speed	5.83	*=5,6,7;	***=8-16;	ns= 2-4
2	Easy recovery from errors	5.80		***=8-16;	ns=1,3,4-7
3	Clarity of answers	5.77		***=8-16;	ns=1,2,4-7
4	Delivering relevant information	5.72		***=8-16;	ns= 1-3, 5-7
5	Error free speech recognition	5.67	*=1;	***=9-16;	ns=2-4,6-8
6	Error free object recognition	5.61	*=1;	***=9-16;	ns=2-5, 7,8
7	System transparency	5.52	*=1;	***=10-16;	ns=2-6, 8, 9
TOTAL	FUNCTIONAL ASPECTS	5.70	*** TOTAL NON FUNCTIONAL		
8	Pleasant voice	5.26		**=1-4,12, 14-16;	ns=5-7,9-11,13
9	Friendly behavior	5.22		**=1-6,14-16;	ns=7,8,10-13
10	Gestures and mimic	5.01		**=1-7,15,16;	ns=8, 9,11-14
11	Polite way of talking	5.01		**=1-7,14-16;	ns=8-10,12,13
12	Nice physical appearance	4.95		**=1-8,15,16;	ns=9-11,13,14
13	Emotion displaying	4.92		**=1-7,15,16;	ns=8-12,14
14	Humorous way of talking	4.72		**=1-9,11,15,16;	ns=10,12,13
15	Gender displaying	4.10		**= 1-14,16;	ns= none
16	Age displaying	3.66		**=1-16;	ns= none
TOTAL	NON-FUNCTIONAL ASPECTS	4.80	*** TOTAL FUNCTIONAL		

[a]*** significant at: p<.005 (9 comparisons), ** at p< .006 (8 comparisons), * at p< .008 (6 comparisons); 'ns' stands for no significant

Table 6.8: *Overall quality scale with cumulated ratings*

6.4.1 Effects within groups

Next, we checked the effects of educational background, age, gender, profession, background knowledge on the evaluation using a Mann-Whitney test.

We found that users with higher educational degrees (Master and PhD) found the interaction with the robot easier (p=.005) as compared with users with lower degrees, such as Bachelor, Diploma etc. Female visitors found the robot's social skills (p=.006), including its capabilities to express emotions (p=.036) and humor (p=.010), better than male participants did. People with some background knowledge or information in robotics rated lower the interaction features (p=.006), especial the object recognition (p=.001) as compared with those without background knowledge. Visitors with speech recognition background or experience rated lower the robot's speech recognition capabilities (p=.033) and the response's speed (p=.042) as compared with visitor with no speech recognition background, who on the other side, rated better the robot's social skills (p=.035) and its ability to express emotion (p=.026).

No significant differences were found for the categories concerning the users' age and profession.

6.4.2 Relationships between users' behavior and evaluation results

Further, we analyzed how the user behavioral categories - gaze, reactions to robot's humor, speech patterns, degree of participation and body posture - relate to the evaluation results gathered from the questionnaires. The categories were inspired from MUMIN coding scheme [196] (the scheme will be presented in details the following section 7.2.2). The data was annotated with two annotators. The percentage of agreement calculated on 10% of the data and revealed a high Krippendorff α value of K=.9365.

Concerning the visitors' gaze behavior we identified the following categories: gazing predominantly at the screen (type 'A'), gazing predominantly at the robot (type 'B'), and mixed gazing at both screen and robot (type 'C'). The gaze behavior annotations did not include interaction sequences where the visitor's attention was intentional guided in a particular direction, e.g. to the screen or to game objects. Results showed that 37.3% of the visitors exhibited a gaze behavior type 'A', 34.7% type 'B' and 28% type 'C'.

Regarding visitors' mimics and reactions to robot's humor we differentiated between 'positive' and 'negative' reactions. 'Positive' reactions were expressed in both verbal and non-verbal form, i.e. interjections, hilarious answers to robot's humor, smiles or laughs. As 'negative' reactions we considered the lack of response to the robot's humor and the display of a general serious attitude. Statistics revealed that 68% of the visitors smiled during the interaction and 50% had a positive reaction to the robot's hilarious talk. According to a chi square (χ^2) test we found significant correlations between gaze behavior type 'A' on one side, and 'negative reactions' to robot's humor (p=.001) and lack of smiling (p=.040) on the other side.

Next, we analyzed visitors' speech behavior focusing on the input shortness, polite markers and whether they approached the assistant while interacting with the robot. Most of the visitors (65.27%) preferred to use keywords instead of sentences (34.73%). Significant correlations were found between both gaze behavior type 'B' and 'C' and the preference for using sentences (p<.001). Only 27.6% used polite markers, e.g. greetings or thanks. 40.8% approached the assistant while interacting with the robot.

Despite predominant answer shortness on visitors' side we distinguished between two types of communicative behavior, i.e. degree of participation among visitors: a highly interactive type and a low-interactive type. A highly interactive type showed initiative in conversation, asking the robot additional questions and using mainly speech to communicate. A low-interactive type used predominantly the touch screen to interact, had no conversation initiative and often relies on assistant's help. The majority of visitors (60%) belonged to the low-interactive type. The low-interactive type correlated significantly with the gaze type 'A', and with the preference for using key-words (p<.001).

The body postures refer to the position in which visitors kept their arms and hands: 9.3% crossed their arms around their body, 2.7% placed a hand on the shoulder, 10.7% kept their hands alongside their body in a rigid position, 1.95% put both hands on their hips, 17.3% locked their hands at the back, 12.65% had their hands busy with bags or other objects, 13.3% hid their hands in their pockets, and 17% held them together in front of the body. Only 15% of the visitors placed their hands on the touch table in an attempt to get closer to the robot. Significant correlations were found between the low-interactive type and the tendency to hide hands behind the back or in the pockets or to keep them rigidly alongside the body (for all χ^2, $p<.001$).

Finally, we investigated how the behavior categories mentioned above related to evaluation patterns found in the questionnaires. According to a Mann-Whitney significance test visitors displaying gaze behavior type 'A' felt more in control of the conversation than visitors with other gaze behavior types (p=.019). On the other side, people displaying gaze behavior type 'C' found the interaction to be easier (p=.020). Visitors, who smiled during the interaction evaluated the robot's ability to express humor better (p=.023). Participants, who reacted to robot's hilarious talk evaluated its ability to express personality much better than those who did not (p=.032). Visitors, who did not pay attention to the human assistant evaluated the robot's ability to express emotion (p=.032) and its capacity to respond fast (p=.044) better than visitors who approached the assistant. People using keywords to communicate with the robot indicated a higher degree of concentration during the interaction, than those who used sentences (p=.010). This indicates that the use of natural interaction patterns, such as full sentences, might have beneficial effects in reducing the cognitive load.

6.4.3 Predicting overall interaction quality

Additionally, we performed an ordinal probit regression analysis to test the predictive power of user personal data and behavioral categories on the overall interaction quality. The best model resulted from combining user's educational degree, their interactive behavior, speech patterns and politeness markers.

The model shows statistical significance (p=.048) and Pseudo R^2 values of 12.8% (Cox and Snell) 14% (Nagelkerke), 5% (McFadden). The model shows that the degree of interactivity seems to be a good predictor for the interaction quality, i.e. the quality increases when users show a more interactive behavior, (p $=.010$). The other parameters used in the model show only statistical trends, but no significant p-values. Since our model was based on a relatively small sample (N=70) its predictive power is limited. However, the results are encouraging for further research.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Quality = 3]	-1,286	,290	19,743	1	,000	-1,854	-,719
	[Quality = 4]	-,091	,240	,143	1	,705	-,562	,380
	[Quality = 5]	1,514	,289	27,508	1	,000	,948	2,079
	[Quality = 6]	2,368	,394	36,192	1	,000	1,597	3,140
Location	Degree_Num	,484	,276	3,060	1	,080	-,058	1,026
	B_Interactive_Value	,932	,360	6,685	1	,010	,225	1,638
	B_speech_Behavior_values	-,668	,346	3,726	1	,054	-1,346	,010
	B_politeness_value	-,490	,341	2,069	1	,150	-1,159	,178

Link function: Probit.

Figure 6.3: Parameter estimates for predicting overall interaction quality

6.5 Summary

In this study we analyzed relationships between a robot's social skills, interaction features and user feelings on one side, and the perceived overall interaction quality, on the other side. Our results showed significant correlations between these three factors and the perceived overall quality, the interaction features showing the highest correlation. We would have expected a stronger relationship between the robot's social skills and the perceived overall interaction quality. Nevertheless, the ability to socialize seems to play an important role, being the second most correlated item with the overall interaction quality.

The robot's speech recognition performance was better ranked than the error logs and total completion time would have predicted. This might be explained, partly by the question formulation bias, partly by a general human tendency to give more positive ratings and partly because the visitors enjoyed the interaction despite errors and long response delays.

The conversational aspects ranking brought us important information that can be used to improve the robot design and set priority decisions: for example, it seems that visitors were more tolerant to errors than to long response latencies. A pleasant voice seems to be more important than a nice physical appearance while humor and gender/age displaying appear to be less important conversational aspects for the interaction quality.

Both, aspects ranking and correlations obtained from the items' evaluation suggest that the overall interaction quality relates more to the robot's ability to lead the interaction (response speed, clarity of answers, interaction easiness) and to appear agreeable (friendly, i.e. sociable, having a pleasant voice), than to its performance accuracy, in terms of speech and object recognition/tracking.

Also, the results of the behavioral analysis are encouraging, since they are consistent with those delivered by the subjective evaluation questionnaire. Such observations can be useful for creating further user profiles and evaluation predictions. However, further research in this area is required.

In the future Olivia's dialogue design would incorporate an error-handling strategy to reduce the robot's response perceptions as being slow. Also, help options and a better system transparency would be integrated to enhance the interaction easiness. Further, adding mimicry to Olivia's face to show emotions, improving her gesture to becoming more natural and making the dialogue script more amusing might increase the robot's perceived ability to socialize.

Our results originated from an exploratory study and therefore, cannot prove causal relationships between the analyzed items. Nevertheless, the study revealed significant item correlations that can be used to improve the current robot design. Their significant impact could be examined in future contrastive laboratory conditions to find statistical evidence.

Chapter 7

Interacting with Olivia’s ’rival’: the human receptionist

In the previous chapter we presented our first experiment performed with the social robot receptionist Olivia. The robot was showcased at the technological exhibition TechFest 2009 and tested with several visitors. The experiment brought us important information on how the visitors interacted with the robot and evaluated its performance and social skills. On the other side, the demo nature of the experiment, i.e. the short interaction time, the robot’s game playing etc., delivered us data that was rather unsuitable for modeling applications involving real receptionist tasks. Such tasks require more extended dialogues and additional domain knowledge. With this purpose in mind we conducted an empirical study in a controlled environment with human participants acting as visitors and receptionists. The goal of the study was to identify multimodal communication patterns evolving in human dialogues carried out between a visitor and a receptionist that could be of potential use for modeling efficient human-robot dialogues in a similar context. As such, the chapter will respond to our third research question (RQ3). Section 7.1 will present a summary of related work, section 7.2 will offer an overview about the experimental settings and the annotation schemes used to annotate the collected data, section 7.3 will report the results gathered from our annotation concerning participants’ details, annotators’ reliability, dialogue act frequency and the analysis of verbal, facial and gestural expressions annotated; the section also contains the results of the interviews conducted with each participant after the study. The last section, 7.4 will present the summary of the chapter.

7.1 A multimodal annotation corpus

Communication in a socially situated context is a highly complex, multi-faceted activity. Understanding and modeling its underlying mechanism as experienced in human face-to-face conversations is essential for the design of efficient spoken dialogue systems, embodied conversational agents or social robots. Here, not only theoretical models but also empirical data gathered from similar situated experiments performed with human participants is needed.

As many research areas are moving their interest focus from single to fully-fledged modalities the number of publicly available multimodal corpora, such as the AMI corpus[1] (containing meeting recordings) or the HUMAINE data base (focusing on recording of emotion) is increasing[2]. However, since our purpose was studying human-human interaction in a specific task-oriented conversation a special corpus was required. Hence, we recorded a video corpus of approximately one and a half hours containing conversational interactions between two human participants acting as receptionist and visitor.

The conversations were based on three specific scenarios. The scenarios tasks were meant to be similar to those encountered in real life by professional receptionists, such as handling appointments, explaining directions, informing about building facilities and assisting with calls (see table 7.1).

Scenario	Tasks receptionist	Tasks visitors
1.	Handling appointments	Inform receptionist about an appointment
2.	Informing about building facilities	Ask details about the sky-garden and the swimming pool Ask for directions on how to get to these locations
3.	Assisting with calls	Ask for assistance to call a staff member Ask for assistance to call a taxi

Table 7.1: *Scenarios for human-human interaction*

The visitors were instructed to perform the tasks as described in the scenarios. For example, in the first scenario the visitor had to inform the receptionist about his appointment with a staff member and ask the receptionist to explain the way to the staff member's office. In the second scenario the receptionist's task was to inform the visitor about two building facilities, the sky-garden and the swimming pool, and to explain how to reach these facilities. In the third scenario the visitor was supposed to request the receptionist's assistance to contact a staff member over the phone and to book a taxi to the airport. Each visitor received the instructions on each particular scenario before entering the observation room.

The receptionists did not receive any scenario, but were asked to behave during the experiment as they would normally do in their daily work. They also received the information material needed to respond to the visitors' requests, such as phone numbers, office locations and facilities descriptions.

Before the experiment both visitors and receptionists received an individual briefing in which they were informed about the goal of the study, that is to say, to collect and analyze task-oriented natural dialogues in order to identify communication patterns that would help to design more efficient human-robot dialogues.

We emphasized that there was no wrong/right way of asking/responding questions and we asked the participants to try to act as naturally as possible within the limitations of the scenario.

We kept the same environment settings as in previous experiment with the robot Olivia (see section 6.3), that means, in front of the receptionist was placed a screen

[1] AMI website: http://corpus.amiproject.org/
[2] HUMAINE website: http://emotion-research.net/

where short information cues and pictures related to possible visitors' questions were displayed. To simplify the receptionist's task the information shown on the screen was manipulated by a wizard located behind a dark glass. The whole experiment was recorded with two cameras: one capturing the visitor, the other one focusing on the receptionist. The dialogues were transcribed and annotated using the ELAN tool[3].

After completing the experiment we interviewed both visitors and receptionists. The visitors were asked to judge the interaction with the receptionists in terms of expectations and receptionist's performance. Additionally, they were asked to make improvements suggestions. The receptionists were asked to judge how realistic the scenarios were and what qualities make a good receptionist. The interviews lasted about 30 minutes for the visitors and ten minutes for the receptionists.

7.2 Annotation schemes

7.2.1 DIT ++

To annotate the verbal statements of our corpus we used the DIT++ schema, as being one of the recent standards in dialogue annotations [61]. The scheme was developed for tagging dialogues acts and it extends the taxonomy of Dynamic Interpretation Theory (DIT) - originally meant for information dialogues [197] - with a number of dialogue act types from DAMSL [198] and other schemes [199], [200] [201]. One important feature of the DIT++ schema is that it supports the marking up of utterances with more than one functional tag. This feature is particular important, as often utterances in dialogues are multifunctional, i.e. they express more than one dialogue act. The following examples should demonstrate the multifunctional nature of the dialogue:

> **Receptionist**: *"Sir, do you have a security pass to enter in the building?"*
> **Visitor**: *"Ahhhh ... no ... I don't have ..."*

In the first utterance the receptionist performs two dialogue acts: 1) assigns the next turn to the visitor and 2) formulates a propositional question. In the second utterance the visitor: 1) accepts the turn, 2) stalls briefly and responds to the question 3) disconfirming and 4) elaborating his answers.

DIT++ contains 11 dimensions, each having several communicative functions: 1-3 are task-oriented while 4-11 are dialogue control related.

1-3. Task (or activity): contains dialogue acts dealing with the task or activity that motivates the dialogue. These can concern information seeking acts (set, propositional or choice questions), information providing acts (confirmation, disconfirmation, agreement, disagreement, inform with rhetorical function, answer elaborations, etc.) and action discussion acts (commissive and directives)

[3]http://www.lat-mpi.eu/tools/elan/

4. Feedback: includes dialogue acts providing or eliciting information about the processing of previous utterances by the current speaker (auto feedback) or the current addressee (allo feedback)

5. Turn management[4] : includes activities for obtaining, keeping, releasing, or assigning the right to speak

6. Time management: refer to those acts managing the use of time in the interaction, such as stalling and pausing statements

7. Contact management: contains dialogue acts, such as contact check and contact indication establishing whether the dialogue partner is present and is paying attention

8. Discourse structuring: contains dialogue acts dealing with topic management, such as opening and closing (sub-) dialogues and topic shifts

9-10. Own and partner communication management: refers to actions performed by the speaker in order to correct errors in his current contribution or to complete/correct current contributions made by another current speaker

11. Social obligations management: refer to dialogue acts for dealing with social conventions such as greeting, introducing oneself, apologizing, thanking, saying good-bye; also downplay responses to these acts are included here

7.2.2 MUMIN

For the annotations of gesture and facial expressions we used the MUMIN coding scheme [196]. The scheme was originally developed to annotate multimodal communication presented in broadcast television videos, but it can also be deployed as a general instrument for studying gestural and mimic expression in interpersonal face-to-face communication. The annotation of facial expression and gestures was done in relation with the corresponding speech transcript, i.e. taking into account possible determinations by the speech, gesture or facial expression produced by the interlocutor. Additionally, we included in our annotations information related to the speakers' body positions.

• The **general face** category refers to the expression on the speaker's face, which can be a smile, laughter, the expression of scowl (frowning, down turning of the mouth corner), or other facial expressions that cannot be put under the mentioned categories. The neutral face expression was not annotated.

[4]This dimension, concerning the assignment of turns has more relevance for conversations involving more than two persons. Since our dialogues are exchanged between only two participants we did not include this dimension in our current analysis.

• For the **eyebrows** we retained movements concerning raising and frowning.

• The **eyes** category was labeled in terms of the eyelid movements. We used for the labeling the following tags: exaggerated opening, referring to a situation when the eyes are wide open, as in the case of surprise; both eyes closed, when the eyes remain closed longer than during the normal blinking with the intention of underline the meaning of a certain communicative act; eyes semi-closed, when the eyes diminish their openness as in short-sightedness cases - this eye expression can occur when the speaker needs to concentrate or tries to emphasize his words.

• Concerning the **gaze**, we concentrated on annotating the direction the speaker was looking in. Gaze is used to control the conversation flow by managing turn regulation, monitoring feedback, but also by expressing emotion, internal information processing or level of attention. Mutual gaze refers to the situation when both dialog partners are looking mutually at each other; 'up'/'down' indicates the direction the person is looking in while 'away' refers to the situation when the person looks away from the interlocutor.

• For the **mouth** category we annotated the position of the mouth related to the facial display. This means we annotated whether the person has his mouth open - as a sign of surprise, whether his lips were protruded (i.e. rounded), or retracted (i.e. when the lips are sucked-in and pressed together). The mouth position 'closed' was not annotated since it is the standard position while not speaking.

• **Head** movements were annotated in terms of: head nods (single (s) or repeated (r)), head shakes (repeated rotation of the head from one side to the other), waggles (repeated side to side head movement, very usual for Indian interlocutors), tilts (head leaning on one side), head moves forwards (head and sometimes trunk moves forwards, often as turn elicit signal), head moves backwards (head and sometimes trunk moves backwards), often as a turn accepting signal. We added one more value to the head movements that we called 'head pointing' - this value refers to a deictic movement in which the head is used to point to an imaginary or real object.

• The **hand** gestures refer to deictic (gestures used to point to a person, object or a certain direction), iconic (gestures used to illustrate physical items), emblematic (gestures with a specific standard meaning, such as waving the hand to mark valediction etc.), beat (rhythmic gestures with no particular meaning but performed to emphasize particular words or even whole speech parts). McNeill [34] also defines a gesture category called, 'metaphoric' (gestures used to explain a concept). However, we decided to make no distinction between iconic and metaphoric gestures since both express an idea - whether abstract or concrete - through gestural similarity.

• Finally, we annotated the body position concerning forward, backward and sideways deviations from the main position held by the speakers.

Modality	Expression type	Value
Facial expression	General face	smile, laughter, scowl, other
	Eyebrow	raise, frown
	Eyes	exaggerated opening, both eyes closed, semi-closed
	Gaze	mutual, up, down, away
	Mouth	open, lips protruded, lips retracted
	Head	nod (s/r), shake, waggle, tilt, forward, backward, point
Gesture	Hand gestures	deictic, iconic, emblematic, beat
Body posture	Body posture	forward, backward, sideways

Table 7.2: *Coding scheme for gestures and facial display*

Table 7.2 presents an overview of the expression type and form analyzed during our study with details on each tag.

7.3 Annotation results

7.3.1 Participants' details

A total of 17 people participated in the study: six were playing the receptionist role (five female and one male), the other eleven participants (eight male and three female) were acting as visitors. Twelve were Chinese, four were Indians and one was Caucasian. All participants (except for one) indicated English as being their most frequently used language in daily communication. The participants were recruited from the staff member working in I^{2}R Singapore. The people appointed as receptionists - except for one- had previous experience (6 months-2 years) working as receptionists. The range of their age varied from 18 to 60 years old: three were between 18-25 years old, three were between 26-30 years old, six were between 31-40 years old, three were in the age range of 41-50 years old and two were between 51-60 years old.

7.3.2 Annotators' reliability

Since the data was annotated with two annotators we calculated the inter-annotators' reliability on 10% of the data. For the calculation we used the Krippendorf's α coefficient. The values obtained were between 0.7-0.9, as shown in table 7.3. According to K. Krippendorff [202] such values are suggesting a good inter-annotators' reliability.

Annotation	Participants	Krippendorf's alpha
Dialogue act	Visitor Receptionist	.7402 .7562
Body position	Visitor Receptionist	.9286 .8751
Facial expression	Visitor Receptionist	.7842 .7692
Gaze direction	Visitor Receptionist	.8536 .7153
Gesture	Visitor Receptionist	.8828 .7004

Table 7.3: *Inter-annotators reliability for the human receptionist corpus*

7.3.3 Dialogue act frequencies

In total, 805 turns[5] were exchanged between visitors and receptionists. The first scenario (handling appointments) was with an average of 19.27 turns the shortest scenario, followed by the third scenario (assisting with calls) with an average of 22.81 turns. The second scenario (informing about the building's facilities) was the longest scenario having an average of 31.09 turns. The scenario's length is explained by the fact that some receptionists often gave more detailed explanations about the building's facilities which in return caused additional questions.

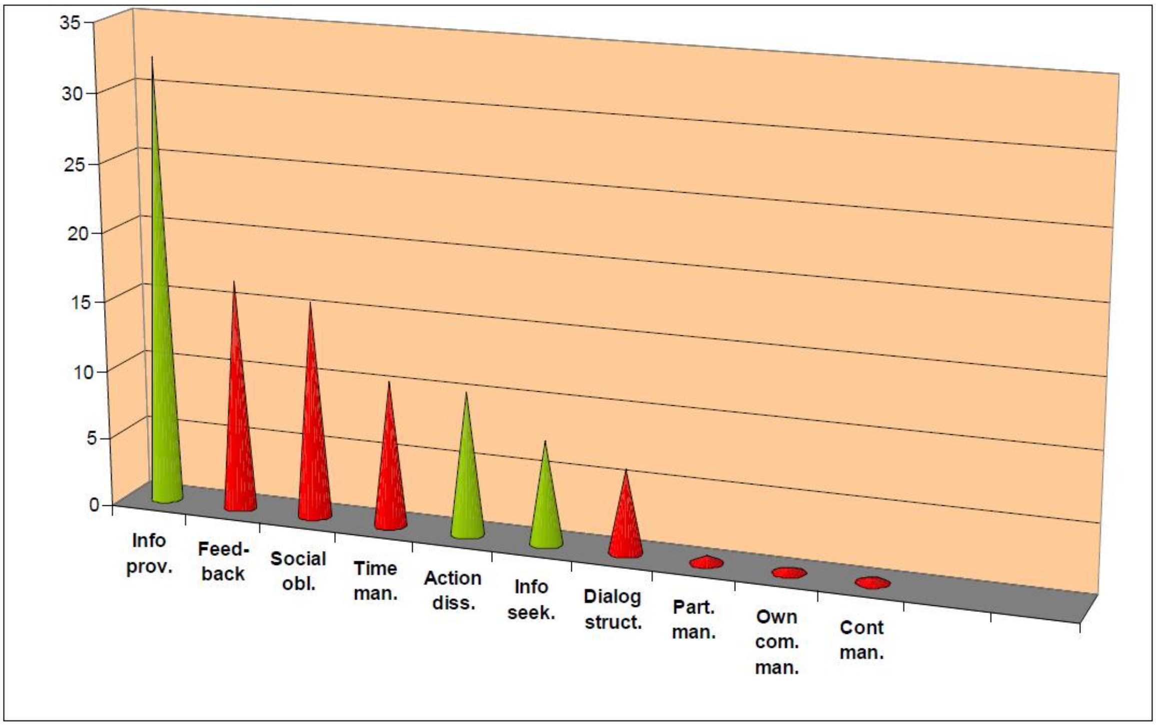

Figure 7.1: Dialogue act dimension frequency distribution

[5]The analysis presented in this paper refers strictly to the turns exchanged between the visitors and the receptionists; the utterances exchanged over the phone between the participants and a third person were not included.

Figure 7.1 depicts the frequency distribution of the dialogue acts exchanged between the receptionist and the visitor during the experiment (the task-oriented categories are marked in green). The figure is a visual and structural representation of an 'invisible' process: a verbal dialogue exchange. Since all three scenarios handle task-oriented dialogues around the same 'receptionist' topic we joined the dialogue acts in a single bar graph. The data gathered can be further used to develop appropriate dialogue strategies for the receptionist tasks, to create more robust language models and to extend the vocabulary of the robotic application with relevant domain words.

Next, we will discuss the components of each category starting in descending frequency order, as follows:

Information providing (32.24%)
The highest frequency distribution was covered by information providing acts. These were composed by the following items:

- Single statement answers (14.46%)[6]; 54.4% of the single statement answers were confirmations (*"yes"*), 1.75% disconfirmations (*"no"*) and 43.85% short one-sentence answer (*"the swimming pool is on the 24th floor"*)

- Informative statements (9.13%); these statements are similar to short one-sentence answers, however, the difference is that they do not precede any question

- Answer elaboration (40%) - answers containing additional or more detailed information about something the speaker had mentioned before

- Explanation (25.57%) - mostly indications on how to reach a certain point of interest in the building.

