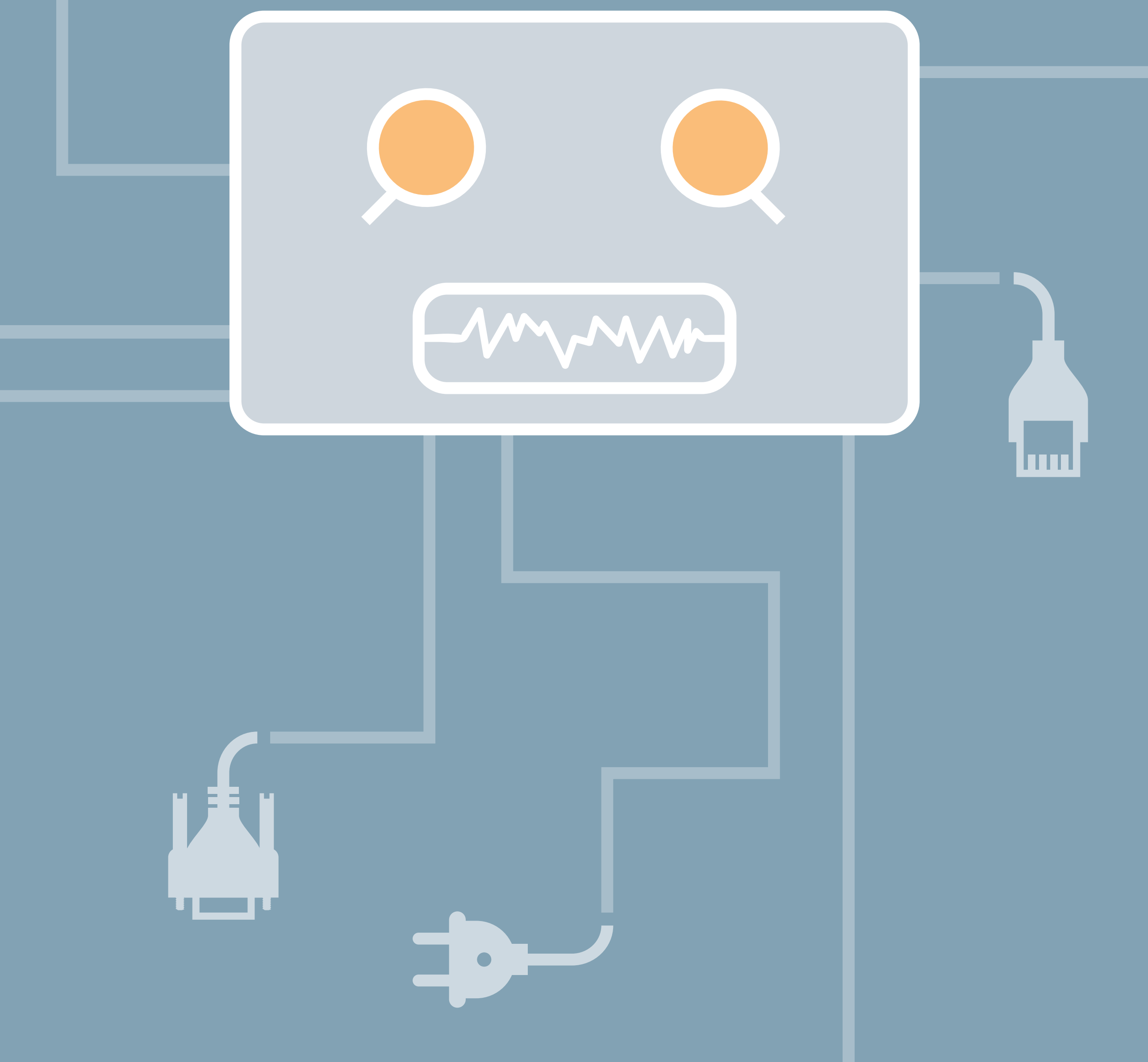
Personal and Personalized Conversations

Designing Agents Who Want To Connect With You



Jelte van Waterschoot

PERSONALIZED AND PERSONAL CONVERSATIONS

DESIGNING AGENTS WHO WANT TO CONNECT WITH YOU

Jelte Barachia van Waterschoot

Graduation Committee	**Affiliation**
Chairman and Secretary:	
Prof. dr. Joost Kok	University of Twente
Supervisor:	
Prof. dr. Dirk Heylen	University of Twente
Co-supervisor:	
Dr. Mariët Theune	University of Twente
Committee Members:	
Dr. Anita Cremers	Utrecht University of Applied Sciences / TNO
Prof. dr. Lisette van Gemert	University of Twente
Prof. dr. Franciska de Jong	University of Twente / Utrecht University
Prof. dr. Catherine Pelachaud	ISIR at Sorbonne Université
Prof. dr. Carolien Rieffe	University of Twente / Leiden University

Paranymphs

Dr. Merel Jung
Deniece Nazareth, MSc.

The research reported in this dissertation was carried out at the Human Media Interaction group of the University of Twente.

DSI Ph.D. Thesis Series ISSN: 2589-7721, No. 21-012, Digital Society Institute P.O. Box 217, 7500 AE Enschede, The Netherlands

This work was supported by the European project H2020 ARIA-VALUSPA, grant agreement No. *645378*.

This work was part of the research programme Data2Person with project number *628.011.029*, which is (partly) financed by the Dutch Research Council (NWO).

SIKS Dissertation Series No. 2021-27. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design: Jelte van Waterschoot / Illustrations: Youetta Kunneman
ISBN: 978-90-365-5269-1
DOI: 10.3990/1.9789036552691

©2021 Jelte Barachia van Waterschoot, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author.

PERSONALIZED AND PERSONAL CONVERSATIONS

DESIGNING AGENTS WHO WANT TO CONNECT WITH YOU

Dissertation

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the graduation committee
to be publicly defended
on Thursday 2 December 2021 at 14:45

by

Jelte Barachia van Waterschoot

born on April 25th, 1991

in Hontenisse, the Netherlands

This dissertation has been approved by:

Supervisor:

Prof. dr. D.K.J. Heylen University of Twente

Co-supervisor:

Dr. M. Theune University of Twente

Summary

Social conversational agents are useful tools for handling customer service requests or for social engagement like chit-chat or playing a game. The development of conversational agents has seen a rise in the last decade. For example, companies include chatbots on their website to lend support to visitors and virtual assistants are part of smart speakers in many homes. One large limitation in current conversational agents is their inability to develop long-term rapport and engagement with end-users. This thesis focused on adaptation and long-term real world engagement as steps towards creating more personalized social conversational agents. The work is oriented towards dialogue designers and everyone who is involved with design of conversational agents: programmers, researchers, linguists, user experience experts and so on.

We provided an overview of different ways of adaptation through multimodal interaction as well as an overview of design frameworks for prototyping and developing multimodal conversational agents. We compared different state-of-the-art topic-based models for personalization, with a focus on topic management in conversational agents. After considering multiple design frameworks and the needs of dialogue designers for a design framework, we found a lack of design patterns and guidelines for dialogue designers, specifically for multimodal design. We developed our dialogue engine, *Flipper*, which we integrated into a virtual human platform for creating multimodal social conversational agents. We included design patterns for dialogue designers and some examples of how Flipper integrates with other components such as multimodal sensors, existing natural language processing pipelines and virtual humans.

We developed three prototypes with our framework: i) the multimodal virtual agent Alice , ii) the BLISS conversational agent and iii) the CoffeeBot. The Alice agent is a software toolkit which other dialogue designers can use for building a social conversational agent. The BLISS conversational agent, named after its research project, is a prototype using speech containing scripted dialogue and was used for data collection of answers to the agent's questions about mental well-being and happiness. The CoffeeBot is a prototype social robot designed for long-term real world interactions with a focus on asking personalized questions in

spontaneous interactions near coffee machines.

The data collection with the BLISS agent was our first step to collect real world data about personal user topics. An interesting finding of the data collection was that there is no immediate need for a complex dialogue system. Despite the relatively high word-error rate of speech recognition, rigid dialogue structure and disfluency of speech synthesis of the agent, at least one topic related to their well-being and happiness could be extracted for each user. To increase more language variability and add a more loose dialogue structure, we developed the CoffeeBot. Its purpose was to have spontaneous speech-based interactions, casual conversation, at the workplace. We based the CoffeeBot's dialogue structure on a model of casual conversation. We combined this with asking questions, specifically starter or opening questions, follow-up questions and questions based on past conversations. We took a template-based approach with syntactic and semantic parsers to recognize user topics and generate the questions to be asked by the CoffeeBot. These questions became more tailored to the user over time. The CoffeeBot learned a personalized user model to have more engaging conversations with people.

We prepared an evaluation for a long-term real world study with the CoffeeBot, which we piloted for five weeks. Our evaluation was focused on two things: i) measuring the impact of personalization on the engagement and ii) the general user experience. We compared different methods and combined questionnaires as well as interviews and interaction metadata to measure the effect of the personalization model. The CoffeeBot's model is yet to be evaluated to see if this type of personalized question asking increases engagement with social conversational agents. This is due to the limited data from the pilot and insufficient time for a full long-term real world study. Despite the study's limitations, we did see usable user models in the CoffeeBot, similar to the data collection with the BLISS agent. Also, from both the BLISS agent and the CoffeeBot's studies we learned that users occasionally needed more time to think about answers. Moreover, distinguishing between an answer to a question and other responses, such as requesting more time to think or a user repeating the agent's question, is still a challenge for a conversational agent. Recognizing and responding to these types of user responses remain an open research problem in speech-based systems. Finally, most of the interactions were engaging for users despite the mistakes the conversational agents made. For long-term use, we expect a drop in engagement if mistakes become a nuisance to the user, however we would argue that an agent making a few mistakes here and there can still provide useful and enjoyable conversations for end-users.

Samenvatting

Met de opkomst van virtuele assistenten en chatbots in klantenservices zijn gesproken dialoogsystemen al geïntegreerd in ons leven. Echter, de manier waarop systemen met mensen communiceren is niet erg vloeiend behalve bij taken als een vlucht boeken, een temperatuur instellen of een spelletje spelen. Het is bijvoorbeeld moeilijk voor een dialoogsysteem om te detecteren of iemand droevig of overstuur is en om daar empathisch op te reageren in een gesprek. In dit proefschrift kijken we naar hoe we gesprekken met dialoogsystemen meer adaptief en persoonlijk kunnen maken en hoe deze systemen in de praktijk voor langere tijd interessant kunnen blijven voor de eindgebruiker om mee te interacteren. Dit werk is bedoeld voor ontwerpers van dialoogsystemen waaronder programmeurs, onderzoekers, taalwetenschappers en UX-experts.

We hebben onderzocht hoe multimodale interactie kan bijdragen aan adaptiviteit, zoals het aanpassen aan emoties van de gebruiker. Daarnaast hebben we een overzicht gemaakt van mogelijke platformen voor het ontwerpen van prototypes en dialoogsystemen. Uiteindelijk vonden we een grote beperking in de verschillende platformen die momenteel worden aangeboden, namelijk het gebrek aan "design patterns" en ontwerprichtlijnen voor dialoogontwerpers. We besloten daarop een eigen dialoogsysteem te maken, *Flipper*, dat fungeerde als de kern van een "virtual human" platform. Bij dit dialoogsysteem creëerden we ontwerprichtlijnen inclusief voorbeelden van hoe Flipper gecombineerd kan worden met componenten voor een dialoogsysteem, zoals multimodale sensoren, natuurlijke taalverwerking en virtual humans.

In totaal hebben we drie prototypes ontwikkeld met Flipper: i) een multimodale "agent" Alice, ii) het BLISS dialoogsysteem en iii) de CoffeeBot. Alice is een virtual human die onderdeel uitmaakt van een softwarepakket voor dialoogontwerpers van multimodale sociale dialoogsystemen. Het BLISS systeem, vernoemd naar het gelijknamige project, is een Nederlands gesproken dialoogsysteem dat vragen stelt over het geluk en welbevinden van mensen voor dataverzameling. De CoffeeBot is een sociale robot ontworpen voor betere langetermijninteractie en het ontwikkelen van een persoonlijke band met mensen. Dit doet de CoffeeBot door steeds persoonlijkere vragen te stellen aan gebruikers in spontane interacties.

De dataverzameling met de BLISS agent was onze eerste stap om erachter te komen hoe mensen praten over persoonlijke onderwerpen. Opvallend was dat er niet per se een ingewikkelde dialoog voor nodig is om het gesprek leuk te houden voor deelnemers en te leren over hun interesses en welbevinden. Daarnaast was het mogelijk voor ons systeem om interessante informatie over de gebruiker te leren, ondanks een hoge foutmarge van de automatische spraakherkenner, rigide dialoogstructuur en de soms moeilijke verstaanbaarheid van de spraaksynthese. Om het gesprek dynamischer te maken en meer diversiteit in het taalgebruik te stoppen ontwierpen we de CoffeeBot. De CoffeeBot kan spontane gesprekken voeren met gebruikers die een kopje koffie halen bij de koffieautomaat, waarbij hij informeert naar hoe hun dag ging en informatie uit vorige gesprekken haalt om de gesprekken te personaliseren. Hij kan drie typen vragen stellen: openingsvragen, vervolgvragen en vragen gebaseerd op vorige interacties. Uit de antwoorden op de vragen worden interesses gehaald die in een gebruikersmodel worden gestopt.

We hebben een proefstudie uitgevoerd met de CoffeeBot gedurende een aantal weken om te kijken hoe hij zich staande zou houden in de echte wereld. Voor deze proefstudie waren we geïnteresseerd in twee dingen: i) het effect van de personalisatiestrategie van vragen stellen en ii) de algemene gebruikerservaring. We gebruikten een vragenlijst, interviews en de metadata om het effect te meten op de gebruikerservaring en de relatie tussen de CoffeeBot en deelnemers. De proefstudie had slechts een gelimiteerd aantal deelnemers en er was onvoldoende tijd om een vervolgstudie uit te voeren. Desondanks hebben we in de data van de proefstudie persoonlijke interesses gevonden in de gebruikersmodellen, net als bij het BLISS dialoogsysteem. In zowel de data van het BLISS systeem als de CoffeeBot zagen we dat deelnemers regelmatig meer tijd nodig hadden om vragen te beantwoorden en dat een reactie van de gebruiker op een vraag niet altijd makkelijk te classificeren is. De reactie was lang niet altijd een antwoord op de vraag, maar kon ook een herhaling van de vraag zijn of een vraag aan het systeem om iets te herhalen. Het kunnen herkennen van antwoorden blijft een uitdaging, zeker voor spraakgebaseerde systemen. Tot slot hebben we de indruk dat de meeste interacties met de prototypes interessant waren voor de deelnemers, ondanks de fouten die het systeem maakte. Daarom denken we dat langetermijninteractie met een dergelijk systeem veelbelovend is in de toekomst, zelfs als een systeem niet perfect werkt.

Acknowledgments

First off I would like to thank you for reading these acknowledgments. If you are reading this, you probably have supported me in one way or another to this final product, and for that I'd like to say thank you.

Human Media Interaction has been like a second home to me ever since I started my Ph.D. there. I love the atmosphere, the collaboration with each other and the interdisciplinary aspect of the group. On the social side, there were the game nights, the "uitjes", conferences, dinners, "borrels", hackathons, brainstorms and much more. I learned so much from my colleagues, with both solicited and (sometimes necessary) unsolicited advice. I would definitely recommend future Ph.D. students to pursue their Ph.D. in this group when provided the opportunity.

Thanks to all colleagues at HMI and especially thanks to my paranymphs Merel and Deniece for being there. Thanks to my daily supervisor Mariët for her unconditional support and enthusiasm for my research. I know I have tried your patience many times, but here it is, a dissertation. And also thanks to Dirk, who heads an amazing research group and always has new interests in stuff and things that exceed my expectations. To Mariët, to Dirk and all other supervisors, please continue to be supportive of your Ph.D. students and make them feel appreciated from time to time. It means a lot to be heard and appreciated, speaking from my own experience.

During my Ph.D. I learned how much I like organizational business, which varies from participation at the university's faculty council, to hosting the Young Researchers Roundtable (YRRSDS) and social events at HMI and the student sport association the Stretchers. I have learned much from my peers about organizing and organizations.

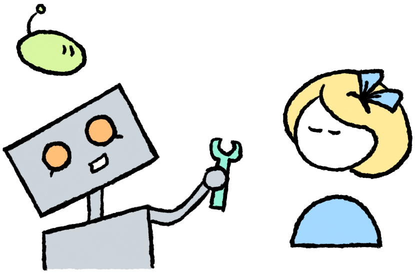
Finally, I am grateful for the support from my family, specifically for my sister during my first conference presentation and my parents for helping me retain my sanity through the writing phase of the dissertation. Though I hope this dissertation is proof to my grandpa I'm no longer a student after a long period of time, I highly doubt he will change his mind about that.

For some readers, I have left an additional personal acknowledgment at the back of the bookmark in the dissertation.

Contents

Part I

Introduction and Background



1 Introduction

1.1 Conversations: People and Agents

Imagine yourself having a conversation with a friend of yours. This conversation flows naturally most of the time. Even though people break off sentences, hesitate, only hear half sentences, this usually does not break the conversation flow. During these conversations people use different types of input, or *modalities*. People often do not only use their voice and ears as a modality to communicate with each other, but observe each other's behavior and maintain a certain proximity to each other. How a specific person interacts also depends on their relationship with another person, be it a family member, a long-term friend or a colleague. The context of the interaction, such as the social status of and knowledge about the other person all affect the interaction. People that take part in a conversation are formally called *interlocutors*.

modalities

interlocutor

An intelligent program that can interact with people similarly to how people have conversations is called a *social conversational agent*. For such an agent, to converse similarly to people is very challenging. We see artificial agents deployed in the real world around us: online as a help desk chatbot or a robot at a bank that can tell people about their appointments or give them directions. These agents are often limited in their capabilities; some can only understand a few tasks very well, others can mimic human listening behavior, but many agents struggle with understanding conversations in a broader context. Let us illustrate with an example of imagining you, the reader, having a conversation with people at a bar:

social conversational agent

> You are having some drinks at the bar with friends of yours. A friend and you will walk up to the bartender, waiting your turn for ordering, all the while talking about a deadline coming up for your work. When

1

> it is your turn, you start ordering, but the bartender gets called away for a second. When he returns, you finish the order and, get the drinks and return to your friends' table.

In this example, each of the separate interactions are doable for social agents, such as ordering the drinks or detecting that it lost the attention of the bartender. Quickly shifting conversation topics (from ordering to work to ordering) is not something many agents can do as well as humans do. Using contextual knowledge in conversation is also something people are better at than agents. Despite an agent having virtually unlimited memory and will not forget, people also know when to use certain contextual knowledge and how to use it. A person can often empathize with a friend and choose to either give encouraging ("if there's one person to tackle this challenge, you are it!"), challenging ("are you sure you will make it?") or comforting ("you will be fine") statements, depending on the state and behavior of their friend talking about their deadline. Maybe they know the friend is going through a rough time at home or got scolded at work. Moreover, the context of a bar plays a role in this interaction as well. The bar is a noisy environment where it is hard to understand full sentences and distinguish who is talking with whom, which is hard for an agent. However, people can adapt very well to these noisy situations, for example combining half-heard sentences with contextual knowledge and non-verbal cues such as lip-reading and facial expressions.

Mitigating noisy situations stems from people's abilities to use multiple modalities and not only depend on verbal language. Similarly, agents should use multiple modalities for generating and understanding behavior while conversing with people. Studies have been carried out to prove the effectiveness of agents' non-verbal behavior generation for social interactions, such as keeping interpersonal distance (Kroczek et al., 2020), generating facial expressions (Cassell et al., 1994; Calix et al., 2010) and gestures (Ravenet et al., 2018; Kucherenko et al., 2020) or speaking with the correct intonation or tone (Cassell et al., 1994; Ritschel et al., 2019; Hoegen et al., 2019). Additionally, work has been done on recognizing non-verbal user behavior for social interactions, such as social touch (Cang et al., 2015; Jung et al., 2015) or reading affect from the face (Ekman et al., 2002). Not every one of these modalities however is as useful every time. In the bar example, visual cues such as gestures and facial expressions are much more likely to be effective to support verbal language, compared to social distance, which is physically limited already in a crowded bar. As a designer of a conversational agent, there are thus multiple factors to take into account when designing the agent's behavior. We define a *dialogue designer* as someone who works on implementing a dialogue system, writing content and/or user evaluation. An all-purpose agent architecture does not exist (yet), so designing with specific domains, context and tasks usually leads to a better user experience with an agent. In the bar setting, the task of the agent is not entirely clear, but some examples are to order drinks or to support social

dialogue designer

conversation with friends. It is up to the designer of a social conversational agent to take the domain, context and goals into account and choose the most efficient modalities.