- Other informative statements: justifications (6.3%), i.e. information delivered to explain why the speaker said something, summarizations (1.52%), rhetorical information (1.52%), clarifications (1%) and conclusions (0.50%)

Feedback (16.70%)
Dialogue acts belonging to the feedback category had the second highest distributions. For the auto feedback we annotated the following categories: non-verbal unspecific (*"mhm"*, (*"aha"*) representing 13% (from the total amount of feedback), positive overall verbal (*"ok"*, *"yeah"*, *"right"* 64.17%), negative non-verbal (*"ah?"*, 0.50%) , attention feedback (12%) consisting in repeating and/or summarizing the interlocutor's statements in order to retain the information and, at the same time, to ensure that the content was correctly understood. The allo feedback was 10.33% (from the total amount of feedback).

Social obligation (15.63%)

[6]The value is calculated from the total amount of information providing acts.

The social obligation dimension is the third largest dimension among our dialogue acts and includes gratitude expression (40%)[7], salutations (35.67%), valediction (17.58%), apologies (4.24%) and self-introduction (2.51%).

We noticed that often most of the thank acts (75%) occurred at the end of the conversations and people used them to replace valediction acts in 38.70% of the cases. We also noticed, that 34.48% of the thank downplaying acts, e.g. *"you are welcome"* were thanks, most probably emphasizing a very polite way of talking.

Greetings, especially salutations mostly occurred in dyads (in 84.37% of the cases) and less as single dialogue act, when only one of the participants greeted (15.62%). In 31% of the cases double greetings were used (*"Hi, good morning!"*). In 18.75% of the cases the receptionist used additional greetings forms, such as *"nice to meet you"* or *"how are you?"*, etc. The valediction acts, i.e. *"good-bye"* occurred in dyads in only 19.34% of the cases; however, in 32% of the cases, additional pre-closing phrases such as, *"take care!"*, *"safe journey!"*, *"have a nice time!"* were performed. Downplaying acts to apologies were not very common (in only 20% of the cases when apologies were presented). Also, self-introduction act were seldom (occurred in only 2.52% of the total amount of social obligation acts performed).

Time management (10.47%)
The time management dimension consist in non-verbal stalling acts (77.34%) and verbal requests for time (7%). 15.65% of the dialogue acts in this dimension were requests for pausing, i.e. indications the speaker need some time to do something.

Action discussion (10.31%)
In the action discussion dimension we found action directive (43.65% cases from the total amount of action discussion), action commissive (30.64%) and action acceptance/rejection responses (25.71%). Among the action directive the most frequent were the requests formulated directly (39.50%) and indirectly (34.50%); 12% were instructions and 14% were suggestions. All commissive actions were offers.

Information seeking (7.04%)
In the information seeking dimension we annotated the total amount of questions that were posed during the entire dialogue. 58.42% were set question (WH- questions) and 41.57% were propositional questions ('yes/no' question). Both visitors and receptionist asked questions. However, as expected the visitors were the participants who asked more frequently: 67.44% of the questions were asked by visitors and 32.55% by the receptionist.

Dialogue structuring (6.39%)
Among the dialogue structuring acts the most frequent were topic introduction statements (81.08%) and pre-closing statements, such as (13.51%). Topic shift acts were less frequent (5.41%). Most of topic introduction statements were *"ok"*[8] state-

[7]The value is calculated from the total amount of social obligation category.

[8]This "ok" statement has a different communicative function than the one used for positive feedback, confirmation or agreements acts.

ments (84.22%) while introductive statements, such as *"I finished another meeting"* were less common (15.78%).

Partner communication management (0.50%)
In the partner communication management we found 2 statements concerning correct-misspeaking acts (the addresses corrected the speaker) and 4 statements concerning the statement competitions (the addressee helps the speaker to complete his statement).

Own communication management (0.40%)
The own communication management dimension contains only five statements concerning own error signalizing. The corrections were usually made without a particular highlighting, i.e. only one statement contained an additional apology. However, to emphasis on the modified value of the statement the correction is repeated: "*it is on your left, ... on your right .. it is on your right*".

Contact management (0.32%)
Contact management dimension counts only four statements regarding contact check (three statements) and contact indication (one statement).

7.3.4 Question-answering pairs in details

The frequency distributions of the dialogue acts is important for providing relevant data for further statistical modeling indicating which dialogue categories are more likely to occur in this particular conversational context. But on the other side, the frequency distribution does not help us understand the finesse and nuance of human talk. Therefore, in this section we will take a deeper look into the core structure of the dialogues focusing on the question/answers pairs exchanged along the scenarios. We will analyze the visitors' and receptionists' statements highlighting problematic issues for automatic labeling and further processing.

Visitor's utterances

<u>Scenario 1</u>
1) Inform receptionist about appointment
When informing the receptionist about their appointment most of the visitors (six persons) just simply stated: *"I have an appointment with* [name]". Three persons explained the motivation behind their visit: *"I am here to look for/to meet/I am looking for/ [NAME]"*. One person used a more polite request form: *"I would like to meet* [name]". Another person used an imperative form: *"I want to meet* [name]". From the data collected we observed that it was not very common for visitors to take the initiative to present or give information about themselves without being asked - only two persons mentioned their names and another one mentioned his status: *"I am a visitor"*. Also, a visitor mentioned his country of origin. Therefore, the receptionist had to ask the visitor's name, usually right from the beginning (*"May/Can I*

have your name sir?"/"May I know who is there on line?"/"And you are?"). Only two receptionists asked the visitor's name while they were already talking with a third person on the phone.

Statement	Function category	How often used
I have an appointment with [..]	info providing; indirect request	6
I am here to meet [...]	info providing (justification); indirect request	3
I would like/want to meet [...]	direct request (+-polite marker)	2

Table 7.4: *Functional categories of dialogue acts used for informing about an appointment*

Table 7.4 presents the functional categories of dialogue acts used for informing about an appointment.

Recommendation 1: Unlike in casual or phone conversations, visitors tend to start talking with the receptionist without introducing themselves. The same applies for the receptionist. In fact, the statistics confirmed the presence of self-introducing acts as being very low (2.52%). On the other side, in the case of appointments the receptionist would need to know the visitor's name. Since visitors at I^2R are often foreigners the use of a touch screen could be advantageous for the visitor to type his name in.

Further, all receptionists informed the visitor about the action they were going to take: they were going to call the staff member to announce the visitor and therefore, they would be pausing; the pausing was also expressed as a direct request (see table 7.5). Only in one case did the receptionist omit to give feedback and just performed the call after asking the visitor's name. We observed that statements concerning the intention to pause and action discussions information can be combined together or can be used independently.

Statement	Function category	How often used
Sure, I will give him a call	action discussion promise	4
Just one minute	pausing	3
Hang on	pausing; direct request	2
Please hold on for a while	pausing; direct request + polite marker	2
Let me call/give a call	pausing	2

Table 7.5: *Functional categories of dialogue acts used for informing about an call*

Recommendation 2: Discussions of own future actions implying a break in the communication flow are a norm for polite verbal behavior. Even if the actions are obvious this kind of statement should not be ignored in a human-robot dialogue scenario (see design guideline 4e) in section 2.4.

Right after finishing the phone call all the professional receptionists (five persons) informed the visitors about the result of their conversation: the staff member was waiting for the visitor on the eighth floor. In two cases the receptionist also indicated the staff member's office number or the fact that the visitor could take the lift to go there. Only the non-professional receptionist started giving directions first and then mentioned the fact that the staff member was waiting for the visitor.

2) **Ask for directions (how to get to the appointment)**
Interestingly, most of the people (eight) did not take the initiative to ask how to get to the person's office - even though it was explicitly indicated in the scenario - expecting the receptionist to address this issue instead. On the other side, most of the receptionists (four persons) did the same thing, that is, they did not start explaining how to reach the staff member's office until the visitor asked explicitly for this information. This behavior illustrates the principle of least collaborative effort, explained in section 2.2.1.

In most of the cases when a visitor did ask for directions they used an information seeking question (*"How can/do I get/go there, How to get to his office?"*); one person used a direct request: *"Can you direct me to his office?"* while another one expressed the question indirectly through an informative statement (*"I am wondering how to get to his office"*)- see table 7.6.

Statement	**Function category**	How often used
How can/do I get/go there?	info seeking (WH question);	3
How to get to his office?	info seeking (WH question);	1
Can you direct me to his office?	(direct) request	1
I am wondering how to get to his office	info seeking; (indirect) request	1

Table 7.6: *Functional categories of dialogue acts used for asking for directions*

Recommendation 3: Anticipating the users' informational needs is the mark of a good receptionist service. However, before providing directions it is advisable to make sure the visitors need the information.

Since the building where the experiment took place has a particular architectonic design, reaching the staff member's office was not a straightforward task, but required more detailed explanation. Here we found two different approaches: two receptionists made use of reference points visible to the visitor (*"Do you see a lift on your right?", "You came by the same way so you go out by the same way"*) while the others just explained step by step how to get to the 8th floor: *"What you need to do is to take the lift up to the 13th floor. You come out of the lift, take another lift ... opposite side ... you know ?...From 13th floor to 8th floor. You take two lifts"*.

When the explanations contained more than three steps most of the receptionists (four persons) summarized or repeated the most important information parts at the end (see design guideline 2b in section 2.4.1). Such repetitions and summarizations of key information have a mnemonic function. If the receptionist omitted to do so the visitor used short check questions to recall and feedback pieces of information relevant to complete the task (*"Level eight?", "I go to the 13th floor, right?"*).

Recommendation 4: It is recommendable for a robot receptionist to inform the visitor where he or she can meet the staff member, for example floor, tower, and so on. Directions should be provided only if requested. The explanations should be given in short sentences (*"you go out and turn right"*). Alternatively, the summarization could be displayed on the screen.

<u>Scenario 2</u>

A. Ask details about sky-garden

Here we identified two dialogue strategies used by visitors to start the conversation about the sky-garden:

1. Asking a direct information seeking question: *"Where is the sky-garden located?"* (three persons)

2. Performing an informative statement about their current state of knowledge regarding the sky-garden (six persons): *"I heard/read in the Internet about the sky-garden"*

The second strategy has a double function: apart from having informative value about the knowledge status of the visitor the statement is a subtle invitation for the receptionist to start his/her explanations about the sky-garden and at the same time a hidden question. If the receptionist did not respond accordingly the visitors often added an explicit question formulated as a direct or indirect request (see table 7.7).

Statement	**Function category**	How often used
Where is the sky-garden?	info seeking (WH question)	3
I heard/ read in the Internet about the sky-garden	info providing - hidden question	6
How do I go/get to the sky-garden?	info seeking (WH question)	2
I don't know how to get there	info seeking - hidden question	1

Table 7.7: *Functional categories of dialogue acts used for asking for building facilities*

The handling of hidden questions consists in differentiating between the literal and pragmatic meaning of the utterance. Statements such as, *"I don't know how to get to the sky-garden"* or *"I heard about the sky-garden"* are literally acknowledgments of the visitor's current state of information about the sky-garden. From a pragmatic point of view the statements are indirect requests for information.

Recommendation 5: Hidden questions need contextual semantic disambiguation. A possible solution could be a response strategy involving explicit feedback on the recognized keywords, such as *'sky-garden'*. The feedback would elicit a follow-up direct question about the sky-garden.

When asked about the sky-garden most of the receptionists (four persons) started their explanations by indicating how many sky-gardens there were in the building and where they were located.

Almost all receptionists (five persons) explained to the visitors how to reach the sky-garden, however the explanation approach was not the same: in some cases the receptionists tried to give more detailed explanations related to the design of the building and how the sky-gardens were linked together (see table 7.8); such explanations were in some cases too long and contained redundant or less relevant information which confused the visitors, leading to additional questions. For example, when the building's towers were mentioned, the visitors got confused about their current location; also the garden linking raised the question why the sky-gardens are separated in three different floors (see table 7.8).

Recommendation 6: Each new informational element introduced in the explanations must be presented in such way that would minimize unexpected questions.

Building design	Sky-garden locations
R: *The sky-garden is on the roof top .. ah not in this building but at the North Connexis tower at the Symbiosis ... floor you can take ...just take the lift up to the top, to the last* **V:** *Symbiosis? How do I get to there?* **R:** *Ok, so you go all the way down.. you keep .. ah ... I mean if you want to go to the North Connexis Tower up to the sky-garden then you keep there .. otherwise opposite you will see Symbiosis. You just proceed your way to the lift and then you press the top button* **V:** *I am confused. What building is this one here?*	**V:***How come its being separated into three different floors?* **R:** *These are the access locations err and because of the err...sky bridge okay? .. And you know the link will only be from the 12th, 18th and the 21st floor. So you have to get there on either from these floors.* **V:** *Umm okay. Is there only one location or three locations?* **R:** *There are three* **V:** *Ah.. okay three? Hm .. I don't quite get it ..* **R:***Yeah you have to gain access over these links via the sky bridge and the sky bridges are on these two.... err few floors*

Table 7.8: *Examples of two inefficient dialogue strategies*

In the other cases the receptionist adopted a more effective response strategy by narrowing down the explanations and asking directly which sky-garden the visitor wanted to see (*"Which level do you want to go to Ma'am?"*(one case). In two cases the receptionist waited to be explicitly asked about it (*"How do I go to the sky-garden that is on the top most floor?"*). This strategy could also be beneficial in a human-robot scenario since it minimizes the risk of confusion by splitting the response into small informational elements (see table 7.9).

Recommendation 7: The answers should handle only one piece of information: giving too many details is not necessarily a good strategy, as it might cause confusion.

Narrowing down 1	Narrowing down 2
V: *Erm, I need to go to the sky-garden.* **R:** *Okay, sky-garden .. there are three levels, level 12, 18 and 21.* **V:** *Erm, the one that has the swimming pool.* **R:** *Okay, the swimming pool is at the level 24.*	**V:** *Ah .. I read from the Internet that there is a sky-garden in this building* **R:***Yeah there are three sky-garden here* **V:** *3 sky-gardens...* **R:** *Yeah at level 12...* **V:** *Ok.* **R:** *"Level 18 and level 21*

Table 7.9: *Examples of two efficient dialogue strategies*

Besides explanation and locations two receptionists added rhetorical information about the beauty of the sky-garden and suggested which one might have the better view. One receptionist also explained why she recommended a certain direction to be followed, as being the shortest and the least complicated.

Recommendation 8: Giving a reason for a particular recommendation might be useful in a human-robot scenario since, intuitively, it could help to build trust: knowing why a certain option is better than another one can make visitors feel more confident.

Visitors also asked additional questions concerning which facilities could be found in the sky-garden or whether the sky-garden could be accessed with a normal visitor pass. In a human-robot scenario such information should be taken into account.

B. Ask details about the swimming pool
Most of the visitors (seven persons) asked information about the swimming pool using a propositional question: *"Is there a swimming pool/facility?"*. Four visitors used a similar hidden question construction: *"I heard there is a swimming pool"*.

Here it is important to notice that even though the question formulation was propositional the answer expected was not only a *'yes'* or a *'no'*: visitors wanted more elaborate information about the facility, such as how to get to the swimming pool, whether they could use the swimming pool, if there was a fee to be paid, how much it would cost or whether the host institution had a membership or discount card that visitors could make use of.

While answering the question all receptionists indicated first the location of the swimming pool and most of them (four persons) explained how to get there without being explicitly asked. Since the swimming pool did not have free access the receptionists additional task was to try to find a solution or to give advice *"you can just go in and let them know that you are there to take a look and they will be happy to show you around"*.

While confronted with a restriction the visitors often asked negative check questions, such as: *"So then I can't go inside?/So I won't be able to use the swimming pool?"*.

Recommendation 9: The facilities' descriptions should be prepared in detail. In case of entry restrictions alternative solutions should be offered to visitors.

During this scenario - which was the longest and richest in question/answer pairs we remarked the receptionist's tendency to put emphasis on particular important informational cues related to the sky-garden or the swimming pool. The emphasis was achieved by accentuating words, raising eyebrows and combining them with iconic deictic or beat gestures (see figure 7.2 (A)) or bringing the head or body forwards while pronouncing a certain word (see figure 7.2 (B)). Additionally, head nods, as well as closing the eyes half-way were used for emphasis.

<u>**Scenario 3**</u>
A. Ask for assistance to call a staff member
Asking for assistance to call a staff member was done by expressing the wish to make a phone call. Additionally, visitors explained the reason behind their request: to thank the staff member for his hospitality, to say good bye, and so on, *"I would*

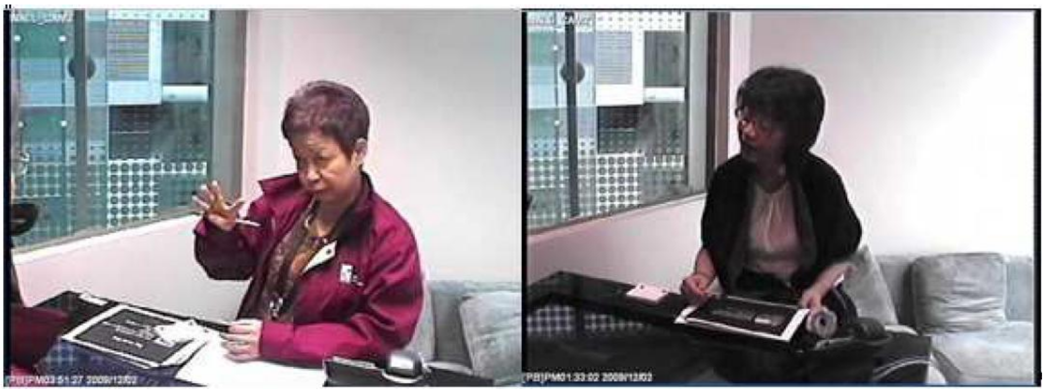

Figure 7.2: Emphasis expression by two receptionists: through gestures left (A) or through body posture forwards right (B)

like/need/ to talk/to contact to Dr. Li Haizhou ..to express my appreciation for his hospitality". Three persons also added to this statement a direct request to be assisted by the receptionist in making a call (*"Can you put me on call/help me contact?", "Would you able to assist me in making a call?"*).

Interestingly ten people out of eleven explained to the receptionist the reason of their request, even though it was not explicitly required by the scenario. A possible explanation could be that in addressing a request people might feel the need to explain their underlying motivation.

When assisting with calls to reach a staff member, the receptionist's answers were straightforward: *"ok, sure", "yes, no problem".* Only two receptionists added an additional statement concerning a break in communication flow, that is to say, they were going to be busy making the phone call.

B. Ask for assistance to call a taxi

While asking for assistance to call a taxi most of the visitors (six persons) used a direct request: *"Can you help me to book a taxi to get to the airport?".* Three people started first with an informative statement to motivate their request *"I need to go to the airport, "I am rushing"*, followed by the help request. Only two persons stated the informative statement without adding any request.

After agreeing to make the phone call most of the receptionists (four persons) asked the visitor to hold on for a while (pausing statement) or even to take a seat on the couch, signalizing it might take time to make the phone call. All receptionists wrote down the taxi number and the estimated arrival time for the visitor suggesting that they should wait in the lobby. One receptionist also asked the visitor to which airport terminal he or she was going while another receptionist added rhetorical information about the quality of the taxi service.

Recommendation 10: For booking a taxi in a human-robot scenario, information on the taxi number and estimated arrival time can be displayed on the screen. Asking additional address details, such as airport terminal is optional.

7.3.5 Facial and gestural expressions

Gaze behavior

Mutual gaze behavior was found to be particularly common in 90.50% of situations where people were exchanging question/answer pairs.

At the beginning of the conversation people also tended to look at each other mutually (90%). However, we found that this behavior was less frequent towards the conversation's end, where mutual gaze occurred in only 48.50% of the cases. Also when a person was referred to - directly or indirectly (e.g. when the receptionist was announcing the visitor's name over the phone) the gaze was found to be mostly mutual (98%).

Gaze shifts occurred often during explanations, especially when visitors had to pay attention to details or to follow the direction pointed by the receptionist (68.50%). In these cases we observed that participants even change their body orientation following the imaginary path pointed by the receptionist (body referencing). Gaze shifts were also common when people couldn't remember a name/word (98%), during feedback (82%) or in uncomfortable situations, such as when correcting a mistake or giving a negative answer (75%).

Hand gestures

Most of the gestures performed (152) during the scenarios were deictic: 33% were gestures pointing the way to a location, 20.40% were pointing gestures to information presented on the screen, 31.57 % were gestures pointing to imaginary objects or people, while 15% were gestures pointing to real, i.e. present in the room, objects or people. We noticed that in 32% of the cases when the receptionist was pointing to a location the visitor repeated the gesture. This gestural repetition had a double function: it gave a visual feedback and, in the same time, it helped the speaker to remember the explanation, i.e. had a mnemonic function. Other gestures were beat (103), iconic (51), emblematic, mostly good-bye gestures (six).

Smiles

In social interactions smiles are often used as markers for polite behavior and readiness to engage in conversations. The more a person smiles the more friendly he or she is perceived to be. In the data we collected we found that the receptionists smiled 1.7 times more often than the visitors did, 64.63% versus 37%, as shown in figure 7.3.

Receptionists smiled in 81.52% of the greeting situations, accompanying the preclosing dialogue acts or thanks at the end of the conversation (63.63%), as shown in figure 7.3. Smiles were also frequently used in 75% of the situations where apologizing or a negative answer occurred.

Visitors smiled often at the end of the conversation while thanking for the receptionist's help (70%). Smiles were less frequently found at the conversation's beginning (36.36%). They also smiled when responding to an apology or giving a negative answer (in 83.33% of the cases). Smiles also were used in other dialogue parts; however, no particular trend could be observed.

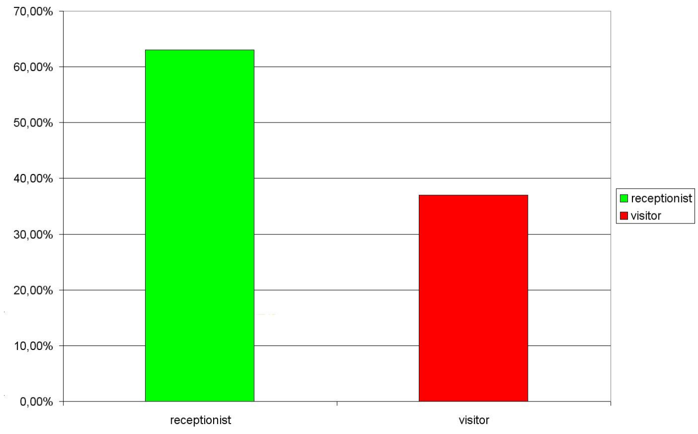

Figure 7.3: Frequency of smiles in our data collection

Head movements
35% of the total amount of head nods annotated were used to accompany verbal feedback. In fact, more than half (55%) of the non-verbal and overall positive feedback categories were accompanied by head nods. Participants also used single head nods to emphasize particularly important information clues (77.38%), to show confirmation or agreement (40.67%) or while asking positive check questions (56%). Single head nods were relatively common accompanying social obligation acts (31.25%), such as thanks act at the end of the conversation, self-introduction and welcoming.
We annotated only three head shakes to express negation and four head movements pointing to some direction. Head nods were used relatively infrequent (9%) as stand-alone gestural feedback intended as an expression of understanding, agreement or a confirmation that they person was paying attention.

Eyebrow raising
Eyebrow raising were used to put emphasis (43%) while correcting a mistake or highlighting important information cues. In some cases (25%) eyebrow raising was used to express readiness for listening to a particular question. In other cases (18%) eyebrow raising was attributed to situations where participants were unsure, i.e. in combinations with check questions or other statements expressing uncertainty. Eyebrow raising was also performed while asking a question (10%).

Frowning
Frowning was not very common in our data. In fact, we annotated only 17 occurrences of frowning. Participants frowned to show they were paying attention to a question (53%), to show concentration while giving explanations (23.52%) or to express uncertainty (two cases, 11.76%). Also, in only two cases frowns were used during the presentation of unpleasant information, such as pointing to mistakes or misunderstandings.

Eyes & mouth
Other patterns, such as open mouth (three occurrences) or eye blinking (six occurrences), closing or widely open (four occurrences) could be observed, but their low occurrence frequency made difficult to relate them clearly to a particular dialogue act.

Body positions
Participants showed a tendency to move their body forwards while starting the conversation showing readiness to engage in conversation (in 50% of the cases) and paying attention (50%) or when starting to explain something (22.64%).
During valediction or after giving explanations both receptionist and visitors tended to move their bodies backwards (18.18% and 57%, respectively).

7.3.6 Interviews with visitors and receptionists

After the experiments we conducted a individual qualitative interview[9] with both visitors and receptionists. During the interviews we asked the visitor to judge the interaction with the receptionists in terms of expectations and receptionist's performance. Additionally, the visitors were asked to make improvement suggestions regarding the interaction with the receptionist and the experiment settings.
The interviews conducted with the receptionist focused on questions related to how realistic the scenarios were, how similar the tasks performed were with those encountered in their daily work and what qualities make a good receptionist. In the end they were also asked to make additional comments and improvement suggestions concerning the experiment.

Visitors
Concerning general expectations regarding interaction with a receptionist most of the visitors indicated they would expect a direct reply to their question and frequent feedback on future actions, especially in situations when the receptionist needs time to perform a call. Further, visitors would expect the receptionist to make them to draw their attention when important information is presented on the screen or when access restrictions are imposed. They would also expect the receptionist to be familiar with the building facilities and surroundings, including shops, restaurants or other places of interest located nearby, such as bus services, subway, etc. Additionally, the receptionist is expected to have some basic knowledge about the staff member working in the building. Other expected attributes of a receptionist were smiling, being nice, good looking, understanding the cultural background of the visitor, adapting to his needs, making an extra step to anticipate his wishes, and well spoken. Also, all visitors indicated they would expect the receptionist to initiate the conversation by greeting, making eye contact, smiling and offering his services.
Most of the visitors were satisfied with the explanations received and found the

[9]The interviews can be found in appendix F.

receptionist quite helpful. However, in terms of performance some differences between the receptionists were found. The visitors were asked to score the receptionists performance on a scale from 1 to 10 and to give a detailed explanation about their reason for their rating. Receptionists who got lower marks were criticized for failing to provide clear directions or for using unknown location denominations as reference points. Criticized was also the fact that many receptionists omitted to inform the visitors about building access restrictions. One receptionist was criticized for speaking too fast and with too strong local accent, a fact that could create understanding problems for foreign visitors. On the other side, receptionists who were fast in providing the information and additionally, took initiative to help the visitor (e.g. in finding a solution on how to access the swimming pool) were better scored.

As improvements, four visitors suggested the use of maps - on paper or displayed on the screen. The map could complement the receptionist's explanations and help visitors to get a better overview of the building.

Some of the most frequently mentioned characteristics of a good receptionist were friendliness - smiling, readiness to help and to answer all the questions, nice voice, fast response, efficient, helpful, polite, good-humored, should recognize the visitor and welcome him back.

Regarding the display screen, the visitors' opinion were split: half of the visitors did not notice it or noticed it only later when the receptionist started making phone calls; they considered the display screen not particular useful, presenting often redundant information, i.e. the same as the receptionist, and distracting the visitors preventing him from paying attention to the receptionist's explanations. In contrast, the other half enjoyed the information presented on the screen, e.g. the staff member's phone numbers and pictures, considering it useful for those who for example don't know how the staff member looks like. Some visitors (three persons) expressed the wish to have an interactive touch screen and criticized the screen layout: the slides were too dark the screen too big and visual information about the building facilities was missing.

Almost all visitors - except for one - criticized the horizontal position of the screen as being the main reason why most of the visitors did not notice it. They suggested changing the screen position to a 45 degree angle.

Receptionists

The receptionists confirmed the scenarios were realistic and the tasks quite similar to those received in their daily work. The real duties of our receptionists were to guide people through the building, receive and transfer calls, make taxi arrangements, offer advice about restaurants and serve tea or coffee.

Among the most important qualities of a receptionist were mentioned the capability to represent the company, to be patient, to be friendly and polite, to make visitors feel comfortable and welcomed, to use some moderate humor or some sentences in the visitor's mother tongue, but not to talk more than required - making the own contribution not more informative than required is also mentioned by Grice in the quantity maxim, see section 2.2.2.

Most of the receptionists (four persons) welcomed the display screen as a useful tool for explanation and displaying difficult visitors' names. They also suggested changing the screen position to a 45 degree angle and using the screen for displaying building and floor maps.

7.4 Summary

In this chapter we have presented an empirical study performed with human actors playing the roles of a receptionist and a visitor. The goal of the study was to determine how people communicate with each other in task-oriented situations and how this knowledge can be used in a human-robot scenario in order to prevent miscommunication, minimize the risk of additional questions and present information in an efficient way. The experiment took place in a controlled environment simulating the conditions in which the robot would later perform.

Receptionists and visitors interacted with each other based on three scenarios. The scenario tasks were similar to those encountered by professional receptionists in their daily work, as later confirmed during the qualitative interview.

To annotate our data we used the DIT++ taxonomy for dialogue acts and the MUMIN coding scheme for mimic, gestures, gaze and body positions.

Concerning the dialogue acts we found that information providing, feedback and social obligation acts had the highest distribution among all dialogue acts exchanged. The most common information providing acts were answer elaborations and explanations while the overall positive feedback statements were the most commonly used feedback form. Among the social obligations thanks acts were the most frequently used; thanks were often used to replace valediction acts at the end of a conversation. In contrast, greetings were always used in dyads at the dialogue's beginning.

The data that we gathered and categorized in this manner can be useful for modeling appropriate dialogue strategies, developing language models and extending the vocabulary with relevant domain words for the robotic application.

Analyzing the dialogue flow we observed that self-introducing acts are seldom performed.

However, knowing the visitor's name in an appointment scenario is critical for the receptionist's task: the receptionist would need to ask the visitor's name increasing the risk of miscommunication, especially because many visitors are foreigners. Here, the use of a touch screen would be advantageous.

Discussions of future actions implying a break in the communication flow should be included in the human-robot dialogue design. The importance of this type of dialogue acts was also highlighted during the interview with the visitors.

Due to a particular building design the receptionist might need to explain the visitor how to get to a certain location. However, before giving the visitors directions it is advisable to ask them first whether they need the information.

The more efficient approach to explaining direction was found to be using visible reference points, such as the lift, the corridor, etc. Receptionists who used unknown reference points (e.g. Symbiosis tower) got caught in dialogue deviations and received lower performance scores. Also, when the explanation contained more than three elements the receptionist repeated or summarized his statements at the end. This approach was welcomed by most of the visitors during the interview.

When visitors performed information seeking acts regarding building's facilities they used not only WH- or propositional questions but also hidden questions disguised in informative statements. Such questions need contextual semantic disambiguation. In a human-robot scenario a possible response strategy could involve an explicit feedback statement on recognized keywords, i.e. "sky-garden". The feedback would elicit a follow-up direct question about the sky-garden. Also, propositional questions should not be treated as such, since the answer expected by visitors would be more than simply a 'yes' or a 'no'; as the interviews confirmed visitors would expect here more elaborate information about the facilities, especially in cases where there are access restrictions. In such cases, the visitors would expect the receptionist to provide a solution. Also, while being confronted with a restriction the visitors often asked negative check questions.

Each new informational element introduced in the explanations should be explained in such a way as to minimize the raising of unexpected questions. Also, answers should handle only one piece of information: giving too many details is not necessarily a good strategy since it might cause confusion (see Grice's maxim of quantity, section 2.2.2). For example, concerning information related to the sky-garden: a good strategy adopted by one of the receptionists to prevent miscommunication was to indicate the access location of the sky-garden asking explicitly which floor the visitor intended to go, in order to shorten further explanations.

Providing a reason for a particular recommendation might be useful in a human-robot scenario since explanations help interlocutors to understand the speakers' intentions. Intuitively, we believe that such a strategy can help build trust.

Concerning gaze behavior, we found that people tend to gaze at each other more often when a conversation starts, when question/answers statements are exchanged or when a person is referred to. Gaze shifts occur when people need to remember information, during feedback, or in uncomfortable situations, such as when giving a negative answer or apologizing. Our findings are in line with other studies in the literature [41], [48].

Most of the gestures performed during our experiment were deictic. We observed that people showed a tendency to repeat deictic gestures in an attempt to give a gestural feedback, but also to retain the direction pointed to.

Smiles were more common for receptionists than for visitors and occurred more frequently at the conversation end or in unpleasant situations, such as apologizing or giving negative answers, as indicated by other studies in the literature [39].

Positive feedback statements were in half of the cases accompanied by head nods. Single head nods were also used to emphasize content. Less frequent were head nods used as 'stand-alone' (i.e. not accompanied by a verbal statement) gestural feedback or agreement.

Eyebrow raising was mostly used to emphasize content such as information cues or mistakes, as indicated in [37]. Frowning and other eye and mouth movements appeared less frequently in our data.

In uncomfortable situations both visitors and receptionists tended to shift the gaze, to smile or to frown.

Both receptionists and visitors showed the tendency to come forwards when starting the conversation or the explanatory part within the conversation, as found in other studies in the literature [40]. They moved their bodies backwards mostly during feedback or valediction acts.

The interviews conducted with the participants confirmed and complemented our observations about the dialogue strategies used in our experiment. Additionally, they provided useful information of what is expected from a good receptionist, such as to have knowledge about the surroundings and the staff member working in the building, to be friendly, good-humored, well spoken, and polite, to speak accent-free English and to initiate the conversation. The interviews also give us feedback on the screen position which should be at an angle of 45 degrees. The screen should also provide touch screen capabilities and additional visual information, such as facilities pictures.

During the experiment we gathered a corpus of dialogues that will serve for developing dialogue strategies for a human robot scenario. The analysis helps us not only to identify efficient dialogue strategies, but also to detect ill-constructed answers that should be avoided in the human-robot scenario, such as unclear or meaningless explanations, answers which do not fit the questions or information that was not requested. Based on our analysis we have developed a set of recommendations for the configuration of the future human robot dialogues. In the following chapter we will show how these recommendations have been incorporated in the dialogue design of a social robot.

Chapter 8

Olivia and Cynthia: effects of empathy, humor and voice pitch

In this chapter we will present another study with the social robot Olivia. The study will explore the effects of two interpersonal skills - empathy and humor - in interaction with a social robot receptionist. We created two robot characters by manipulating the voice pitch: Olivia, the more exuberant, empathic and humorous robot with a higher voice pitch and Cynthia, the calmer, more serious and more reserved robot with a lower voice pitch. The manipulation also allowed us to study the effects of the voice pitch on the user perception of the robots. Thus, the chapter will contribute to our first research questions concerning the effects of voice characteristics (voice pitch, RQ 1) and language features (empathy and humor, RQ 2) on the evaluation of a conversational interface. The content of this chapter was published in [203].
This chapter is structured as follows: section 8.1 will offer a short introduction to the study, section 8.2. will present an overview of relevant background work concerning empathy, voice pitch and humor effects, section 8.3 will discuss the experiment design while and 8.4 will focus on the questionnaire design. Section 8.5 will discuss the results gathered from our study. Finally, section 8.6 will end the chapter with a summary of the main results.

8.1 Social robot receptionists

Social robots are becoming increasingly popular as receptionists and office companions. Wakamaru [204], a yellow midget robot, developed by Mitsubishi, can greet in four languages and is able to escort visitors to different destinations within a building. Wakamaru has been on the market since September 2005 and in the meantime has even found a 'job' at a temporary staffing agency in Japan as dispatch worker. Ms. Saya [205] is a humanoid robot who 'worked' as a receptionist for Tokyo University of Science for a period of two years. Ms. Saya was able to do some basic conversation, guiding visitors to location of interests and luring university graduates to sign up to courses.

Other notable examples are Valerie [206], Tank [207] and Hala [208], all social robot receptionists developed within the Roboceptionist Project[1] at Carnegie Mellon University. These robots were placed at the entrance of the building giving directions, answering phone calls, looking up weather forecasts and giving information about local events, using English or even Arabic (Hala). Apart from their conversational capabilities, the robots represent different characters, each one having a different personal story revealed to visitors through daily interaction over a longer period of time.

The purpose of these robots is not just to inform, but also to be agreeable to the visitors, encouraging them to return in the future. This approach is adopted in real life by human receptionists who are responsible for giving customers a good first impression of the organization they represent.

What qualities are essential for being a good receptionist? Apart from being efficient, multitasking, flexible or computer literate, most of the online adverts for receptionist jobs emphasize excellent interpersonal communication and customer service skills as absolutely essential requirements. Thus, a receptionist should be able to communicate effectively with the visitors, to listen actively to their problems showing empathy and providing help, to be polite and friendly and to possess an appropriate sense of humor. Most of these qualities were confirmed during the interviews with the human participants from our previous experiment (see section 7.3.6).

In this chapter we will explore the effects of two interpersonal skills, empathy and humor, in interaction with a social robot receptionist. We chose these two social abilities for our experiment on one side, because of their proven positive effects in human social interactions as promoting satisfying and healthy interpersonal relationships [209] and, on the other side, because they have not yet been studied in combination with a social robot in a receptionist scenario.

Following the Roboceptionist example, we created two robot characters by manipulating the voice pitch: Olivia, the more exuberant, empathic and humorous robot with a higher voice pitch and Cynthia, the calmer, more serious and more reserved robot with a lower voice pitch.

The experiment was divided into three parts: the first part focused on exploring the robot's empathic reactions on users (scenario 1), the second part was only concerned with the voice pitch manipulation (scenario 2 and 3), while the third part studied the effects of humor in combination with voice pitch manipulation (scenario 4 and 5).

8.2 Related work

8.2.1 Empathy

Empathy is described as the capacity to react emotionally when perceiving another person experiencing, or about to experience an emotion [210]. Thus, expressing

[1] http://roboceptionist.org/project.htm

empathy involves taking perspective, understanding nonverbal cues, sensitivity to other's affective state and communicating feelings of care and desire to help appropriately [17].

Research regarding empathic agents is divided into two categories: one concerns agents simulating empathic behavior towards users - as the case in our experiment, the other concerns agents fostering empathic feelings in users [211].

Many studies in the past have shown the overall positive effects achieved by agents and robots expressing empathy towards users. The study by Brave et al. [212] showed that modeling empathic emotion in an agent was found to increase the positive ratings concerning its likeability and trustworthiness. Additionally, the empathic agent was also perceived as more caring and more supportive. Other studies have also reported that expression of empathy in agents reduces frustration and stress [213], increases user's satisfaction [214] [215], comfort [216], engagement [214] and performance in task achievement [217]. Pereira et al. [211] experimented with an iCat robot displaying empathic and neutral behaviors towards chess players. The results showed that the chess player exposed to the robot's empathic behavior perceived the robot more as friend than the other player did. On the other side, other studies Cramer et al. [218] and Ochs et al. [219] showed that the empathy expression produced positive effects only in situations when it correctly matched the users' affective state. Hence, incongruent emotional responses can lead to negative user rating of the agent.

However, a full overview of the empathy effects on user's perceptions of the robot is not available yet. Factors such as interaction environment, the subject's culture, interaction style or task type might play a role in the way empathy is perceived and reflected on the agent's/robot's ratings [220].

8.2.2 Humor

Humor is pervasive in human social relationships [221] and is one of the most common ways to produce a positive affect in others. Research studies have shown that innocent humour increases likeability and interpersonal attraction [222], [16], boosts friendship and trust [223] and fosters social cohesion [224].

Humor can be spontaneous or it can also be used deliberately, not only in conversations with friends but also in more formal environments[225]. For example, the use of humor in teaching stimulates students' attention, promotes comprehension, retention of information and more positive feelings towards the content [226]. Also, in the work place, humor turned out to be a successful method to alleviate stress [227], reduce tensions, encourage creativity [228] and improve team-work and socialization [229]. The presence of humor - which seems to be correlated with more expressive, self-confident and extrovert personalities [230] - was proven to increase the trust in online negotiations [231]. Even large corporations such as IBM, EASTMAN Kodak and AT&T hired humor experts to help improve teamwork, stimulate creativity and motivate employees[232].

Despite the positive effects of humor in work situations, scientifically proven by decades of research, the HCI field holds a rather negative view about the use of humor in interfaces. The general trend in interaction design is to develop interfaces that increase task performance on one side, and minimize task duration, learning time and error rate, on the other side [233]. Thus, the use of humor would just distract the users from their tasks, increasing the total competition time and therefore, would not fit in the line of thoughts mentioned above.

However, an exception seems to be the artificial intelligence and natural language processing communities where computational humor is a well-established branch. Here, the research is split in two directions: the first one concerns the automatic generation and/or recognition of jokes and humor. The progress in this direction is relatively slow due to the complexity of the humor phenomenon, which is highly reliant on the context. Some successful implementations of language based humor interfaces already exist, [234], [235] but the development of fully automatic humorous machines that spontaneously produce jokes and react to humor might still take a while.

Studies belonging to the second research direction investigate the effects of preprogrammed humor in interaction with artificial entities. Morkes et al. [236] showed that participants who received humorous comments during the interaction with a computer, rated the system as more likeable, more competent and much more cooperative. They also smiled and laughed more, behaving much more socially during the experiment. Their experiment demonstrated that, contrary to the traditional HCI view, humor enhances the users' experience with the system without distracting them from their tasks. These findings also suggested positive effects of humor on the overall system usability.

The study of Huan and Szafir [237] investigated the effect of humor in an agent-based educational context. Their results showed that humor improved significantly the instructor likeability regardless of whether they were human or robotic. However, no effects were found on the content memorization. Another study by Dybala et al. [238] performed with conversational agents showed that test users evaluated the humorous agent as more human-like, funny and likeable. In general, the humorous agent received much higher scores as compared with the non-humorous one.

Some other studies, such as the one by Babu et al. [239], suggest that the use of humor by a virtual receptionist agent may play a major role in engaging users in social conversations, since 50% of the social conversations between users and the agent Marve contained jokes.

8.2.3 Voice pitch

In human social groups, sensitivity to voice and language cues has always played a critical role in evolutionary history [124]. In face-to-face communication, the voice conveys the intended message and simultaneously contains highly relevant cues for social interactions. Such cues point to the speaker's gender, age, personality, emotional state or place of origin and enable socially intelligent individuals to decide who to like, who to trust and who to mate with.

A very important characteristic of the voice is voice pitch: how high or low the voice is. It is determined by the fundamental frequency. Average pitch values for male and female voices are respectively 120 Hz and 210 Hz [240].

Voice pitch was found to have influence on the perceived attractiveness of a person. The study by Riding et al. [241] showed that women found men with high-pitched voices significantly less attractive than men with medium or low-pitched voices. For female voices opposite results were found. On the other side, the study by Colin and Missing [242] showed that men found high-pitched female voices more attractive. According to Jones [243] women also show a preference for high-pitched female voices. Further, voices rated higher in attractiveness are associated with more favorable impressions of overall personality and even with a higher degree of similarity between the test participant and the rated voice [14].

A very recent study [244] associating voice pitch with the retention of content in long term memory found that both high- and low-pitched voices led to better results than medium-pitched voices. This result was independent of whether the speech samples represented natural or manipulated voices.

Pitch, pitch range, volume, and speech rate are the four fundamental characteristics of the voice that indicate personality [245]. People who talk loudly and rapidly, in a high pitch, using a wide pitch range are most likely extrovert, while introvert people often speak slowly with a soft, deep, monotone voice [124]. These four voice characteristics appeared to be sufficient to model the personality of synthetic voices [124], [246]. In [124] the test subjects identified the 'personality' behind the synthetic voice correctly and even used this knowledge to guide their feelings and behavior towards the product represented by the voice.

The voice pitch is also an indicator of a person's maturity [247]. Both for female and male voices, a lower pitch was found to correspond to a higher maturity. Moreover, a lower pitch was associated with lower levels of stress and more positive emotions [248]. A voice with a higher pitch, on the other hand, was perceived as more emotional and immature, indicating greater levels of emotional instability, weakness and psychological tension, again for both male and female voices [249]. Impression formation studies investigating the voice alone have found that more mature voices result in impressions of lower warmth and agreeableness [249] but higher dominance and assertiveness [250].

Surprisingly, research in the field of Human Robot Interaction (HRI) hardly ever focused on the psychological effects of the voice and the voice pitch. One study [251] investigated how people approached a robot when its voice was female, male, highly synthesized (robotic) or when no voice was involved. They found that approach distances were significantly higher when the voice was highly synthesized. Another study [252] investigated the effects of manipulation of affect expression in a robot's voice on task performance. The results showed that voice affect expression motivated people to perform better on a joint task when the robot was present in the environment. Apart from these two studies there are no others known to the authors dealing with voice manipulation in HRI.

8.3 Experiment design

8.3.1 Designing empathic reactions: I can feel what you feel!

Empathic reactions emerge as a consequence of interpreting the user's internal affective state. In our experiment, we tried to artificially induce an affective state by asking the user to imagine herself in the following situation: she/he is a famous scientist traveling from China on business purposes; on the way from the airport to the I^2R building the user experiences a quite unpleasant event: in the taxi she/he forgets a bag containing all his/her important documents, such as the passport, wallet, mobile phone and secret work information sheets. Tired and upset the user runs to the lobby asking for help.

There is, of course, no guarantee that the scenario would induce the desired effect in the user. But using a scenario creates the same conditions for all test participants and translates the participants - at least mentally if not emotionally- in a situation where an empathic response from the robot would not be perceived as inappropriate. Thus, it should also overcome possible problems encountered in [218] and [219] concerning incongruent emotional responses leading to a negative user rating of the agent.

A widely popular model of emotions among computer scientists is the OCC model [253]. The model describes a hierarchy of 22 emotion types. In our scenario we used two of them: feeling 'sad' for the visitor's loss - which triggers the robot's prompt reaction to help recovering the bag from the taxi - and feeling 'happy' once the bag is found (see table 8.1). Hence, the robot response strategy is to mimic the user's affective state (parallel empathy) and to offer immediately help (reactive empathy)[254].

Ideally, the expression of empathy involves three communication channels: facial, vocal-nonverbal and verbal [255]. Unfortunately, the robot we used for our experiments had a synthetic fabric face, which made the display of any facial expression impossible. Thus, our empathy manipulation had to rely only on verbal and vocal nonverbal cues. However, research has shown that it is possible to model empathy through verbal input only: Prendinger and Ishizuka [256] implemented empathic behaviors in a lifelike character companion attached to a web-based application with successful results. The companion aimed to help users to feel less stressed during a job interview by providing real-time, empathic, verbal feedback based on the user's elicited emotions.

For the robot's voice we used the TTS from Loquendo [2]. The TTS enables the robot to perform various vocal nonverbal cues, such as sighs, coughs, laughs and several interjections. To design the robot's empathic reactions we used sighs **(1)** and interjections in lowering tone **(2)** -to express sadness- and in rising tone **(3)**- to express happiness (surprise). We also varied the intonation using pre-built TTS features to stress certain words **(4)** with the purpose of giving them additional affective value (happy, sad, surprised, unsure, etc.).

[2]For more details see section 8.3.3

In terms of verbal empathic communication during the interaction, the robot expressed its 'feelings' of being sorry for the bag loss **(5)**, took the visitor's perspective while commenting on the event **(6)** and showed its 'understanding' concerning the visitor's internal affective state **(7)**. Further, the robot offered to call the taxi company assuring the visitor to do its best to find the bag **(8)**. When the bag was found the robot expressed joy **(9)** and announced the good news with a happy tone **(10)**. At the end of the conversation the robot once again reassured its satisfaction about the finding the lost bag **(11)**.

In the non-empathic version, the robot used a neutral voice without voice pitch variation, did not take the visitor's perspective nor any initiative to help and did not make use of affect to express joy or sadness. Thus, the dialogue between the visitor and the robot was in this version much shorter - for comparison see the table 8.1 below. Based on the results of past studies (see section 8.2.1) we speculate that

Empathic speech	**Non-empathic speech**
Visitor: I've just arrived and I forgot my bag in the taxi... **Robot**: Oh..**(1)** Ah..**(2)** Your bag is gone**(4)**!	**Visitor**: I've lost my bag in the taxi on the way here **Robot**: Okay
Visitor: Yes! **Robot**: I am so sorry to hear that **(5)**!	**Visitor**: Yeah .. ah.. **Robot**: I see your bag is gone!
Visitor: Yeah ... **Robot**: This is really terrible! There is nothing worse than to lose your bag together with your whole documents **(6)**!	**Visitor**: Right..so..ah ..I wish to get it back! -/-
Visitor: Indeed! **Robot**: You must be very worried **(7)**! Should I call the taxi company to help you find the bag? **Visitor**: Yes, that would be really very good...	-/- -/-
Robot: What taxi was that? **Visitor**: It was a blue one.	**Robot**: What taxi was that? **Visitor**: I think, it was a COMFORT taxi.
Robot: Was it a blue colour COMFORT taxi? **Visitor**: Yes, yes it was!	**Robot**: Was it a blue colour COMFORT taxi? **Visitor**: Yeah ...
Robot: Ok. Just give me a minute, please. I will try my best to find you bag **(8)**... *[calls]*	**Robot**: Ok. Just give me a minute, please. *[calls]*
Robot *(on call)*: [...] You found the driver **4)**? Oh **(2)**.. great **(4)**! He will bring the bag right now? Good **(4)**! Thank you so much **(4)**! Bye-bye!	**Robot** *(on call)*: [...] You found the driver? Okay. Thank you! Bye!
Robot: Yes **(4)**!I have good news **(10)**! **Visitor**: Oh!	**Robot**: Okay, okay! **Visitor**: Yes?
Robot: They found your bag! The taxi driver will be here in 2 min with your bag! **Visitor:** Thank you so much! **Robot**: You are welcome! I am happy that we managed to find your bag **(11)**!	**Robot**: They found your bag! The taxi driver will be here in two minutes with your bag. **Visitor**: Thank you so much! **Robot**: You are welcome!

Table 8.1: *Empathic vs. non-empathic dialogue*

an empathic social robot would be generally higher rated in terms of likeability, trustworthiness, behavior and social skills. Consequently, we expect the empathic robot to induce users more positive feelings towards the interaction and the robot self. This could eventually lead to better evaluation results in terms of enjoyment and overall interaction quality.

8.3.2 Designing humor: let's laugh a bit!

Humor is a powerful tool, however not every type of humor is advisable. For example sarcastic and irreverent humor can be perceived as hostile, especially when delivered by a female voice [124]. Ethnical, racial and sexist jokes have offensive connotations, intellectual jokes might not be always understood by all people, while self-deprecating jokes can affect the system image. Also dark, vulgar or toilet humor should be avoided, since it suggests an interlocutor with rather low class attributes [124]. The only type of humor recommendable is the innocent humor, also defined as inoffensive and light humor.

For our experiment, we chose two jokes of innocent humor type (see table 8.2). The first joke is a punning riddle -a question-answer joke based on the pun- while the second joke is based on incongruity, on a surprise element, i.e. the question gets an unexpected, out of place answer.

Joke 1	Joke 2
Robot: ... but you better don't go there *(i.e. to the sky-garden)* in the evening because there are a lot of mosquitoes attracted by the pond. But, if I have to choose between flies and mosquitoes definitively prefer mosquitoes. Do you know why? **Visitor**: No .. why ? **Robot**: Because mosquito can fly, but fly cannot mosquito! He, he, he *(laughing)* **Visitor**: He, he, he *(laughing)*	**Visitor**: Do you know if they have chicken feet[a] in the canteen? **Robot**: Hm ...I cannot tell you if they have chicken feet: they always have their shoes on! He, he, he *(laughing)* **Visitor**: Ah he, he, he (*laughing*) you are funny! **Robot**: Of course they have the best chicken feet in whole Singapore, at stall no. 4!

[a]Chicken feet is a Chinese delicacy dish

Table 8.2: *Jokes told by the robot during the experiment (scenario 5)*

We expect the humor to contribute to the robot's likeability, improving the users' perception of its social skills, including friendliness, helpfulness and trustworthiness. We also expect the robot to be perceived as more extrovert. Furthermore, we expect an increase of more positive user feelings towards the robot which will be reflected in their ratings concerning the overall enjoyment and interaction quality.

8.3.3 Voice pitch manipulation

Due to resource limitations we used for our experiment only one robot: Olivia[3]. However, to study the effects of different variable manipulations within the same group of participants two different robot characters were obviously needed. Thus, we 'created' a second robot character by modifying the voice pitch. We called the new character Cynthia.

To generate the robot's voices the TTS engine from Loquendo was used. The engine allows pitch variations between 0 and 100. A normal inflected voice would have a pitch of 50. Highly animated voices would display pitch values greater than 50 while a pitch value of 0 would produce a flat, monotone voice. For our experiment we chose 'Allison', a U.S. English accented voice. We set the pitch value for Cynthia's voice at 20 and for Olivia's voice at 70. All other voice parameters were tuned to the same values: timbre=50, speech rate=40, volume=50.

Additionally, to make the difference between the two robot characters even more explicit, we added two pigtails and a pink band to Cynthia's fabric hair (see figure 8.1).