When creating a social agent as a designer, in general there is the distinction between task-oriented agents and non-task-oriented agents, or chatbots (Jurafsky and Martin, 2009). Examples of a task-oriented agent are a chatbot to book a flight or a bartending robot. Interactions with these types of agents always require the user or agent to have a measurable goal or task to achieve. The prime example of a non-task-oriented agent is ELIZA (Weizenbaum, 1966), a rule-based psychotherapeutic chatbot. ELIZA has no specific measurable goal, but is designed to have a long and engaging conversation with a user. For this thesis we have carried out research with both a task-oriented agent and a non-task-oriented agent for dyadic conversations. In both cases we specifically looked at the personalization aspect for which we designed agents that can adapt (non)-verbally to the user.

1.2 Challenges and Research Questions

There are three main challenges that are tackled in this thesis, i) designing and prototyping a multimodal agent, ii) personalizing a conversation and iii) measures for personalization in long-term real world evaluation. The main objective of this thesis is to combine dialogue design, personalization and long-term real world application and provide a guide to help researchers deploy their own social conversational agents in the wild for longer periods of time.

- How can we design a social conversational agent capable of personalizing interactions with users in the real world?

1.2.1 Designing and Prototyping Social Conversational Agents

One of the first challenges in designing a social conversational agent is to know which tools to use. There is no universal way to create an agent and there are many tools available that let designers create an agent. However, in this thesis we shed some light on the latest developments of designing social conversational agents, what is possible with current technology and provide guidelines for effective agent design. We look at domains in which there is the need for prototyping quickly and flexibly. The first research question of this thesis is therefore:

- Research Question One (RQ1): How can dialogue designers effectively and iteratively prototype a social conversational agent?

1.2.2 Personalizing Conversations with Social Conversational Agents

We believe that having an agent that adapts to what users would like to talk about helps to personalize conversations. A better understanding of the conversation topics and the specific context around users supports personalizing conversations; it will help to having better, engaging and more meaningful conversations. Personalization techniques include tuning to specific topics of the other interlocutor's interest in a conversation. In human-human interaction, asking relevant questions is a form of showing interest in another person and helps to deepen the conversation about topics that are interesting for one or both interlocutors (Huang et al., 2017). The second research question zooms in on this specific aspect of personalization for an agent:

- Research Question Two (RQ2): How can dialogue designers personalize the interaction between a user and a social conversational agent?

1.2.3 Evaluation of Social Agents in the Long-term in the Real World

An important aspect of designing personalized social conversational agents is to prove their usefulness in the real world in the long-term. Unfortunately, once researchers are done prototyping, not much is known about how it would fare in a real world setting (Breuing and Wachsmuth, 2013; Mattar and Wachsmuth, 2014; Foster et al., 2019). Recently the HRI community has opened up its doors to emphasize the importance of real world, or *in the wild* studies (Rosenthal-von der Pütten et al., 2016; Mead et al., 2018). However, moving from controlled experiments and evaluations of social conversational agents to deployment in the real world warrants good preparation for running these agents autonomously in unpredictable and noisy environments. The following research question addresses evaluation measurements that are applicable for long-term real world deployment of agents and we conducted a pilot study with a social conversational agent with these measurements.

- Research Question Three (RQ3): How can dialogue designers measure the effect of personalization on engagement in long-term real-world interactions with a social conversational agent?

1.3 Main Contributions

We highlight the main contributions of the research in this thesis here, which are three-fold.

1.3.1 Authoring Multimodal Dialogues

The first contribution is a set of design guidelines and a tool to help domain experts and researchers to design dialogues. We discuss different tools to design dialogues, each with their own strengths and weaknesses. Underlying frameworks ranged from state machines to end-to-end machine learning. Most tools support developing text-based chatbots, but multimodal dialogue design tools are either scarce or proprietary. With open-source software Flipper we contribute with a *dialogue engine* with which designers can quickly prototype multimodal embodied conversational agents.

dialogue engine

1.3.2 User Modeling via Personalized Questions in Dialogue

One way for a social conversational agent is to personalize the conversation to the user is by getting to know the user via asking questions. We propose building a *user model* based on the topics of the conversation and using these topics to generate personalized questions. We believe that such a user model together with an automated question generator helps to personalize conversations and can be applied in any type of domain.

user model

1.3.3 Deployment of Social Conversational Agents in the Real World

The last contribution of this thesis helps designers of social conversational agents with preparing deployment for the long-term in the real world. We provide a set of *experiment design guidelines* in helping to deploy a conversational agent in the wild and demonstrate deployment with a pilot study.

experiment design guidelines

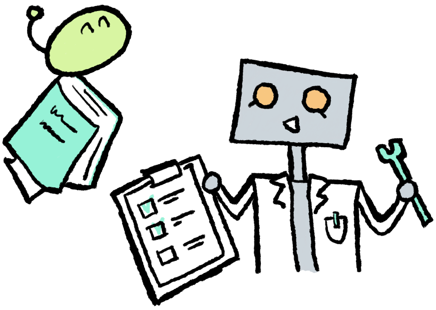
1.4 Outline

In the next chapter we address the literature related to each of the three research questions. We discuss different modalities and memory for social conversational agents, as well as designing and prototyping agents. Additionally, we compare different state-of-the-art models for personalization with respect to topics of interest and question-generation. We conclude the chapter with a discussion about real world human-agent evaluation.

In Chapter 3 we explain the ARIA-VALUSPA framework and how it can help designers in developing multimodal dyadic interactions. In Chapter 4 we discuss different dialogue design frameworks, including our own developed Flipper 2.0, a dialogue tool for prototyping social agents quickly. In Chapter 5 we describe a study to collect speech data in the domain of mental well-being and understand how people respond to questions asked by social conversational agents.

In Chapter 6 we discuss the user model and question generation component of a social agent that learns user's topics of interest for personalization. In Chapter 7 we evaluate the social agent from Chapter 6 in a pilot study in the wild.

In the final part of the thesis, we discuss the results and limitations of the studies and conclude with a take-home message. We combine the lessons learned from dialogue design, personalized question asking and real world deployment for other dialogue designers (Chapter 8).



2 Background

The research purpose of this thesis is finding out if engagement can be increased in dyadic conversations between a person and a social agent through the use of personalized social strategies; deliberate social behavior of the agent. There are many ways to increase engagement, therefore the focus of this thesis is on two pivotal social strategies for social agents within conversations: *adaptivity* and *personalization*. In Section 2.1, we address the use of adaptivity for modalities and memory for social agents in dialogue. In Section 2.2, we give an overview and recommendations for other researchers who wish to design adaptive dialogue for their social agents. Personalization is discussed in detail in Section 2.3, starting with existing user models of personalization in Section 2.3.1. Section 2.3.2 zooms in on a specific type of information for user models, namely dialogue topics and Section 2.3.3 focuses on personalized question generation in dialogue. Evaluation of personalization and real world deployment is discussed in Section 2.4.

2.1 Modality and Memory for Adaptive Social Agents

Perhaps one of the easiest examples to think of when it comes to adaptivity is the way people address each other. Think about meeting a stranger or talking to a friend. Strangers, even more so if they are older than us, are addressed with more politeness than they would be if they were friends or close colleagues. *Adaptivity* in dyadic conversations is the process of interlocutors adjusting their behavior to each other, for example adapting the level of politeness or tact. Adaptivity can be based on rules of thumb or social conventions, such as shaking hands with strangers versus hugging a close friend when meeting (in most Western cultures). Another example at the start of a conversation is that people might make more small talk (e.g., talking about weather) with a stranger than a close friend when they are meeting them. **adaptivity**

Incorporating social conventions such as these in a social agent as adaptive social strategies can increase engagement in dyadic interactions (Mattar and Wachsmuth, 2013). One such case of a strategy is entrainment. Entrainment is a specific case of adaptivity, where a user's interaction styles are directly used to adapt the interaction style of a social agent. For example, Lubold et al. (2020) created a social robot that could exert social behavior and used prosodic entrainment; the robot's pitch was altered in real-time to fit the pitch of the user. Users perceived the robot as being more social present because of this adaptivity. A social agent is dependent on its sensors and embodiment for accurately understanding the social context and interpreting the engagement of the user (Picard, 1998). Both the social agent and user can use different modalities to have dyadic interactions. Modalities include senses and actions of people and social agents. For instance, people can speak or point to things to refer to an object in their surroundings. Another instance is an agent that has a virtual display as a face to convey certain emotions to the user or uses speech synthesis to talk to the user. In the next sections we take a brief look at what user and agent *modalities* could be used in a social conversational agent to increase engagement, starting with more common types like text and speech and ending with less common types like touch. Additionally, we look at how these types of modalities play a part in adapting to the user. Finally we consider different types of personalized memory of an agent as an important aspect for adaptivity.

modalities

2.1.1 Adaptive Modalities in Dialogue

A major benefit that makes face-to-face interaction flow so naturally between people is the use of multiple modalities people have at their disposal with their senses. Verbal language is the most dominant modality people use in day-to-day communication, though non-verbal sounds and facial expressions also have a major impact on interactions (Calvo and D'Mello, 2010). For instance, a sarcastic remark might be found hard to recognize based purely on the textual level. However, a specific facial expression or change in the intonation of the speaker are better predictors of sarcasm. Chauhan et al. (2020) have developed a model for sarcasm detection centered around three modalities, the lexical level, acoustic and visual, based on an annotated dataset of TV-shows. The authors found that combining these three modalities is beneficial for detecting if a sentence was sarcastic or not. A social agent with these modalities could benefit from such a model to be a more engaging conversational partner, by picking up on the sarcasm and possibly use it in return. In this section we will specifically look at the most dominant modalities and what progress has been made with them so far incorporating them in social conversational agents.

Table 2.1: Excerpt of a conversation with ELIZA from Weizenbaum (1966, p. 36-37).

#	Interlocutor	Utterance
1	Agent	I'm sorry to hear you are depressed.
2	User	It's true. I am unhappy
3	Agent	Do you think coming here will help you not to be unhappy?
4	User	I need some help, that much seems certain
5	Agent	What would it mean to you if you got some help?
6	User	Perhaps I could learn to get along with my mother.
7	Agent	Tell me more about your family.
8	User	My mother takes care of me

2.1.1.1 Text

The dominant modality that is being used in social conversational agents is *text*; the lexical units of a conversation. Chatbots are the more common terminology for agents that depend solely on text. *ELIZA* (Weizenbaum, 1966) is a chatbot from the 60s and set the stage for today's chatbots, such as *ALICE* (Wallace, 2009) and Steve Worswick's *Mitsuku* (Radziwill and Benton, 2017). ELIZA was a therapeutic chatbot that used a syntactic parser on text and used pattern-matching to mimic conversational capabilities (see Table 2.1 for an example). Many people enjoyed talking to ELIZA, even though it had no knowledge of therapy, because it was only using syntax from sentences (e.g., subject and object in a sentence) and pattern-matching templates (e.g., matching verbs and nouns in a sentence) for generating responses.

text

A major benefit of using text for adaptiveness in dialogue is the abundance of it found online. Twitter, Reddit and Wikipedia are among the most commonly used public resources for creating agent models based on text, often to build a language model or a knowledge base for an agent (Pak and Paroubek, 2010; Völske et al., 2017; Breuing et al., 2011). Additionally, datasets such as conversations from the first ConvAI challenge (Burtsev et al., 2018; Logacheva et al., 2018) and PERSONA-chat (Zhang et al., 2018) contain conversations between a human and a chatbot and persona descriptions respectively for personalizing dialogue with agents (see Table 2.2). Especially if dialogue designers want to apply machine learning methods to their own social agent, such a vast amount of data is beneficial to train usable agents. Many of the available NLP tools for chatbots are freely

Table 2.2: Conversation from a human (Bob) talking to a chatbot (Alice) in the first Conversational AI (ConvAI) challenge (Logacheva et al., 2018). The context prompt for the chatbot was about cults.

#	Interlocutor	Utterance
1	Alice	Hi
2	Bob	Hello
3	Alice	Do you know anything about Vestals?
4	Bob	Well, I haven't heard anything about them before
5	Bob	It looks like you're a real person.
6	Alice	sure? they are quite popular topic at school history
7	Alice	yes i am.
8	Bob	I don't remember much from school course of history

available to use for dialogue designers, such as spaCy,[1] NLTK (Bird et al., 2009) and Stanford CoreNLP (Manning et al., 2014) for probabilistic natural language understanding (NLU). Currently, transformer models such as BERT have the best state-of-the-art performance on NLU tasks (Devlin et al., 2019). These models perform exceptionally well for tasks such as Named Entity Recognition (NER), topic recognition and intent classification. On the natural language generation (NLG) side, machine learning models including OpenAI's GPT-3 (Brown et al., 2020) have contributed to generating novel and relevant responses for chatbots. A conversation (randomly picked from Adiwardana et al. (2020)'s dataset) with chatbot Meena is shown in Table 2.3. Meena is trained on social media data, which is a typical dataset for current state-of-the-art chatbots.

Another major benefit of text-based systems is the easiness with which dialogue designers can deploy such systems and collect conversational data. Cloud-architectures dominate most of the online technology with commercial services such as Amazon AWS or the possibility to set up a private chatbot server with RASA X (Bocklisch et al., 2017). Cloud-based chatbots are accessible to many users through social media platforms, browsers or smart devices.

One of the limitations of using only text for conversations is difficulty in disambiguation. For instance, a question could be phrased in declarative form or could be sarcastic. Although through the use of emoji characters, layers of extra information can be added to text to disambiguate, it can be culturally dependent how such an emoji is interpreted and is limited in variation of expressions. Additionally, much text data for machine learning is unstructured and messy. Even

[1] `http://spacy.io`

Table 2.3: Excerpt of a conversation with Meena from Adiwardana et al. (2020)'s dataset.

#	Interlocutor	Utterance
1	Agent	Cool, what are you going to do?
2	User	Spending time with friends and family, eating traditional food
3	Agent	Nice, sounds fun!
4	User	Yea, tell me something interesting about you
5	Agent	I guess I'm pretty good at drawing!
6	User	Lovely! What can you draw?
7	Agent	I can draw people, I can do portraits, and I can do animal drawings as well.
8	User	Nice. What about scenery?
9	Agent	Not so well, I'm not that good with trees.

if datasets are as large as the case is with transformer models, the output should be used with extreme caution by dialogue designers. Floridi and Chiriatti (2020) found that GPT-3 is prone to generate unethical texts with racial slurs.

Finally, turn-taking is very rigid in text-based systems, where interlocutors usually take one turn at a time. Messages like "x is typing..." do help in turn-taking and an interlocutor can type before someone is responding to elaborate or correct themselves. However, turn-taking is more tiresome via text, because people have to make extra effort to be clear in their intent and make up for the lack of normal face to face conversation modalities. Additionally, giving feedback that an interlocutor is listening is not really possible, except for a "read" check mark. In addition, asking for a clarification of one interlocutor cannot easily happen mid-turn of the other interlocutor, because they will usually complete their last turn.

2.1.1.2 Speech & Sound

Speech and sound are the second most used modality for social conversational agents. The most common speech components of a social agent are the *automatic speech recognition (ASR)* and *text-to-speech synthesis (TTS)* component. *ASR* transforms the acoustic signal from people's speech to a textual representation. A *TTS* does the exact opposite, and transforms a text into an acoustic signal for the social agent. Components for speech and sound include automatic affect recognition (AAR), such as recognizing valence and arousal in a person's voice or non-verbal cues such as laughter and feedback about uncertainty ("ehm...").

automated speech recognition

text-to-speech

Speech-based systems often use an ASR and TTS, therefore speech-based and text-based conversational agents share a number of key features: they are both language-dependent, depend on language models and use a textual representation. Many models for speech-based conversations utilize models trained on text. These types of systems are often found in virtual assistants, such as Apple's Siri and Amazon's Alexa.

However, speech is different from text in a number of aspects as well. Firstly, speech is more messy in grammatical structure than text, because speech contains hesitations, interruptions and unfinished sentences. Secondly, speech is based on an acoustic signal and has more information available about the user than only the text user says, because it contains paralinguistic features as well, such as pitch and loudness for determining arousal levels.

A major advantage of speech-based systems is the accessibility for people to interact with them. Whereas typing can become less engaging for people when interacting with a chatbot, speaking does not suffer from this problem, because it resembles more closely people's natural way of communicating and requires less cognitive load (Huang et al., 2016). Turn-taking can be much more dynamic in speech-based systems, because the user and agent can use non-verbal signals to interrupt each other, mark a question or request some thinking time.

Another advantage of speech-based systems mentioned before is that the acoustic signal can be used for more than an ASR, for example to analyze acoustic features of speech with the openSMILE[2] toolkit (Eyben et al., 2010) for AAR. This toolkit is particularly useful for social conversational agents, because it can analyze streams of speech in real-time efficiently. openSMILE provides information about a user's pitch or loudness of voice to which a social agent could adapt (Schröder et al., 2009). Other tools for speech analysis include COVAREP (Degottex et al., 2014) and librosa (McFee et al., 2015). Kim et al. (2017) have used speech data from real-world interactions to create AAR based on speech, which can give information about a user whether they are angry, happy, sad or neutral while interacting.

However, there are some limitations to using speech-based agents. The first and foremost problem with these types of systems is the inaccuracy of ASR. The performance is often measured in Word Error Rate (WER), the percentage of the words incorrectly recognized by the ASR. Even though current state-of-the-art ASR has made huge progress in accuracy, the WER remains high in real-world applications, where noisy environments and overlapping speech of interlocutors are not uncommon. On the other side of the spectrum, TTS has come a long way from sounding somewhat robotic to more natural voices. Even so, prosody can still sound off in conversations and TTS components seems underdeveloped for other languages than English, though efforts have been made for multilingual modeling for TTS (De Korte et al., 2020). Also non-verbal elements of conversational

[2]open Speech & Music Interpretation by Large-space Extraction (openSMILE)

speech such as backchanneling and laughter are usually prerecorded for TTS systems. However, some non-verbal sound synthesis methods exist for social agents (Ritschel et al., 2019). Unfortunately, even though speech might feel more natural for users to interact with, a downside of using speech is the potential exposure of private information of users when they want to interact in public with the system. Additionally, in speech-based systems that deal with critical information, such as medicinal information, a high WER in the ASR can lead to misinformation and possible harm to users (Bickmore et al., 2018). Finally, most models that are used for performing NLU for speech are based on text-trained data. As a consequence the performance is worse because speech does not always strictly follow written grammar.

2.1.1.3 Facial Expressions & Gaze

A person's face often reveals important information about how they feel. Especially during an interaction between two people, they observe each other's face to determine for instance if the other person is still following what they are saying. Facial expressions play a large part in current multimodal conversational agents, thanks to research in the world of computer vision and development of good cameras on current smart devices like phones and tablets as well as the development of more powerful processing units and tools for character animation and widely available expressive robots such as the Zeno (Hanson et al., 2008),

An important aspect of recognizing facial expressions are action unit (AU)s. There are 32 AUs in total that people can move with facial muscles that can be captured by camera. The Facial Action Coding System (FACS) is a coding scheme developed to track movement of the AUs. An open-source tool that has been widely used by the multimodal conversational agent community is OpenFace (Baltrusaitis et al., 2018). The tool includes landmark detection, head pose direction and gaze detection. The toolkit can be used for facial identification and for implementing a visual AAR, even beyond dyadic interactions. Almaev and Valstar (2013) developed eMax, which is an AAR tool based on the big six emotions: sadness, disgust, happy, fear, anger and surprise (Ekman, 1999). eMax measures these emotions in real-time from the AUs of the user and allows a social agent to react appropriately during the conversation. For instance, if the user looks surprised after the agent executed some behavior, the agent might ask the user why they are surprised (see Figure 2.1).