[3]Technical details about Olivia can be found in section 6.2.2

Figure 8.1: *Olivia (left) and Cynthia (right)*

The differences in look and voice were explained as being related to different software modules used for Olivia and Cynthia. Such simple visual modifications, even if minimal, are required in order to keep the robot character consistent. Literature studies in the past warned about the negative effects of multiple voice personalities associated with the same system [124]. Thus, an experiment design in which we would alternate voice and look in a 2x2 matrix should be avoided, since it would confuse the participants about the robot's personality inducing unwanted negative effects.

Following the literature findings, our hypothesis was that a high-pitched 'female' robot would be perceived as generally more attractive, but also more emotional and possibly more extrovert. Additionally, the high pitch robot would achieve higher ratings concerning its perceived personality appeal and degree of perceived similarity with the test participants. Further, we hypothesized that an increased level of attractiveness and personal appeal would lead to a better interaction evaluation in terms of enjoyment and overall quality. On the other hand, we would expect the low pitch robot to be perceived as less pleasant but more assertive, stronger, leading to more positive users' feelings towards the robot and decreasing the interaction stress.

8.3.4 Prompts design

In the previous chapter we presented a study performed with human receptionists. The study allowed us to gather several observations that we could use to design the human-robot dialogue in the current experiment. In the following we will present some examples in which we incorporated the recommendations made in chapter 7, as well as the prompt design guidelines from section 2.4. There is a partial overlap between the guidelines and the recommendations, since both originated from user experience studies.

Recommendation 1: *Provide an adequate input option for the visitor's name* (see also section 2.4.2, guideline 4)

Due to a tight schedule[4] the display could not be replaced with a touch screen. Therefore, we used the same screen, but changed the slide design adding more pictures about the building facilities and floor structures. The lack of a touch screen did not allow the visitor to type in his name during the appointment scenario. Thus, the robot announced the visitor to the staff member using the following sentence:
Robot: "*You have a visitor here. She/he is waiting here at the reception area.*"
The visitor's name was not announced: however, in the dialogue design we considered the visitor's gender.

Recommendation 2: *Signalize breaks in the communication flow*
We incorporated dialogue acts signalizing a break in the communication flow, i.e. pausing statements and also a feedback on the robot's future action:
Robot: *"Please give me a minute. I have to check whether he* [the staff member] *is in his office. I will call him right now".*

Recommendation 3 and 6: *Ensure the visitor needs directions (3); always explain each new element introduced in the description (6)* (see also section 2.4.1, guideline 2b)
Directly after the phone call the robot informed the visitor that the staff member was waiting for him/her and asked whether s/he knew the way. The robot also included a new element in the description, but explained how this element, i.e. the South tower related to the current location.
Robot: "*Ok sir. Dr. Tan will receive you at level 8, in the South tower - the same tower where we are now. Do you know how to get there?*"

Recommendation 4: *When giving directions use short explicative steps. Information summarization should be printed on the screen.* (see also section 2.4.1, guideline 2b)
If asked for directions the robot used short explicative steps and visible reference points to direct the visitor. The most important steps (marked in bold) to follow were printed on the screen :
Robot: *"Okay, when you **go out** here you **turn right** and **go to the glass door**. Then you will see the lift. Take then the **lift up to level 13**. Then **get off, cross oppositely** and take another **lift down to level 8**. Mr. Tan will wait for you at the entrance at level eight."*

Recommendation 5: *Handling hidden questions with explicit feedback statements.* (see also section 2.4.1, guideline 2b)
In the case of hidden question the dialogue strategy incorporates an explicit feedback statement:
Robot: *"Do you want to go to the sky-garden?"*

Recommendation 7: *Answers should handle only one piece of information.*
The information concerning the sky-gardens locations, including pictures, was pre-

[4]Between the human participants study and the current experiment there was a time slot of only two weeks.

sented on the screen. The robot presented the access locations without mentioning tower names or complicated building's linking.
Visitor: *"How do I get to the sky-garden?"*
Robot: *"From here you can take the lift to go either level 12, level 18 or level 24."*

Recommendation 8: *Provide a reason for your recommendation.*
When asked which garden should be visited first the robot provided a reason for its recommendation.
Robot: *"The sky-garden on level 24 has a better view and it is located next to the swimming pool."*

Recommendation 9: *Offer solutions to restrictions.*
Since the access to the floor is restricted to working personnel the robot informed the visitor about the restriction and provided additional information to prevent the visitor waiting at the door:
Robot: *"Just in case he is not there I will give you his phone number. You should write down the number, just in case [number is printed on the screen]".*

Recommendation 10 (*Print taxi number and arrival time*) was not incorporated since in this experiment we did not have a taxi booking scenario.

8.3.5 Design of gestures, body movements and head turns

Since the robot's face was made of fabrics our receptionist could not display any facial expression. However, the robot compensated the lack of mimic through body movements, head turns and gesture.

When the visitor came into the observation room the robot lifted up its head, moved its upper body backwards and set its arms aside signalizing it had noticed the visitor's presence and it was ready to engage in conversation. When the visitor started talking the robot brought its arms together and shifted its upper body forwards to show it was paying attention to the visitor.

During the conversation the robot balanced its body slightly left-right creating the impression of adjusting its position. For confirmation the robot used head nods along with verbal statements.

In the scenarios when phone calls were performed the robot pushed its upper body forwards and bowed its head like the human receptionist to avoid "eye contact" while being busy on the phone.

When the robot was referring to the visitor while informing the staff about the visit it lifted up the head to simulate "eye contact" with the visitor, such as in the human receptionist scenario.

When information cues were presented on the table the robot position shifted backwards a bit, opened its arms and bowed its head to "gaze" at the screen. Also the robot's gaze shifted from looking at the screen and looking at the visitor.

The robot used deictic gestures to point to cues of interest on the screen, beat gestures to express empathy with the visitor when her/his bag got lost and referred to the swimming pool using iconic gestures, such as rotating arms, as a place where one can swim. Finally, the robot moved its upper body backwards to suggest the conversation was approaching the end, made a cute head tilt and waved its hand while saying ”good-bye” (emblematic gesture).

8.3.6 Experimental set-up

As mentioned before, the experiment was performed along 5 scenarios. In the first experiment part (scenario 1) we chose to split the participants into two groups: one half interacted with an empathic robot character (group A) and half with a non-empathic robot(group B). The decision was taken in order to avoid repeated exposure to an emotional stimulus -in this case the bag loss- which would create habituation and consequently, effect diminution [257]. For this particular scenario we chose only one robot character (Olivia) to play both roles. We did this because the groups were independent, i.e. they were not interacting twice with the same robot, thus, no inconsistent behaviors could affect the participants' perceptions. Also, additional voice pitch effects that could interfere with the expression and perception of empathy could be avoided.

In the second experiment part the test participants interacted with Olivia (scenario 2) and Cynthia (scenario 3). The robots' task in this scenario pair was to assist the participants in getting an appointment with an I^2R staff member. The scenarios were identical, apart from the staff's name.

Finally, in the third experiment round Cynthia (scenario 4) and Olivia (scenario 5) informed the participants about buildings amenities, such as the sky-garden, the swimming pool and the staff canteen. The information presented in the scenarios was identical. The only difference was that in scenario 5 two jokes were added - see table 8.3 for a better scenario overview.

Scenario	Variable	Topic
1	high voice pitch +/- empathy	calling taxi
2	high voice pitch	arranging appointments + giving direction
3	low voice pitch	arranging appointments + giving direction
4	low voice pitch - humor	informing building amenities
5	high voice pitch + humor	informing building amenities

Table 8.3: *Overview experiment scenarios*

The scenario order was randomized to exclude any potential biases that might arise from being exposed to one particular robot character before the other.

The robot was placed in a small observation room designed for usability experiments. Attached to the robot was a screen were information on name, office location, phone number, staff's picture or other building related issues could be displayed.

To ensure homogeneity between the trials we substituted the speech recognition module by a human wizard. The prompts were prepared in advance and played back by the wizard during the experiment. Equipped with headphones and sitting behind a dark screen the wizard team could see and hear the participants interacting with the robot.

Before entering the room the participants received a short briefing on the experiment procedure and the scenario for the current trial. After each interaction with the robot the participants were given a questionnaire to fill in.

The experiment was recorded by two cameras placed behind the robot and on the side.

8.4 Questionnaire design

In order to test how the variable manipulations affected users' perception of the robot and of the interaction we needed a questionnaire[5] that addressed these two questions.

If we take a closer look at human robot interaction from the perspective of our taxonomy of **QoCI** (described in section 3.3.2) we can identify the following factors that could shape the human's perceptions about the robot and eventually influence the robot's evaluations:

- Robot appearance appeal - how the robot presents itself to the outside world, i.e. how it looks, talks, behaves, expresses personality -(hedonic aspects)

- Task appeal - how enjoyable (hedonic aspect) and how effective, efficient and easy (pragmatic aspect) it is to interact with the robot

- Content appeal - how attractive is the content presented (hedonic aspect), how interesting (hedonic aspect) and good (pragmatic aspect) are the answers delivered

The effect produced by the interplay of functional and non-functional interaction aspects is reflected on the users' feelings during the experiment, as described in the same taxonomy of **QoCI** in section 3.3.2. We measured user feelings also measured in the experiments described in chapter 5 and 6.

Finally, a fifth dimension was included in the questionnaire which directly links the robot's personality and behavior with our variable manipulation concerning the empathy and the humor. We call this factor robot social skills.

Since there is no single questionnaire that includes all these factors, we designed our own survey using AttrakDiff [114], SASSI [113], ITU MOS [115] - these questionnaires are described in section 3.4.2.

We also incorporated in our survey items originating from the four Gricean conversational maxims [49] which address basic principles for an effective communication.

The questionnaire was built, similar to AttrakDiff - from which most of the items are taken - using semantic differentials on 7 point bipolar scales.

[5]The questionnaire used in this experiment can be found in appendix G.

8.4.1 Robot appearance appeal

The appeal of the robot's appearance addresses 4 different aspects comprised in the following subscales: look, voice, speaking style and behavior. Additionally, there is a fifth component - the robot's personality- which includes variables meant to assess different personality dimensions that could not be subsumed in one single subscale.

The robot's look subscale contains 5 items: presentable/unpresentable, professional/unprofessional, pleasant/unpleasant, inviting/rejecting and overall impression (AttrakDiff).

For the robot's voice appeal we used 5 - out of 8 - items from the ITU MOS questionnaire which measures the quality of TTS voices. These are: pronunciation, articulation, speaking rate, pleasantness and overall impression. Two remaining items, listening effort and comprehension problems are referred to on a further subscale addressing the answer quality, while a third one -acceptance- was reformulated into another item referring to the matching degree between the robot's voice and look.

For the third subscale, the speaking style, we used the AttrakDiff word pair stylish/tacky split to two different concepts that are more appropriate for a conversational speech style: polite/impolite and elegant/rough. The subscale also contains a 3rd item concerning the overall speaking style impression.

The robot's behavior scale includes mostly word pairs from AttrakDiff questionnaire, such as boring/funny, likeable/disagreeable. The word pair sympathetic/ unsympathetic was split in two more precise concepts: friendly/unfriendly and receptive/unreceptive. Additionally, we added another word pair especially relevant for a social robot receptionist: helpful/unhelpful. The scale also contains a statement about the overall robot's behavior impression.

Regarding the robot's personality, we included 4 questions inspired from the work of Kahn and De Angeli [258] concerning personality traits dealing with potency, social and emotional competence: extrovert/introvert, rational/emotional, strong/weak and assertive/submissive. The personality 'module' also included an overall statement about overall impression about robot's personality is.

All items related with the robot's appeal subscales hold only hedonic values, meaning that they belong to the non-functional aspect of the interaction.

8.4.2 Task appeal

The task appeal includes functional (i.e. with pragmatic aspect) as well as non-functional (i.e. with hedonic aspect) interaction aspects grouped in two subscales. The non-functional subscale refers to the task enjoyment and subsumes the word pairs: new/common, undemanding/challenging and simple/complicated (AttrakDiff). The functional aspect refers to interaction features emerging from performing the tasks. The interaction features originated from the SASSI questionnaire and concern: speech accuracy, interaction ease, transparency, speed, flexibility and usefulness.

8.4.3 Content appeal

For the non-functional aspects of the content appeal we chose 2 statements concerning the content interest (exciting/lame) and the content presentation (original/conventional). Both word pairs were taken from the AttrakDiff questionnaire. For the functional aspects we built up a subscale regarding the quality of the content delivered. The subscale includes variable pairs inspired from the Gricean conversational maxims and concerns the answers' informativeness, trustworthiness, relevance and clarity.

8.4.4 User feelings

This subscale contains items addressing the user feeling during the interaction with the robot. The following word pairs were included: motivated/unmotivated (AttrakDiff); annoyed/pleased, tense/calm, confident/insecure, frustrated/satisfied comfortable/awkward, in control/out of control (SASSI).

8.4.5 Robot social skills

The subscale was used in one of our previous studies [100] and contains 4 items referring to the robot's capabilities to socialize with humans and to express personality traits, emotions and humor.

8.4.6 Overall judgments and other personal details

The questionnaire also includes global ratings concerning the robot's overall aesthetic appeal (look, voice, speaking style) overall degree of entertainment, overall enjoyment and overall interaction quality. Additionally, we asked the users about the similarity degree between the robot and themselves and about how much they liked the robot in each particular scenario context. At the beginning of the questionnaire the users filled in their details concerning: gender, age, work background, knowledge/experience with social robots and speech recognition applications, expectations regarding the robot and personality type. The personality type was determined via an online Myers Briggs personality test[6]. This particular data was collected with the purpose of determining whether there are interaction effects between users' and robots' (perceived) personality.

8.5 Results and discussion

The experiment took one hour for each participant to be completed. The study was conducted over a period of three days. A total of 28 people participated in this experiment. 75% were people recruited from outside I^2R and they paid for participating

[6] http://www.humanmetrics.com/cgi-win/jtypes2.asp

in the experiment. The other 25% were I^2R staff members. The questionnaires of one participant could not be considered due to several wizard errors.

The majority of users (77.80%) were under 30 years while 22.20% were between 31-41 years. 63% were male and 37% were female. Concerning the ethnic groups the majority (81.48%) were Chinese, 9.40% Malay, 8.12% European/North American and one person was Indian.

44% of the participants had no knowledge of social robots, 45.20% had seen and/or read about them. A small percentage (7.40%) had interacted with a social robot while one person was involved in robotic design.

66.70% of the participants had no experience with speech recognition systems. 33.30% indicated having some occasional experience with speech recognition built-in applications for mobile phones, PC software, social robots or web interfaces but they use it relatively seldom; only one person reported a frequent use. However, half of the people who used speech recognition devices did it for real purposes, such as airport enquiries, database searches, ordering a pizza or interacting with speech handicapped people. The other half did it for testing or hobby purposes. Four people reported having had good experience with such devices, one was neutral and three reported negative experiences.

46.10% of the participants had high expectations, 19.20% had standard or medium expectations while 34.7% indicated to have low expectations.

Some of the test users indicated they would expect the robot to be able to recognize speech and gestures (21.72%), and to respond well (26.05%) and fast (2 people) to standard questions. Others expected the robot to speak clearly and to express a proper body language (2 people). While some participants expected the robot to be somewhat restricted to basic tasks and having limited capabilities (17.37%), others expected the robot to act as in real life (2 people) having a friendly disposition (2 people), a nice voice tone (2 people) and even to be able to understand the feelings of the users (2 people).

Concerning the users' personality profiles 44.40% of the participants were extrovert while 55.60% were introvert.

8.5.1 Scale reliability

Firstly, we performed a reliability analysis on our scales to verify the internal consistency of the items. All our scales show relatively high Cronbach values, between 0.760-0.903. Table 8.4 presents the Cronbach values for the robot's look (α=.856), robot's voice (α=.801), robot's speaking style (α=.782) and robot's behavior appeal (α=.903). Only one variable was removed from the robot's look subscale - 'professional look', which shows a low correlation (α=.214) with the other subscale items.

The subscales addressing the task and content appeal, as well as the user's feelings and robot's social skills also show high internal consistency: α=.760 for interaction features, α=.795 for content quality, α=.856) for task appeal, α=.846 for user feelings and α=.825 for robot's social skills (see table 8.5 and table 8.6).

Look	Items	Item-total correlation r	Speaking style	Items	Item-total correlation r
Items: 4 $\alpha = .856$	1. Presentable	.717	Items: 3 $\alpha = .782$	1. Elegance	.680
	2. Pleasant	.741		2. Politeness	.611
	3. Inviting	.557		3. Overall impression	.597
	4. Overall impression	.797			
Removed	_*Professional_	.214			
Voice	**Items**	**Item-total correlation r**	**Behavior**	**Items**	**Item-total correlation r**
Items: 6 $\alpha = .801$	1. Articulation	.544	Items: 6 $\alpha = .903$	1. Funny	.701
	2. Pronunciation	.568		2. Friendly	.782
	3. Match look	.488		3. Likeable	.843
	4. Speaking rate	.728		4. Receptive	.668
	5. Pleasantness	.407		5. Helpful	.681
	6. Overall impression	.642		6. Overall impression	.784

Table 8.4: Robot's appeal on look, voice, speaking style and behavior subscales

Content quality	Items	Item-total correlation r	Tasks	Items	Item-total correlation r
Items: 4 $\alpha = .795$	1. Trustworthiness	.471	Items: 3 $\alpha = .856$	1. New	.641
	2. Informativeness	.715		2. Challening	.888
	3. Relevance	.617		3. Simple	.889
	4. Clarity	.635			
Interaction features	**Items**	**Item-total correlation r**	**User feelings**	**Items**	**Item-total correlation r**
Items: 6 $\alpha = .760$	1. Speech	.523	Items: 7 $\alpha = .846$	1. Motivated	.575
	2. Transparency	.474		2. Pleased	.687
	3. Ease	.463		3. Calm	.606
	4. Fastness	.496		4. Confident	.645
	5. Flexibility	.590		5. Comfortable	.662
	6. Usefulness	.528		6. Satisfied	.632
				7. In control	.451

Table 8.5: Content quality, interaction features, tasks and user feelings subscales

Social skills	Items	Item-total correlation r
Items: 4 $\alpha = .825$	1. Express emotions	.621
	2. Express personality	.754
	3. Socialize	.399
	4. Express humor	.561

Table 8.6: Robot's social skills subscale

8.5.2 Empathy

No visible effects were found on our scale constructs between the groups, according to a Mann-Whitney test. However, on the level of each scale, we discovered significant differences between the two groups. Firstly, the robot was perceived indeed as being more receptive, i.e. more empathetic (p=.039)[7] to visitors' problems, more emotional (p=.002) in expression in group A (exposed to the empathic robot character) as compared to group B (exposed to the non-empathic robot character). Also, the overall behavior of the empathic robot was better rated by group A than by group B (p=.048). This shows that our manipulation might have been successful. Also, the participants from group A felt more confident interacting with the robot (p=.033) and the interaction appeared to them to be easier (p=.037).

On the other side, despite our relatively successful manipulation, our results indicated just minor evaluation effects: only the interaction was perceived as being easier, which probably led test users to feel more confident. This result can be directly linked to the robot's reactive empathic behavior - taking the initiative to help - and consequently might have influenced the users' perceptions of the interaction easiness.

Our manipulation did not trigger better evaluation results in terms of robot likeability, trustworthiness or social skills and users - apart from feeling more confident - did not have more positive feelings towards the empathic robot, as we expected. One explanation might be that exposure to empathy expression might need longer a time in order to achieve 'visible' results. The scenarios duration for both groups was below 2 minutes, i.e. 1.92 minutes for group A (empathy) and 1.42 minutes for group B (non-empathy) while most literature studies indicated a scenario durations of at least 5 minutes exposure per trial.

Another explanation could be that, even though the scenario intended to transpose the users in a particular mental situation, it did not achieve a change in their affective state. Thus, the empathy expression failed to achieve its main target: improving the evaluation results.

8.5.3 Voice pitch

To detect differences in ratings between the two robot characters, we performed a Wilcoxon signed rank test. The results show that our hypothesis concerning the attractiveness of the high pitch robot (Olivia) was correct: she was perceived as having significantly more appeal in terms of voice, (p=.038)[8] and overall aesthetic appeal (p=.017). The test participants also found her to have a more appealing behavior ($p<.001$), better social skills (p=.012) and a more pleasant overall personality, (p=.029), thus to be more agreeable than Cynthia. As expected, Olivia also appeared to be a more extrovert robot (p=.031) and users found her much more like themselves (p=.009). On the other hand the low pitch robot (Cynthia) was

[7]Median and U-values are indicated in appendix H

[8]Median and Z-values are indicated in appendix H.

perceived as being stronger (p=.046) but, contrary to our prediction, not more assertive. Also, Olivia was the one inducing more positive feelings in users (p=.027), and not Cynthia, as we expected.

The interaction features subscale was better rated in the 'Olivia' condition (p= .044). The content presented by the high pitch robot - even though identical (apart from the staff's name) with the one presented by its low pitch counterpart - was perceived as more exciting (p=.002). In general, the interaction with Olivia was considered much more entertaining (p=.002). As predicted, the overall enjoyment experienced with Olivia during the interaction was higher ranked (p=.006), as was the overall interaction quality (p=.001).

No significant differences could be found for the content presentation, content excitement, user feelings, tasks appeal, answer quality, or robot's look - the last one most probably due to the minimal difference between Olivia and Cynthia's look.

8.5.4 Humor

The ability to express humor was in scenario 5 perceived as being the highest in scenario 5 (p<.001), according to a Wilcoxon post-hock analysis with Bonferroni correction (p=.012). Most of the participants (81%) had a visible reaction (through smile and laughs) to the jokes. Thus, we can assume that our manipulation was successful.

Since the humor was manipulated together with the voice pitch, similar effects as in the previous part were expected. Thus, we found the same preferences concerning the robot's behavior (p=.001), voice appeal (p=.001), overall appearance (p=.002) and social skills (p<.001). Also, as before, significant difference concerning the interaction features (p=.005), overall entertainment (p=.001), overall enjoyment (p=0.28) and interaction quality (p=.042).

However, apart from the above results, there are also others that were not found in the previous experimental part and that could be linked alone to the humor manipulation. For example, speaking style of the humorous robot was considered much more appealing (p=.005). Also the tasks to accomplish with the robot were considered more appealing (p=.032). The robot personality appeared to be stronger (p=.008) but also more emotional (p<.001).

The humorous robot was also perceived as being more friendly (p=.001) and as having a more extrovert personality (p=.011). However, it was not considered more helpful or more trustworthy and neither the user feelings towards the robot were more positive nor the robot's likeability improved, as we expected. Furthermore, no other differences were found concerning the answer quality, overall look, content excitement, content presentation, assertiveness, overall hedonic quality and similarity degree between users and robot.

Anecdotal Results

More than half of the people who participated in the study came back to complain that the famous stall no. 4, mentioned by the robot, did not sell chicken feet. The stall was, of course our invention and we did not expect anybody to check the veracity of the information provided in our experiment. However, our mistake was that we did not into account the fact that eating is a national sport in the Singaporean gourmet society. On the other side, this result confirms once again the CASA paradigm [18], proving that our participants took the robot seriously.

8.5.5 Effects within groups

Next, we checked the effects of gender, age, expectations, personality, work and knowledge/experience background on the evaluation using a Mann-Whitney test. Additionally, three annotators analyzed the user behavior. The categories annotated similar to those described in 6.3.2 concerning speech patterns (dialogue acts), body postures, facial expressions (focusing on smiles), gaze direction and gestures. Degree of participation was relatively similar for all test participants due to a controlled experiment environment, i.e. the use of scenarios.

The percentage of agreement was calculated on 10% of the data. The table below shows reliable Krippendorf alpha values (between 0.7-0.8).

Annotation	**Krippendorf's alpha**
Dialogue acts	.8729
Body	.7301
Facial expression	.7769
Gaze direction	.8216
Gestures	.7468

Table 8.7: *Inter-annotators reliability for the human-robot corpus*

Gender effects

Male participants felt significantly more comfortable with the empathic robot than female did (p=.012). Men also scored the voice of the non-empathic robot significantly better (p=.019) and indicated more positive feeling towards it (p=.028). On the other hand, female participants found the non-empathic robot more useful (p=.019).

Male participants also rated the overall personality appeal (p=.013) of the high pitch robot(Olivia) much better than the female participants. They also indicated significantly higher values for the overall user feelings (p=.029) and overall enjoyment (p=.013). Additionally, they found the interaction more transparent (p=.036). On the other hand, female participants assessed Olivia's personality as being more extrovert than male participants did(p=.006). Male participants also found the low pitch robot's (Cynthia) look significantly more professional ($p<.001$), her voice matched the face better (p=.039) and her speech was faster (p=.025). The content presented by Cynthia appeared to them significantly more interesting (p=.008) and her overall behavior more appealing (p=.045). The men's higher ratings for both robots might be explained by a general stereotypes tendency which predicts men's preference for female robots, also found by other studies [259].
On the other side, female participants found the answers of the humorous robot

more informative (p=.019) and more clear (p=.012) than male participants. Additionally, females also found the humorous robot as having a better ability to socialize (p=.027).

Work background

Users having a non-IT work background found the empathic robot as being more extrovert compared with IT background users (p=.008). They also rated the social skills of the humorous robot much better (p=.047).

Background in speech

Users with no experience using speech recognition devices found the empathic robot more emotional as compared with experienced users (p=.018). On the other hand, experienced users had more positive feelings towards the empathic robot (p=.028). Interestingly, the users with no experience using speech recognition devices rated both robot characters significantly better concerning their abilities to socialize. This result was constant along the entire experiment: scenario 1 (empathic robot, p=.034, non-empathic robot, (p=.050), scenario 2 (p=0.12), scenario 3 (p=.002), scenario 4 (p=.008), scenario 5 (p=.010). This finding suggests that more experienced users tend to be less open in perceiving the robot as social entity.

Expectations

Users with higher expectation towards the robot found the high pitch robot to be more flexible (p=.036). On the other hand, users with low expectations found the speaking style of the low pitch robot (p=.023) and answer quality (p=.003) better. This outcome might be explained by the fact that the majority (71%) of the users with high expectations kept their questions strictly limited to the scenario settings, while almost half of the users with low expectations (48%) asked some additional questions (most probably trying to explore the robot's capabilities). Most 'unexpected' questions caused some delay - since not all could be prepared in advance by the wizard team- and might have caused the impression of an inflexible answer behavior. On the other hand, the answers improved the information richness and therefore, the answer quality could have been perceived as being better. As for the speaking style we have no explanation, except maybe for the fact that the test participants who interacted longer with the robot could have been more aware of its polite and elegant style of speaking.

Personality profile

Finally, introvert users found the interaction with the low pitch robot much easier (p=.012) and her behavior significantly more receptive (p<.001) than extrovert users. Since the low pitch robot Cynthia was perceived as being a more introvert robot, our result confirms to the findings of Nass and Reeves [8] concerning the attraction effects between human and computers and/or voice with similar personalities.