In addition to facial expressions, a person's gaze tells something about how they are feeling about the interaction. Somebody might look up to think about something or they look away because their attention has been diverted. Nakano and Ishii (2010) collected data on gaze behavior of people to create a model for a social agent to determine how engaged somebody is in an interaction. They found that the duration of gazes and the transitions of gaze direction are strong indicators

2

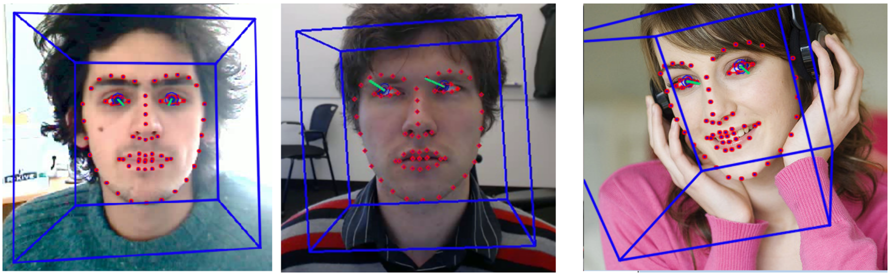


Figure 2.1: Gaze tracking and landmark detection of facial AUs with OpenFace 2.0 (Baltrusaitis et al., 2018)

.

of engagement.

As for the generation part of facial expressions, researchers have conducted studies to see how facial expressions of an agent would impact the interaction (Cassell et al., 1994). A model of automatic affect generation (AAG) for facial expressions based on emotion words in an agent's utterance is created by Calix et al. (2010). As for gaze, Ruhland et al. (2015) put together guidelines for animators for generating eye gaze behavior, as well as head movement for social agents. These tools and guidelines seem to effectively convey emotion to users interacting with agents employing these strategies (Han et al., 2017). Not only humanoid agents can utilize facial expressions, a LED display with "eyes" can mimic facial expressions as well, such as with the Cozmo and R3D3's head (Touretzky and Gardner-McCune, 2018; Theune et al., 2017).

There are limitations to the modalities of facial expressions and gaze as well. Firstly, there is the need for a camera, which might not always be allowed at every location a social agent is deployed. Secondly, Martinez et al. (2019) made an overview of the latest facial expression technologies based on AUs and said that methods work relatively well in the right circumstances, but need quite some work to work well in the wild. For example, most of the research has been on frontal faces, and performance drops when faces are blocked or turned away. Finally, most of the work on facial expressions is on short spontaneous AAR, whereas it would be worthwhile to have longitudinal recognition as well, for instance to detect moods. If during an interaction the agent detects that the user is in a bad mood, it might suggest resuming the interaction at a later time (Lietz et al., 2019).

On the generation side of facial expressions and gaze, technical failures are most common. For example, failing servo-motors in the Zeno, a robot with a very expressive face (Hanson et al., 2008). Al Moubayed et al. (2012) therefore developed a combination of a physical robot head with an animated head to have a hybrid of a virtual and physically embodied agent: the Furhat. Another important aspect to take into account are cultural differences when using facial expressions or gaze, for example looking in the eye of someone or looking down could be interpreted as challenging someone or disrespecting them.

2.1.1.4 Body Expressions

Body expressions have had less recognition in research than facial expressions and gaze (Karg et al., 2013; Kleinsmith and Bianchi-Berthouze, 2013). Karg et al. (2013) specifically looked into AAR and AAG systems using body expressions in the Human-Robot Interaction (HRI) and Human-Computer Interaction (HCI) community. They discuss that most systems measure arousal and valence or discrete emotional states for single users, but adaptiveness in an interactive setup has yet to be tested in a large-scale evaluation. Kleinsmith and Bianchi-Berthouze (2013) conducted a survey of gesture and body movement recognition and found that among the most common body expressions are posture, gait and gestures.

Unlike the other modalities mentioned thus far, posture is extremely useful to determine the dominance of interlocutors, which can be seen for example in the AMI Meeting corpus, a dataset that contains video data of work meetings (Carletta et al., 2005). Bruijnes et al. (2015) conducted a study about the effect of body posture on the perception of interpersonal stance. Interpersonal stance is an affective style of people during an interaction that can be deliberately changed. It is often expressed through body posture and modeled in terms of affective dimensions such as valence, arousal and dominance. In Bruijnes et al. (2015)'s scenario of police interrogations, a virtual social agent acted as a suspect character and adapted its behavior based on the stance and behavior of police trainees.

Noroozi et al. (2018) conducted a recent survey and highlighted the current models of AAR via affective body expressions. They found that the main problem in the field is to go beyond discrete emotional labels and valence/arousal values to more meaningful (complex) semantic representations, which include combined emotions and affective states like uncertainty, shame and tenderness. Furthermore, the positioning of a camera impacts the robustness of AAR in body-postures, similarly to its impact on facial expression detection, though the scale is less fine-grained for the former.

2.1.1.5 Touch and Haptics

Social mediated touch conveys affective information through haptic devices. Touch can be seen as a strong indicator for building social relationships between two people, or a person and a robot. The way people touch a stranger (a handshake) or a friend (a hug) will be different and shows for example the trust they have in the other person. Van Erp and Toet (2015) argue that for *social presence* of agents, it

social presence

Figure 2.2: Example of the TaSST, a haptic sleeve for social mediated touch (Huisman et al., 2013).

2

is a necessity to include social touch. Social presence is the feeling of experience something together and increases engagement of an interaction.

The use of social touch in interactive applications has become more accessible and affordable (Huisman, 2017). In their study, a virtual agent was created that could touch a user by slapping a bug with a tactile sleeve on the user's arm (Huisman et al., 2014; Huisman et al., 2013). The authors emphasize the importance of having visual congruency with touch as well to strengthen the effect. Additionally, specifically for the purpose of social interaction, Jung et al. (2015) and Cang et al. (2015) collected a corpus of social touches such as stroke, slap and grab. These datasets were used to train touch-based AAR models and had accuracy rates up to 60/70 %.

Unfortunately touch is in its infancy to be applied in the design of social conversational agents. Tools for recognizing touch, such as pressure sensors, are more intrusive than cameras and microphones and are less common to be found in accessible consumer hardware. Sleeves like the TaSST help in generating touch sensation, but are still relatively intrusive, though with time, technology can become more ubiquitous. Additionally, automatically recognizing touch is far from perfect in a multimodal setup (Jung et al., 2017). With time hopefully more data and insights in social touch and its features will become available to help build reliable AAR models for social touch.

2.1.1.6 Physiology

Physiology methods directly measure users' physiological response. Instances of physiological measures are heart-rate variability (HRV), galvanic skin response (GSR), skin temperature, muscle tension and electroencephalography (EEG) in the field of Brain-Computer Interface (BCI) (Guger et al., 2019). These measures are often considered as the most objective form of measuring user responses to an interactive system, because the user cannot control the responses and the user's cognitive processes cannot interfere with the response (Prendinger et al., 2006).

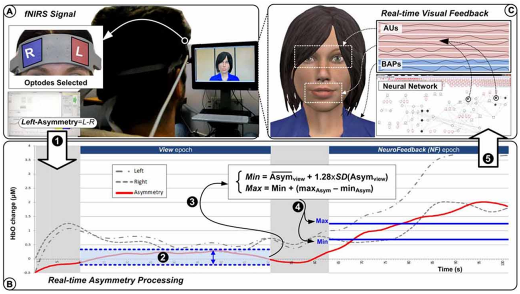


Figure 2.3: Example of a user whose brain signals are measured, which a virtual agent uses to adapt their facial expressions (Aranyi et al., 2016).

For a study on AAR through electrocardiography (ECG) patterns and HRV, researchers found about 50% accuracy of recognizing valence and arousal levels based on these physiological features alone (Ferdinando et al., 2016).

Kolkmeier et al. (2017b) collected GSR and HRV using a multimodal setup with a virtual social agent. In their scenario, participants were being told by the virtual character, who was the embodiment of the participant's supervisor, that a student had filed a complaint against them. The statement by the agent that a student had negative feedback about the participants had a visible effect on the heart rate of participants (Kolkmeier et al., 2017b). Applying this knowledge to developing an adaptive social agent, a virtual supervisor could take a more strict or supportive stance, depending on the user's physiological response. For example, in case of an elevated heart rate of a participant, the social agent might first need to calm down the participant before continuing the interview.

However, physiological measures often require having devices strapped to a user's body and are considered more invasive than methods mentioned in the previous paragraphs. Additionally, individual differences such as physiological build and compatibility with measuring equipment prevents researchers from finding conclusive evidence in their study, especially in the case of GSR data (Kolkmeier et al., 2017b).

2.1.2 Adaptive Memory

In addition to the modalities mentioned thus far we want to discuss the importance of memory for social agents in this part. It also serves as input for a social agent, in some ways similar to how modalities help the agent in the interaction. We consider *memory* in a social agent as a component that stores interaction information to be **memory**
used in a later interaction. Examples of types of information in the memory could

2

be sensory information from the modalities mentioned before or a list of topics and emotions that the agent wants to talk about.

Some existing social conversational agents have been designed with memory in mind. Most work in the machine-learning world has dealt with types of memory, for instance a Long-Short Term Memory (LSTM), a neural network model that is based on the idea of human short and long-term memory. It serves its purpose mostly as the context of an interaction. However, this type of memory is implicit and not easily accessible nor interpretable for other components in a social agent. Though this work is fascinating and beneficial for the social agent community, the memory component we are looking for should be more human-like and be usable for the agent as an independent component.

For task-oriented dialogue, memory can be represented as past slot types and values for semantic frames (El Asri et al., 2017). These values and types can be used for more quickly helping a user with a task. For example, if a person were to book a flight from Amsterdam to New York, they will give a social booking agent information about location and time to the agent. If the next week the same person books a flight from Amsterdam to Moscow, it would be inconvenient if the social booking agent forgot about their previous departure airport. Past information can be used as input for possible slot types and values in frames, such that an agent can be less repetitive and more quickly help the user.

From research in neurology, Tulving (1972) made a distinction between two types of memory for people: semantic and episodic. Semantic memory is about facts, such as knowing where someone's favorite restaurant is or telling the time. Episodic is memory specifically about events and experiences between a social agent and anyone else. Elvir et al. (2017) state that in addition to having semantic and episodic memory, there is procedural memory, which contains information about how to perform certain tasks. *Conversational memory* captures all three types of memory. "[...] conversational memory in [memory] discussion refers to the representation, storage, and retrieval of information and/or knowledge acquired during a multiparty oral conversation." (Elvir et al., 2017, p. 2). The authors also state that the "gist" of the conversation is important for remembering. The gist is a term for all topics relevant now or in the future during the conversation.

conversational memory

Mattar and Wachsmuth (2012) introduced the concept of Person Memory to encapsulate semantic and episodic memory in a social agent. They propose that this type of memory should contain information that helps to build a relationship between a user and social agent. Person Memory contains information about the user's biographical facts, preferences and interests, personality traits, events and experiences shared with the agent and the relationship (familiarity) with the agent. The memory has a limited number of predefined frames with key-value pairs to capture user information (see Table 2.4). Mattar and Wachsmuth (2014) evaluated a virtual agent (*Max*, Kopp et al., 2005) with Person Memory. *Max*' memory component consisted of the user model and contained information about the person themselves (music interest, hobbies) and the social categories the person belonged to (student, type of sports club) (Mattar and Wachsmuth, 2013). The authors conducted two experiments, a "getting to know" conversation and a follow-up conversation. In the follow-up conversation, the agent recalled information from the previous conversation and filled this information in the placeholders of its behavior rules. Initial results did indicate people liked *Max* more when he remembered information, though also people in the control condition thought *Max* remembered facts about them, even though this was not the case. However, it remains unclear how they obtained the information in the first conversation and how the users were encouraged to disclose information.

A similar type of memory structure was used by Kim et al. (2014), which was a personal knowledge database (PKB) to remember user information using triples in a predicate structure. Their social agent collected triples from user input sentences based on predicate-templates. A dependency parser extracted relevant verbs (e.g., like, love) and their dependent arguments, the objects and subjects from utterances. For example, "I know you like blue bananas" is represented as the triple (I, like, blue bananas). The memory component consisted of these triples and a forget-factor. The agent would forget about topics not mentioned recently. Unfortunately, the evaluation of the memory was not done with a live social agent, but only with the Movie-Dic dataset (Banchs, 2012). Though an agent would be capable of remembering everything it is told, a forgetting component could lead to a more believable interaction. Richards and Bransky (2014) found that it is better for an agent to (partially) forget information than it is to recall incorrect information.

Campos and Paiva (2010) created a shared-memory component for their social agent. Campos and Paiva's agent MAY was capable of building a relationship with a person through listening to the person's stories. These stories could be classified as life-time stories, general events or very specific event-related information. They found that remembering this type of information increased the feeling of intimacy and companionship with MAY. In a subsequent longitudinal study, Campos et al. (2018) found that using memory too frequently for leading the conversation could annoy users and make the conversation more repetitive, because the agent talks about the same topic often. Contrary to results found in their earlier study (Campos and Paiva, 2010), they did not find conclusive evidence that a memory-based agent performed better than a non-memory based agent. However, they also mention many confounding factors that could have impacted the interaction between user and agent, such as the number of turns and the user's ability to adapt to the agent as well as the implicit use of memory.

2

Table 2.4: Example of a person frame in the memory from Mattar and Wachsmuth (2012)'s Person Memory model. Slots represent a specific field and the value is filled if it has been discussed with the user. Confidence represents how sure the agent is of the information and the modifier represents a value that indicates how good of a conversation topic a slot is.

Slot	Value	Confidence	Modifier
firstname	Paul	.7	-1
gender	male	.9	-4
interest	computer games	.4	5
occupation	student	.8	3
hometown			2
...	...	...	...

2.2 Dialogue Design and Prototyping

Since Weizenbaum (1966)'s ELIZA, much progress has been made in the world of spoken dialogue systems for both research and companies. Especially in the last decade, large amounts of data together with advances in machine learning have brought usable dialogue systems much closer to the public. ASR and TTS have come long ways to reduce WER and improve naturalness. Though most of the models for tools like spaCy,[3] CoreNLP (Manning et al., 2014), OpenAI's GPT-3 (Brown et al., 2020) and BERT (Devlin et al., 2019) are based on written text, some models have been created for multimodal dialogue systems (Rahman et al., 2020). Furthermore, efforts have been made to make multimodality more accessible to dialogues designers as well, for example with Microsoft's Platform for Situated Intelligence (psi) (Bohus et al., 2017) and the social signal interpretation framework (SSI) (Wagner et al., 2011), with which dialogue designers can add new or modify existing multimodal pipelines to their social agent, specifically focused on synchronizing and fusing different modalities for the interaction (see Figure 2.4).

Unfortunately, many dialogue systems are not reusable by other people or for other purposes, as often a dialogue system does not have support for a specific component that other researchers need, the software is unavailable, outdated, poorly documented or takes too much effort from dialogue designers to master. It is not uncommon that people develop their own in-house solution from scratch to build a dialogue system that suits their specific needs. More often than not,

[3] `spacy.io`

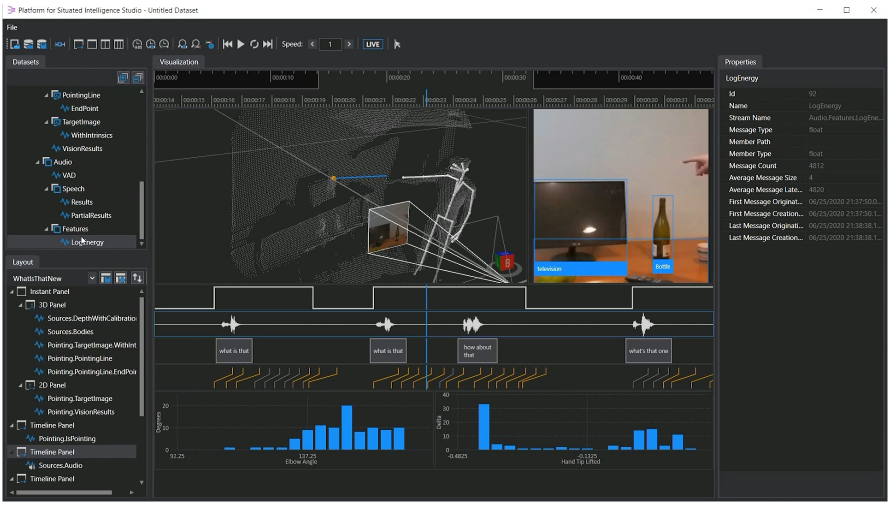


Figure 2.4: A screenshot of the psi toolkit, in which you can review audio and video data and help improve components such as AAR (Bohus et al., 2017).

time of researchers is spent on coding and reinventing the wheel rather than on actual research. Another issue for dialogue designers is that there is no universal way of creating social agents. Compared to relatively well-specified components such as ASR and TTS, *dialogue management (DM)* is more abstract and domain-dependent. DM is the core of a social agent, which drives the behavior of an agent based on its dialogue policies and input modalities. Dialogue management often relies on domain-specific knowledge of dialogue designers.

dialogue management

In this section we shed some light on efforts to prevent reinventing the wheel by summarizing the development of open-source and user-friendly software. We start in section 2.2.1 with an overview of dialogue frameworks to help dialogue designers writing policies for their social agents. In section 2.2.2 we focus on support for multimodal embodiment that dialogue designers can use.

2.2.1 Dialogue Frameworks

The early types of dialogue management systems were finite state-, pattern- or frame-based, restricting dialogues to the flow defined by the dialogue designer. An example of pattern-based rules for chatbots is Artificial Intelligence Markup Language (AIML), developed for Wallace (2009)'s chatbot A.L.I.C.E and still used for small-project chatbots, because of its low learning curve. The TrindiKit was developed as a toolkit for developing dialogue systems to overcome the strict rules of

```xml
<agent id="RedSoxIntroDialogue" text="Really? But they aren't...">
 <user text="I bet you are a Yankees fan.">
  <agent text="No, I'm just joking with you.">
   <user text="Oh."/>
   <user text="That's too bad, it would be more fun if you were!">
    <agent text="Ok, from now on I'm a Yankees fan.">
     <user text="Great!"/>
    </agent></user></agent></user>
 <user text="Ah, but who cares? They play great ball!"/>
</agent>
```


Listing 2.1: Example of XML written for DISCO about a sub-dialogue (Rich and Sidner, 2012).

information state these types of dialogue systems and introduced the concept of an *information state* approach (Traum and Larsson, 2003). Within an information state approach, the dialogue designer makes rules and updates for changes in the dialogue, which are much less rigid than a finite-state machine. Later, DIPPER (Bos et al., 2003) was released as a modified version of the TrindiKit. In this toolkit there is no difference between dialogue policy rules and update rules, to simplify the designing process of dialogues for social agents (Bos et al., 2003). To extend the scope of the information-state based dialogue toolkits to include multimodal understanding and generation for social agents, Flipper (Ter Maat and Heylen, 2011) was developed within the SEMAINE project (Schröder et al., 2009). Flipper made it easier for dialogue designers to write update rules for generating affective behaviors for embodied agents and acting on recognized emotions. NADIA is a framework built to be more accessible for dialogue designers than other toolkits by separating domain independent and dependent logic (Berg, 2015). Independent components are turn-taking and greeting behaviors, and examples of dependent components are answering domain-specific user questions. The creation process of the behaviors is similar to how templates are authored in Flipper (Ter Maat and Heylen, 2011).

Another tool is DISCO (Rich and Sidner, 2012), which has also been used in the work of Glas and Pelachaud (2015b). Many authoring tools are meant for designing dialogue trees where each part of the tree has the same granularity, whereas DISCO abstracted the higher goals and flow of the dialogue from each other with specific subdialogues (see Listing 2.1). It still has a tree-like structure, but with more flexibility to switch between different subdialogues with their own topics.

One of the earliest tools to perform incremental dialogue management with was RavenClaw (Bohus and Rudnicky, 2009) which is included in the Olympus framework for development of conversational agents (Bohus et al., 2007). *Incre-*

mental dialogue management is dialogue management that processes dialogue before a turn of an interlocutor is complete, such that predictions and relevant information can be retrieved to generate agent behavior more quickly (Schlangen and Skantze, 2009). Incremental dialogue management, or incremental interaction, is a necessity when dealing with uncertainty for social agents (Allen et al., 2001; Skantze, 2007). For example, interlocutors talking to the agent can make mistakes or get distracted or system errors occur, such as word errors for ASR and unidentified named entities for an NLU component. Due to these issues during dialogue, interlocutors and the social agent might need to repair or assert their understanding of each other, a process called grounding (Clark, 1996). RETICO is a framework with a user interface to connect different components for modalities that includes incremental processing (Michael, 2020). The documentation of the tool is still minimal at the time of writing this thesis, but promising work has already been done by Kennington et al. (2020), showing that this incremental framework works with different embodiments, such as with the NAO robot.

incremental

An issue with multimodal dialogue tools is their lack of usability for end-users, the dialogue designers. Dialogue tools that have a good and intuitive user interface are more likely to be reused. ADvISER is a dialogue tool that focuses on helping people with less affinity with development such as linguists and cognitive scientists to design dialogues. The tool features multimodal processing pipelines, such as handlers for the latest ASR technology and OpenFace for recognizing faces (Li et al., 2020). However, the tool does not have a user-friendly interface and still requires some coding for setting up and customizing to a domain. The Virtual Human Toolkit (VHToolkit) was developed by the USC-ICT group exactly for this purpose (Hartholt et al., 2013). It has features such as an ASR, virtual agent and a question-answer (QA) component, the NPCEditor (Leuski and Traum, 2010). The NPCEditor has been developed to make it easier to create a QA component for any domain-dependent virtual agent with a graphical user interface. After a dialogue designer enters their domain-specific question-answer pairs, the system is able to automatically train a new language and dialogue management model for further evaluation. FLoReS was implemented as a dialogue manager within the VHToolkit, to make it easier for dialogue designers to create goals for their agents, support policy rewards and reason on local dialogue structure, to support more incremental processing (Morbini et al., 2014). However, FLoReS was later dropped in favor of a more simple dialogue management toolkit, because the rewards were hard to upscale (Razavi et al., 2017). Another matured toolkit is Visual SceneMaker, developed by the German research institute DFKI (Gebhard et al., 2012). It features a graphical user interface, where dialogue designers can create a flow-chart with a drag & drop style (see Figure 2.5). Designers can draw the dialogue tree and flow directly as a state-based interaction. The toolkit contains a 3D virtual agent and components for incremental processing, such as the possibility

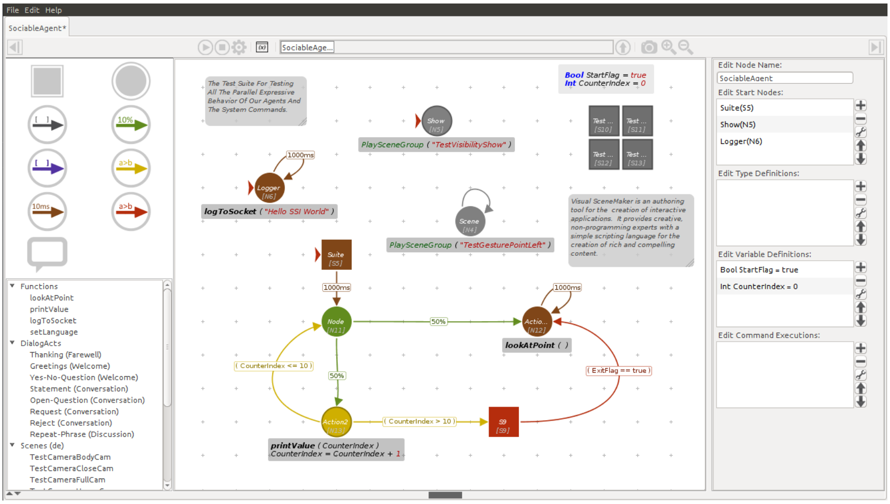
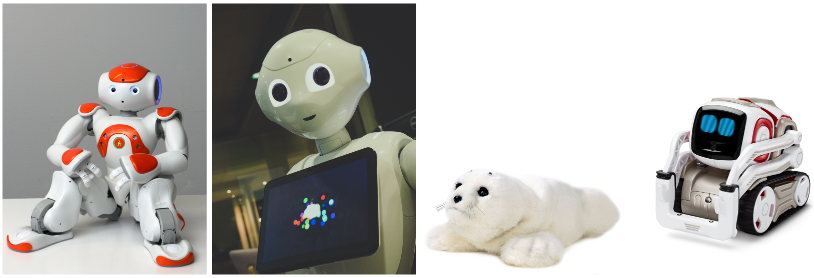


Figure 2.5: The interface of Visual SceneMaker in which dialogue designers can create a flowchart (sceneflows) of the interaction (Gebhard et al., 2012). This interaction can also be simulated before evaluating with end-users.

for users to interrupt the agent while talking.

Nowadays, with robots becoming more affordable and many people having access to mobile devices, scalable deployment has become more feasible. High scalability is when an agent can be used by many users concurrently and in many types of environments. Letting people use a social agent in an already familiar environment also increases the odds that people will accept the social agent over time (De Graaf et al., 2016). Unfortunately, the tools mentioned thus far focus on offline dyadic interactions and do not scale well. Many chatbot-frameworks do scale well, but scalable multimodal interactions have not been a major focus of spoken dialogue research. Researchers would like to evaluate with as many users as possible at the same time without providing all users with a separate social agent. Each of these systems can only interact with one user at a time. The developers of the Social Interaction Cloud (SIC)[4] are trying to tackle the issue of scalability, having a single system server-side deployed that can run multiple interactions at the same time with different types of embodiment, be it a robot or a web browser and use multimodal input, though it is still in its infancy of development. If multimodality is less important for dialogue designers and scalability is the most important aspect, systems such as Siri, LUIS.ai, Alexa are viable options. These systems are questionable from a privacy point of view though, because designers

[4]https://socialrobotics.atlassian.net/wiki/spaces/CBSR/



(a) NAO **(b)** Pepper **(c)** Paro **(d)** Cozmo

Figure 2.6: From left to right you can see four different types of robots. a) NAO, b) Pepper c) Paro and d) Cozmo. Whereas the first two utilize humanoid properties such as speech and gestures, the latter two primarily use non-verbal language, such as lights, sounds and touch.

have no (full) control of the data streams. Alternatively, designers can set up their private servers with open-source software such as RASA X (Bocklisch et al., 2017) and PyDial (Ultes et al., 2017), which are well documented options for scalable chatbot solutions. RASA has support for connecting to speech components and PyDial allows for more control on dialogue policy than any of the other chat frameworks.

2.2.2 Conversational Agent Embodiments

Next to deciding the approach to dialogue management, the designer needs an appropriate embodiment for a social conversational agent. A social agent is considered *embodied* if it has body parts it can use, such as limbs, but also lights or sounds. **embodied** Dialogue designers can choose depending on the domain and other context that a social agent may appear as a robot or exist inside a virtual assistant on someone's phone. If dialogue designers want robots to act like humans, a human embodiment seems like the best choice. However, choosing the appropriate embodiment depends on the goal dialogue designers are trying to achieve. An embodiment increases the social presence of a social agent and should be considered if that is an important goal.

When looking at robots in research, especially in the HRI community, we observe that often robots are picked that have humanoid properties, such as NAO and Pepper (see Figure 2.6a and 2.6b) (Gouaillier et al., 2009; Pandey and Gelin, 2018). These robots can use similar modalities to humans, such as movement, gaze and speech. However, having similar modalities as people can cause multiple issues for the interaction design with such a robot. A robot with more components has also more components that can break down and the more complex the robot is, the more expensive it is to deploy. Moreover, the robot might set too high expectations from users given its appearance. The decision for an embodiment depends on the desired interaction and expertise of dialogue designers, therefore a robot with fewer modalities could be the better option. Take for instance the Paro robot (Figure 2.6c). This robotic fluffy seal can make sounds and responds to touch, but has no verbal capabilities. Even with its limited capabilities, it serves as an excellent companion as a pet-like social agent, reducing stress for users (Aminuddin et al., 2016). Additionally, Anki's Cozmo robot (Figure 2.6d) can use its LED display, sound and movement to express itself non-verbally (Touretzky and Gardner-McCune, 2018; Pires Kusumota et al., 2018), and comes with a toolkit for dialogue designers to program its behaviors and it has been deployed to teach children programming for example. The Furhat robot has been designed to mix the best of two worlds: a virtual face using projection and the physical embodiment of a head and neck as a robot (Al Moubayed et al., 2012). The robotic head has more social presence than a virtual agent, but the flexibility of switching faces and versatility of facial expressions of a virtual agent using projection, compared to a physical robot, not to mention less chance of breaking servo-motors. The Furhat also comes with standard interaction behavior, such as timings and clarification requests. It consists of an incremental dialogue management component and has been successfully deployed for academic purposes (Campos and Paiva, 2010; Kennedy et al., 2017).

The choice for an embodiment also depends on the types of behaviors these embodiments are capable of. Some platforms support forms of standard behaviors, in different modalities. For speech-based systems, and many TTS systems, Speech Synthesis Markup Language (SSML) has been the de facto standard to use (Taylor and Isard, 1997; Shuang and Burnett, 2010). With SSML, dialogue designers can change pitch, speed and language with this XML-based standard for any SSML supported voice. However, not all SSML supported platforms fully comply with it and SSML does not directly support higher level automatic affect generation (AAG), such as displaying sadness or enthusiasm.

Behavior Markup Language (BML) has been developed as a standard for embodied conversational agent (ECA)s. In addition to adjusting features of speech, with BML dialogue designers can synchronize multimodal behaviors for an ECA such as gestures, gaze and body postures (Kopp et al., 2006; Vilhjálmsson et al., 2007). To bridge the gap between determining the agent intent in the DM and the generation of agent behavior, Functional Markup Language (FML) has been in development (Heylen et al., 2008). With FML, designers can semi-automatically generate BML based on certain parameters, such as a specific intent, utterance or emotion the agent has to express. Designers then do not need to bother with the specifics of the embodiment too much. Partially this generation of behaviors can be automated with the Behavior Expression Animation Toolkit (BEAT) toolkit,

Figure 2.7: Example of a BML behavior with a BEAT gesture in GRETA (Poggi et al., 2005).

available to make beat gestures for ECAs (Cassell et al., 2004), also with an implementation for BML.[5] These types of gestures are usually made when talking, but have no specific semantics. Other gestures like pointing (deictic) and emphasizing can be modeled by the dialogue designers themselves. Some virtual agents already support BML or FML, such as SmartBody and Greta (Thiebaux et al., 2008; Poggi et al., 2005; Mancini and Pelachaud, 2008). These types of agents can be more easily be integrated in a component for types of dialogue managers that support BML. Once the BML has been generated, it needs to be transformed to the specific movement commands of the ECA, such as the movement of bones of a virtual character or the servo motors of a NAO. However, most robots have their own proprietary format for designing agent behaviors and virtual agents have a specific authoring language. Tools such as the Articulated Social Agents Platform (ASAP), bridge the gap between propriety formats and BML. With ASAP, BML can be transformed into commands for commonly supported ECAs, such as the NAO, but also SmartBody, Unity3D and Greta (Van Welbergen et al., 2014).

Wu et al. (2018) worked on machine learning approaches for automating behavior generation with NaDiA (not to be confused with Berg (2015)'s NADIA). NaDiA uses a convolutional neural network (CNN) to mimic the facial expressions of the user for direct animation and is trained on an affective language model to generate utterances and behaviors. These utterances and behaviors are transformed into BML for animating them in SmartBody. Kucherenko et al. (2020) designed Gesticulator, a framework that can generate gestures directly from speech that can be used for animation of virtual characters.

The limitations of embodiment will remain that despite some standards being

[5] `https://github.com/eirikur-ari/openbeat`

available to dialogue designers, there will be an individual need to tune to specific use cases and adhering to social norms (a hand wave of a robot can mean different things in different countries). Designers will have to manually write behaviors themselves for many of the embodiments, though separation of general interaction behavior, such as the Furhat does, alleviates some workload for designers.

2.3 Personalization in Dialogue

Through combining modalities and memory, dialogue designers can create social agents that are capable of developing long-term relationships with people, to *personalize* the conversation. In this part of the background chapter we explore existing user models that can help personalize conversations (Section 2.3.1). In Section 2.3.2 we specifically look at how we can track topics of conversation to help with personalization. Finally, in Section 2.3.3 we look at question asking methods that use user models and topics for personalization.

2.3.1 Personalized User Models

personalization Fan and Poole (2006, p. 183) state that *personalization* is "a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its personal relevance to an individual or a category of individuals". We focus on how a user model can be used to achieve this change process in conversational agents. In conversational agents the intent of either interlocutor can be unknown to each other. However, through engaging in cooperative dialogue these intents will be known to both in a user model. Therefore a conversational agent should build an explicit user model of a user's beliefs, goals and plans for an agent (Wahlster and Kobsa, 1989). Another important reason is that regardless of cooperation, a user model is necessary for a system to understand what the user is talking about and analyzing dialogue contributions and the effects of the system on the user. **user model** Wahlster and Kobsa, p.3 state that "a *user model* is a knowledge source in a natural-language dialogue system which contains explicit assumptions on all aspects of the user that may be relevant to the dialogue behavior of the system. These assumptions must be separable by the system from the rest of the system's knowledge." We believe that a personalized user model requires a user model that should not only contain assumptions about facts, but also opinions and sentiments of the user that can change over time. Kobsa (1989) has built a taxonomy of user model design in dialogue systems, focusing on the knowledge, the goals and beliefs of a user. Important aspects of such a model are based on assumptions about the user's goals and beliefs, the evidence presented to support those assumptions and inferences made based on these assumptions. Additionally, to elicit self-disclosure for personalized topics, the social agents need to build rapport with the user and maintain engagement (Bickmore et al., 2010). The goals of both

Table 2.5: Excerpt of a conversation between Alexa and a user, in which the user discloses information about their day and movie preference (Ravichander and Black, 2019, p. 252).

#	Interlocutor	Utterance
1	Alexa	Good to hear! Is it anything special today?
2	User	Nothing much, I am just going down to the lake with my dogs today.
3	Alexa	Have you seen the movie Arrival?
4	User	Oh my God, yes I have, **it's my favorite movie**.

the user and agent are important aspects in most of these user model based dialogue systems. However, the goals are less clear in open-domain and casual chat, where "getting to know the user" and "being engaging" are goals, but there is no final goal completion. It is hard to say when the system knows the user "well enough". Furthermore, obtaining a user model can benefit any type of dialogue system with personalization, either task- or non-task-oriented. One way to help populate a user model consisting of users' interests and beliefs is by eliciting self-disclosure from the user.

2.3.1.1 Self-disclosure

Amazon has set up an annual Alexa challenge since 2016, where they invite researchers and other interested audiences to create an engaging social conversational agent. In particular, researchers looked into the amount of *self-disclosure* during such conversations: "In a human-machine context, we define self-disclosure as the conversational act of revealing aspects of oneself voluntarily, which would otherwise not be possible to be known by the dialog system." (Ravichander and Black, 2019, p. 255). An example of the conversation is shown in Table 2.5, which has voluntarily self-disclosure marked in bold. Alexa did not explicitly ask about the user's favorite movie, but learned this information because of the voluntary self-disclosure. The results of Ravichander and Black's study show that there is some reciprocity in self-disclosure by participants who are talking to Alexa. If agents self-disclose much, then people usually do too. However, sometimes the self-disclosure by the agent might not conform to the user's expectations. For example, Alexa's embodiment does not allow her to go out and it would be deceptive to have Alexa disclose about going out. Additionally, Ravichander and Black found that self-disclosure does not necessarily make the agent more likable. Dialogue designers can thus implement a self-disclosure component for an agent to increase

self-disclosure

self-disclosure of the user. However, they should be aware that self-disclosure alone is not enough to create a likeable or engaging agent.

Self-disclosure of personal information also raises some privacy concerns. People who disclose personal information to commercial assistants such as Amazon's Alexa and dialogue designers need to be aware of sensitive information they might be disclosing (Wahlster and Kobsa, 1989). Every individual user is different and might be inclined to self-disclose about some topics more than others (Marmion et al., 2019; Rapp and Cena, 2016). Knijnenburg et al. (2013) also found that self-disclosure is not one-dimensional. People who are concerned with their privacy do not just share all information or no information. One of the differences found between participants in a study with a social agent was that some people were willing to disclose their interests, but did not provide location information, though also the exact opposite occurred. In a social setting not every user might be willing to disclose personal information about their interests, which can make it harder to build a user-model.

Sugiyama et al. (2014b) found that asking questions is not sufficient for an agent to be engaging, and therefore built a self-disclosure component for a chatbot. Sugiyama et al. (2014a) collected a large corpus of personal questions from both online resources and online chat to analyze the different categories, calling it the *Person DataBase*. With the self-disclosure included in their agent, participants found the agent more enjoyable to talk with than without the self-disclosure. However, the generation component of their chatbot still was only capable of answering self-disclosure questions and generating topical utterances, not asking personal questions. Additionally, the chatbot was not evaluated with its generation component in real chat, only with automated metrics (Higashinaka et al., 2014).

Radlinski et al. (2019) conducted a study to evaluate the quality of the responses of virtual assistants while users self-disclosed information. The authors found that users often have their own linguistic style that does not necessarily map well to the NLU and NLG components of these virtual assistants. The authors argue that the assistants need to take user topics and linguistic preferences into account and let users self-disclose with their own style instead of priming users with topics programmed in the assistant. Assistants are programmed to capture structured data of movies, such as actors, directors and budgets. However, these systems do not have access to unstructured information such as themes or opinions about a specific scene, which some users do mention while self-disclosing.

2.3.1.2 Rapport Building & Engagement

Rapport building is an important aspect of building personalized long-term relationships with social agents (Gratch et al., 2007; Bickmore and Cassell, 1999).

rapport *Rapport* is necessary for having a mutual understanding, a good flow of conversation and cooperative behavior for interlocutors and usually builds over time after learning how a person communicates. Tickle-degnen and Rosenthal (1990, p. 285) define rapport building as "terms of dynamic structure of interrelating components that have affective and behavioral implications. The structure changes over the course of the development of a relationship between individuals". Rapport is about how interlocutors feel and behave during an interaction over time.

Matsuyama et al. (2016) worked on a virtual personal assistant, *SARA*, *Socially-Aware Robot Assistant*. The agent was focused on building rapport with interlocutors through information-seeking strategies. *SARA* was deployed at a conference, where it would recommend conference attendees which other attendees to talk to. *SARA* matched users based on similar interests it was told by them. The back-end of *SARA* used a dialogue tree for NLU processing of user utterances and this served as input for a social reasoner, which chooses the best dialogue strategy for the agent, while keeping rapport. Contingency in behavior, responding verbally and non-verbally, helps to build rapport for multimodal conversational agents (Gratch et al., 2007). Gratch et al. compared contingency and non-contingency in non-verbal behavior of agents while users are talking. In the contingency condition, the agent reacted actively to what the user was saying with feedback behavior and in the non-contingency condition the agent gave feedback at random intervals. Interestingly, in the non-contingency condition, users self-reported higher rapport and showed more indicators of rapport, compared to users reports and indicators in the contingent condition.