User behavior

People who looked predominantly downwards, i.e. to the screen and not to the robot, gave the low pitch robot (Cynthia) significantly lower rating for the voice appeal (scenario 3: p=.007) and social skills (scenario 4: p=.041, especially for the ability to express emotions, p=.010). They found the robot to be much less like themselves (scenario 4: p=.011) as compared with users with mixed or forwards (to the robot) gaze behavior. People who smiled more found the high pitch robot (Olivia) to have much better social skills (scenario 5: p=.036). Also, people who brought their upper body forwards rated the overall interaction quality with Cynthia better (scenario 3: p=.022, (scenario: 4 p=.019). Additionally, they gave Cynthia better rating (scenario 4) for overall enjoyment (p=.004), overall entertainment (p=.028) and overall comfort (p=.018).

No effects

No particular effects were found for different age groups. Additionally, no strong patterns could be found in our data for building user groups based on similar gestures and speech patterns - most of users used sentences to communicate with the robot.

8.6 Summary

The main focus of this study was to analyze the relationships between empathy, voice pitch and humor, on one side and the way people perceive and rate a social robot receptionist, on the other side. Our ultimate goal was to determine how our variable manipulations influence the ratings concerning the robot and the interaction quality.

Two robot characters were created: Olivia, the more exuberant, empathic and humorous robot with a higher voice pitch and Cynthia, the calmer, more serious and more reserved robot with a lower voice pitch. The voice pitch was varied by manipulating the TTS settings.

The experiment was divided into three parts: in the first part we examined the robot's empathic reactions on users (scenario 1), in the second part we focused only on voice pitch manipulation (scenarios 2 and 3), while in the third part we studied the effect of humor in combination with voice pitch manipulations (scenario 4 and 5).

The scenario order was randomized to exclude any potential biases that might arise from being exposed to one particular robot character before the other.

We designed most of the robot's prompts, gestures, body postures and head moves designed based on the data collected from the chapter 7. Also, the experiment took place in the same environment settings as the one described in chapter 7.

To measure the impact of our variable manipulation on the user ratings, we constructed a questionnaire based on the questionnaires AttrakDiff, SASSI, ITU MOS. Additionally, we incorporated items originating from the Gricean maxims concerning basic principles for effective communication.

Concerning the users' perception of empathy, our manipulation seemed to be successful, i.e. the robot was indeed perceived as being more empathic by group A, although only had a minor affect on the evaluation results. This might be explained partly because of the short exposure to the empathic stimulus and partly because of the stimulus intensity. In other words, the scenario setting did not achieve the desired effect of transposing the user in the required affective state. Future research should reconsider the scenario settings, enlarge the exposure to the empathic stimulus and incorporate additional facial reactions to increase empathy effect.

The manipulation of voice pitch showed visible effects on how users perceived the robots and the interaction quality: Olivia, the higher pitched robot, was better rated in terms of overall appearance, behavior and personality; users also felt more positive feelings towards Olivia and rated the interaction concerning both hedonic and pragmatic quality better as compared with the low pitch robot, Cynthia.

As indicated by the questionnaires, the ability to express humor was highest in the scenario containing jokes, which confirmed that our manipulation was successful. However, the effects of humor were studied in combination with the voice pitch. Combining the voice pitch manipulation with the humor had the advantage of creating a more homogeneous personality profile and allowed us to collect more data related to the effects of the voice pitch. On the other side, the results obtained cannot be entirely related to the humor effects since the voice pitch had also an inherent influence in the manipulation. But, since the second experiment deals with voice pitch effects only though comparisons we hoped to be able to detect effects caused by the voice pitch on one side and by humor on the other side.

The results reinforced our findings concerning the effects of the voice pitch since the same ratings preference were maintained concerning the high pitch robot - but at the same time some additional effects were found. These effects could be linked to the humor manipulation and concerned the robot's personality, that is to say, the humorous robot was rated as more emotional and more appealing in its style speaking. Additionally, the tasks, which were identical in both scenarios appeared to be more appealing while interacting with the humorous robot.

Concerning effects within groups, we found that male participants in general, gave better ratings to both robots as compared with females, confirming stereotypes found by other studies which predicts men's preference for female robots. More experienced users with IT background and/or knowledge in speech recognition devices rated the social skills of both robots significantly lower as compared with non-experienced users. This finding suggests that more experienced users tend to be less open in perceiving the robots as social entities. Participants with high expectations found the high pitch robot more flexible. Participants with low expectations appreciated the speaking style and the answer quality of the low pitched robot better.

Also, introvert participants found the low pitched robot much easier to interact with and more receptive to their needs.

The major results of the study are related to the manipulation of the voice pitch: our results demonstrated the high impact of the voice pitch on robot's attractiveness, which further influenced the evaluation results of the entire interaction. We believe that choosing the right voice pitch should be a priority in social robot design. Finally, through this study, we would like to stress the enormous importance of the voice in human robot interaction and to encourage further research on this topic.

Chapter 9

Conclusions

This chapter concludes the dissertation. Section 9.1 provides a summary of the contributions of the thesis. Section 9.2 presents the results gathered from our studies in light of the research questions formulated in section 1.2. Section 9.3 highlights some of the most important lessons gained from our studies, while section 9.4 suggests general directions for future work.

9.1 Summary of our contribution

There are four main contributions that are relevant for the HCI community: theoretical, methodological, empirical and design related. We will describe these as follows:

Contribution 1 (theoretical) - refers to a newly compiled set of guidelines for dialogue and interaction design for spoken conversational interfaces from the reviewed literature (chapter 2) and a taxonomy of conversational interaction quality focusing on hedonic and pragmatic quality aspects (chapter 3). Both contributions build on theoretical work presented in the domain literature. The guidelines were used to design the dialogue with a voice user interface (chapter 4) and a social robot (chapter 8). The taxonomy was used in all our empirical studies (chapter 4, 5, 6, 8) in order to guide evaluations.

Contribution 2 (methodological) - refers to a novel approach to evaluating the adequacy of conversational structures implemented in conversational interfaces using a new concept: *verbal affordance*. The concept of affordance is used in interaction design with a focus on visual elements and regards verbal units as simple cognitive support for the graphics. We define the concept of *verbal affordance* as the correct orientation of each communicative action towards a preceding action, that is to say, an action that leads to a certain response. Since verbal affordances are abstract, that is, they are immaterial, we use the conversational protocol as a constraint. What does it mean? The shape of an object represents the physical constraint that usually indicates how the object can be used, i.e. the object's affordance. For example, the form of a glass indicates what we can do with the glass (e.g. drinking, holding objects, etc) but also what we cannot do with it (e.g. biking). In conversational interactions, the 'object' is a verbal statement with no physical shape and no physical constraint as such. In order to understand what a verbal statement affords, we need a set of rules that we call conversational protocol. This protocol can be seen as a grammar for social interaction that prescribes what a verbal sentence logically affords, that is to say, what kind of reply we should give when we hear a certain verbal statement. We showed how verbal affordances can be used in conversational interfaces 1) to detect inefficient constructions leading to disruptions in the dialogue flow; 2) to spot unnecessary functions and; 3) to provide important insights on a system's ease-of-use (chapter 5).

Contribution 3 (empirical) - represents the main contribution of this dissertation and refers to the effects that voice characteristics and language features can have on the evaluation of a conversational interface, in particular on the overall interaction quality. We used accent, consistency with physical appearance and pitch as voice characteristics (chapters 4, 5, 8). Language features were implemented as social skills, empathy and humor (chapters 6, 8). Using subjective questionnaires, user behavior analysis and interaction parameter measurements, we showed how users' ratings can be influenced by manipulating these design aspects.

Contribution 4 (design related) - there are two design-related contributions:
• the first contribution refers to the design of conversational interfaces based on human dialogue interactions. We showed how to design efficient human robot dialogues in a task-oriented scenario using multimodal analysis of human-human dialogues (chapter 7).
• the second contribution refers to a novel application: a voice enabled user manual for mobile phones. We showed how to design an application prototype using written instructions and how to improve the dialogue script using human spoken dialogues (chapter 4).

9.2 Research questions

The research presented in this thesis focused on the design and evaluation of conversational interfaces for task-oriented dialogues using speech as the main interaction modality. Since speech plays a central role in interaction with such interfaces, two of our research questions addressed design aspects related to this modality.

RQ1: What impact do voice characteristics, such as voice pitch, voice accent and voice consistency with physical look have on the evaluation of a conversational interface?

In order to address this question, we performed three studies. The first study examined the impact of voice accents (Singaporean Standard English versus British Standard English), on the evaluation results of a voice user interface application for mobile phone users (chapter 4). The second study looked at the way in which users evaluated three embodied conversational agents of a QA system displaying feminine, masculine and gender ambiguous traits; the gender ambiguity was implemented as a combination of female voice and (rather) masculine look, that is to say as voice inconsistency with physical look (chapter 5). The third study was performed with two social robots using high pitch and low pitch voices to communicate with the users; the aim of the study was to determine how the voice pitch influenced people's perception with regard to the robot and the interaction quality (chapter 8).

All of the studies showed that the voice characteristics seem to have a strong impact on the system evaluation results. While the configurations of all three systems were kept identical during the experiments, the users found significant differences between the compared interfaces, as listed below:

1) the British system was ranked higher in terms of politeness, voice quality and ease of dialogue
2) the agents displaying consistent voice and look (in particular the feminine agent) were preferred in terms of interaction enjoyment; users also had more positive feelings towards the female agent Anna
3) the high pitch robot was ranked higher in terms of voice appeal, behavior, personality, overall enjoyment and interaction quality; users also had more positive feelings towards the high pitch robot

The explanations for the preferences are diverse (e.g. the British accent belongs to a highly esteemed culture in Singapore, gender labeling lightens cognitive load, while high pitch voices are associated with more attractive look and more sociable personalities), but they all converge to the same conclusion: users react emotionally to voice characteristics and this reaction has consequences on their judgment.

RQ 2: What impact do social skills, empathy and humor (implemented as language features) have on the evaluation of a conversational interface?

We conducted two experiments with a social robot in order to respond to this question. In chapter 6, we performed a first uncontrolled experiment with the social robot Olivia, with the aim of exploring relationships between the robot's social skills, user behavior and overall interaction. In chapter 8, the effect of humor and empathy were tested in a controlled experiment on users' perceptions of the robot and interaction quality.

The first experiment showed that the robot's ability to socialize was the second highest variable (among a total of 15 variables) correlated with the overall interaction quality. In the second experiment, humor was found to influence the ratings for the robot's speaking style appeal and tasks appeal; moreover, the robot personality appeared to be stronger while also being more emotional. The effects of empathy were found to be minor. The interaction with the empathic robot was perceived as being easier; further, no other significant differences were found.

The ability to socialize and the expression of humor seem to influence the users' ratings concerning hedonic quality aspects and, to a certain degree, pragmatic quality aspects. The effects were, however, less strong as compared with those achieved by voice characteristics. This might be explained by the fact that voice characteristics address sensorial perceptions and are therefore perceived and processed faster by users. In contrast, language features, such as empathy, humor and social skills target the users' feelings of sympathy towards the interface and as such, their perception evolves over time, requiring greater exposure in order to show effects. Another explanation could be related to the fact that such features are more difficult to model as compared with voice characteristics, which can be more easily tuned.

RQ 3: Which communicative interaction patterns are relevant for task-oriented human-human interaction with potential applicability in human-machine interaction?

In chapter 7, we performed an empirical study with human actors playing the roles of a receptionist and a visitor. The goal of the study was to determine how people communicate with each other in task-oriented situations and to explore how this knowledge can be used in a human-robot scenario in order to prevent miscommunication, to minimize the risk of additional questions and to present information in an efficient way. The study outcome revealed several interaction patterns that can be summarized as follows:

- use a touch screen to let visitors input their names, to display building maps, taxi booking information or summarization of route description
- announce breaks in the communication flow, e.g. if the receptionist needs to make a phone call
- do not provide more information than required
- split information into short sentences
- if no question (e.g. WH-, propositional, request etc) is detected, provide explicit feedback on recognized keywords
- prevent unexpected follow-up questions by:
  – avoiding the use of new, and possibly unknown, informational elements without prior explanations (e.g. new building name, locations)
  – providing a reason for a particular recommendation; this strategy can additionally help to build trust
- answers should handle only one question, even though users might ask two different questions at the time

- descriptions regarding orientation in the building should start using reference points visible to users
- in the event that information about a location with restricted access is requested, inform users about the restriction and provide alternative solutions if possible
- if the information provided (e.g. directions in the building) contains more than three explicative sequence (i.e. dialogue acts) a short summarization should be provided

RQ 4: How can we use human communicative interaction patterns to test and enhance conversational interfaces?

In this thesis, we used human communicative interaction patterns in chapter 4 and in chapter 5.

In chapter 4, we designed the dialogue script with the voice user interface using written manual instructions, in addition to real questions and answers posted on websites by real mobile phone users. Once configured, the dialogue was tested and enhanced using real dialogues between test participants and a tester. The participants received a list with the questions related to sms messaging, while the tester received a list with the designed prompts. Both users and tester talked over a simulated phone line. They were instructed to imagine themselves in a help desk scenario where people call to ask questions and are attended to by a help-desk employee. The conversations were recorded and additional questions asked by the test users were noted. Other possible question/answer pairs were explored and the dialogue was enhanced to a final version.

In chapter 5, we used human communicative interaction pattern together with the affordance concept to test the adequacy of conversational structure implemented in a multimodal question answering system. The analysis was performed by comparing the system's conversational structures with those occurring in natural human conversations. The communicative interaction patterns were grouped in a conversational protocol consisting of conversation initialization, body and termination. The protocol was indicative of which verbal statements are afforded in a particular situation.

9.3 Take away messages

Apart from specific contributions and answers to research questions, the experiments performed in this thesis brought some additional insights with relevance for the HCI community that might need further articulation. These insights represent the lessons we learned from our experiments, that is to say, the take away message for other researchers and practitioners. Below, we list the most important observations we gathered from our experiments.

1) Widely recognized stereotypes, such as 'similarity attracts' might not be applicable in situations where cultural and psychological biases interfere; as such, they might even have detrimental effects for the acceptance of the interface by users. We showed in chapter 4 that users preferred a foreign accent to their own and how this preference was reflected in their ratings.

2) Shortcuts are poor options. Often computer applications display avatars or agents whose physical appearance does not point to any particular gender. This look is intentionally created by designers such that both male and female users can relate to the character. In chapter 5, we found some evidence showing that gender ambiguity can be disturbing and that users tend to prefer gender marked agents (in particular, the female one).

3) Subjective assessment of interaction parameters, such as speed or recognition accuracy might not be correlated with their objective measurements. In chapter 6, we found that the speech recognition performance and the speed was ranked better than the error logs and total completion time would have predicted. This might be explained, partly by the question formulation bias, partly by a general human tendency to give more positive ratings in order to please the interviewer, and partly because the visitors enjoyed the interaction despite errors and long response delay.

4) Voice characteristics are easier to manipulate and produce stronger effects on users' ratings as compared to language features. This was explained in the previous section.

5) Response speed seems to be key factor in the evaluation of conversational interfaces. The speed was mentioned in the top most influential factors on the interaction quality, being considered even more important than speech and object recognition (chapter 6). This finding could indicate that users tend to be more tolerant of errors but less willing to wait too long for a response. In addition, significant correlations between speed and interaction quality were found in both experiments with the social robot (chapter 6 and 8), while in chapter 5 the speed was mentioned as one of reasons for preferring the agent Anna.

6) More experienced users tend to be less open to perceiving a robot as a social entity. We found that users with IT backgrounds and/or knowledge in speech recognition devices rated the social skills of the robots (chapter 6 and 8) significantly lower as compared with non-experienced users. This was consistent in both experiments with the social robot.

7) Human interaction patterns are rich sources of inspiration for designing, enhancing and evaluating human-machine dialogues, as explained in the previous section of this chapter.

9.4 Future research

In this subsection we discuss some possible research avenues and suggest questions we would like to address in the future.

In chapter 3, we proposed a taxonomy of quality of conversational interactions. The taxonomy was built on relevant domain literature. Future research work could extend our taxonomy, establish new relationships between the elements, assign weights and rank priorities.

In chapter 4, we presented a study concerning the effects of voice accents on the evaluation of a mobile phone VUI. Most of our participants (72%) were Singaporean, belonging to the Chinese community. Thus, future research could include a more balanced user sample consisting of all representative ethnic groups in the population sample and a more nuanced questionnaire in order to confirm our findings.

In chapter 5, we presented a pilot study performed with 8 test participants concerning the effects of androgynous voice and look on the overall interaction quality. The study showed encouraging results for our hypothesis concerning the human preference for interacting with consistently gender labeled entities. Future experiments could continue the research on this topic by conducting studies with a larger number of participants and an additional set of agents displaying similar gender-ambiguous vs. gender-marked voice and look characteristics in order to gain statistical evidence for our hypothesis.

In chapter 6, we presented an experiment with a social robot performed in real life conditions. One interesting observation we made during the data analysis was the fact (also mentioned in the previous section) that the speech recognition performance was ranked as being better by users than the error logs and total competition time would have predicted. This mismatch between objective measurements and subjective perceptions demonstrates the difficulty of making accurate evaluation predictions based on data collected from the log files. Future research can determine the levels of tolerance users have for errors concerning speech or object recognition in relation to the overall interaction quality.

In chapter 8, we performed a study in which we manipulated voice pitch, empathy and humor. Since humor showed steady correlations with the interaction quality in our experiments, we believe that it can have generally positive effects on the evaluation of a speech based interactive system. Due to experimental circumstances, the effects of humor were studied in combination with the voice pitch. Further experiments should address the humor manipulation separately from other influence sources. Additionally, the influence of humor could also be studied in combination with other types of speech interfaces, such as pure voice user interfaces, virtual embodied agents or question answering systems. With regard to empathy, while our manipulation was successful, it achieved only a minor effect on the robot's evaluation results and did not show a particular correlation trend with the overall interaction quality. However, showing empathy in social situations proved to have positive effects towards the interlocutor in many experiments conducted in the past. Therefore, future research could to reconsider the scenario settings, using a different approach to induce empathy effects, like for example, in a game setting.

Additionally, the exposure to the empathic stimulus can be enlarged and combined with additional expression modalities, such as mimics for example.

In our experiment described in chapter 8, we proposed a questionnaire designed to assess the hedonic and pragmatic quality of a (multimodal) speech based system - in this case, a social robot. Since there are no standardized questionnaires to evaluate the quality of interaction, we used several questionnaires from the literature to build our own. The questionnaire was already tested during our experiments in chapter 8. However, in future the questionnaire could be extended and validated within further experiments.

We found certain matching patterns between users' ratings and behavioral categories measured for mimics, degree of involvement, speech pattern (keywords versus sentences). Therefore, such observations, if detected automatically, could be useful for user modeling or as predictors for evaluation results. Some first attempts were made in chapter 6: we performed an ordinal probit regression analysis to test the predictive power of user personal data and behavioral categories on the overall interaction quality. The best model resulted from combining user's educational degree, their interactive behavior, speech patterns and politeness markers. First experiments showed encouraging results, however further research is needed to confirm the adequacy of these categories as predictors by enlarging the data set and refining the annotations.

Bibliography

[1] Wikipedia. History of robots. http://en.wikipedia.org/wiki/History_of_robots, 2011.

[2] J. Needham. *Science and Civilisation in China. History of Scientific Thought*, volume 2. Cambridge University Press, New York, USA, 1991.

[3] C. G. Kratzenstein. Sur la raissance de la formation des voyelles. *Journal of Physics*, 21:358–380, 1782.

[4] Ch. Wheatstone. *The Scientific Papers of Sir Charles Wheatstone*. Taylor and Francis, London, 1879.

[5] B. H. Juang and L. R. Rabiner. *Automatic speech recognitiona brief history of the technology*. Elsevier Encyclopedia of Language and Linguistics, 2005.

[6] M. Mori. *The Buddha in the Robot*. Tuttle Co., 1982.

[7] D. Thomson and J. Wisowaty. Confusion in natural language services. In *Proc. ESCA Workshop on Interactive Dialogue in Multi-Modal Systems*, pages 189–196, Kloster Irsee, 1999.

[8] B. Reeves and C. Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, New York, USA, 1996.

[9] N. Bernsen, L. Dybkjaer, and H. Dybkjaer. Cooperativity in human-machine and human-human spoken dialogue. *Discourse Processes*, 21(2), 1996.

[10] V. Zue. Conversational interfaces: Advances and challenges. In *5th European Conference on Speech Communication and Technology*, pages 9–18, 2000.

[11] E. Hutchins. Metaphors for interface design. In M.M. Taylor, F. Neel, and D.G. Bouwhuis, editors, *The Structure of Multimodal Dialogue*. Elsevier Science Publishers, New York, 1989.

[12] S. Oviatt. Multimodal interfaces. In J. Jacko and A. Sears, editors, *Handbook of Human-Computer Interaction*. Lawrence Erlbaum, Lawrence Erlbaum, 2002.

[13] M. Anisfeld, N. Bogo, and W.E. Lambert. Evaluational reactions to accented English speech. *Journal of Abnormal and Social Psychology*, 65(4):223–231, 1962.

[14] M. Zuckerman and K. Miyake. The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, 17(2):119–135, 1993.

[15] C. Nass, K. Isbister, and E. Lee. Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 374–402. MIT Press, Cambridge, MA, USA, 2000.

[16] A. Cann, L.G. Calhoun, and J.S. Banks. On the role of humor appreciation in interpersonal attraction: Its no joking matter. *Journal of Humor Research*, 10(1):77–89, 1979.

[17] A. P. Goldstein and G. Y. Michaels. *Empathy: development, training, and consequences*. Erlbaum Associates, Hillsdale, New York, USA, 1985.

[18] C. Nass, J. Steuer, and E.R. Tauber. Computers are social actors. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, pages 72–78. ACM, 1994.

[19] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Chicago, and London, 1949.

[20] W. Schramm. How communication works. In Wilbur Schramm, editor, *The Process and Effects of Communication*, pages 3–26. University of Illinois Press, Urbana, 1954.

[21] J.N. Cappella and C. Pelachaud. Rules for responsive robots: Using interaction to build virtual interaction. In H.T. Reis, M.A. Fitzpatrick, and A.L. Vangelisti, editors, *Stability and Change in Relationships*. Cambridge University Press, UK, 2002.

[22] T.A. Dijk. Introduction: Discourse analysis as a new cross-discipline. In T. A. van Dijk, editor, *Handbook of Discourse Analysis*, pages 1–10. Academic Press, Orlando, 1985.

[23] D. Schiffrin. *Approaches to discourse*. Blackwell, Oxford, 1991.

[24] H. Sacks, E.A. Schegloff, and G. Jefferson. A simplest semantics for the organization of turn taking in conversation. *Language*, 50(4):696–735, 1974.

[25] E.A. Schegloff and H. Sacks. Opening up closings. In *Semiotica 8: 4*, pages 289–327, New Orleans,USA, 1973.

[26] L. Ahrenberg, A. Joensson, and N. Dahlbaeck. Discourse representation and discourse management for natural language interfaces. In *Proc. of the Second Nordic Conference on Text Comprehension in Man and Machine*, Taeby, Sweden, 1990.

[27] H.H. Clark and S.E. Brennan. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*. APA, 1991.

[28] P. Dillenbourg, D. Traum, and D. Schneider. Grounding in multi-modal task-oriented collaboration. In *Proc. of the European Conference on AI in Education*, 1996.

[29] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[30] T. Holtgraves. *Language as social action: social phychology and language use*. Lawrence Erlbaum Associates, New Jersey, 2008.

[31] A. Green. *Designing and Evaluating Human-Robot Communication*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2009.

[32] A. Kendon. *Gesture–Visible Action as Utterance*. Cambridge University Press, Cambridge, 2004.

[33] D. Efron. *Gesture and Environment*. Kings Crown Pres, Morningside Heights, NY, 1941.

[34] D. McNeill. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press, Chicago, 1992.

[35] L.P. Morency, C.L. Sidner, C. Lee, and T. Darrell. Head gestures for perceptual interfaces: The role of context in improving recognition. *Artif. Intell.*, 171(8-9):568–585, 2007.

[36] H. van Dam. *Dialogue Acts in GUIs*. PhD thesis, Technische Universiteit Eindhoven, The Netherlands, 2006.

[37] M.L. Flecha-Garcia. *Eyebrow raising in dialogue: discourse structure, utterance function, and pitch accents*. PhD thesis, University of Edinburgh, UK, 2004.

[38] L. J. Brunner. Smiles can be back channels. *Journal of Personality and Social Psychology*, 37:728–734, 1979.

[39] P. Ekman and D. Keltner. Nonverbal communication: Where nature meets culture. In U. C. Segerstrale and P. Molnar, editors, *The social function of 'smile' and 'laughter': Variations across primate species and societies*, pages 27–46. Lawrence Erlbaum, Hillsdale, New Jersey, 1997.

[40] S. P. Gill, M. Kawamori, Y. Katagiri, and A. Shimojima. Pragmatics of body moves. In *Proc. of the Third International Cognitive Technology Conference (CT99)*, San Francisco, USA, 1999. Networked Minds.

[41] M. Argyle and M. Cook. *Gaze and mutual gaze*. Cambridge University Press, UK, 1976.

[42] C. Goodwin. *Conversational Organization*. Academic Press, NY, 1981.

[43] M.H. Goodwin and C. Goodwin. Gesture and co-participation in the activity of searching for a word. *Semiotica*, 62:51–72, 1986.

[44] I. Poggi, C. Pelachaud, and F. de Rosis. Eye communication in a conversational 3d synthetic agent. *Special Issue on Behavior Planning for Life-Like Characters and Avatars. AI Communications*, 13(3):169–181, 2000.

[45] D. Heylen. A closer look at gaze. In C. Pelachaud, E. Andre, S. Kopp, and Z.M. Ruttkay, editors, *Creating Bonds with Embodied Conversational Agents*, pages 3–9. University of Utrecht, 2005.

[46] Y. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *Proc. of Association for Computational Linguistics*, Sapporo, Japan, 2003.

[47] J. Cassell, O. Torres, and S. Prevost. Turn taking vs. discourse structure. *Machine Conversations*, pages 143–154, 1999.

[48] D. Heylen. Head gestures, gaze and the principles of conversational structure. *I. J. Humanoid Robotics*, 3(3):241–267, 2006.

[49] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, San Diego, CA, 1975.

[50] J. Austin. *How to do Things with Words*. Clarendon, Oxford, 1962.

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

[52] J. Jaworowska. Speech act theory. `http://instructional1.calstatela.edu/lkamhis/tesl565_sp04/troy/spchact.htm`, 2011. Accessed June 20, 2011.

[53] M. Cohen, J. Giangola, and J. Balogh. *Voice User Interface Design*. Addison Wesley, New York, 2004.

[54] J. R. Searle. A taxonomy of illocutionary acts. *Language, Mind, and Knowledge*, 7, 1975.

[55] B. J. Grosz and C.L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

[56] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

[57] D. R. Traum and E. A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599, 1992. Special Issue on Non-literal language.

[58] D. G. Novick. *Control of mixed-initiative discourse through metalocutionary acts: a computational model*. PhD thesis, University of Oregon, USA, 1988.

[59] D. J. Litman and J. F. Allen. A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200, 1987.

[60] H.C. Bunt. Dynamic interpretation and dialogue theory. In M. Taylor, D. G. Bouwhuis, and F. Neel, editors, *The Structure of Multimodal Dialogue*, volume 2, pages 81–150. John Benjamins, Amsterdam, 1999.

[61] H. Bunt, J. Alexandersson, J. Carletta, J.W. Choe, A.C. Fang, K. Hasida, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, and D. Traum. Towards an ISO standard for dialogue act annotation. In N. Calzolari, K.Choukri, B.Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).

[62] H. Bunt. Dimensions in dialogue act annotations. In *5th. International Conference on Language Resources and Evaluation (LREC)*, Genova, Italy, 2006.

[63] D.R. Traum. Conversational agency: The trains-93 dialogue manager. In *Proc. of the Twente Workshop in Language Technology: Dialogue Management in Natural Language Systems (TWLT11)*, pages 1–11, Enschede, The Netherlands, 1996.

[64] N.O. Bernsen, H. Dybkjaer, and L. Dybkjaer. *Designing interactive speech systems. From first ideas to user testing*. Springer Verlag, NY, 1998.

[65] S. Gamm and R. Haeb-Umback. User interface design of voice controlled consumer electronics. *Philips Journal of Research*, 49(4), 1995.

[66] S. Rosset, S. Bennacef, and L. Lamel. Design strategies for spoken language dialog systems. In *Proc. of the European Conference on Speech Technology, EuroSpeech*, pages 1535–1538, Budapest, 1999.

[67] S. Moeller. *Quality of Telephone-based Spoken Dialog Systems*. Springer Verlag, New York USA, 2005.

[68] N. Fraser and N. Gilbert. Effects of system voice quality on user utterances in speech dialogue systems. In *EUROSPEECH*, pages 57–60, 1991.

[69] A. Hauptmann and A. Rudnicky. Talking to computers: An empirical investigation. *International Journal of Man-Machine Studies*, 28(6):583–604, 1988.

[70] L. Karsenty. Shifting the design philosophy of spoken natural language dialogue: From invisible to transparent systems. *International Journal of Speech Technology*, 5(2):147–157, 2002.

[71] A. Hjalmarsson. *Human interaction as a model for spoken dialogue system behaviour*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2009.

[72] L. Dybkjaer and N. O. Bernsen. Optimising the usability of spoken language dialogue systems. *Natural Language Engineering*, 6:243–272, 2002.

[73] T. Falck, S. Gamm, and A. Kerner. Multimodal dialogues make feature phones easier to use. In *Proc. Applications of Speech Technology*, pages 125–128, Budapest, 1993.

[74] S. Moeller, K. Engelbrecht, C. Kuehnel, I. Wechsung, and B. Weiss. A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. In *Proc. QoMEX*, pages 7–12, 2009.

[75] D. Needham. BSI - Quality and Standards. `http://wwww.bsieducation.org/Education/downloads/ijqs/teaching-materials/stand-alone/Lecture-Materials-Concept-of-Quality.doc`, 2011.

[76] V.E. Sower. *Essentials of Quality with Cases and Experiential Exercises*. John Wiley and Sons, USA, 2010.

[77] E. Weekley. *An Etymological Dictionary of Modern English*. Dover Publications, New York, 1976.

[78] The American Heritage. Dictionary of the english language. `http://http://dictionary.reference.com/`, 2011.

[79] J. L. Ackrill. *Aristotle's Categories and De Interpretatione*. Clarendon Press, Oxford, clarendon aristotle series edition, 1961.

[80] G. Galilei. *The Assayer*. University of Pennsylvania Press, USA, 1960.

[81] R. Descartes. *Meditations on First Philosophy*. Cambridge University Press, Cambridge, 1996.

[82] J. Locke. *An Essay Concerning Human Understanding*. Prometheus Books, USA, 1994.

[83] M. Tye. The subjective qualities of experience. *Mind*, 9(5):1–17, 1986.

[84] W.A. Shewhart. *Economic control of quality of manufactured product*. Van Nostrand Company, USA-New York, 1931.

[85] J.M. Juran. The quality trilogy. *Quality Progress*, 9(8):19–24, 1986.

[86] A.V. Feigenbaum. Total quality control. *Harvard Business Review*, 34(6):93–101, 1956.

[87] A. Parasuraman, V. Zeithaml, and L. Berry. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 61 (1):12–40, 1988.

[88] ASQ. Basic concepts. `http://asq.org/glossary/q.html`, 2008.

[89] D.A. Garvin. What does 'product quality' really mean? *MIT Sloan Management Review*, 26 (1):25–45, 1984.

[90] ISO 9000 quality management systems fundamentals and vocabulary, 2005. International Standard.

[91] M. Walton. *Essentials of Quality with Cases and Experiential Exercises*. Berkley Publishing Group, New York, USA, 1986.

[92] G. Taguchi, M. El Sayed, and C. Hsaing. *Quality engineering and production systems*. McGraw-Hill, New York, 1989.

[93] P. Crosby. *Quality is Free*. McGraw-Hill, New York, 1979.

[94] D.A. Garvin. Competing on the eight dimensions of quality. *Harvard Business review*, 65 (6):101–109, 1987.

[95] W. Stevenson. *Production Operations Management*. McGraw-Hill, New York, 1999.

[96] C. Chakrapani. *How to Measure Service Quality and Customer Satisfaction*. American Marketing Association, Chicago, 1998.

[97] R. Norman. *Service Management Strategy and Leadership*. Wiley, New York, 1984.

[98] J. Evans and W. Lindsay. The management and control of quality. *Journal of Retailing*, 61 (1):12–40, 1988.

[99] S. Moeller. Messung und Vorhersage der Effizienz bei der Interaktion mit Sprachdialogdiensten. In *Fortschritte der Akustik - DAGA 2006*, page 463464, Braunschweig, 2006.

[100] A.I. Niculescu, E.M.A.G. van Dijk, A. Nijholt, D.K. Limbu, S.L. See, and A.H.Y. Wong. Socializing with Olivia, the youngest robot receptionist outside the lab. In S.S. Ge, H. Li, J.-J. Cabibihan, and Y.K. Tan, editors, *Proc. of the 2nd International Conference on Social Robotics, ICSR 2010*, volume 6414 of *Lecture Notes in Artificial Intelligence*, pages 50–62, Berlin, 2010. Springer Verlag.

[101] U. Jekosch. *Voice and Speech Quality Perception. Assessment and Evaluation*. Springer, Berlin, 2005.

[102] S. Moeller. Evaluating telephone-based interactive systems. In *ASIDE-2005*, 2005.

[103] N.O. Bernsen. Multimodality in language and speech systems from theory to design support tool. In B. Granstroem, D. House, and I. Karlsson, editors, *Multimodality in Language and Speech Systems*, page 93148. Kluwer Academic Publishers, Dordrecht, 2002.

[104] M. Hassenzahl, A. Platz, M. Burmester, and K. Lehner. Hedonic and ergonomic quality aspects determine software's appeal. In *Proc. CHI 2000*, pages 201–208, Den Haag, The Netherlands, 2000. Guilford.

[105] ISO 9241-11: 1998 ergonomic requirements for office work with visual display terminals part 11: Guidance on usability, 1998. International Standard.

[106] ITU-T Rec. P.851. Subjective quality evaluation of telephone services based on spoken dialogue systems, 2003. International Telecommunication Union.

[107] Wikipedia. Hedonism. `http://en.wikipedia.org/wiki/Hedonism`. Accessed June 24, 2011.

[108] V. Huta. Linking peoples' pursuit of eudaimonia and hedonia with characteristics of their parents: Parenting styles, verbally endorsed values, and role modeling. *Journal of Happiness Studies*, pages 1–15, 2011.

[109] G. Li. Human and humanoid don't match: consistency preference and impact on users' trust. In *Proc. of INTERACT'03*, pages 161–167, 2003.

[110] M. Walker, D. Litman, C. Kamm, and A. Abella. Paradise: A general framework for evaluating spoken dialogue agents. In *Proc. of the 35th Annual Meeting of the Association of Computational Linguistics (ACL/EACL)*, pages 271–280, 1997.

[111] E. Frokjaer, M. Hertzum, and K. Hornbaek. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proc. of CHI 2000 Human Factors in Computing Systems*, pages 345–352, New York, 2000. ACM Press.

[112] N. Fraser. Assessment of interactive systems. In D. Gibbon, R. Moore, and R. Winski, editors, *Handbook on Standards and Resources for Spoken Language Systems*, pages 564–615. Mouton de Gruyter, Berlin, 1997.

[113] K.S. Hone and R. Graham. Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, 6 (3-4):287–303, 2000.

[114] M. Hassenzahl, M. Burmester, and F. Koller. Attrakdiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualitaet. In *Mensch und Computer '03*, page 187196, 2003.

[115] ITU-T Recommendation P.85. Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice out-put devices, 1994.

[116] J. Nielsen and R. L. Mack. *Usability Inspection Methods*. John Wiley and Sons, USA-New York, 1994.

[117] C. Wharton, J. Rieman, C. Lewis, and P. Polson. The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R. Mack, editors, *Usability Inspection Methods*, pages 105–140. Wiley, New York, USA, 1994.

[118] M.A. Walker, C.A. Kamm, and D.J. Litman. Towards developing general models of usability. *Natural Language Engineering*, 6(3-4):363–377, 2000.

[119] K.P. Engelbrecht and S. Moeller. Pragmatic usage of linear regression models for the prediction of user judgments. In *Proc. of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 291–294, Antwerp, 2007.

[120] N. Beringer, U. Kartal, K. Louka, F. Schiel, and U. Tuerk. PROMISE - A procedure for multimodal interactive system evaluation. In *Proc. of the LREC Workshop on Multimodal Resources and Multimodal Systems Evaluation*, pages 77–80, Las Palmas, 2002.

[121] A.I. Niculescu, G.M. White, S.L. Swee, R.U. Waloejo, and Y. Kawaguchi. Impact of English regional accents on user acceptance of voice user interfaces. In *Proc. of the 5th Nordic conference on Human-computer interaction*, volume 358, pages 523–526, New York, 2008. ACM.

[122] Y. Gu and J. E. Gilbert. The development of a voice system that interacts with a student information database. In *Proc. of the 42nd Annual Southeast Regional Conference*, volume 358, pages 248–252, New York, 2004. ACM.

[123] D. Mecanovic and H. Shi. Voice user interface design for a telephone application using voicexml. In *Proc. of 7th Asia-Pacific Web Conference Web*, Shanghai, China, 2005.

[124] C. Nass and S. Brave. *Wired for speech. How Voice Activates and Advances the Human-Computer Relationship*. Cambridge MIT Press, USA, 2005.

[125] E.L. Low and A. Brown. *English in Singapore: An Introduction*. McGraw-Hill Education, Asia, 2005.

[126] A. Brown. *Singapore English in a nutshell: an alphabetical description of its features*. Federal Publications, Singapore, 2000.

[127] D. Birch and J. McPhail. The impact of accented speech in international advertisement. *Global Business Languages*, (2):91–105, 1997.

[128] H. Giles. Patterns of evaluation in reactions to RP, South Welsh and Somerset accented speech. *British Journal of Social and Clinical Psychology*, 10:280–281, 1971.

[129] R. Luhman. Appalachian English stereotypes: Language attitudes in Kentucky. *Language in Society*, 19:331–348, 1990.

[130] B. L Brown, H. Giles, and J. N. Thakerar. Speaker evaluation as a function of speech rate, accent and context. *Language and Communication*, 5(3):207–220, 1985.

[131] G. E. Belch and M. A. Belch. *Introduction to Advertising and Promotion: An Integrated Marketing Communications Perspective*. The Dunmore Printing Company Ltd, McGraw-Hill Inc, Boston, 1993.

[132] J. N. P. Francis. When in Rome? The effects of cultural adaptation on intercultural business negotiations. *Journal of International Business Studies*, 5(3):403–428, 1991.

[133] A.F. Gupta. Contact features of Singapore colloquial English. In K. Bolton and H. Kwok, editors, *Sociolinguistics Today: International Perspectives, Routledge*, pages 323–345. Arnold, London and New York, 1992.

[134] D. Detering. *Singapore English*. Edinburgh University Press, UK, 2007.

[135] H. Giles. Communication effectiveness as a function of accented speech. *Speech Monographs*, 40:330–331, 1973.

[136] W.S.W. Lim. Development and culture in Singapore and beyond. *SOJOURN, Journal of Social Issues in Southeast Asia*, 14:249–261, 1999.

[137] J. Edwards. Language attitudes and implications among English speakers. In E. B. Ryan and H. G. Giles, editors, *Attitudes Towards Language Variations*, volume 2, pages 20–33. Arnold, London, 1982.

[138] D. Bayard. *Kiwitalk: Sociolinguistics and New Zealand Society*. The Dunmore Printing Company Ltd, Palmerston North, New Zealand, 1995.

[139] I. Huygens and G. M. Vaughan. Language attitudes, ethnicity and social class in New Zealand. *Journal of Multilingual and Multicultural Development*, 4:207–223, 1983.

[140] A.I. Niculescu. Affordances in conversational interactions with multimodal QA systems. In A. Holzinger, editor, *Proc. of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008*, volume 5298/2008 of *LNCS*, pages 221–236, Germany, 2008. Springer Verlag.

[141] A.I. Niculescu, F. van der Sluis, and A. Nijholt. Feminity, masculinity and androgyny: How humans perceive the gender of anthropomorphic agents. In C. Stephanides, editor, *HCI International. 13th International Conference on Human-Computer Interaction*, Proc. and Posters DVD, pages 628–632, Heidelberg, 2009. Springer Verlag.

[142] A.I. Niculescu, D.H.W. Hofs, E.M.A.G. van Dijk, and A. Nijholt. How the agent's gender influence users' evaluation of a qa system. In *Proc. of the International Conference on User Science and Engineering 2010*, pages 5–9, USA, 2010. IEEE.

[143] J.L. Vicedo and Mollina D. Open-domain question-answering technology: State of the art and future trends, 2001.

[144] B.F. Green, A.K. Wolf, C. Chomsky, and K. Laughery. Baseball: An automatic question answerer. In *Proc. of the Western Joint Computer Conference 19*, pages 219–224, 1961. Reprinted in Grosz et al. (1986).

[145] W. A. Woods, R.M. Kaplan, and B.N. Webber. The lunar sciences natural language information system: Final report. Technical Report BBN Report 2378, Cambridge, MA, 1961.

[146] L. Hirschman and R. Gaizauskas. Natural language question answering: The view from here. *Natural Language Engineering*, 7(4):275–300, 2001.

[147] E.M. Voorhees. The evaluation of question answering systems: Lessons learned from the trec qa track. In *Proc. of the Question Answering: Strategy and Resources Workshop at LREC-2002*, 2002.

[148] S. Quarteroni and S. Manandhar. Designing an interactive open-domain question answering system. *Natural Language Engineering*, 15:73–95, January 2009.

[149] R. op den Akker, H. Bunt, S. Keizer, and B. van Schooten. From question answering to spoken dialogue: Towards an information search assistant for interactive multi-modal information extraction. In *Proc. of Interspeech*, Lissbon, Portugal, 2005.

[150] S. Small and T. Strzalkowski. HITIQA: high-quality intelligence through interactive question answering. *Natural Language Engineering*, 15(1):31–54, 2009.

[151] M. Theune, B. van Schooten, R. op den Akker, W. Bosma, D. Hofs, A. Nijholt, E. Krahmer, C. van Hooijdonk, and E. Marsi. Questions, pictures, answers: introducing pictures in question-answering systems. In *Proc. of 10th. International Symposium on Social Communication*, Santiago de Cuba, Cuba, 2007.

[152] D. Hofs, B. van Schooten, and R. op den Akker. The IMIX demonstrator: An information search assistant for the medical domain. In A. van den Bosch and G. Bouma, editors, *Interactive Multi-modal Question-Answering. Theory and Applications of Natural Language Processing*. Springer-Verlag, 2011.

[153] J. Forlizz, J. Zimmerman, J. Mancuso, and S. Kwak. How interface agents affect interaction between humans and computers. In *Proc. of Designing Pleasurable Products and Interfaces*, pages 209–221. ACM Press, 2007.

[154] D.A. Norman. *The Psychology of Everyday Things*. Basic Book Inc., New York, USA, 1988.

[155] J.J. Gibson. The theory of affordances. In R. Shaw and J. Bransford, editors, *Perceiving, Acting, and Knowing*. Hillsdale, New York, USA, 1977.

[156] J.J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, USA, 1979.

[157] J. Mcgrenere and W. Ho. Affordances: Clarifying and evolving a concept. In *Proc. of Graphics Interface 2000*, pages 179–186. Basic Books, 2000.

[158] M. Oliver. The problem with affordance. *E-Learning*, 2(4):405–411, 2005.

[159] D.A. Norman. Affordance, conventions, and design. *Interactions*, 6(3):38–43, 1999.

[160] D.A. Norman. *Memory and Attention: an introduction to human information processing*. Wiley & Sons, London, U.K., 1969.

[161] W.W. Gaver. Technology affordances. In *Proc. of CHI*, pages 79–84, New Orleans,USA, 1991.

[162] R.H. Hartson. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22:315–338, 2003.

[163] L. Vainio, E. Symes, R. Ellis, M. Tucker, and G. Ottoboni. On the relations between action planning, object identification, and motor representations of observed actions and objects. *Cognition*, 108(2):444–465, August 2008.

[164] J.G. Sheridan and G. Kortuem. Affordance-based design of physical interfaces for ubiquitous environments. In *Proc. of Graphics Interface 2000*, Lecture Notes in Computer Science, pages 183–199, Berlin, Germany, 2006. Springer.

[165] L. Ping, A. Eikman, W. Kealy, and W. Qian. On the relations between action planning, object identification, and motor representations of observed actions and objects. *Academic Radiology*, 13(12):1542–1552, 2006.

[166] H. Bussmann. *Lexikon der Sprachwissenschaft*. Kroener Verlag, Stuttgart, Germany, 3rd. edition, 2002.

[167] J. Cassell, T. Bickmore, H. Vilhjalmsson, and H. Yan. More than just a pretty face: Affordances of embodiment. In *Proc. of 2000 International Conference on Intelligent User Interfaces*, New Orleans, USA, 2000.

[168] M.A. Hogg and G.M Vaughan. *Social psychology*. Pearson Prentice Hall, Harlow, England, 2005.

[169] A. De Angeli and S. Brahnam. Sex stereotypes and conversational agents. In *Proc. of Gender and Interaction: real and virtual women in a male world*, Venice, Italy, 2006.

[170] S. T. Fiske and S.E. Taylor. *Social Cognition. From Brains to Culture*. McGraw-Hill, New York, USA, 2008.

[171] J. Zimmerman, E. Ayoob, J. Forlizzi, and M. McQuaid. Putting a face on embodied interface agents. In *Proc. of Designing Pleasurable Products and Interfaces*, pages 233–248, The Netherlands, 2005. Eindhoven Technical University Press.

[172] A.L. Baylor, R.B. Rosenberg-Kima, and E.A. Plant. Interface agents as social models: the impact of appearance on females' attitude toward engineering. In *CHI Extended Abstracts'06*, pages 526–531, 2006.

[173] R. Catrambone, J. Stasko, and J. Xiao. Anthropomorphic agents as a ui paradigm: experimental findings and a framework for research. Technical Report GIT-GVU-02-10, Georgia Institute of Technology, 2002.

[174] A.I. Niculescu, E.M.A.G. van Dijk, A. Nijholt, S.L. See, and H. Li. How humans behave and evaluate a social robot in real-environment settings. In W.P. Brinkman and M. Neerincx, editors, *Proc. of the 28th European Conference on Cognitive Ergonomics, (ECCE)*, pages 351–352, Delft, 2010. Mediamatica.

[175] H. Asoh, S. Hayamizu, I. Hara, Y. Motomura, and T. Akaho, S.and Matsui. Socially embedded learning of the office-conversant mobile robot Jijo-2. In *Proc. of 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 880–885, 1997.

[176] W. Burgard, A. B. Cremers, D. Fox, D. Haehnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. The interactive museum tour-guide robot. In *Proc. National Conference on Artificial Intelligence (AAAI)*, 1998.

[177] J. Schulte, C. Rosenberg, and S. Thrun. Spontaneous, short-term interaction with mobile robots. In *Proc. of IEEE Int. Conference on Robotics and Automation*, pages 658–663, 1999.

[178] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, and N. Hagita. Humanoid robots as a passive-social medium - a field experiment at a train station. In *Proc. of the ACM/IEEE international conference on Human robot interaction*, pages 137–144, 2007.

[179] S. Sabanovic, M. P. Michalowski, and R. Simmons. Robots in the wild observing human-robot social interaction outside the lab. In *Proc. of AMC*, pages 596 – 601, 2006.

[180] M.E. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J.T. Matthews, J. Dunbar-Jacobs, C. McCarthy, S. Thrun, M. Montemerlo, J. Pineau, and N. Roy. Pearl: A mobile robotic assistant to the elderly. In Edmonton, editor, *In AAAI Workshop*, pages 713–718, 2002.

[181] T. Kanda, T. Hirano, and D. Eaton. Interactive robots as social partners and peer tutors for children: A field trial. *Human Computer Interaction*, 19:61–84, 2004.

[182] M. Fujita. Aibo: Toward the era of digital creatures. *The International Journal of Robotics Research*, 20:781–794, October 2001.

[183] K. Wada, T. Shibata, T. Saito, and K. Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. In *International Conference on Intelligent Robots and Systems (IEEE)*, 2002.

[184] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots: Concepts, design, and applications. *Robotics and Autonomous Systems*, 42 (3-4):143–166, 2003.

[185] R.R. McCrae and P.T. Costa. A five-factor theory of personality. In L. Pervin and O.P. John, editors, *Handbook of personality: Theory and research*. Guilford, New York, USA, 1999.

[186] H.J. Eysenck and S.B.G. Eysenck. *Manual of the Eysenck personality inventory*. University of London Press, London, UK, 1964.

[187] F.M. Gresham and S.N. Elliot. *Social skills rating system manual*. American Guidance Service, London, UK, 1990.

[188] M. Heerink, B.J.A. Kroese, B.J. Wielinga, and V. Evers. Studying the acceptance of a robotic agent by elderly users. *Int. Journal of Assistive Robotics and Mechatronics*, 7(3):33–34, 2006.

[189] J. A. Yip and R. A. Martin. Sense of humor, emotional intelligence, and social competence. *Journal of Research in Personality*, 40(6):1202–1208, 2006.

[190] T. Watanabe, M. Okubo, and H. Ogawa. A speech driven embodied interaction robots system for human communication support. In *Proc. of Int. Conf. on Systems, Man, and Cybernetics*, 2000.

[191] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proc. of IROS*, Barcelona, 2005.

[192] B.G. Tabachnick and L.S. Fidell. *Using multivariate statistics*. Allyn and Bacon, Boston, 2001.

[193] J. Jaffe and S. Feldstein. *Rhythms of Dialogue*. Academic Press, New York, 1970.

[194] M. L. McLaughlin and M. J. Cody. Awkward silences: behavioral antecedents and consequences of the conversational lapse. *Human Communication Research*, 8(1):299–316, 1982.

[195] R.M. Worcester and T.R. Burns. A statistical examination of the relative precision of verbal scales. *Journal of Market Research Society*, 17,(3):181–197, 1975.

[196] J. Allwood, K. Cerrato, L. Jokinen, C. Navarretta, and P. Paggio. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3-4):273–287, 2007.

[197] H. Bunt. Context and dialogue control. *THINK Quarterly*, 3, 1994.

[198] J. Allen and M. Core. Draft of DAMSL: Dialogue act markup in several layers. `http://www.ahds.ac.uk/guides/linguistic-corpora/bibliography.htm`, 1997. Accessed July 8, 2011.

[199] J. Allen, G. Schubert, L. Ferguson, P. Heeman, C.H. Hwang, T. Kato, N. Light, B. Martin, B. Miller, M. Poesio, and D. Traum. The trains project: A case study in defining a conversational planning agent. Technical Report 532, University of Rochester, Computer Science Department, 1994.

[200] J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. HCRC dialogue structure coding manual. Technical Report HCRC/TR-82, University of Rochester, Computer Science Department, 1996.

[201] J. Alexandersson, B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz, and M. Siegel. Dialogue acts in VERBMOBIL-2. Technical Report Verbmobil Report 226, Saarbruecken: DFKI, 1998.

[202] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage Publications, Ltd, London, 1980.

[203] A.I. Niculescu, E.M.A.G. van Dijk, A. Nijholt, and S.L. See. The influence of voice pitch on the evaluation of a social robot receptionist. In *Proc.of the International Conference on User Science and Engineering 2011*, USA, 2011. IEEE.

[204] Mitsubishi. Communication robot wakamaru. `http://www.mhi.co.jp/en/products/detail/wakamaru.html/`. Accessed November, 2011.

[205] S. Levenstein. Robot receptionist saya - if looks could kill. `http://inventorspot.com/articles/robot_receptionist_saya_if_looks_could_kill_22078/`. Accessed November, 2011.

[206] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A.C. Schultz, and J. Wang. Designing robots for long-term social interaction. In *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, pages 2199–2204, 2005.

[207] M.K. Lee and M. Makatchev. How do people talk with a robot? An analysis of human-robot dialogues in the real world. In *Proc. of CHI*, pages 3769–3774, Boston, 2009.

[208] M. Makatchev, I.A. Fanaswala, A.A. Abdulsalam, B. Browning, W.M. Ghazzawi, M. Sakr, and R. Simmons. Dialogue patterns of an Arabic robot receptionist. In *Proc. of the 5th ACM/IEEE international conference on Human-robot interaction*, page 167168. ACM/IEEE, March 2010.

[209] W.P. Hampes. The relation between humor styles and empathy. *Europe's Journal of Psychology*, 6(3):34–45, 1994.

[210] L. Wispe. *History of the concept of empathy*. Cambridge University Press, New York, USA, 1987.