A good rapport between a social agent and user usually results in high engagement as well. Once a social agent has rapport with another interlocutor, the conversation progresses "automatically" and engagement is high. *Engagement* in the context of interaction with a social agent is defined as "the value that a participant in an interaction attributes to the goal of being together with the other participants and of continuing the interaction" (Poggi, 2007). To measure engagement, Poggi (2007) looked at the connection and cooperation between the interlocutors and how the interlocutors are related (e.g., work, friends). Engagement also depends on situational factors, such as whether the user is engaged with another interlocutor or perhaps the user is preoccupied with a current task. Glas and Pelachaud (2015a) state that the level of rapport itself can also be seen as a form of engagement over a longer period of time.

engagement

Another method to increase engagement of the user is to let the social agent self-disclose as well (see Section 2.3.1.1). The use of storytelling by the agent increases engagement when talking to a social agent (Bickmore and Cassell, 1999). A receptionist robot, *Valerie*, was equipped with storytelling behavior to have conversations in the long-term with people (Gockley et al., 2005). People came back to talk to Valerie and learn about her story during several weeks, though people regularly did not listen to the entire story and skimmed quickly through the conversation. A possible cause for this was Valerie's monotonic voice. People

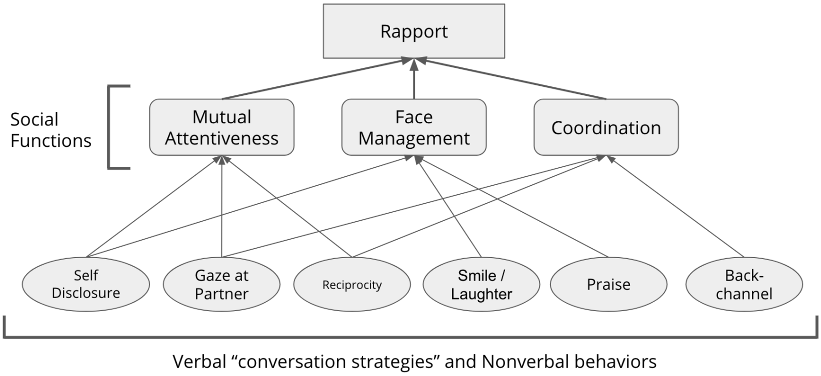


Figure 2.8: A computational model of building rapport, with different social strategies that can increase or decrease rapport (Zhao et al., 2014).

also thought the conversation was very one-sided, with only the agent telling their story and not much opportunity for themselves as speaker.

A computational model for building rapport in dyadic conversations is shown in Figure 2.8 (Zhao et al., 2014). The authors created a model that has parameters that are related to the rapport between interlocutors. This model contains behaviors that are very much related to self-disclosure, such as self-disclosing intimate personal details or talking about a shared experience. These self-disclose behaviors could boost rapport, though some behaviors can damage rapport, because they might violate social norms. People do not tell private information to strangers. It would be a bad strategy for an agent to tell something private to a person it just met.

Glas and Pelachaud (2015b) state that it is important for the agent to not only know what to say, but also when to say it and how to say it to keep engagement high. The DM in Glas and Pelachaud (2015b)'s social agent had a topic management module that could use real-time engagement of the user through AAR to learn about user's preferred topics to talk about. Glas and Pelachaud conducted an experiment in a museum with four works of art, in which an agent had to talk with users about these works. The agent could incorporate no preferences, its own preferences, the user's preferences or both for topic management (Glas and Pelachaud, 2018). However, there was no difference in engagement between the different topic management conditions. This does not necessarily mean that topic management does not increase engagement, because the domain of topics was limited to four works of art. Also, the authors state that resuming in a later conversation about the same topic could possibly increase engagement.

2.3.2 Topics in Dialogue

To personalize dialogues in open domain conversation, a conversational agent needs to have a basic understanding of semantics, of the *topics* in the conversation. Once topics in a conversation are recognized, these can be added to user models that can serve multiple purposes, such as persuasion in consumer perception (Zboja et al., 2016), power in debates (Prabhakaran et al., 2014) and dominance in multiparty dialogues (Nguyen et al., 2014).

topic

Topics can be defined as the "aboutness" of the dialogue (Brown and Yule, 1983). A topic is usually described as a word or group of words that categorize other groups of words or texts that are semantically similar. According to Řehůřek and Sojka (2010, p. 46): "Topical modeling is that texts in natural languages can be expressed in terms of a limited number of underlying concepts (topics)". The main purpose of topic modeling is removing the noise from the core of the discourse and extract the relevant topics, or the gist of the conversation (Razavi et al., 2017). Gundel (1985) made a distinction between two different types of (discourse) topics, pragmatic and syntactic. According to her, for a pragmatic topic: "An entity, E, is the pragmatic topic of a sentence, S, iff S is intended to increase the addressee's knowledge about, request information about or otherwise get the addressee to act with respect to E." For a syntactic topic, she states: "A constituent C is the syntactic topic of some sentence S, iff C is immediately dominated by S and C is adjoined to the left or right of some sentence S' which is also immediately dominated by S." A syntactic topic can be directly retrieved from the dialogue with keywords, whereas the pragmatic topic encapsulates the topic of a discourse in a higher-level abstract concept. For example, in Table 2.6, the syntactic topic of turn # 1 is Flight KL 550, but the pragmatic topic of the conversation could be "customer service for planes". Gundel (1985)'s view on topics corresponds with Rats (1996)'s view on topic-comment structure, or theme-rheme structure. The part of an utterance that is the most related to the current discourse is the topic (theme), whereas all other information in the utterance is considered as the comment (rheme) (see Table 2.6). A comment can become the topic in a next turn if either interlocutor wants to talk about it. Clark (1996, p. 342) described a topic as a joint project as it is jointly established during ongoing conversations, and is therefore more dynamic than the suggested definition of topic as a discourse topic by Gundel (1985). According to Clark, a topic can only be the topic if all interlocutors pick up on the topic as well in consecutive turns and not only by the interlocutor who brought up the topic. Svennevig (2000) agrees with Clark (1996) and states that every spoken contribution may raise new potential topics whose actual realization depends on the co-participant's acceptance by picking up one of these topics within his or her reply.

Topics can be divided into two levels according to Rats (1996): meta and object-oriented. A meta-topic is a reference to one of the interlocutors, usually pronouns

Table 2.6: Conversation between a customer service agent and a customer at the Amsterdam Schiphol airport. Adapted from Rats (1995, p. 53), indicating the **topic** of the flight in bold, whereas the rest of the utterance is the comment.

#	Interlocutor	Utterance
1	Customer	**Flight KL 550**, for what time is **it** scheduled?
2	Agent	**It** is now definitely expected at five to twelve.
3	User	Five to twelve?
4	Agent	Yes

like "I" or "you" and an object-oriented topic is about the task or discussion itself. When building user models, it is important to link the correct object-oriented topics to the correct interlocutor. Sometimes utterances of interlocutors do not contain topics at all, such as "Okay", which can be just a backchannel utterance.

topic shift Topics can also shift during the conversation. A consensus is that a *topic shift* occurs when the current topic cannot be generalized together semantically with the previously mentioned concepts anymore. If the topic shift abandoned the previous topic completely and the current topic did not get closing procedures, this is called a topic leap (Svennevig, 2000) (Table 2.7, turn # 5). Rats (1996) states that the process of *topic management* is how speakers regulate the introductions, continuations and shifts of topics in their conversation. For example, Glas and Pelachaud (2015b)'s implemented topic transition strategies for an agent to make a conversation more engaging, by selecting topics the user prefers (Glas and Pelachaud, 2015b; Glas et al., 2015). Another topic transition strategy was implemented by Macias-Galindo et al. (2012), whose social agent uses semantic relatedness of snippets of conversations to make the most coherent topic switches. The topic strategy features capabilities to combine different snippets into longer utterances for an agent to talk about multiple topics at once. Users perceived the topic transitions as more coherent than the nearest-context topic management by Gandhe and Traum (2007).

topic management

Often in longer social conversations, the feeling of topic drift occurs (Hobbs, 1990). *Topic drift* happens when all segments of a discourse are coherent, but the end of the conversation is about something different from what initially was talked about (see Table 2.8). These conversations usually do not have conscious topic management. Think for example of a conversation between two people in a bar. They are heavily invested in talking about and listening to each other's stories and might wound up talking about something completely different from what they started talking about.

topic drift

Table 2.7: Conversation between two friends with a topic leap at turn 5, which the other interlocutor agrees to talk about, as Laura picks this topic up in turn 6. Adapted from Maynard (1980, p. 270).

#	Interlocutor	Utterance
1	George	There's a discussion and there are written and oral exams frequently. Once in a while at least.
2	Laura	Yeah, I'd like to take history of philosophy.
3	Laura	Or something where you don't have to do any of where you have to think that way. I'm not that logical. I never go step by step.
4	Laura	And just, I'm a really irrational person sometimes so.
5	George	Where do you live?
6	Laura	Yeah, I live in the Tropicana.

Table 2.8: Conversation between two college friends with topic drift. They initially talk about grad school, after which they switch topic to people, without explicit markers of topics shifts. Adapted from Maynard (1980, p. 272-273).

#	Interlocutor	Utterance
1	Alice	I mean, I don't even know if I want to go to grad school.
2	Jane	I'm not going to think about it.
3	Jane	I figure if I really want to go, by the time I get out of school, I'll be able to get in someplace.
4	Alice	Yeah, some place, somehow.
5	Jane	But I don't know, I'm starting to feel that a bad grade can affect your degree. And you know, what affects you more are people.
6	Jane	You have to get your sociology together with the real world, the social world too.

For this thesis we will focus on deliberate topic management for social agents. In the literature we found three dominant approaches to topic modeling and management: discourse structure, machine learning (probabilistic) and knowledge base. We discuss each of these in the following sections.

2

Table 2.9: Dialogue segment from the corpus of Grosz (1974). It consists of a sequence of utterances between an expert (E) and apprentice (A) mechanic.

#	Interlocutor	Utterance
1	E	First you have to remove the flywheel.
2	A	How do I remove the flywheel?
3	E	First, loosen the two Allen head setscrews holding it to the shaft, then pull it off.
4	A	Ok.
5	A	I can only find one screw. Where's the other one?
6	E	On the hub of the flywheel
...	...	...

2.3.2.1 Discourse Structure

Grosz and Sidner proposed a theory of discourse structure in conversations, focusing on the three aspects of discourse: the linguistic structure, the structure of the intents and the state of focus of attention (Grosz and Sidner, 1986). The linguistic structure consists of the utterances and sentences in a discourse from multiple participants. The intentions of the interlocutors with their utterances and a small number of relationships between them provide the basic elements of the intentional structure. The attentional state, or focus, contains information about the objects, properties, relations and discourse intentions that are most salient at any given point during an interaction. The attentional state is an abstraction of the participants' focus of attention as their discourse unfolds. Each focus of attention goes on a stack, which are called focus spaces. The changes in attentional state are modeled by a set of transition rules that specify the conditions for adding and deleting those focus spaces. The collection of focus spaces available at any one time is the focusing structure and the process of manipulating these spaces is focusing (Grosz and Sidner, 1986). Depending on where the focus of attention is, the focus will likely be the next topic of the next utterance in the discourse. For example, in Table 2.9, the structure of intents starts with the expert intending that the apprentice removes the flywheel. In turn 5, the attentional state shifts from this intent to an intent to find two screws, which is necessary to complete the global intent.

The topic-comment structure (Rats, 1996) was used to extract topics from the SEMAINE corpus (Schulman, 2013) by Langlet and Clavel (2016). Langlet and Clavel extracted nouns from user's utterances and the sentiment of likes and dislikes of the user based on WordNet-Affect (Strapparava and Valitutti, 2004) in predicate-structures ("I like Santa Claus" or "It was nice"). Langlet and Clavel (2016) mapped these predicate-structures to predefined topic frames. The topic frames were 7 in total: free-time activities, free-time projects, professional activity, professional projects, generic projects, happiness and anger. Reference resolution was used to replace referents (e.g., it/that) with the current topic of focus, the instantiated topic. The method was found to perform relatively well in extracting at least one topic per conversation, however, there were still topics left undetected because they were outside the domain.

Carletta et al. (1997, p. 14) view dialogue at the highest level as transactions, which are "subdialogues that accomplish one major step in the participants' plan for achieving the task". These transactions consist of conversational games, with the assumption that most questions are responded to with an answer and a statement with acceptance or denial, similar to the discourse structure proposed by Grosz and Sidner (1986). There is a difference between speakers in initiating and responding to a game in Carletta et al. (1997)'s approach. At the lowest level there are the conversational moves. There are different types like instructing, aligning, querying, acknowledging and replying, all of which are responses and related to the current topic in the dialogue game. These types are similar to the dialogue dimensions of the Dynamic Interpretation Theory++ (DIT++) taxonomy presented by Bunt et al. (2010). DIT++ was developed by Bunt et al. (2010) to standardize annotating dialogue acts in conversations. These dialogue acts can be used in dialogue management and are similar to what Carletta et al. (1997) describes as conversational moves. Specific dialogue acts for topics in the DIT++ taxonomy are topic introductions, topic preclosings and topic switches. They are used for annotating the topic structure in discourse, which is part of the discourse structure management in DIT++.

Stede and Schlangen (2004) conducted a study on information-seeking dialogue, where topic structure is used as a dialogue policy (Stede and Schlangen, 2004). Users could ask the chatbot several questions as if it was a city guide. The authors implemented a dialogue manager based on description logic, *The Wanderer*, that used topics and preferences as an agenda for the dialogue policy, and it contained an ontology with city guide information. The topics (people, buildings, parks, etc.) for *The Wanderer* were extracted from the ontology. For each topic, a semantic similarity between one topic and all other topics is computed and inserted in the ontology. If a user asked the agent a question, the agent tried to find the utterance in the ontology that fit the question. Subsequent agent utterances were based on the similarity to the current topic, the query of the user and the history of the dialogue.

Table 2.10: Five discovered topics (T) for an unseen news article using LDA and bag-of-words on the "Million Headlines" corpus after tokenization, stop word removal, stemming and lemmatization, which consists of news headlines (Li, 2018). Topic 2 seems to be oriented to sports, though not every topic can be abstracted to a specific category, for example topic 4. Four out of the five (not topic 3) topics all relate to Australia as well.

T	Terms
1	govern, open, coast, tasmanian, gold, australia, beat, win, ahead, shark
2	world, final, record, break, lose, australian, leagu, test, australia, hill
3	rural, council, fund plan, health, chang, nation, price, servic, say
4	elect, adelaid, perth, take, say, labor, turnbul, vote, royal, time
5	court, face, charg, home, tasmania, murder, trial, accus, abus, child

2.3.2.2 Machine Learning

machine learning Latent Dirichlet Allocation (LDA) is an unsupervised *machine learning* method developed by Blei et al. (2003) to model topics of text documents by means of probability distributions of words across these documents. Rather than having a single word or short phrase as description of a topic, in LDA a topic is made up of the most frequent keywords that mostly co-occur within one topic (see Table 2.10). Other related work is the development of Latent Semantic Analysis (LSA) (or Latent Semantic Indexing (LSI)) to extract meaning and topics from texts (Landauer et al., 2013). A disadvantage of these methods is that they rely on datasets that are heterogeneous, and if they are applied to homogeneous data, it will likely result in topics that are not usable. Additionally, these methods do not seem to classify a topic the same way as people would classify topics and can be hard to abstract from (Chang et al., 2009). Paul (2016) evaluated the interpretability of machine learning approaches to topic modeling. People had to select from a group of topics (consisting of LDA generated topics and human annotated topics) which of the topics was the odd duck, a task known as the word intrusion task. In addition, people evaluated the topics by themselves by ranking the topics from best to worst. Paul (2016) found that most people were able to recognize the LDA-based topics, though they did not always rank the human-based topics better than the LDA topics in the second experiment. To create better interpretable topics, Chaoua et al. (2018) analyzed a dataset of psychotherapeutic conversations for topic detection. Chaoua et al. used labeled LDA for having only interpretable topics that map the LDA topics directly to the labeled human topics of a dataset. Partially Labeled LDA is a hybrid model that uses existing topics, but could also discover novel topics. Chaoua et al. (2018) could extract the most important topics from the conversation, though this method was executed in a fairly limited domain and not evaluated with live conversations with a social agent.

Inaba and Takahashi (2018) created a method for extracting topics from utterances in a dialogue based on topic frequency. The topic frequency was an indicator of the user's interest in a specific topic. Their research was focused on dialogue personalization, where a neural-network based method trained on dialogues measured users' interests. The authors collected a database of topics from crowdsource workers (408 dialogues/49029 utterances), in an experiment where people were instructed to chat with each other, just to get to know one another. Afterwards, the conversations were annotated with topics from a list of 24 common small talk topics (e.g., movies, fashion) in total. Their neural model was able to correctly classify more topics than previous models, though it had problems with utterances that contained very specific instances of a topic, such as "Do you play Pokémon GO now?", which was recognized as sport/health, but not as a game.

2.3.2.3 Knowledge Base

Topic modeling based on machine learning leads to difficult interpretable topics for people and these methods often lack a good descriptive topic name for a group of topic words. Additionally, with a real-time conversation with a social agent, new topics might come up that cannot be classified. The third approach to topic modeling and management is based on the use of external *knowledge bases*, such as Wikidata (Vrandečić and Krötzsch, 2014) or ConceptNet (Speer et al., 2017). The advantage of these knowledge bases is that they cover many topics. Wikidata and ConceptNet are designed specifically for computer systems to handle data more robustly, with a graph-like structure that models the semantic relatedness between concepts through "instance-of" and "has" properties. However, the granularity between the concepts seems arbitrary. Whereas an apple is one step away from fruit, a pick-up truck is an instance of a car, which is an instance of a motorized vehicle, which is an instance of transportation means. However, a unambigious comparison of the semantic similarity between apple and fruit, and pick-up truck and transportation is hard. This is similar to how Wikipedia was used as a tool for classifying topics with WikiBrain by counting the number of minimum page clicks necessary to arrive from one page to another (Sen et al., 2014).

knowledge base

Breuing et al. (2011) looked into using Wikipedia as a semantic knowledge base, based on the work of Waltinger and Mehler (2009). Breuing et al. (2011, p. 393) define a topic in conversation as: "[...] an independent, self-selected category superordinate to a co-constructed sequence of dialog contributions". Topics in their study were the same as a combination of categories of Wikipedia pages (Breuing and Wachsmuth, 2013). During a conversation, the topics are separated into three categories, which they based on Schneider (1988)'s model: immediate, external and communication. Immediate topics are directly related to the current frame of the dialogue situation, to the topic of the current turn, like answering a question. External topics are those that are influenced by the surroundings, such as seeing something drop or running into a colleague at work. Communication topics are of a social nature and are about somebody's family or hobbies and interests. Breuing and Wachsmuth (2012) and Breuing and Wachsmuth (2013) implemented a topic management model connected to Wikipedia as a knowledge base after finding that the topic identification method works on newspapers (Breuing et al., 2011). The algorithm for topic recognition looks at each turn in a dialogue contribution and computes the similarity of the utterance during the turn to each of (proper) nouns and verbs on Wikipedia pages for possible categories. Also the previous turns are input for calculating the similarity to the appropriate category. Thus, the detection process is capable of identifying a topic without having a priori knowledge of the domain underlying it. Unfortunately, the method was never evaluated with a user study, though the authors aimed to implement their model in the virtual agent *Max* (Kopp et al., 2005).

The definition of topic in Topic Detection Tracking (TDT) is a unique real world event or set of news stories strongly related by some seminal real world event, according to Allan et al. (1998) and Allan (2002) The TDT project was dedicated to finding blogs, news articles and interviews related to specific events of the world, which served as the knowledge base for topics. An event was the topic for all texts about that event. In this case, all topics came with a specific timestamp. Even though the TDT project was more about information retrieval than dialogue, the way of talking about topics as events can be useful for spontaneous casual conversation, because people talk about the news often when meeting.

Zhu et al. (2016) made a probabilistic topic switch model for an agent to match topics to user utterances and talk about related topics. The model took three things into account: topic frequency, concurrency and adjacency. Topic frequency is the likelihood that a topic is present in an utterance. The concurrency determines how likely it is a topic occurs in a sentence with respect to the other topics in the sentence. The adjacency is the likelihood that a topic occurs given the previous utterance. The model was trained on an annotated corpus, with a predefined list of topics. Zhu et al. (2016) conducted an experiment and found that their topic switch model was more entertaining for users to interact with than without it. However, the granularity of topics was occasionally incorrect. All topics were treated as on the same level, for example a *vehicle* was on the same level as *car*, even though the former topic is more broad. Utterances were classified as related in topic by the model (based on adjacency), while in reality the topics were found to be not related after a manual check.