[211] A. Pereira, I. Leite, S. Mascarenhas, S. Martinho, and A. Paiva. Using empathy to improve human-robot relationships. In Maarten H. Lamers and Fons J. Verbeek, editors, *HRPR*, volume 59 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 130–138. Springer, 2010.

[212] S. Brave, C. Nass, and K. Hutchinson. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. J. Hum.-Comput. Stud.*, pages 161–178, 2005.

[213] K. Hone. Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 2(2):227–245, 2006.

[214] J. Klein, Y. Moon, and R.W. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, pages 119–140, 2001.

[215] H. Prendinger, J. Mori, and M. Ishizuka. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *Int. J. Hum.-Comput. Stud.*, pages 231–245, 2005.

[216] T.W. Bickmore and D. Schulman. Practical approaches to comforting users with relational agents. In *CHI Extended Abstracts'07*, pages 2291–2296, 2007.

[217] T. Partala and V. Surakka. The effects of affective interventions in human-computer interaction. *Interacting with Computers*, pages 295–309, 2004.

[218] H.S.M. Cramer, J. Goddijn, B.J. Wielinga, and V. Evers. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *Proc. of HRI'10*, pages 141–142, 2010.

[219] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine interaction. In *Proc. of AAMAS (1)'08*, pages 89–96, 2008.

[220] V. Evers, A. Winterboer, G. Pavlin, and F.A.C. Groen. The evaluation of empathy, autonomy and touch to inform the design of an environmental monitoring robot. In *Proc. of ICSR'10*, pages 285–294, 2010.

[221] R.A. Martin. *The Psychology of Humor: An Integrative Approach*. Elsevier Academic Press, USA, 2007.

[222] C. P. Wilson. *Jokes: Form, content, use and function*. London Academic Press, 1979.

[223] W.P. Hampes. The relationship between humor and trust. *Journal of Humor Research*, 12(3):253–259, 1999.

[224] R. Middleton and J. Moland. Humor in negro and white subcultures: A study of jokes among university students. *American Sociological Review*, 24:61–69, 1959.

[225] A. Nijholt. Conversational agents and the construction of humorous acts. In T. Nishida, editor, *Conversational Informatics: An Engineering Approach*, pages 21–47. John Wiley & Sons, Chicester, England, 2007.

[226] A. Ziv. *Personality and sense of humor*. Springer, 1984.

[227] D. F. Roy. Banana time: Job satisfaction and informal interaction. *Human Organization*, 18:158–168, 1960.

[228] L. G. Bolman and T. E. Deal. What makes a team work? *Organizational Dynamics*, 21(2):34–44, 1992.

[229] R.W. Clouse and K. L. Spurgeon. Corporate analysis of humor. *Journal of Human Behavior*, 32 (3-4):1–24, 1995.

[230] W. Ruch. Corporate analysis of humor. *Personality and Individual Differences*, 16:89–102, 1994.

[231] T. R. Kurtzberg, Ch. E. Naquin, and L.Y. Belkin. Humor as a relationship-building tool in online negotiations. *International Journal of Conflict Management*, 20(4):377–397, 2009.

[232] D. E Gibson. Humor consulting: Laughs for power and profit in organizations. *Humor*, 7 (4):403–428, 1994.

[233] B. Shneiderman. *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley, Menlo Park, CA, 1998.

[234] D. Loehr. An intergartion of a pun generator with a natural language robot. In *Proc. of the International Workshop on Computational Humor*, pages 161–172, Twente, the Netherlands, 1996.

[235] H.W. Tinholt and A. Nijholt. Computational humour: Utilizing cross-reference ambiguity for conversational jokes. In F. Masulli, S. Mitra, and G. Pasi, editors, *7th International Workshop on Fuzzy Logic and Applications (WILF 2007)*, volume 4578 of *Lecture Notes in Artificial Intelligence*, pages 477–483, Berlin, July 2007. Springer Verlag.

[236] J. Morkes, H.K. Kernal, and C. Nass. Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of SRCT theory. *Human Computer Interaction*, 14:395–435, December 1999.

[237] C.M. Huan and D. Szafir. No joke: Examining the use of humor in computer-mediated learning. `http://pages.cs.wisc.edu/~dszafir/resources/Humor\%20in\%20Computer-Mediated\%20Learning.pdf/`, 2001. Unpublished material.

[238] P. Dybala, M. Ptaszynski, R. Rzepka, and K. Araki. Humoroids: conversational agents that induce positive emotions with humor. In *AAMAS (2)*, pages 1171–1172, 2009.

[239] S. Babu, S. Schmugge, T. Barnes, and L. F. Hodges. What would you like to talk about? An evaluation of social conversations with a virtual receptionist. In *IVA*, pages 169–180, 2006.

[240] H. Traunmueller and A. Eriksson. The frequency range of the voice fundamental in the speech of male and female adults. Manuscript, 1994.

[241] D. Riding, D. Lonsdale, and B. Brown. The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women. *Journal of Nonverbal Behavior*, 30:55–61, 2006.

[242] S.A. Collins and C. Missing. Vocal and visual attractiveness are related in women. *Vocal and visual attractiveness are related in women*, 65:997–1004, 2003.

[243] B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovi. Integrating cues of social interest and voice pitch in mens preferences for womens voices. *Biology Letters*, 4:192–194, 2008.

[244] H. Helfrich and P. Weidenbecher. Impact of voice pitch on text memory. *Swiss Journal of Psychology*, 70(2):85–93, 2011.

[245] W. Apple, L. A. Streeter, and R.M. Krauss. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37:715–727, 1979.

[246] J. Trouvain, S. Schmidt, M. Schroeder, M. Schmitz, and W.J. Barry. Modelling personality features by changing prosody in synthetic speech. In *Proc. of Speech Prosody*, Dresden, Germany, 1972.

[247] J. M. Montepare and L. Zebrowitz-McArthur. Perceptions of adults with child-like voices in two cultures. *Journal of Experimental Social Psychology*, 23:331–349, 1987.

[248] G. Fairbanks. Recent experimental investigations of vocal pitch in speech. In G. Fairbanks, editor, *Experimental phonetics: Selected articles*, pages 161–170. University of Illinois Press, USA, 1996.

[249] K. R. Scherer. Personality markers in speech. In K.R.Scherer and H. Giles, editors, *Social markers in speech*, pages 147–209. Cambridge University Press, New York, USA, 1979.

[250] M. Zuckerman, K. Miyake, and C. S. Elkin. Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality*, 29:253–272, 1995.

[251] M.L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. te Boekhorst. Human approach distances to a mechanical-looking robot with different robot voice styles. In *Proc. of 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 707–712, Muenchen, Germany, 2008.

[252] M. Scheutz and P. Schermerhorn. Affective goal and task selection for social robots. In J. Vallverd and D. Casacuberta, editors, *The Handbook of Research on Synthetic Emotions and Sociable Robotics*, pages 74–87. IGI Global, 2009.

[253] A. Ortony, G. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge University Press, United Kingdom, 1988.

[254] N. Bee, E. Andre, T. Vogt, and P. Gebhard. First ideas on the use of affective cues in an empathic computer-based companion. In *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 5009–5014, Budapest, Hungary, 2009.

[255] X. Li, C. Watson, A. Igic, and B. MacDonald. Expressive speech for a virtual talking head. In *Proc. of the Australasian Conference on Robotics and Automation (ACRA)*, pages 5009–5014, Sydney, Australia, 2009.

[256] H. Prendinger and M. Ishizuka. The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, pages 267–285, 2005.

[257] A. Oehman, A. Hamm, and K. Hugdahl. Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In J.T. Cacioppo, L.G. Tassinary, and G.G. Bernston, editors, *Handbook of psychophysiology*, pages 533–575. Cambridge University Press, Cambridge, UK, 2000.

[258] R. Khan and A. De Angeli. The attractiveness stereotype in the evaluation of embodied conversational agents. *INTERACT (1)*, pages 85–97, 2009.

[259] M. Siegel, C. Breazeal, and M.I. Norton. Persuasive robotics: The influence of robot gender on human behavior. In *Proc. of IROS*, pages 2563–2568, Dresden, Germany, 2009.

Summary

This dissertation focuses on the design and evaluation of speech-based conversational interfaces for task-oriented dialogues. Conversational interfaces are software programs enabling interaction with computer devices through natural language dialogue. Even though processing conversational speech is a challenging problem, mainly because users' spoken language can be extremely variable, the speech modality remains an attractive option because of its naturalness: speech is learned since childhood, that means, users neither need to learn nor to adapt to the designer's interaction style. Also, speech can be very useful in situations when users cannot use other input modalities (e.g. while driving, accessing the interfaces over the phone, using pocket size devices or when impaired). The technological growth of the past decades brought conversational interfaces to a level of maturity which allows widespread application. Examples include interactive information systems, in-car applications, smart environments, media guides, training and educational systems, social robots, and so on. Therefore, the design and evaluation of conversational interfaces towards achieving a better interaction quality are as crucial as ever.

The dissertation is organized in two parts: part I presents the theoretical background concerning design and evaluation approaches (chapter 2-3) while part II focuses on experimental studies (4-8) .

Chapter 2 is concerned with various theoretical and practical modeling approaches of human communication with applicability in the design of conversational interfaces. The theories refer to the core structure of dialogues and help in understanding how verbal and non-verbal exchange occurs in natural circumstances. The chapter contributes with a newly compiled set of guidelines for dialogue and interaction design for spoken conversational interfaces from the reviewed literature.

Chapter 3 reports on the evaluation of conversational interfaces from the perspective of interaction quality. Since the notion of quality is central in this work the chapter shows an overview of several definition approaches. Further, the chapter deals with evaluation methods and taxonomies of quality aspects. The contribution of the chapter consists of a taxonomy of conversational interaction quality focusing on hedonic and pragmatic quality aspects. The taxonomy war built on theoretical work presented in the domain literature and was used to guide the evaluations performed in the empirical chapters.

Chapter 4 presents an empirical study concerning the impact of voice accent. The study demonstrated that the voice accent is a critical design issue of VUIs as it strongly affects the user perceptions of other system features, such as dialogue

structure or voice quality. Additionally, the chapter contributes with the prototype design of a novel application: a voice enabled user manual for mobile phones. We showed how to design a prototype by creating and implementing a dialogue script from written instruction and how to improve the script using human spoken dialogues.

Chapter 5 contributes with a novel approach of evaluating the adequacy of conversational structures implemented in conversational interface using a new concept: *verbal affordance*. The concept was used to evaluate the adequacy of dialogue structures implemented in a multimodal question answering system. The chapter contains an additional study concerning the appearance of the agent attached to the system. The study showed the effects of gender-marked vs. gender ambiguous agent's look on the system evaluation. The gender ambiguity was implemented as a combination between a female voice with a rather masculine look, thus as voice inconsistency with physical look. The results showed that users tend to prefer gender marked agents and the preference has consequences on the system evaluation.

Chapter 6 addresses the evaluation of a social robot in an open uncontrolled environment. We used languages features (in combination with gestures and body movements) to design the social skills of a robot receptionist. The study explored relationships between the robot's social skills and the way users reacted and evaluated the robot. Additionally, the study aimed to determine additional important interaction quality features with potential general validity.

Chapter 7 presents an empirical study performed with participants acting as visitors and receptionists in a controlled environment. The goal of the study was to identify multimodal communication patterns evolving in human-human task oriented dialogues that can be of potential use for modeling efficient human-robot dialogues. The study outcome consists in a set of 10 recommendations for modeling task-oriented dialogues with a receptionist.

Chapter 8 explores the effects of empathy, humor and voice pitch manipulations in interaction with a social robot receptionist. Our ultimate goal was to determine how the variable manipulations influence users' perception concerning the robot and the interaction quality. The design of the dialogues was based on the recommendations and data collected from the study performed in chapter 7. Results showed that among the variables manipulated, the voice pitch has the strongest effects on the evaluation while the empathy achieves only minor effects.

Finally, we present the general conclusions of this thesis including the discussion of the main results, design implication for conversational interfaces and avenues for future work.

Rezumat

Lucrarea de faţă, prezentată ca teză de doctorat, se ocupă de proiectarea şi evaluarea interfeţelor conversaţionale bazate pe vorbire pentru dialogurile cu scop practic. Interfeţele conversaţionale sunt programe software care permit interacţiunea cu calculatorul printr-un dialog natural. Chiar dacă prelucrarea conversaţiei reprezintă o provocare, în principal pentru că limbajul vorbit al utilizatorilor poate fi extrem de variat, modalitatea de exprimare rămâne o opţiune atractivă datorită naturaleţei sale: vorbirea se învaţă din copilărie ceea ce înseamnă că utilizatorii nu au nevoie nici să înveţe şi nici să se adapteze la stilul de interacţiune al programatorului. De asemenea, vorbirea poate fi foarte utilă în situaţiile în care utilizatorii nu pot folosi alte modalităţi de comunicare (de exemplu la volan, în accesarea interfeţelor prin telefon, în folosirea dispozitivelor de mici dimensiuni sau în cazul unor dizabilităţi).

Dezvoltarea tehnologică din ultimele decenii a ridicat interfeţele conversaţionale la un nivel de maturitate care permite aplicarea lor pe scară largă. Exemplele includ sisteme interactive de informare, aplicaţii la bordul autovehiculelor, medii inteligente, ghiduri media, sisteme de formare şi educaţionale, roboţi sociali şi aşa mai departe. Prin urmare, proiectarea şi evaluarea interfeţelor conversaţionale pentru realizarea unei interacţiuni de o mai bună calitate sunt mai importante ca oricând.

Teza este organizată în două părţi: prima parte prezintă fundamentele teoretice ale proiectării şi evaluării interfeţelor conversaţionale (capitolele 2–3), iar partea a doua are în vedere studii experimentale (4–8).

Capitolul al 2-lea tratează despre diferite abordări teoretice şi practice de modelare a comunicării umane, cu aplicabilitate în proiectarea de interfeţe conversaţionale. Teoriile se referă la structura de bază a dialogurilor şi ajută la înţelegerea modului în care se produc comunicarea verbală şi comunicarea non-verbal în circumstanţe naturale. Capitolul aduce contribuţii noi în domeniu, bazate pe literatura de specialitate, prin formularea unui set de linii directoare privind dialogul şi programarea de interacţiune destinată interfeţelor conversaţionale vorbite.

Capitolul al 3-lea se ocupă de evaluarea interfeţelor conversaţionale din perspectiva *calităţii* interacţiunii. Deoarece *calitatea* este un obiectiv principal în această lucrare, capitolul prezintă o privire de ansamblu asupra mai multor definiţii date acestei noţiuni. Apoi urmează o analiză a metodelor de evaluare şi o prezentare a taxonomiilor privind aspectele legate de calitate. Contribuţia capitolului constă într-o taxonomie a calităţii interacţiunii conversaţionale care se concentrează pe calitatea hedonică şi pragmatică. Taxonomia este construită pe conceptele teoretice prezentate în literatura de specialitate şi este utilizată ca îndrumar în evaluările efectuate în capitolele empirice.

Capitolul al 4-lea prezintă un studiu empiric referitor la impactul pe care îl are accentul vocii. Studiul a demonstrat că accentul vocii reprezintă o problemă critică în proiectarea VUI, deoarece afectează puternic percepţiile celorlalte caracteristici ale sistemului, cum ar fi structura dialogului sau calitatea vocii. În plus, capitolul oferă proiectarea prototip al unei noi aplicaţii: un manual al utilizatorului cu activare vocală pentru telefoane mobile. Scopul a fost să se arate felul în care se proiectează un prototip prin crearea şi punerea în aplicare a unui script de dialog cu pornire de la instrucţiuni scrise şi cum se poate îmbunătăţi scriptul folosind dialoguri umane vorbite.

Capitolul al 5-lea oferă o abordare nouă în evaluarea caracterului adecvat al structurilor conversaţionale implementate în interfaţa conversaţională prin utilizarea unui concept nou: *intuitivitatea verbală*. Conceptul a fost utilizat pentru a evalua gradul de adecvare a structurilor dialogale implementate într-un sistem multimodal de răspunsuri la întrebări. Capitolul conţine un studiu suplimentar cu privire la aspectul „agentului“ ataşat sistemului. Studiul arată efectele genului ambiguu al agentului asupra evaluarii sistemului. Ambiguitatea de gen a fost implementată ca o combinaţie între o voce feminină şi un aspect fizic relativ masculin, sub forma contradicţiei dintre voce şi aspectul fizic. Rezultatele arată că utilizatorii tind să prefere agenţii cu gen univoc şi că această preferinţa are consecinţe asupra sistemului de evaluare.

Capitolul al 6-lea se referă la evaluarea unui robot social în cadrul unui experiment necontrolat. Am folosit caracteristici lingvistice (în asociere cu gesturi şi mişcări ale corpului) pentru a proiecta aptitudinile sociale ale unui robot recepţioner. Studiul explorează relaţiile dintre aptitudinile sociale ale robotului şi felul în care utilizatorii au reacţionat şi au evaluat robotul. În plus, studiul a urmărit şi determinarea altor caracteristici calitative interactive importante, cu o posibilă valabilitate generală.

Capitolul al 7-lea prezintă un studiu empiric realizat cu participanţi „vizitatori“ şi recepţioneri în cadrul unui experiment controlat. Scopul studiului este acela de a identifica tipare de comunicare multimodale care evoluează în dialogurile om-om cu scopuri practice şi care ar putea fi utile în modelarea de dialoguri eficiente om-robot. Rezultatul studiului constă într-un set de 10 recomandări pentru modelarea dialogurilor cu scop practic cu un recepţioner.

Capitolul al 8-lea explorează efectele empatiei, ale umorului şi ale manipulării înălţimii vocii în interacţiunea cu un robotul social recepţioner. Scopul nostru final este de a determina modul în care manipulările variabilelor influenţează percepţia utilizatorilor cu privire la robot şi calitatea interacţiunii. Proiectarea dialogurilor s-a bazat pe recomandările şi datele colectate din studiul efectuat în capitolul al 7-lea. Rezultatele arată că printre variabilele manipulate, înălţimea (frecvenţa) vocii are cel mai puternic efect asupra evaluării, în timp ce empatia produce doar efecte minore.

În cele din urmă prezentăm concluziile generale ale acestei teze, inclusiv principalele rezultatele, implicaţiile proiectării pentru interfeţele conversaţionale şi perspectivele pentru cercetări viitoare.

Appendix A - Chapter 4

Pronunciation features of SSE

This is a summarization of the main pronunciation differences between Singapore Standard English (SSE) and British Standard English (BSE) as presented in the work of A. Brown [126].

Vowels

The largest difference between Singapore English and many other varieties of English is that there is generally no difference between long and short vowels. Singaporeans are inclined to shorten the long vowels considerably: ‘see‘, would be pronounced [sI]. The following pairs or sets of words would be homophones in SSE:

a) Marry / merry [e]
b) pull / pool [U]
c) cot / caught [V]
d) come / calm [Q]
e) sit / seat [I]

Diphthongs

Singaporeans have the tendency to pronounce diphthongs, such as [ou] or [ei] like a single vowel, [o] or [e]. This kind of variation occurs in words like ’home’, ’so’, ’hope’ (for [ou]) and ’take’, ’made’ (for [ei]).

Consonants

1. Some speakers do not make the distinction between the following initial and finally pairs (technically known as voiceless and voiced consonants respectively):

Initial sounds:

a) [p] and [b] sounds (both sounding like [p]), in words like ’park’/’bark’;
b) [t] and [d] sound (both sounding like [t]), in word like ’tan’/’Dan’;
c) [k] and [g] sounds (both sounding like [k]), in words like ’come’/’gum’;
d) [ch] and [j] sounds (both sounding like [ch]), in words like ’cheap’/’jeep’;

Final sounds:

a) [p] and [b] sounds (both sounding like [p]), in words like ’rip’/’rib’;
b) [t] and [d] sound (both sounding like [t]), in word like ’seat’/’seed’;
c) [k] and [g] sounds (both sounding like [k]), in words like ’pick’/’pig’;
d) [ch] and [j] sounds (both sounding like [ch]), in words like ’rich’/’ridge’;
d) [f] and [v] sounds (both sounding like [f]), in words like ’leaf’/’leave’;
e) [s] and [z] sounds (both sounding like [s]), in words, like ’rice’/’rise’;

2. Most speakers lack a clear distinction between the following initial and final sounds:

Initial sounds:

a) [t] and [th] (both sounding like [t]), like ’taught’/’think’;
b) [d] and [th] (both sounding like [d]), like ’day’/’they’;

Final sounds:

a) [th] and [f] sounds (both sounding like [f]), in words like ’death’ and ’deaf’;
b) [th] and [v] sounds (both sounding like [v]), in words like ’breathe’ and ’brave’;

3. Many speakers omit the following final sounds:

a) [l] sound, so that ’drawl’ and ’temple’ sound identical ’to draw’ and temper’;
b) [n] sounds, so that ’brown’ resembles ’brow’;
c) [t] and [d] sounds; this has obvious consequences for [ed] endings: supposed to’ becomes ’suppose to’.

Appendix B - Chapter 4

Questionnaire

PART I

I. PERSONAL DETAILS

1.1 Age range**:** ☐ < 25 ☐ 26-30 ☐ 31- 34 ☐ 35- 44 ☐ 45-54 ☐> 54

1.2 Gender: ☐ female ☐male

1.3 Mother tongue: ☐ Chinese ☐ Malay ☐ Indian ☐ Other:___________

1.4 Most frequent language/s used in daily communication _____________

1.5 Educational background: ☐ technical ☐ non-technical

1.6 Your highest educational qualification: ☐ Bachelor ☐ Master ☐ PhD ☐ Others ___

II. EXPERIENCE WITH THE SYSTEM

2. 1 How do you rate the overall quality of the voice:

1	2	3	4	5
☐	☐	☐	☐	☐
Excellent	Good	Fair	Poor	Bad

2.2 Did you notice any anomalies in the pronunciation?

1	2	3	4	5
☐	☐	☐	☐	☐
Yes very annoying	Yes annoying	Yes slightly annoying	Yes but not annoying	No

2.3 The average speaking rate of delivery was:

1	2	3	4	5
☐	☐	☐	☐	☐
Much faster than preferred	Faster than preferred	Preferred	Slower than preferred	Much lower than preferred

2.4 How would you describe the voice?

1	2	3	4	5
☐	☐	☐	☐	☐
Very unpleasant	Unpleasant	Fair	Pleasant	Very pleasant

2.5 How would you rate the system's reactions in terms of politeness?

1	2	3	4	5
☐	☐	☐	☐	☐
Very Polite	Polite	So-So	Impolite	Very impolite

2.6 The handling of the system using spoken dialogues was easy:

1	2	3	4	5
□	□	□	□	□
Strongly agree	Agree	Neither agree nor disagree	disagree	strongly disagree

2.7 The system delivered trustworthy information:

1	2	3	4	5
□	□	□	□	□
Strongly agree	Agree	Neither agree nor disagree	disagree	strongly disagree

---

PART II
III. FURTHER REMARKS

3.1 Do you think a voice user manual for mobile phones is useful?

□ Yes   □ Maybe   □ No

3.2 In the future do you think you would use it?

□ Yes   □ Maybe   □ No

3.3 You enjoyed the dialogues with the voice user manual:

1	2	3	4	5
□	□	□	□	□
Strongly agree	Agree	Neither agree nor disagree	disagree	strongly disagree

3.4 What is your overall impression about the interaction with the voice user manual?

____

____

3.4.1 What did you like about the application (positive remarks) ?

____

____

3.4.2 What didn’t you like about the application (negative remarks)?

____

____

3.5. Improvement suggestions, if any

____

____

Appendix C - Chapter 5

<u>Quantitative questionnaire</u>

I. PERSONAL DETAILS

1.1 Age range: ☐ < 25 ☐ 26-30 ☐ 31- 34 ☐ 35-44 ☐45-54 ☐>54

1.2 Gender: ☐ female ☐ male

1.3 Educational background: ☐ technical ☐ non-technical

II. EXPERIENCE WITH THE SYSTEM

2.1 The level of concentration required when performing the task with the system was:

Very low — Very high

2.2 How easy was to use the system?

Very easy — Very difficult

2.3 How flexible was the system?

Very flexible — Very inflexible

2.4 How accurate was the system?

Very accurate — Very inaccurate

2.5 The system way of expression was :

Very clear — Very unclear

2.6 I enjoyed interacting with the system

Not at all — Very much

2.7 Overall I felt comfortable interacting with the system

Very comfortable　　　　Very uncomfortable

2.8 I felt calm using the system

Very calm　　　　Very tense

2.9 I felt confident using the system

Not at all　　　　Very confident

2.10 Rate the overall interaction quality of the system

Extremely poor　　　　Excellent

Appendix D - Chapter 5

Qualitative interview

PART I

1.1 Imagine you would have a medical question: where would you look for an answer?

1.2 Which are the situations in which you might need medical information?

1.3 Do you know what a 'question-answering' system is?

**if no: short explanation & find and example; continue with 4.0*

**if yes:*

3.1 Could you describe it?

3.2 Have you ever used such a system?

**if not go to 4.0*

**if yes*:

3.2.1 How often?

3.2.2 With which purpose?

1.4 Do you think that such system might be useful for somebody looking for medical information?

1.5 Do you think you might use such a system?

1.6 Imagine you have access to such a QA system for medical questions:

6.1 How do you imagine the dialog/conversation with the system?

6.2 What kind of information and in which form are you expecting from the system?

6.3 Do you think you can handle the system? (*do you expect it to be easy/difficult?*)

6.4 How should the system assist you in order to make you a happy costumer?

1.7 Which advantages/disadvantages might such a system have?

1.8 If you could choose what would you prefer?

1.8.1 To stick to facilities that you know?

1.8.2 To work with a system that challenges you?

PART II

2.1 After you perform the test with these three systems could please tell me if the interaction was similar to the way you imagined it?

7.1 If not, why not? What was different?

2.2 Is there a system you preferred? Which system did you like most? And why?

2.3 Could you please tell me – resuming – what did you like about the interaction?

2.4. Could you please tell me – resuming – what you did not like about the interaction?

2.5 Did you have communication problems?

2.5.1 If yes, could you please specify where exactly and with which system?

2.6 Was it clear to you from the beginning how to approach the system? Was it intuitive?
Or maybe you had to read too much text on the screen before getting started?

2.7 Did you have the feeling the system understood your question? I mean ,did you have the feeling you were using the right words?

2.8 During the interaction did the system make clear enough how you were expected to put your next questions? I mean, did you know what to do next?

2.9 Where there pauses in the dialogue?
- 9.1 If yes, how often and with which system?
  - 9.1.1 Were you upset /irritated by the pauses?
  - 9.1.2 Were the system's reactions fast enough?