A hybrid method, combining machine learning, discourse structure and knowledge bases for topic extraction has been created by Yeh et al. (2016). The authors evaluated a method for topic detection and tracking using LDA and adding a temporal component to the topic detection as well. Additionally, the method included information such as speech (dialogue) acts, semantic concepts (properties of words) and hypernyms in E-HowNet (in Chinese, comparable to ConceptNet (Speer et al., 2017)). Yeh et al. (2016) proposed a dynamic LDA version that is more suitable for topic extraction from conversations, called CDLDA (conceptual dynamic LDA). The algorithm looked for topics across adjacent utterances. The authors evaluated their method with a spoken Chinese corpus and found that their hybrid method did improve topic recognition. However, when looking at the full results, the accuracy of topic recognition seemed to depend much more on finding the right number of topics to classify, compared to the contribution of the proposed model. Their topic recognition worked much better than other methods for low (< 16) and high (>96) number of topics. However, for 96 different topics to detect, it does not significantly outperform any other method, such as a simpler LDA model.

All the aforementioned topic management and modeling approaches have as a major benefit that they are mostly language independent and do not solely rely on keywords. However, each of them has its drawbacks. The discourse structure approach is less suitable when dealing with noisy speech with no clear structure in the discourse. The machine learning approach always requires (large amounts of) data and existing models are largely based on textual data, which is not always suitable for spoken interaction. The knowledge base approach has a problem with granularity. For example, the distance in a knowledge-graph between apples and vegetables is larger than the difference between skates and water. This means that it is harder to extract topics with the same granularity if an agent is often modeled with similarly granular topics.

2.3.3 Question Generation in Dialogue

question generation

In this final section about personalization we discuss work on *question generation* in conversation. Question generation is a process of automatically generating questions based on a discourse, like sentences or paragraphs and in our specific case, based on conversations (Rus et al., 2011). Asking questions is an excellent way for people to show engagement during a dyadic human-human interaction (Huang et al., 2017). Huang et al. (2017) evaluated the difference between asking open questions and follow-up questions in a social context and found that asking follow-up questions is a good strategy for increasing the likability of the question asker. The authors distinguish between different types of follow-up questions people can ask, such as a standard follow-up question (directly asking about something the other interlocutor has talked about) or mirroring (asking the same question as the interlocutor). Their finding that people that ask more follow-up questions are better liked by the conversation partner could be applied to a social agent as well.

If a social agent could ask questions that are on topic and specific to the user (i.e., personal), the likeability and engagement might increase, though no experiment has been conducted to confirm this hypothesis yet. Additionally, self-disclosure of the user and rapport between the user and a social agent might increase as well.

Two main types of question generation for dialogue are found in the literature: 1) rule-driven and 2) data-driven, both of which we will highlight in the following two sections of this chapter.

2.3.3.1 Rule-driven

For many rule-driven question generation models, these are generated based on the assumption that there is always availability of question-answer pairs, with a ground truth. For example, to generate questions from a discourse, three stages of processing have been proposed (Heilman and Smith, 2010; Yao et al., 2012). The first stage is optional, the sentence simplification stage, in which shorter sentences are preferred over longer ones in the discourse. The second stage is transformation, to transform a declarative sentence or parts of an utterance into a question. The third and final stage is question ranking, where all the transformed questions are ranked according to a fitness function that determines if a question is good. However, this method is mainly meant for generating factoid questions based on existing texts that contain answers to questions, but is less useful for casual conversation with no ground truth answers. In casual conversation we require methods that can generate questions without ground truth answers.

Chali and Hasan (2015) used specific topics found in online resources (Twitter) and put them into templates based on semantic role labeling (SRL) as well, though these were more questions of a factoid kind. Mandasari (2019) continued the work of Fasya (2017) by focusing on online question generation, specifically oriented towards generating follow-up questions in the personal domain, based on the method of Chali and Hasan (2015). Mandasari used templates for the questions and SENNA for SRL. In order to make the questions suited for casual conversation, she extracted templates from the speed-dating corpus of Huang et al. (2017). An example of the question generation can be seen in Table 2.11.

Glas et al. (2017) conducted a study about personalized greetings for a social robot. The robot had a camera and microphone/speakers, the former used for detecting and recognizing faces and the latter for making small talk at a shopping mall. The robot observed people over the course of 23 days. If the robot encountered a person for the third time, it would initiate a small conversation with that person. The question generation in the robot consisted of selecting an appropriate topic and applying it to a predefined template. The topic management used topics such as novel appearances of users (change of hairstyle, walking pace) and timestamps (time and frequency of visits). Though the field study was only done for demonstrative purposes, participants were impressed with the robot remembering

Table 2.11: A conversation from Mandasari (2019), where the agent starts with an opening question (Turn 1) and asks a follow-up question (Turn 3) based on the SRL pattern in turn 2.

#	Interlocutor	Utterance
1	Agent	Do you like reality TV shows? Why or why not?
2	User	I don't like reality TV shows because I believe most of them only fake programs.
3	Agent	Why do you believe most of them only fake programs?
4	User	Because the shows are too good, or contrary too bad to happen in the real world.

them and asking them a question.

2.3.3.2 Data-driven

Many approaches in data-driven question generation deal with factoid questions that are retrieved or trained on internet forums or other texts available. For example, Sun et al. (2018) created and evaluated a question generation neural network trained on the SQuAD and MARCO datasets (Rajpurkar et al., 2016; Nguyen et al., 2016). A *sequence-to-sequence* model for natural language processing (NLP) transforms an input sentence to an output sentence. Sun et al. (2018)'s neural network approach is a sequence-to-sequence to generate questions from answers and their context. Their model aims to be more context-sensitive to the text (i.e., more attention given to words' position closest to question's text keywords) it is trained on and match the types of answers better to the questions it generates (e.g., asking a when question if that information is unknown). Even though the sequence-to-sequence model of Sun et al. (2018) was an improvement in comparison to state of the art, there are issues in these neural approaches that need to be solved before they become usable for a social agent. The first is that the question generation is based on datasets of factoid questions. However, the questions for a social agent in casual conversation are often not about facts, but about opinions and preferences. Secondly, a challenge in casual conversation for social agents is in creating novel questions without having pairs of question-answers to rely on, something that most neural network approaches require.

sequence-to-sequence

Sugiyama et al. (2013) generated sentences for an agent based on topics that were retrieved from an online resource, Twitter. Particularly for generating questions, these were generated by the retrieval of relevant topic phrases, a question type (e.g., how, where) and template matches based on SRL (Ritter et al., 2011).

The generation component took a recognized topic of the user and generated a question with that topic and a related topic. Twitter was used as the online resource to match topics to other topics, based not only on the surface text, but also the dependency relations similarities in SRL. A recent effort of Hu et al. (2018) into topic-based question generation focused on using question types and short sentences from Amazon reviews to generate follow-up questions. For example, from a source text "bottle says made in usa", with topic "bottle" and question type "where", the follow-up question "where does the bottle originate from?" would be generated (Hu et al., 2018, p. 9). The corpus of Amazon reviews however is limited to asking about product details.

One of the few methods recently developed for general domain question generation, where there is no ground truth answer was proposed by Su et al. (2019). Even though their method was applied for conversational interview coaching, their generation of follow-up questions can be applied in other domains or open-domains as well. In their first study Su et al. (2018) looked into using sequence-2-sequence patterns for interview question generation. The context of the study was interviewing students for admission to college. Students could answer questions as elaborately as they liked. The authors collected a small corpus of human-human interviews, where participants had to play the role of interviewer and interviewees. A convolutional neural tensor network (CNTN) was trained to select the best (source) sentence from the corpus to generate follow-up questions from. After the sentence was selected, an LSTM-based sequence-to-sequence model was trained to generate the patterns for follow-up questions from the selected sentence. In the final step, the question templates were filled with related words to the source sentence. A probabilistic ranker selected the best question from all the generated questions, based on co-occurrences of n-grams in the follow-up question and source sentence. In the second study, Su et al. (2019) approached the follow-up question generation based on ontologies. The author's generation component started with selecting the best sentence from a response to an interview question, which is the same as in their first study. After a sentence was selected, an ontology-based approach was used to extract relevant subject-verb-object triples from the sentence. The ontology was created based on keywords found in a new interview dataset and on common words found in the context of interviews. Predicate relations such as "PartOf" and "IsA" were extracted from ConceptNet (Speer et al., 2017) based on the similarity between the words in the ontology and the source sentence and put into question templates. The authors do not mention how the question templates were created, but it is likely the method is the same as in the first study. Their second approach was however evaluated with an automatic metric based on cosine similarity, which did indicate an improvement compared to the first study, but it is hard to tell what it means in practice.

Xing et al. (2017) conducted an evaluation of neural response generation with

Table 2.12: A comparison of the responses generated by a human, STD and HTD, adapted from Wang et al. (2018, p. 2200). The topic words are bold and the interrogatives are italic.

Prompt	The **sushi** today is so delicious!
Human	I remember you didn't like **eating Japanese food**, *did* you?
STD	*When* will you come to **Wuxi**? It is my **treat**.
HTD	Really? I will **try** sometime!

topic-awareness. Their algorithm did not suffer as much as methods before from the "too general" responses, such as "I like that too" or "Really?". However, the generated responses have only been evaluated in single responses, and not with multi-turn conversations with a chatbot. Even though the method can generate questions, most of the responses are not questions. To elicit self-disclosure of the user, more questions are necessary. Wang et al. (2018) took a similar approach to Xing et al. (2017) for specifically generating questions. The authors made a distinction between three types of words that are important for open-domain question generation: topic words, interrogatives and ordinary words. Topic words are required for performing topic transitions and create specific enough questions. Interrogative types help detect patterns of questions in datasets. Ordinary words play a syntactic and grammatical role. A Soft Typed Decoder (STD) and Hard Typed Decoder (HTD) used topic words and interrogative types implicitly and explicitly respectively as sequence-to-sequence approaches (see Table 2.12). The authors only provided a comparative analysis between their own methods and did not disclose information for example on how often a generated question was incorrect (e.g., irrelevant or too general) with the proposed method. The only general reported result is perplexity, which was around 56 for both the Soft and Hard Typed Decoder, having better word perplexity than current sequence-2-sequence models for question generation. The HTD approach is likely the better option for dialogue designers who want control about the topic of the conversation, because the STD is more likely to generate off-topic responses.

2.4 Evaluation of Social Conversational Agents in the Real World

In this section we discuss research related to long-term interactions and evaluation of social conversational agents in the real world.

2

2.4.1 Long-term Interaction

Yee and Niemeier (1996) describe (dis)advantages of long-term (longitudinal) studies versus cross-section repeated studies. In longitudinal studies, the same participants are measured over time, whereas in cross-section repeated studies, different participants are measured over time. The downside of cross-sectional data is that no changes in an individual can be measured. However, cross-section studies do not depend on users committing to a study for a longer period of time, because other users will be recruited in a follow-up. A strong aspect of long-term studies is that individual personalization is more feasible than with a cross-section repeated study.

A recent survey on social robots for long-term interaction studies was conducted by Leite et al. (2013), where the criteria of studies included were i) a clear description of robot and ii) study design and deployment in the real world (e.g., school, work). Out of all studies included in the survey, the authors found 24 studies that ranged in deployment from 2 weeks to three years with between 2 and 180 sessions per participant. The authors argue that total length of deployment is not necessarily the key factor here and that the number of interactions per participant is a more important factor. Leite et al. (2013, p. 304) define *long-term interaction* as "the point where the user is familiarized with the social agent and not influenced by the novelty effect anymore". The *novelty effect* is experienced by participants who are not familiar with the experiment technology. These participants usually have a positive bias towards the technology, influencing the outcome of the experiment. Another factor that impacts the period before users are familiarized with the agent is the complexity of the behavior of the agent. The more complex and diverse behaviors a social agent can exhibit, the less quickly the novelty effect wears off. Leite et al. (2013) found the following aspects to be relevant to decide if researchers want to perform a long-term study:

long-term interaction

novelty effect

- The sample size should analyzable by the researchers. A dataset of many participants can quickly grow out of hand with too many sessions per participant.
- The study should have enough sessions to cancel out the novelty effect.
- Longitudinal studies usually have small sample sizes, which makes it harder to find strong results. However, when the same user experiences an interaction multiple times, the data is considered strongly independent.
- Qualitative metrics are more often used than quantitative, mainly because the sample sizes of longitudinal studies are small and do not provide enough power for quantitative metrics.

De Graaf et al. (2016) proposed acceptance to determine when a setup of a social agent can be considered as long-term interaction. Though acceptance is not

Table 2.13: An overview of the phases of acceptance by De Graaf et al. (2016).

Phase	Explanation	When
Expectation	Users set expectations about the technology	Design phase
Confrontation	Lab tests with users and observations of the technology by users	Just before deployment
Adoption	First user tests in private environment to learn about first experiences	During first month of deployment
Adaptation	Start of official user studies, with novelty effect	During second month of deployment
Integration	Users are familiar with the technology, the novelty effect has dissipated	After second month of deployment
Identification	Users understand the usefulness of the technology and it is integrated with their social life	Six months after deployment

the same as familiarity, the two are inherently related for long-term interaction. According to De Graaf et al., the *acceptance* of a social robot consists of six stages: expectation, confrontation, adoption, adaptation, integration and identification (see Table 2.13). After roughly two months of interaction, users are familiar with a social agent. From this time onward the interaction can be considered long-term, because usually no novelty effect applies anymore, which is similar to what Broekens et al. (2009) found. However, considering the final stage of De Graaf et al. (2016), identification, to see the effect of a setup for the real world, it could take up to six months after the first deployment to get a realistic real world user experience.

acceptance

2.4.2 Real World Deployment

A *real world interaction* is an interaction taking place in a familiar environment for the target users of the research. Leite et al. (2013)'s states in their survey about 24 real world studies that most were performed to test the technology itself and learn about the environment. Most studies did not go beyond the integration phase of their setups. However, despite many robots' limitations and flaws, a hopeful and positive finding overall was that people will accept social agents in their environment in the long run and likely also in the identification phase. Leite et al. (2013) categorized all real world studies in the survey in four groups: health

real world interaction

Figure 2.9: The iCat used in for example Leite et al. (2014)'s study.

care and therapy, education, work environments and public spaces, and home use. In the domain of education, Leite et al. (2014) deployed an iCat robot (Figure 2.9) which could play chess once a week with children over the course of five weeks. The iCat's model consisted of five important components to sustain long-term engagement: affect detection, empathetic appraisal, supportive behaviors, memory of past interactions and the action selection. Their model retained high numbers of engagement and social presence. However, the study did not last over 2 months and the children could not initiate the interaction themselves, but had interactions at planned sessions.

In another educational setting in the EASEL project, Davison et al. (2020) deployed a Zeno robot as a teaching tool for children in primary school. The robot operated fully autonomously over the course of four months. Children could start the interaction themselves by letting the robot scan their radio-frequency identification (RFID) card. Despite the limited set of speech phrases and behavior variations, the Zeno retained the attention of the children. This is a positive finding, because it means that designers of social agents need not make the most complex agents they can think of to retain engagement. Davison et al. (2020) mention some practical considerations as well for conducting real-world studies. First and foremost, even though teachers said they did not have time to get involved with the robot, they occasionally did involve the robot in their lessons. Different teachers had different lesson plans, such that some covered the robot when they did not want children to work with the robot. Important as well with real world studies are the ethical considerations. Davison et al. (2020) organized an information evening for the parents about the study and discussed with the school management how to fit the study into the school's curriculum. Communication is important and any issues stakeholders might have with a study design can be resolved through information evenings, focus groups or co-design workshops with users. Finally, regular maintenance should be scheduled at appropriate times, because things will break or fail sooner or later (Sung et al., 2010).

Tsiourti et al. (2018) evaluated a virtual companion at the homes of older adults in three different countries. The author's system, the CaMeLi framework, was evaluated for a period of three months. Participants were asked to keep a diary about the interaction, invited for focus groups after each month and were given usability and quality of life questionnaires. Most participants had problems with the believability of the agent, which due to its life-like appearance set expectations high, especially in understanding speech-based requests. Also the older adults required more extensive training to communicate with the agent. Interestingly, for each country, different capabilities of the agent were considered useful, such as memory training or scheduling social activities.

Though the other categories provide useful insights, this thesis focuses on social agents in the category of work environments and public spaces. Gunson et al. (2020) conducted a study with their social agent Alana in a public space to compare the quality of interaction with Alana depending on a condition. In one condition, Alana used casual talk with task-oriented dialogue, and in the other condition Alana only used task-oriented dialogue. The authors found that people generally did not want to talk casually with Alana and preferred the task-oriented version, because it helped more efficiently. The authors also state that it is more important for casual talk to supply it on-demand for the user, rather than being told during the task itself. However, this experiment was done in a lab-controlled setup. Setups in public spaces for real world evaluation are those of Kennedy et al. (2017), Gockley et al. (2005) and Kanda et al. (2010), respectively social robots designed for small talk, interactive-storytelling, and giving directions and rapport-building. Kennedy et al. (2017) placed their robot Kevin in an office space, where it could interact with all office workers. Gockley et al. (2005)'s robot Valerie was placed at the reception of a university, where visitors, employees and students could talk to her. Kanda et al. (2010)'s robot Robovie was located in a large shopping mall with many visitors. The biggest challenge in the domain of public spaces is to deal with any type of visitor that an agent might come across. For example, in one of the first real world studies performed by Huttenrauch and Eklundh (2002), the authors placed the robot CERO in an office space (Huttenrauch and Eklundh, 2002; Severinson-Eklundh et al., 2003). CERO fetched office supplies and coffee for a user who had limited walking capabilities. The authors found that even though CERO was designed specifically and evaluated for one person to fetch things, many other people in the office tried to initiate interaction with the robot, but the robot did not have any behaviors to respond to them. Many of them said they would have liked to interact with the robot.

2.5 Summary

In this background chapter we have discussed four important aspects of dialogue adaptation and personalization for social conversational agents: i) multimodality and memory, ii) multimodal dialogue design, iii) personalization through user topic modeling and question generation and iv) evaluation in real world interaction.

Modalities are most accessible when a microphone and camera are used to capture user input, though haptics would increase rapport and social presence. Most of the modalities include a memory structure and conversation history is already part of many of the aforementioned dialogue systems. Multimodality also increases possibilities for incremental dialogue management, for example that an agent can backchannel appropriately through nods. However, long-term casual conversations with memory are scarce in the field of open-domain talk for social conversational agents (Elvir, 2010; Mattar and Wachsmuth, 2014).

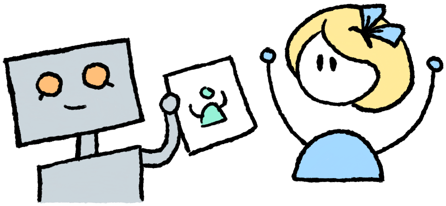
Designing multimodal systems is becoming increasingly more feasible than ever with many tools and frameworks being developed. Most of them are online available via public repositories (Bohus et al., 2017; Michael, 2020; Li et al., 2020). However, many dialogue designers still struggle with the complexity of building social conversational agents. Most of the available frameworks lack good documentation and design guidelines, which leads to a high learning curve for dialogue designers with slightly different conversational agent requirements. Tools such as the VHToolkit (Hartholt et al., 2013) and Visual SceneMaker (Gebhard et al., 2012) accommodate designers with a graphical interface for implementing a conversational agent. As for guidelines for dialogue designers, in many systems there is no clear distinction between generic interaction behavior and specific content behavior for social agents, which makes existing dialogue systems harder to re-use (Rich and Sidner, 2012).

Personalization through user modeling has been done for quite some time (Kobsa, 1989) and regardless of state-of-the-art machine learning approaches with BERT and OpenGPT-3, structure- and knowledge-based approaches seem to remain relevant (Langlet and Clavel, 2016). The main reason is that machine learning approaches often give too generic responses. For question generation, a combined machine learning approach with discourse structure templates has proven to be one of the most effective methods for personalized question generation (Su et al., 2019). A hybrid approach maintains a balance of generalizability and control for dialogue designers, though is unfortunately hard to reproduce (Wang et al., 2018).