2.10 How nice/appealing did the interface look?

2.11 What do you think about the embodied characters/talking heads? (*show picture*)
- 2. 11.1 Did they appear friendly? Nice? Trustworthy?
- 2. 11.2 Which one did you like most?
- 2. 11.3 What didn't you like about the heads? How disturbing was **that*?
- 2. 11.4 Did the heads look appropriate for their job? Which one was the most appropriate?
- 2. 11.5 Were the heads useful?

2. 12 Did you understand the heads' voice? Were the voices clear?
- 2. 12.1 Were they pleasant?
- 2. 12.2 Which one was more pleasant?

2. 13 Where the heads talking politely to you?

2. 14 Where their statements nicely formulated?

2.15 Did you like the way the information was presented to you?
- 2. 15.1 Was the information well distributed, easy to read?
- 2. 15.2 Was the font big enough ? (*easy to read*)

2.16 Was the content presented by the systems interesting to you?
- 2.16.1 Was it new?
- 2.16.2 Was it understandable?
- 2.16.3 Was the information complete?
- 2.16.4 Was the content relevant to the question you asked?

2.17 Was it fun interacting with this kind of system?

2.18 How did you feel during the interaction?
- 2. 18.1 Was it stressful? Tiring? Or did you feel relaxed?

2.19 Were the tasks difficult to accomplish?

2.20 Did you find the info you were looking for?

2.21 Did it take long to finally get the answer you were looking for?

2.22 Did the systems make mistakes? If yes do you remember where exactly?
- 2. 22.1 Could the system solve the problem reasonably well?
- 2. 22.2 Was it also easy to "recover" from such a mistake?

2.23 Did the systems appear to be as delivering trustworthy information?

2.24 Did the systems appear flexible enough?

2.25 Did you get feedback from the systems?
- 2.25.1 Did you understand that feedback?

2.26 During the whole test did you have the feeling you were in control of the situation?

2.27 Did you generally enjoy interacting with the systems?

2.28 Would you use the services of such systems? Would you recommend these systems to other people?

2.29 Now if you compare the 3 systems A, B, C :

29.1 Which one did you like most?
29.2 Which one was more fun?
29.3 With which one you enjoy it more?
29.4 Which one appeared more flexible?

2.30 Is there something else that you would like to add about your experience with the systems?

Appendix E - Chapter 6

Quantitative questionnaire

SECTION 1: PERSONAL DETAILS

1.1 Age range: ☐ < 18 ☐ 18-25 ☐ 26-30 ☐31-40 ☐ 41- 50 ☐ > 50

1.2 Gender: ☐ female ☐male

1.3 Ethnical group: ☐ Chinese ☐ Malay ☐ Indian ☐ Other:_________

1.4 Most frequent language/s used in daily communication _____________

1.5 Work area:
☐ Social science ☐ Business ☐ IT & Engineering ☐ Others :________________

1.6 Highest educational degree: _____________

SECTION 2: BACKGROUND KNOWLEDGE

2.1 My knowledge in social robotic:
☐ Design and Development ☐ I have seen social robots ☐ None
☐ I have interact with social robots ☐ I have read about social robots

2.2 Did you ever interact with a device that recognize speech and/or gesture?
☐ yes ☐ no
If *"yes"* :

2.2.1 What kind of device? ______________

2.2.2 How often? ______________

2.2.3 What was the purpose?____________________

SECTION 3: YOUR ROBOT EXPERIENCE

3.1 Interacting with the robot was easy

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.2 I enjoyed the interaction with the robot

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.3 I felt comfortable interacting with the robot

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.4 The information delivered by the robot was useful:

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.5 A high level of concentration was required while performing a task with the robot:

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.6 The robot reacted flexible:

	1	2	3	4	5	
strongly disagree	☐	☐	☐	☐	☐	strongly agree

3.7 During the interaction I felt:

	1	2	3	4	5	
4.1 very tense	☐	☐	☐	☐	☐	very calm

	1	2	3	4	5	
4.2 very frustrated	☐	☐	☐	☐	☐	very satisfied

4. Please rate the following robot capabilities:

4.1 able to recognize speech:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.2 able to recognize objects:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.3 able to follow objects:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.4 able to express emotions:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.5. able to use natural gestures:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.6. able to exhibit personality traits:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.7 able to socialize with human:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.8. able to be express humor

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

4.9. able to respond fast

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

5. Please rate the overall interaction quality:

	1	2	3	4	5	
very bad	☐	☐	☐	☐	☐	very good

6. Please rank the following robot features according to their importance for the conversational interaction quality:

Example: not important at all **|-----|-----|-----|--X--|-----|-----|-----|** *extremely important*

6.1 Pleasant voice:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.2 Nice physical appearance:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.3 Gestures and mimic while talking:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.4 Emotion displaying:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.5 Gender displaying:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.6 Age displaying:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.7 Polite way of talking:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.8 Humorous way of talking:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.9 Clear way of talking:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.10 Fast response:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.11 Delivering relevant information:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.12 Displaying a friendly behavior:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.13 Making clearly what the robot can/cannot do:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.14 Speech recognition without errors:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.15 Object recognition without errors:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.16 Easy recovering from errors:
not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

6.17 Others : _____________

not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

____

not important at all |-----|-----|-----|-----|-----|-----|-----| extremely important

7. Please rank which part of the robot attracted your attention the most (1- most attractive)

	Rank
8.1 Arm	☐
8.2 Head	☐
8.3 Body	☐
8.4 Others: _________	☐
_________	☐

8. Reason(s) for the rank: _______________

9. Would you accept a robot to work in the following areas (multiple-choice)?

☐ Office (Receptionist)
☐ Restaurant Waiter/Waitress)
☐ City (Patrolling – no weapons)
☐ Hospital (Carrying patient)
☐ Hotel (Carrying your luggage)
☐ Others:______________

SECTION 4: COMMENTS AND SUGGESTIONS

10. What did you like/dislike about the robot?

Positive features

____

____

____

Negative features

____

____

____

11. Please suggest a future conversation topic(s) between you and the current robot:

____

____

12. Please write down some example questions related to the topic(s) you proposed:

____

____

13. Improvement suggestions or other comments:__________________________________

____

____

Appendix F - Chapter 7

Interview - Visitor

1.0 General impressions about the receptionist
1.1 What was your general impression about the interaction with the receptionist?

1.2 Was the receptionist helpful?

2.0 Understanding the style of interaction
2.1Were the explanations given by the receptionist clear? (*did you have any communication problems, misunderstandings, etc.?)*

2.2 Were the answers relevant to your questions?

2.3 Were the answers informative?

2.4 Did you like the way the receptionist explained what you had to do? (*if not*, *do you maybe have suggestions for improvements for the receptionist?*)

2.5 Were you satisfied with the receptionist's service? Give a rating from 1 to 10!

2.6 What did/didn't you like about the receptionist?

3.0 About the display counter
3.1 Was the information presented on the display counter useful ?

3.2 What did/didn't you like about the display?

3.4 Do you have any suggestions for improvements? (for example, something that you missed and could be useful to be presented on the display?)

4.0 Visitors' expectations towards a good receptionist
4.1 What would you expect from a good receptionist?

Let's take an example: you are entering a building and at the front door you see a receptionist:

4.2 What do you expect to happen?

4.3 How should the receptionist behave when a person is approaching? (*gesture & talk & address*)

4.4 What kind of greetings should a good receptionist use?

4.5. What kind of information should a receptionist deliver in general?

4.6. What kind of knowledge is desirable for a good receptionist, in your opinion?

5.0 Future use of a robot receptionist

5.1 What would you think about a company having a robot receptionist? Will your impression of the company change?

5.2 What are the advantages of having a robot receptionist?

5.3 What are the tasks the robot receptionist will be able to perform?

5.3 What other conversation topics would you like to have with a robot receptionist ? (some example, such as family, sports, travel, etc.)

Interview – Receptionist

1. What tasks do you think a receptionist has to do – from own experience or the experience of others?

2. How different from your previous experience were the tasks you had to perform today?

3. What qualities should a good receptionist have?

4. How do you think we can improve the interaction with the visitors in the experiment?

Appendix G - Chapter 8

SECTION 1: PERSONAL DETAILS

1.1 Age range: □ < 18 □ 18-25 □ 26-30 □31-40 □ 41- 50 □ > 50

1.2. Gender: □ female □ male

1.3 Ethnical group: □ Chinese □ Malay □ Indian □ Other:_________

1.4 Most frequent language/s used in daily communication _____________

1.5 Work area:
□ Social science □ Business □ IT & Engineering □ Others :________________

1.6 Highest educational degree: _____________

SECTION 2: BACKGROUND KNOWLEDGE

2.1 My knowledge in social robotic:
□ Design and Development □ I have seen social robots □ None
□ I have interact with social robots □ I have read about social robots

2.2 Did you ever interact with a device that recognize speech and/or gesture?
□ yes □ no
If "*yes*" :

2.2.1 What kind of device? ______________

2.2.2 How often? ______________

2.2.3 With which purpose?____________________

2.2.2 How often? ___________________________

2.2.3 With which purpose?_____________________
____

2.2.4 Was it a good experience? ______________

If „no " 2..2.4.1. Why not? ____________________

2.3 What expectation do you have from a social robot, (i.e. do you have high/low expectation)?

____

____

____

____

<u>I Receptionist appearance</u>

<u>Look</u>

presentable ☐ ☐ ☐ ☐ ☐ ☐ ☐ unacceptable
professional ☐ ☐ ☐ ☐ ☐ ☐ ☐ unprofessional
pleasant ☐ ☐ ☐ ☐ ☐ ☐ ☐ unpleasant
intelligent ☐ ☐ ☐ ☐ ☐ ☐ ☐ unintelligent
inviting ☐ ☐ ☐ ☐ ☐ ☐ ☐ rejecting

Please rate the overall **look:**
appealing ☐ ☐ ☐ ☐ ☐ ☐ ☐ unappealing

<u>Voice</u>

clear ☐ ☐ ☐ ☐ ☐ ☐ ☐ unclear
accurate pronunciation ☐ ☐ ☐ ☐ ☐ ☐ ☐ highly accented pronunciation
matches look ☐ ☐ ☐ ☐ ☐ ☐ ☐ doesn't match look
pleasant ☐ ☐ ☐ ☐ ☐ ☐ ☐ unpleasant

Please rate the overall **voice**:
appealing ☐ ☐ ☐ ☐ ☐ ☐ ☐ unappealing

<u>Speaking style</u>

elegant ☐ ☐ ☐ ☐ ☐ ☐ ☐ rough
polite ☐ ☐ ☐ ☐ ☐ ☐ ☐ impolite
fast ☐ ☐ ☐ ☐ ☐ ☐ ☐ slow
loud ☐ ☐ ☐ ☐ ☐ ☐ ☐ quite

Please rate the overall **speaking style**:
appealing ☐ ☐ ☐ ☐ ☐ ☐ ☐ unappealing

Please rate the overall appearance of the robot in terms of look , voice and speaking style
appealing ☐ ☐ ☐ ☐ ☐ ☐ ☐ unappealing

<u>II Content & task related ratings</u>

Content presented was:
exciting ☐ ☐ ☐ ☐ ☐ ☐ ☐ lane

Content presentation was :
conventional ☐ ☐ ☐ ☐ ☐ ☐ ☐ original

Tasks to complete were
common ☐ ☐ ☐ ☐ ☐ ☐ ☐ new
undemanding ☐ ☐ ☐ ☐ ☐ ☐ ☐ challenging
easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ difficult

Please rate the overall degree of entertainment while performing the tasks with the robot
entertaining ☐ ☐ ☐ ☐ ☐ ☐ ☐ dull

III. Receptionist behavior

The receptionist's behavior was
humorous ☐ ☐ ☐ ☐ ☐ ☐ ☐ boring
friendly ☐ ☐ ☐ ☐ ☐ ☐ ☐ unfriendly
likeable ☐ ☐ ☐ ☐ ☐ ☐ ☐ unlikeable
receptive/empathic ☐ ☐ ☐ ☐ ☐ ☐ ☐ unreceptive/indifferent
helpful ☐ ☐ ☐ ☐ ☐ ☐ ☐ unhelpful

Please rate the overall impression about the receptionist's **behavior:**

appealing □ □ □ □ □ □ □ unappealing

The receptionist's personality was:

emotional □ □ □ □ □ □ □ rational
weak □ □ □ □ □ □ □ strong
submissive □ □ □ □ □ □ □ assertive
introvert □ □ □ □ □ □ □ extrovert

Please rate the overall impression about the receptionist's **personality**

appealing □ □ □ □ □ □ □ unappealing

IV. Your feelings during the interaction with the receptionist:

motivated □ □ □ □ □ □ □ discouraged
annoyed □ □ □ □ □ □ □ pleased
calm □ □ □ □ □ □ □ tense
confident □ □ □ □ □ □ □ insecure
awkward □ □ □ □ □ □ □ comfortable
frustrated □ □ □ □ □ □ □ satisfied
in control □ □ □ □ □ □ □ controlled by the robot

V. Olivia in this scenario was

a lot like me □ □ □ □ □ □ □ not at all like me
flexible □ □ □ □ □ □ □ inflexible

VI Interaction features

Interacting with the robot was:

easy □ □ □ □ □ □ □ difficult
fast □ □ □ □ □ □ □ slow

The answers to my questions were

informative □ □ □ □ □ □ □ not informative
relevant □ □ □ □ □ □ □ not relevant
clear □ □ □ □ □ □ □ unclear
trustable □ □ □ □ □ □ □ not trustable
useful □ □ □ □ □ □ □ not useful

I know in very moment how to talk with the robot

Strongly agree □ □ □ □ □ □ □ strongly disagree

VII Please rate the following robot capabilities:

speech accuracy **excellent** □ □ □ □ □ □ □ **very poor**
able to express emotions **excellent** □ □ □ □ □ □ □ **very poor**
able to use natural gestures **excellent** □ □ □ □ □ □ □ **very poor**
able to exhibit personality traits **excellent** □ □ □ □ □ □ □ **very poor**
able to socialize with humans **excellent** □ □ □ □ □ □ □ **very poor**
able to be express humor **excellent** □ □ □ □ □ □ □ **very poor**

Please rate the overall degree of enjoyment**:**

high □ □ □ □ □ □ □ low

Rate how much you like the robot in this scenario:

very much □ □ □ □ □ □ □ not at all

Rate the overall quality of the interaction

excellent □ □ □ □ □ □ □ very poor

Appendix H - Chapter 8

Statistical data

Item	Empathic robot Median	Non-empathic Robot Median	U-values	p-values
Receptive	6	5	50.5	.039
Emotional/Rational	6	4	30.5	.002
Overall behavior	6	5	50	.048
Confident	5	4	48	.033
Interaction ease	6	5	50	.037

Table: *Results empathic/non-empathic robot*

Item	Low-pitch robot Median	High-pitch robot Median	Z-values	p-values
Robot’s voice	4	5	-2.080	.038
Overall aesthetic appeal	4	5	-2.392	.017
Robot’s behavior	4	5	-3.988	.000
Overall personality	4	5	-2.181	.029
Extrovert/Introvert	4	4	-2.156	.031
Rational/Emotional	6	5	-3.315	.001
Strong/Weak	4	4	-2.006	.046
A lot like me	3	3	-2.600	.009
User feelings	3.87	4.57	-2.213	.027
Interaction features	4.50	4.83	-2.014	.044
Content presentation	3	4	-3.123	.002
Overall entertainment	4	5	-3.035	.002
Overall enjoyment	3	5	-3.318	.001
Interaction quality	4	5	-3.274	.001

Table: *Results high-pitch/low-pitch robot*

Item	Humorous robot Median	Non- humorous robot Median	Z-values	p-values
Robot’s behavior	5.33	4.58	-3.229	.001
Robot’s voice	5.50	5	-3.330	.001
Robot social skills	5	3	-3.986	.000
Interaction features	4.7	5.2	-2.362	.005
Overall enjoyment	5	4	-3.348	.028
Interaction quality	5	5	-2.029	.042
Speaking style	5.33	5	-2.824	.005
Tasks	4.66	5.50	-2.148	.032
Strong/Weak	4.5	4	-2.653	.008
Emotional/Rational	3	5	-3.724	.000
Friendly	6	5	-3.214	.001
Extrovert/Introvert	5	4	-2.527	.011

Table: *Results humorous robot/non-humorous robot*

APPENDIX H - CHAPTER 8

Item	Condition / Scenario	Female Median	Male Median	U-values	p-values
Comfortable	empathic /1	5	3	4	.012
Robot's voice	non-empathic/1	4.1	5.91	4.5	.019
User feeling	non-empathic/1	3	4.57	3	.028
Useful	non-empathic/1	5	7	4.5	.019
Overall personality	high-pitch/2	4	5	31	.013
User feelings	high-pitch/2	4	5	36	.029
Overall enjoyment	high-pitch/2	4	5	31.5	.013
Transparency	high-pitch/2	4	6	27	.036
Extrovert/Introvert	high-pitch/2	5	4	27	.006
Professional look	low-pitch/3	3	5	6.5	.000
Look matches face	low-pitch/3	2	3	23.5	.039
Speech rate	low-pitch/3	2	4	35	.025
Content presentation	low-pitch/3	2	4	28	.008
Overall behavior	low-pitch/3	3	4	39	.045
Informativeness	humorous/5	7	6	38	.019
Clear answer	humorous/5	6.50	5.50	33.5	.012
Able to socialize	humorous/5	6	4	38.5	.027

Table: *Results within groups: females vs. male participants*

Item	Condition / Scenario	Non-IT Median	IT Median	U-values	p-values
Extrovert/Introvert	empathic/1	5	3.5	4.5 .	008
Robot social skills	humorous/5	5.2	4	36.5	.047

Table: *Results within groups: participants with non-IT vs. IT work background*

Item	Condition/ Scenario	ASR non-experienced Median	ASR experienced Median	U-values	p-values
Emotional/Rational	empathic/1	5.50	3	6	.018
User feelings	empathic/1	3.64	4.64	7	.028
Able to socialize	empathic/1	4.50	3	8	034
Able to socialize	non-empathic/1	4.50	2	4	.050
Able to socialize	high-pitch/2	5	3	33	.012
Able to socialize	low-pitch/3	5	3	28	.008
Able to socialize	non-humorous/4s	5	3	28	.008
Able to socialize	humorous/5	5	3	30	.010

Table: *Results within groups: ASR devices non-experienced participants vs. ASR experienced*

Item	Condition/ Scenario	High expectation Median	Low expectation Median	U-values	p-values
Flexible	high-pitch/2	5	4	43	.036
Robot's speaking style	low-pitch/3	4.16	5	40	.023
Answer quality	low-pitch /3	4.75	5.50	40	.031

Table: *Results within groups :participants with high vs. low expectations*

Item	Condition/ Scenario	Extrovert Median	Introvert Median	U-values	p-values
Interaction easiness	low-pitch/3	5	6	39	.012
Receptive	low-pitch/3	5	3	15	.000

Table: *Results within groups: participants with extrovert vs. introvert personality*

Item	Condition / Scenario	Gaze downwards Median	Gaze forwards/ mixed Median	U-values	p-values
Voice appeal	low-pitch /3	3	4	19	.007
Social skills	low-pitch/4	2	3	35.5	.041
Ability emotions	low-pitch/4	2	3	22.00	.004
A lot like me	low-pitch/4	2	3.50	27.50	.011

Table: *Results within user behavior group: gaze downward vs. gaze forwards/mixed*

Item	Condition / Scenario	Smile Median	No smile - Median	U-values	p-values
Social skills	humorous/5	5.3	4.5	40.5	.036

Table: *Results within user behavior group: smile vs. no smile*

Item	Condition / Scenario	Upper body forwards Median	No movement Median	U-values	p-values
Overall interaction quality	low-pitch/3	3	4	29	.022
Overall interaction quality	non-humorous/4	3	5	28.5	.019
Overall comfort	low pitch/ 4	3	4	26.5	.018
Overall enjoyment	low-pitch/4	3	5	23.5	.004

Table: *Results within user behavior group: upper body forwards vs. no movement forwards*

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D.-students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series.

2011-49 Andreea Niculescu (UT), *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality.*
2011-48 Mark ter Maat (UT), *Response Selection and Turn-taking for a Sensitive Artificial Listening Agen.*
2011-47 Azizi Bin Ab Aziz(VU), *Exploring Computational Models for Intelligent Support of Persons with Depression.*
2011-46 Beibei Hu (TUD), *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work.*
2011-45 Herman Stehouwer (UvT), *Statistical Language Models for Alternative Sequence Selection.*
2011-44 Boris Reuderink (UT), *Robust Brain-Computer Interfaces.*
2011-43 Henk van der Schuur (UU), *Process Improvement through Software Operation Knowledge.*
2011-42 Michal Sindlar (UU), *Explaining Behavior through Mental State Attribution.*
2011-41 Luan Ibraimi (UT), *Cryptographically Enforced Distributed Data Access Control.*
2011-40 Viktor Clerc (VU), *Architectural Knowledge Management in Global Software Development.*
2011-39 Joost Westra (UU), *Organizing Adaptation using Agents in Serious Games.*
2011-38 Nyree Lemmens (UM), *Bee-inspired Distributed Optimization.*
2011-37 Adriana Burlutiu (RUN), *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference.*
2011-36 Erik van der Spek (UU), *Experiments in serious game design: a cognitive approach.*
2011-35 Maaike Harbers (UU), *Explaining Agent Behavior in Virtual Training.*
2011-34 Paolo Turrini (UU), *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations.*
2011-33 Tom van der Weide (UU), *Arguing to Motivate Decisions.*
2011-32 Nees-Jan van Eck (EUR), *Methodological Advances in Bibliometric Mapping of Science.*
2011-31 Ludo Waltman (EUR), *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.*
2011-30 Egon van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions.*
2011-29 Faisal Kamiran (TUE), *Discrimination-aware Classification.*
2011-28 Rianne Kaptein(UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure.*
2011-27 Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns.*
2011-26 Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots.*
2011-25 Syed Waqar ul Qounain Jaffry (VU)), *Analysis and Validation of Models for Trust Dynamics.*
2011-24 Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.*
2011-23 Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media.*
2011-22 Junte Zhang (UVA), *System Evaluation of Archival Description and Access.*
2011-21 Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems.*
2011-20 Qing Gu (VU), *Guiding service-oriented software engineering - A view-based approach.*
2011-19 Ellen Rusman (OU), *The Mind ' s Eye on Personal Profiles.*
2011-18 Mark Ponsen (UM), *Strategic Decision-Making in complex games.*
2011-17 Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness.*
2011-16 Maarten Schadd (UM), *Selective Search in Games of Different Complexity.*
2011-15 Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evidence for Information Retrieval.*
2011-14 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets.*
2011-13 Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling.*
2011-12 Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining.*
2011-11 Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective.*
2011-10 Bart Bogaert (UvT), *Cloud Content Contention.*
2011-09 Tim de Jong (OU), *Contextualised Mobile Media for Learning.*
2011-08 Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues.*
2011-07 Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction.*
2011-06 Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage.*

2011-05 Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.*.
2011-04 Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference.*
2011-03 Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems.*
2011-02 Nick Tinnemeier(UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language.*
2011-01 Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models.*
2010-53 Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access.*
2010-52 Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention.*
2010-51 Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources.*
2010-50 Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives.*
2010-49 Jahn-Takeshi Saito (UM), *Solving difficult game positions.*
2010-48 Withdrawn, .
2010-47 Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples.*
2010-46 Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment.*
2010-45 Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services.*
2010-44 Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain.*
2010-43 Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies.*
2010-42 Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach.*
2010-41 Guillaume Chaslot (UM), *Monte-Carlo Tree Search.*
2010-40 Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web.*
2010-39 Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents.*
2010-38 Dirk Fahland (TUE), *From Scenarios to components.*
2010-37 Niels Lohmann (TUE), *Correctness of services and their composition.*
2010-36 Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification.*
2010-35 Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval.*
2010-34 Teduh Dirgahayu (UT), *Interaction Design in Service Compositions.*
2010-33 Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval.*
2010-32 Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems.*
2010-31 Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web.*
2010-30 Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval.*
2010-29 Stratos Idreos(CWI), *Database Cracking: Towards Auto-tuning Database Kernels.*
2010-28 Arne Koopman (UU), *Characteristic Relational Patterns.*
2010-27 Marten Voulon (UL), *Automatisch contracteren.*
2010-26 Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines.*
2010-25 Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective.*
2010-24 Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies.*
2010-23 Bas Steunebrink (UU), *The Logical Structure of Emotions.*
2010-22 Michiel Hildebrand (CWI), *End-user Support for Access to Heterogeneous Linked Data.*
2010-21 Harold van Heerde (UT), *Privacy-aware data management by means of data degradation.*
2010-20 Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative.*
2010-19 Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems.*
2010-18 Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation.*
2010-17 Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications.*
2010-16 Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice.*
2010-15 Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models.*
2010-14 Sander van Splunter (VU), *Automated Web Service Reconfiguration.*
2010-13 Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques.*
2010-12 Susan van den Braak (UU), *Sensemaking software for crime analysis.*
2010-11 Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning.*
2010-10 Rebecca Ong (UL), *Mobile Communication and Protection of Children.*
2010-09 Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging.*
2010-08 Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments.*
2010-07 Wim Fikkert (UT), *Gesture interaction at a Distance.*
2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI.*
2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems.*
2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments.*
2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents.*
2010-02 Ingo Wassink (UT), *Work flows in Life Science.*
2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter.*
2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion.*
2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful.*
2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations.*
2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients.*
2009-42 Toine Bogers (UvT), *Recommender Systems for Social Bookmarking.*
2009-41 Igor Berezhnyy (UvT), *Digital Analysis of Paintings.*
2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language.*
2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets.*
2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a*

metadata ecology for learning resources in a multilingual context.
2009-37 Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks.*
2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks.*
2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling.*
2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach.*
2009-33 Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?.*
2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors.*
2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text.*
2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage.*
2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications.*
2009-28 Sander Evers (UT), *Sensor Data Management with Probabilistic Models.*
2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web.*
2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services.*
2009-25 Alex van Ballegooij (CWI), *"RAM: Array Database Management through Relational Mapping".*
2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations.*
2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment.*
2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence.*
2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification.*
2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making.*
2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets.*
2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System.*
2009-17 Laurens van der Maaten (UvT), *Feature Extraction from Visual Data.*
2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess.*
2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense.*
2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA).*
2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems.*
2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin), *Operating Guidelines for Services.*
2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law & the Semantic Web.*
2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications.*
2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems.*
2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments.*
2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion.*
2009-06 Muhammad Subianto (UU), *Understanding Classification.*
2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality.*
2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering.*
2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT.*
2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques.*
2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models.*