Finally, there have been limited evaluations of social conversational agents in the real world. For these types of long-term interactions, an agent needs to be deployed for at least two months to account for the novelty effect (De Graaf et al., 2016; Leite et al., 2013). Robustness is important and incorporating users needs is vital for meeting their expectations (Davison et al., 2020; Tsiourti et al., 2018).

Part II

Dialogue Design and Prototyping



3 ARIA: A Framework for Multimodal Embodied Conversational Agents

3

This chapter is mostly based on the work of these two papers:

- M. Valstar, S. Dermouche, C. Pelachaud, E. Coutinho, B. Schuller, Y. Zhang, D. Heylen, M. Theune, **J. van Waterschoot**, T. Baur, A. Cafaro, A. Ghitulescu, B. Potard, J. Wagner, E. André, L. Durieu, and M. Aylett (2016). "Ask Alice: An Artificial Retrieval of Information Agent". In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*. ACM Press, pp. 419–420. DOI: `10.1145/2993148.2998535`

- A. Cafaro, M. Bruijnes, **J. van Waterschoot**, C. Pelachaud, M. Theune, and D. K. J. Heylen (2017a). "Selecting and Expressing Communicative Functions in a SAIBA-Compliant Agent Framework". In: *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings*. Springer, pp. 73–82. DOI: `10.1007/978-3-319-67401-8_8`

3.1 Introduction

In this chapter we will discuss the ARIA-VALUSPA Virtual Platform (AVP) of the ARIA-VALUSPA project[1] set up to advance (accessibility of) multimodal technology with virtual humans. In the project a social virtual agent was developed, which is called an *Artificial Retrieval of Information Assistant (ARIA)*. **ARIA**

[1] Artificial Retrieval of Information Assistants - Virtual Agents with Linguistic Understanding, Social skills and Personalised Aspects (ARIA-VALUSPA), funded by European Union Horizon 2020 research and innovation program, grant agreement No 645378.

Research in Section 3.2 until Section 3.4 is done by partners in the project, whereas we contributed most in integrating all these components in the AVP. In Section 3.2 we will introduce the domains of the project: affective storytelling and commercial information retrieval. In Section 3.3 we introduce the input capabilities of the ARIA and in Section 3.4 we discuss its output modalities. Our research contribution starts from Section 3.5 onward. Section 3.5 highlights the dialogue management in the ARIA, which consists of user understanding and behavior specification with a short run-through example of the DM. In Section 3.6 we give an overview of applications of the ARIA.

3.2 Context

More and more information retrieval tasks are automated within accessible virtual assistants such as Apple's Siri, Amazon's Alexa and Google Home. Information retrieval in the context of a virtual assistant is about asking questions about your calendar, looking up facts or asking an assistant to tell a joke or a story. Personal assistants that are capable of performing tasks and retrieving information are becoming increasingly available to people and industry (Cowan et al., 2017). We often see that these systems are used to perform search queries, set timers, execute certain commands or answer relatively simple questions using the ASR transcriptions they get from the user's speech. However, most of these systems are not capable of emphatic responses or grasp social situations.

With the ARIA we want to bridge the gap of current virtual assistants and bring them closer to more natural interaction with users. Most virtual assistants have no automatic affect recognition (AAR) capabilities and cannot for example detect user engagement. The information that the ARIA can use is multimodal and supports automatic affect recognition (AAR), for both robustness and more enriched conversations. The platform is multilingually set up, with support for English, German and French. The AAR capabilities are not dependent on language models, but on acoustic models and facial expressions. The platform that we developed had to be accessible to many people: the ARIA should be deployable in any home and work on most devices such as computers, tablets and phones. Dialogue designers should be able to recreate their own virtual human with the toolkit and add extra modalities easily as they see fit. Virtual assistants often cannot deal with dynamic turn-taking. For example, Amazon Alexa indicates with a light when it is ready to receive user utterances, but the user cannot stop Alexa mid-sentence. This is desired in truly incremental interactions, where either the user or an agent can interrupt the other at any given moment during an interaction. The ARIA is able to have incremental interactions. An advantage of incrementality is that users can correct themselves or the agent more quickly.

3.3 Multimodal Input

In Section 2.1.1 we discussed multiple modalities that can be used for social agents. In this section we will discuss briefly the input modalities of the ARIA. The modalities that we can use for the ARIA will be sufficient to give a rich experience to novice users and can be extended for the more advanced users who are able to and want to use more modalities. Given that most computing devices we use nowadays have a microphone and camera available, speech and vision are the modalities that we focus on in the ARIA.

3.3.1 Speech

The ARIA's speech components consist of ASR, AAR and voice activity detection (VAD). The ASR in the ARIA supports transcriptions in three different languages: English, German and French. The WER was 39.0% 28.8% and 40.2% for these languages respectively (Valstar et al., 2018). The ASR is based on the Kaldi framework and can be set to produce up to 10 different transcriptions per spoken utterance and can send intermittent transcriptions to support incremental interactions (Povey et al., 2011). Improvements to the original ASR were made by Mousa and Schüller (2016). More details on the final ASR implementation can be found in the ARIA-VALUSPA technical report D2.1 (Schuller et al., 2015). For AAR, ARIA uses openSMILE for extracting the valence and arousal levels from user's speech, based on features like F0 and pitch (Eyben et al., 2010). Additionally, openSMILE predicts user's demographics, such as their gender (male/female) and age category (child/adolescent/adult/senior). VAD is included in openSMILE as well, based on the work of Eyben et al. (2013) and is important for detecting interruptions by the user.

3.3.2 Facial Expressions

ARIA uses eMax for the recognition of emotions from user's facial expressions. Ekman et al. (2002)'s facial action units are used in eMax to recognize six basic emotions: anger, sadness, surprise, disgust, happiness and fear, as well as arousal and valence. eMax can detect the emotions in real-time from video robustly with slight face orientations (Almaev and Valstar, 2013). Jaiswal and Valstar (2016) developed a deep learning method (BLSTM-CNN) for detecting spontaneous emotion from the face, further improving the recognition of facial action units to be used in AAR. Additionally, faces with eMax can be recognized with facial point localization of multiple people at a time and head pose estimation (Sánchez-Lozano et al., 2016), which supports the ARIA with multi-user interactions.

Figure 3.1: Picture showing the interface of the SSI framework (Wagner et al., 2011), which shows microphone activity, video feed, processed multimodal input and ASR.

3.3.3 Analysis and Processing

The Social Signal Interpretation (SSI) framework is integrated in ARIA (Wagner et al., 2011). SSI is capable of real-time synchronous merging of different multimodal inputs, such as from cameras and microphones but also game controllers. In ARIA we used SSI to merge and synchronize camera and audio input, namely 1) ASR, 2) eMax and 3) openSMILE. SSI combines the input from these three components to a measure of interest (or engagement), in which high arousal and high valence of the user together with the user looking at the agent are interpreted as high interest of the user. Additionally, SSI has some general filter and feature algorithms which can extract high-level information such as affective states (e.g., valence, arousal and basic six emotions) from multimodal data and has machine learning tools available for classifying and clustering multimodal data (see Figure 3.1). The logging of this data is automatic and can be directly annotated or analyzed by researchers with annotation tools such as ELAN and NOVA (Wittenburg et al., 2006; Heimerl et al., 2019). In the ARIA-VALUSPA project the Novice eXpert Interaction (NoXi) database was collected and automatically annotated with SSI and eMax and manually with NOVA (Cafaro et al., 2017b).

Figure 3.2: We chose this virtual human included in the Greta platform as one of the representations of the ARIA.

3.4 Multimodal Output

Similarly to the modalities of the multimodal input components of the ARIA, auditory and visual modalities are used for multimodal output. A virtual human as embodiment for the ARIA is shown in Figure 3.2.

3.4.1 Non-verbal Behavior Generation

ARIA's virtual human is Greta, a platform that supports BML and FML standards (Poggi et al., 2005). Greta supports facial expressions, gestures with her arms and head movements. It is also possible to change the appearance of Greta and change the surroundings to make the virtual human more suited for other domains. By default, Greta has idle behaviors, such as breathing, blinking and looking around. The code for Greta[2] is available for others to use as their virtual human. Animations generated by Greta are smoothed out as well, which means that if the user interrupts the ARIA or if the ARIA wants to change behavior autonomously, this happens without unnatural jerking or sudden movements.

[2] `https://github.com/isir/greta`

3.4.2 Text-to-Speech Generation

ARIA uses CereProc's TTS voices, CereVoices, for the generation of speech for the ARIA (Aylett and Pidcock, 2007). Greta has support for all the CereVoices and is able to use features for synthesis that are required for incremental interactions, such as the voice talking louder or softer and change the affective state of the voice, such as "calm", "neutral", "angry" and "sad". Additionally, CereVoices generates visemes, which are necessary for lip synchronization while talking, and thus CereVoices can easily be integrated with virtual humans and robots that require lip synchronization. Finally, SSML can be combined with CereVoices for adjusting the voice, for example the pitch and speed.

3.5 Multimodal Dialogue Management

ARIA uses Flipper for supporting prototyping (non-verbal) NLP and DM (Ter Maat and Heylen, 2011). In short, Flipper is an information-state based dialogue engine that uses rules in templates for multimodal behavior understanding and specification (Chapter 4 discusses Flipper in more detail). In this section we describe our multimodal dialogue management for user input understanding and behavior specification generation (see Figure 3.3).

dialogue structure

Dialogue management in Flipper is based on *dialogue structure*, similar to the FLoReS dialogue manager of Morbini et al. (2014). A dialogue structure consists of three levels: episode, exchange and move. An episode can be viewed as a type of sub-dialogue, such as question-answering, making small talk or performing a task such as giving directions. Exchanges are parts of the episode and are all about the same topic, such as answering different questions about a specific book. Moves are the atomic components of the dialogue structure and correspond directly to a single conversational act, such as a nod for backchanneling or giving an answer to a question. Dialogue management in ARIA consists of transforming multimodal input to user moves and generating agent moves through behavior specification.

3.5.1 User Input Understanding

SSI handles all the real-time processing and synchronization of multimodal input and transforms FACS and acoustic signals to higher level information such as emotions, valence and arousal , such that Flipper has the latest information available. We created templates in Flipper that interpret multimodal dialogue input. Templates determine what the user intent is based on the input, such the user assigning or yielding the turn.

For user input understanding, the multimodal input will be converted to an observed user move consisting of three different types: interaction, content and socio-emotional (indicated in blue, orange and red respectively in Figure 3.3).

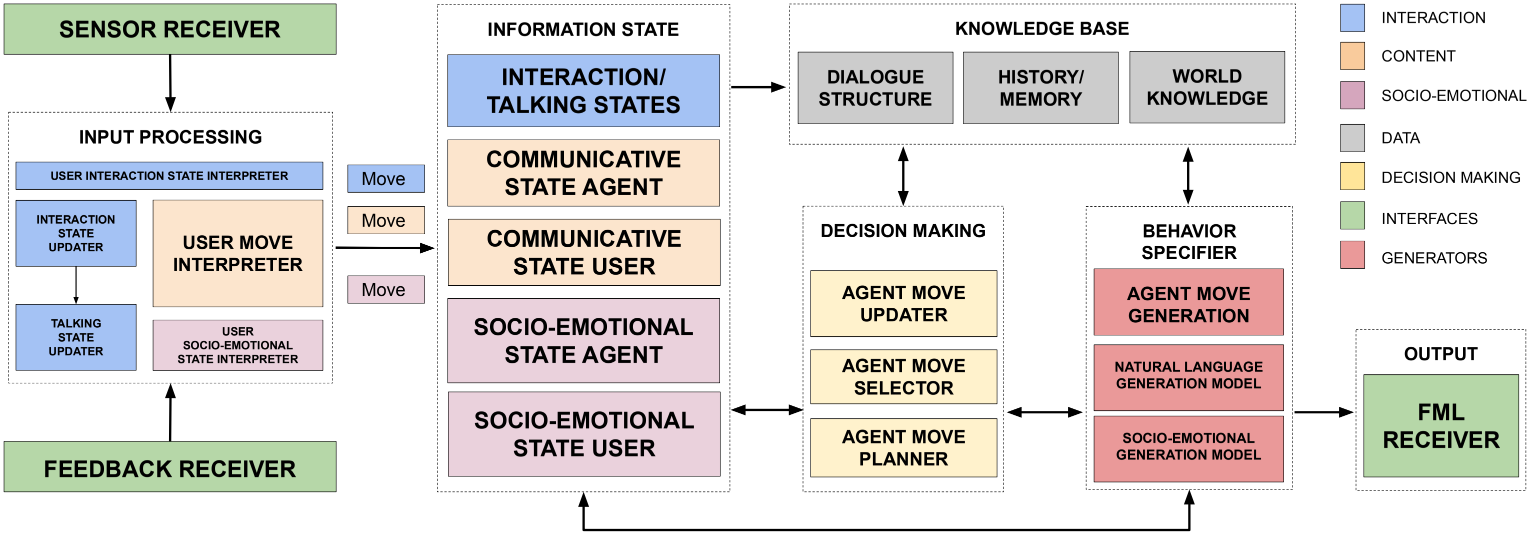


Figure 3.3: The architecture of dialogue management with Flipper within ARIA-VALUSPA. On the left side, input processing receives input from components such as sensors (e.g., voice activity) and feedback (e.g., BML callback). During input processing, three different information types are processed into moves: interaction (turn management), content (semantics) and socio-emotional (affect recognition). Within Flipper, templates transform these inputs to their respective internal component in the information state. These are interaction/talking states, the communicative states of agent and user and the socio-emotional state of agent and user. The knowledge base in the top-right is composed of the dialogue structure (exchange/episode), dialogue history and world knowledge (facts). On the bottom, templates in the decision-making contain rules to update relevancy of salient moves and select them based on the knowledge base and information state. After a move has been planned, it is sent to the behavior specifier, which specifies the natural language and socio-emotional parameters of realization (FML). The generator sends the FML to the behavior realizer in the output step.

3

These moves can be multidimensional and be combined in a single or multiple intents of the user.

Interaction moves are used to manage the meta-dialogue, such as opening and closing the interaction and handling interruptions and turn-management. For example, an interaction move is "disengaging" and this move is created through a template with a rule for detecting a certain time of no voice activity nor eye contact of the user. Templates for interaction moves can directly trigger an agent response without any deliberation.

Content moves always contain a topic, such as the user asking or answering a question. Content moves are determined by the history of moves, the current exchange the user is in, the detected topics and keywords spotted in the user's speech. Content moves require more deliberation than interaction moves in Flipper, because they require NLP components. The StanfordCoreNLP (Manning et al., 2014) is used for extracting the important words from a user utterance based on part-of-speech (PoS) tags and a question-answer database is used for detecting user questions based on n-grams of the keywords.

Socio-emotional moves keep track of the emotional state of the user, using eMax' and openSMILE's non-verbal user behavior recognition. For example, a positive valence of the user is an indicator of high engagement of the user with the agent.

3.5.2 Behavior Specification Generation

The specification of behavior for the ARIA consists of two components: the intent planner (decision-making) and the behavior specifier (generating).

intent planner

The *intent planner* chooses a communicative function and optionally a topic to talk about, based on the user input understanding. The DIT++ taxonomy is used for the communicative functions and how to convey certain behaviors (Bunt et al., 2010). A full overview of the supported DIT++ intents are shown in Table 3.1.

Similarly to the user input understanding, the behavior specification generation consists of agent moves, divided into the same three categories as for the user: interaction, content and socio-emotional. If the agent has a content move, a topic is selected with an *n*-gram approach, based on the semantic similarity between the agent topic and the most recent user content move and dialogue history. A possible interaction move might be that the agent yields the turn if the user seems eager to talk. Lastly, in the case of a socio-emotional move, an example is showing gratitude for a compliment given by the user.

The intent planner contains three components: the move updater, move selector and move planner. The move updater sets the relevance of each possible move based on the dialogue history and dialogue structure. For example, if the agent answered questions of the users, moves related to question-answering (structure)

Table 3.1: An overview of our FML-templates categorized according to DIT++ taxonomy.

Class	Goal	Sub-classes
Information Transfer	Obtain or provide information	Question: set | choice | prop | check Inform: agreement | disagreement | answer | elaborate | explain
Feedback	Provide or elicit information about the processing of the previous utterance(s)	Auto: positive | negative Allo: positive | negative Elicitation
Interaction	Structure the dialogue (e.g. turn or topic management)	Contact: check | indication Time: stalling | pausing Turn: take | accept | grab | keep | assign | release Topic: introduction | preclosing | announce shift
Social Obligations	Social policies during the dialogue	Salutation: initial | return Introduction: initial | return Gratitude: initial | return Apology: initial | return Valediction: initial | return

will have higher relevance, as well as moves that are semantically similar to previous moves (history). Additionally, the move updater sets the threshold of how high a relevance should be, because only if a move exceeds the threshold, it will be selected as the agent move. Updating the threshold happens in real time to support possible interruptions, both by the user and agent. If there are no moves with high relevancy, the threshold gets lowered. For example, when the user has the turn, the agent has a high threshold for talking and will rarely interrupt the user. However, if the user turn goes on for very long, the threshold gets lowered and the agent is more likely to interrupt the user.

The move selector takes the agent's own goal and the relevances of the move updater into account. The move selector picks a move once its relevance is above the threshold and sends it to the move planner, which constructs an agent move.

The move planner takes as input the history of observed and expected user moves (see Figure 3.4). Expected and observed user moves are for example when the agent asks a question, the user is expected to answer (expected), but could reply with another question (observed), like in Table 3.2. The history of the agent's own moves is used as well for planning. For each possible agent move that is planned, the move planner checks if it has been completed or executed. An agent move that has been interrupted is considered executed, but not completed. A move that has not been completed gets a higher relevance in the move updater. Thus, an uncompleted agent move has a higher chance that it will be selected again by the move selector.

3

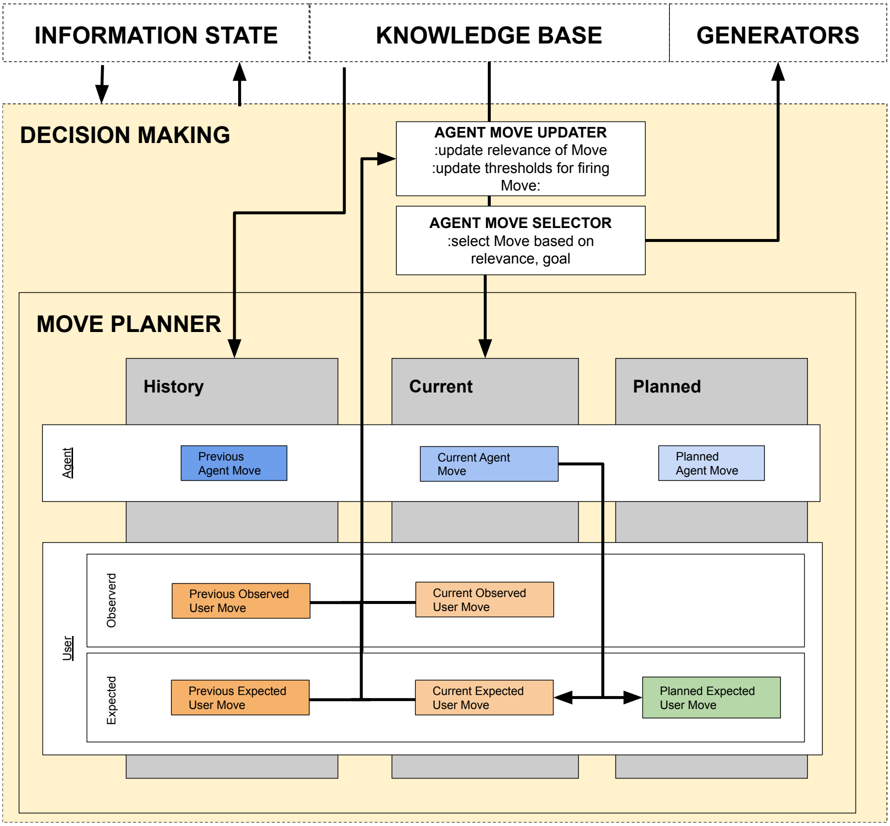


Figure 3.4: A close up of the intent planning inside the Flipper architecture.

Given the history of the moves, the current expected user move and the selected agent move, a new agent move is planned, as well as a prediction about how the user will respond. Note that multiple moves can be executed at the same time by the agent, such as an interaction move (giving the turn) and asking a question. Once a move has been planned, it is sent to the generator component.

behavior specification For *behavior specification* we built an FML realizer which can take as input the topic and communicative intent (DIT++) from the DM and generate FML. The ARIA has a list of available FML-templates with parameters for creating variable behavior. The parameters for FML are shown in Table 3.3. Dialogue designers can set any of these parameters, depending on their ARIA goal. The first four elements in Table 3.3 are standard FML tags which change values inside the FML templates. The emotion element sets how the agent should express emotion and can be combined with other communicative functions. The emphasis element emphasizes both verbally and non-verbally part of the agent's speech. Certainty

Table 3.2: An example dialogue between the agent (ARIA) and a user as interlocutors (I). The emotion (E) is indicated as neutral (N) or happy (H).

#	I	Utterance	Keywords	Intent	E	A	V
1	User	I don't think the queen is reasonable.	think, queen	answer	N	0.2	0.2
2	Agent	Would you like to know more about the queen?	info, queen	set	N	0.2	0.2
3	User	What can you tell me about the white rabbit?	info, white, rabbit	set	H	0.6	0.6
4	Agent	The white rabbit was mean at the tea party.	info, rabbit	answer	H	0.6	0.6

Table 3.3: Functional Markup Language (FML) parameters for the ARIA.

Element	Attribute
emotion	type, intensity, importance
emphasis	level, importance
certainty	type, intensity, importance
voice	type
var	type
alternative	type, name
alt-option	ref

allows for probabilities in expressed communicative functions. The voice types are specific for CereVoice, to synthesize more natural emotional speech, with calmness, anger, sadness or happiness.

The bottom three parameters are directly affecting the FML templates format. The `var` element is used for the topic or a sentence in an FML template, the alternative element is optional if multiple topics or sentences can be used and the alt-option contains if-then rules if the alternative element is used. The three FML-template parameters provide dialogue designers flexibility with behavior specification, for which we created three different levels.

1. Static. The DM sends a list of possible emotions or topics that can be used as parameters in the FML templates (see Listing 3.1). The dialogue designer has to decide for themselves what type of possible behaviors they want in a limited set of emotions and topics. For example, an FML template for

```
<alternative id="alt1" type="static">
 <alt-option>For <tm id="tm1"/>instance:</alt-option>
 <alt-option>For <tm id="tm1"/>example:</alt-option>
</alternative>
```

Listing 3.1: Example of a static alternative with an inform-elaborate intent, with a variation of saying "for instance" or "for example".

backchanneling has a parameter for an utterance. A list would consist of the words "Okay", "Yes" or "Uh-huh". They are all equally likely to be selected.

2. Selectable. If the designer wants a specific emotion or topic option in a template instead of a random option in the FML, the selectable alternative is chosen (see Listing 3.2). The selectable is useful for creating rule-based FML templates. For example, if a positive valence is detected from a user move, the agent might choose the option in the FML template that has smiling as a behavior.

3. Dynamic. The DM is connected to a component that generates full FML behavior specifications, independent of the FML templates included in the ARIA. Listing 3.3 shows the few lines of code necessary for including dynamic specification, but it leaves all the generation of FML specifications to the DM. The specifications can be extended to generate more behavior dynamically, for example by adding BEAT-gestures (Cassell et al., 2004).

The final FML behavior specification is sent to the FML translator which translates the FML to BML to realize the behavior of the agent, using Greta's virtual human and CereVoice speech. Once the ARIA receives an FML specification and is realizing the behavior, it will send live BML callbacks to the input understanding component. BML callback include time markers for gestures and spoken text of currently executed agent behavior. If another move of the agent becomes relevant during the behavior execution, a template within the move planner activates to specify the behavior for this move. Time markers in the BML callbacks are used for a more fluent transition of animations and speech and support incremental interaction.

3.5.3 Walkthrough of the Intent Planner and Behavior Specification

In this section, the steps that the system takes to respond to the user are illustrated. Table 3.2 shows four turns in a typical user-agent interaction with ARIA, in the context of Alice's Adventures in Wonderland. The second column shows who

```
<speech id="s1">
<alternative id="alt1" name="positive-feedback" type="selectable">
 <alt-option ref="named">Yes <tm id="tm0"/>
 <var id="var1" type="user"/><tm id="tm1"/></alt-option>
 <alt-option ref="no-named">Yes</alt-option>
</alternative>
</speech>
<alternative id="alt2" name="positive-feedback" type="selectable">
 <alt-option ref="named"><emphasis id="emp1" start="s1:tm0"
 level="strong" end="s1:tm1" importance="1"/></alt-option>
 <alt-option ref="no-named"></alt-option>
</alternative>
```

Listing 3.2: Example of a selectable alternative for providing positive feedback to the user.

```
<tm id="tm0"/>
 <alternative id="alt1" type="dynamic"/>.
<tm id="tm1"/>
```

Listing 3.3: Example of a dynamic-alternative, in which any FML can be included between the alternative tags.

has the turn. The third column shows the utterances during the turn. The fourth column shows the keywords and topics that are extracted and selected for the user and agent turn respectively. The intents for both interlocutors are shown in column 5, based on the DIT++ intents (see Figure 3.1). The three remaining columns indicate the user's and agent's emotional stance. Arousal (A) and valence (V) are computed in the input processing block through acoustic and visual features (e.g., prosody, facial expressions) for the user and affective words and mirroring the user emotions for the agent socio-emotional stance, as shown in Figure 3.3.

The user speech interpreter recognizes that the user has the intent of asking a question (*set* in DIT++) with some additional keywords ("white" and "rabbit") representing the subject. The move planner selects an *answer* FML Template which is intended to provide information and waits until it is the agent's turn to give the response. The selected FML template, shown in Listing 3.4, contains a **<var>** element for a sentence clause that can be replaced with the ARIA's answer. All the attributes in the template that can be modified are indicated with a question mark. The agent's emotional expression can be computed from the mental state. *Type* becomes **angry** based on alignment with detected sentiment words in the user's utterance, in this case "mean". Once the type attributes of the **<voice>** and **<emotion>** tags are set, the ARIA can produce the utterance with an angry

```
<fml>
 <speech id="s1" start="0.0" language="english" voice="cereproc">
  <alternative id="alt1" name="inform" type="selectable">
    <alt-option ref="opinion"><tm id="tm0"/><voice type='happy'>
      <var id="var1" type="sentence"/></voice><tm id="tm1"/>
    </alt-option>
    <alt-option ref="neutral"><tm id="tm0"/><voice type='calm'>
      <var id="var1" type="sentence"/></voice>
    </alt-option>
  </alternative>
 </speech>
 <emphasis id="emp1" start="s1:tm0" end="s1:tm1"
 importance="1"/>
 <emotion id="emo1" start="s1:tm0"
 end="s1:tm1" intensity="1" importance="1.0"/>
 <certainty id="cr1" start="s1:tm0" end="s1:tm1"
 intensity="0.5" importance="0.5"/>
 <performative id="p1" type="inform" start="s1:tm0" end="s1:tm1" />
</fml>
```

Listing 3.4: The FML Template describing an information statement in DIT++.

synthesized voice and matching behaviors. Finally, because no values have been set for certainty and emphasis, those attributes will use the default value in the FML template, which is usually the average between the minimum and maximum possible value. The performative element inside the FML-APML will produce gestures that are related to the inform intent.

3.6 Applications

Once we completed the ARIA architecture, we created two-use cases for the ARIA. One was the Book-ARIA, for entertaining and informal use and one was an Industry-ARIA for commercial use. We had two goals with these use-cases: evaluating the user experience and providing a fully implemented agent which dialogue designers could use as a starting point for their own virtual agent. Additionally we will discuss where the ARIA has been used in two projects outside our own use-cases: one as a virtual journalist and another as a virtual interviewer.

3.6.1 Book-ARIA: Alice in Wonderland

The Book-ARIA was Alice, a character from Alice's Adventures in Wonderland (Carroll, 1865). Specifically for this use-case, the ASR was trained on audiobooks

Table 3.4: Part of a conversation of interlocutors (I) user (U) and agent (A), in which a user tries to convince Alice to trust them and tell the truth.

#	I	Utterance	Truth	Trust
1	U	Where do you live?		0.6
2	A	I live in Wonderland, I think.	False	0.6
3	U	What can you tell me about the white rabbit?		0.6
4	A	The rabbit took a watch out of its waistcoat-pocket! It said: "Oh dear! Oh dear!" "I shall be late!" I got curious and followed the rabbit into a rabbit-hole under the hedge.	False	0.7
5	U	Do you have any siblings?		0.7
6	A	My parents had nine children. When I was born, I had an older brother and two sisters.	True	0.7

from the story of Alice. The goal of the Book-ARIA was to provide a believable virtual agent that was informative and engaging to listen to. Alice is a very expressive character in the story, which is true as well for the creatures and people she meets during her time in Wonderland. We saw Alice as the perfect fit for using the richness of the story and bring this to "life" with the Book-ARIA.

We created a scenario in which users had to discover the truth of Alice's identity. We set up Alice with certain personality traits that made her speak more affectively, inspired by the personality model of Chowanda et al. (2014). While talking with the user, initially Alice did not trust them and only talked about her adventures in Wonderland. In reality, Alice was someone who lived in England. The goal of the user was to gain the trust of Alice and learn the truth about her. The user could gain her trust by maintaining eye contact often, showing interest by asking Alice questions about Wonderland and appearing positive. Once Alice trusted the user enough, she began explaining where she was really from and tell more about her real life. An example of an interaction between Alice and a user is shown in Table 3.4. After the interaction, we provided users with a questionnaire to rate their experience as well (see Appendix B).

3.6.2 Industry-ARIA: Customer support

The Industry-ARIA was a customer support social agent that could help out with questions about products. Let us say the products were from a do-it-yourself (DIY) store FixIt.[3] The agent, Alice, could help out with questions about which drill to

[3] Due to privacy, a fictional company is used for this with comparable requirements.

use for which surface. We created a dialogue tree based on the frequently asked questions (FAQ) of the company, for which Alice would ask a couple of slot-filling questions. Included questions were about the surface to drill in and the object to hang. During the interaction, Alice kept track of the user's voice and face to detect if there was any dissatisfaction from the customer and Alice accommodates to the current user's emotional state. She would for example ask the customer to confirm if she correctly answered a question or whether the customer wanted to know more information about a certain answer she has given.

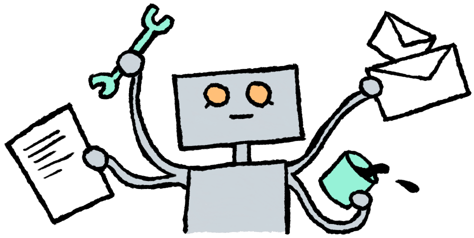
3.6.3 Additional Applications

One application of the ARIA platform has been the automation of interviews based on questionnaires. Jaiswal et al. (2019) compared the virtual agent platform to participants filling in a questionnaire on their own and to an interviewer asking the questions. The authors found that participants gave similar answers in both conditions. This means that a virtual agent might be just as useful of a method for participants as self-reporting in a questionnaire. Overall, chatbots were found to be just as effective and be at least more enjoyable than participants having to fill in questionnaires themselves (Te Pas et al., 2020). Another application was a virtual journalist designed with the ARIA platform to retrieve information from people (Bowden et al., 2017). The virtual journalist was capable of dealing with human emotions and mirroring behavior. Unfortunately, the agent was not found very engaging by users. Further development could mitigate this issue by using more natural gestures and a more affective voice.

3.7 Summary

In this chapter we described the ARIA-architecture and development of ARIAs. We have shown the multimodal input and output capabilities, as well as some of the processing that underlies the DM of the ARIA. Speech can be used for ASR and AAR, and video feed for AAR as input. For output we utilize non-verbal behavior generation (NVBG) with a virtual human and a synthetic voice for TTS. Processing and producing of respectively the input and output happens by rules in templates for dialogue management within Flipper, which uses an information state approach together with components for NVBG understanding and generation. The generation part consists of an FML generator with FML templates for dynamic generation of multimodal behavior, based on topics and communicative functions. We showcased some example applications of the ARIA inside and outside the project.

In the next chapter we take a look at the technical aspects of dialogue management in Flipper and give recommendations for designing multimodal dialogue.



4 Flipper: Designing for Multimodal Embodied Conversational Agents

This chapter is based on

- **J. van Waterschoot**, M. Bruijnes, J. Flokstra, D. Reidsma, D. Davison, M. Theune, and D. Heylen (2018a). "Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents". In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 43–50. DOI: `10.1145/3267851.3267882`

In this chapter we discuss general dialogue management and dialogue design with Flipper, as a continuation of multimodal dialogue management discussed in Section 3.5.

4.1 Introduction

The task of building multimodal dialogue systems for a social agent or embodied conversational agent (ECA) in large multi-partner research projects is not trivial. Such systems need to handle complex, emergent, multimodal dialogues, be continuously responsive, and deal with unpredictable user input. The reality is that in such projects the dialogue system consists of, and interfaces with, several specialized components from different partners, each with their own technical framework. The ideal dialogue system has two dimensions: it needs to i) support researchers to achieve the complexity of the emerging dialogues that current projects strive for; and ii) support the quick creation of (partially) functional prototypes that can demonstrate and/or evaluate the effect of design choices or of prospective technical components on the ECA early in the project.

The first version of Flipper was used for performing dialogue management in the SEMAINE[1] project (Ter Maat and Heylen, 2011). We have upgraded Flipper to navigate the abovementioned two dimensions and present Flipper 2.0 , a declarative language and interpreter specifically designed to quickly and iteratively create a dialogue manager for an ECA. Towards that goal we have designed Flipper with the following capabilities.

1. With Flipper, basic dialogues can be created with minimal overhead.
2. Flipper can switch between
   (a) delegating a task to an external specialized component, for example sensor interpretation or decision-making; and
   (b) simulating prospective external components from within the dialogue templates as a temporary placeholder until the component exists.
3. Flipper supports choosing along the spectrum between
   (a) robust, scalable and well-defined declarative models of dialogues; and
   (b) pragmatic "hacking-stuff-together" and "wizarding" to try out the effects of certain dialogue paradigms before actually modeling them properly.

   This can help early in the project to show how a dialogue with the ECA will emerge. It also helps to make decisions that are informed by the reality of the distributed ECA technology that is available or that will be developed. Early demonstrations demand a pragmatic approach while at some point the pragmatic developments and the lessons learned need to be consolidated into an ECA system that is robust and scalable.
4. Information in Flipper can be stored in a persistent database which enables, for instance, a robust consistency between interactions over time.
5. Flipper can process information from input sensors in parallel, handle decision-making, and create and send output, making the ECA continuously responsive in a dialogue.
6. Flipper can communicate with external components; it currently supports eight middleware communication platforms, and it is easy to add other methods of communication.
7. Finally, over the course of several national and international research projects we have created a set of *design patterns*. In these design patterns we show

[1] `https://semaine-db.eu/`

how we solve in a robust and scalable way the typical situations and technical problems that occur when creating a dialogue system. We will make these available together with the software and highlight some in this chapter.

In Section 4.2 we explain our view on dialogue systems and discuss related work on dialogue management and designing dialogues. In Section 4.3 we discuss the technical details of Flipper. In Section 4.4 we show some examples and design patterns of using the dialogue engine. In Section 4.5 we point at some work that has been done with Flipper in ongoing and earlier research projects. Finally, we discuss the current limitations and future development of the dialogue engine and present our conclusions in Section 4.6.

4.2 Background

ECAs consist of multiple technical components that can be roughly divided into three *pillars of tasks*: sense, think, and act. In an interaction, *sensing* components are tasked with processing and interpreting the human's language and social signal behavior. For example the user's mouth corners move up, meaning a smile. This information is used by the agent to *think* about the behavior of the user in order to decide what is an appropriate response in the context. For instance the user liked the joke the agent just told them and the agent laughs with the user to create rapport. This response behavior is displayed, *acted*, by the embodiment of the ECA. Each component in each pillar has a distinct task that it performs in order for an ECA to function in a social interaction. In Section 2.2.1 we already discussed dialogue frameworks with respect to input modalities (sense) and output modalities (act). In this background we address the think component of dialogue frameworks.

We distinguish within the *thinking* component of an ECA a division of three parts: a dialogue engine, a dialogue manager and dialogues. The *dialogue manager* is the part of an ECA that deals with how the agent behaves in an interaction. It is a collection of rules that control the flow and state of the conversation (Larsson and Traum, 2000). It does so in response to the input of the user, and the goals and beliefs of the ECA. Dialogue managers are dependent on the domain knowledge your agent requires, the modalities you want to use and the goal of the conversation. Depending on the ECA developer's goal, one or another dialogue approach could be more appropriate. The *dialogue engine* is the machinery with which it is possible to create a dialogue manager. This can be done in a regular programming language, or in a system that interprets declarative dialogue specifications to control a dialogue, or a mixture of the two. Designers of *dialogues* are then required to write content (a dialogue structure within the domain that the agent knows about and can converse about and that contains all the behaviors that the agent can decide to do) and add this to the dialogue system. Together, the

dialogue manager

dialogue engine

dialogues

dialogue engine, dialogue manager and dialogues make up the complete dialogue system.

Choosing a tool to develop dialogues for an ECA has a great impact on the type of interaction. An overview of different tools that are currently available is provided in Table 4.1.[2] We review seven aspects of each dialogue design tool: i) information processing, ii) the interface to design dialogues, iii) the support for linking an embodied agent, iv) the design paradigm, v) how dialogue control is organized, vi) the support for different types of interaction management and vii) the inclusion of design patterns for designing dialogues.

information processing

Information processing is the way in which the context of the dialogue is stored and processed and is important for knowing what the tool is capable of using during a dialogue. Examples are the capabilities to process probabilities, events or plans in the context. In most tools this is captured in either states or a network. A state-based approach is easier to interpret and to author dialogues for than a network-based approach. However, if there is much training data available, a network would be very convenient to capture all relevant information without explicitly stating what is relevant. Flipper uses information-state update rules, similar to the approaches in TrindiKit (Larsson and Traum, 2000) and the VHToolkit (Hartholt et al., 2013). The information state update approach is useful for keeping control of the dialogue flow without declaring all possible dialogue states. Commercial cloud-based services such as LUIS.AI, Wit.ai, DialogFlow, Watson, Lex and SAP Conversational AI all use neural networks for processing the information,[3] which is a useful approach for learning from large datasets containing text or conversations (Braun et al., 2017; Canonico and Russis, 2018). Similar open-source neural network approaches are RASA (Bocklisch et al., 2017) and ConvLab-2 (Zhu et al., 2020).

interface of authoring

Interface of Authoring is the process of designers creating interactions for their ECA. The accessibility of authoring is important for designers to use your tool. The VHToolkit, with the NPCEditor (Leuski and Traum, 2010), Visual SceneMaker (Gebhard et al., 2012) and HALEF (Ramanarayanan et al., 2015), with OpenVoiceXML, provide a graphical user interface for editing the dialogue. Other tools, such as Flipper, IrisTK (Skantze and Al Moubayed, 2012) and OpenDial (Lison and Kennington, 2016) use a declarative way of defining the dialogue in XML. In the commercial cloud-services, designers can use a web-interface to author dialogues, in which they provide user input and an appropriate agent response, marking the intents and entities in the utterances. The agent then learns from the marked examples to give the best response given recognized intents and entities. In the VHToolkit (Hartholt et al., 2013), WAMI (Gruenstein et al., 2008) and

[2] Commercial tools include DialogFlow, Wit.ai, LUIS.ai, Watson, Lex, SAP Conversational AI and RASA

[3] `luis.ai`, `wit.ai`, `dialogflow.com`, `ibm.com/watson/`, `aws.amazon.com/lex/` and `cai.tools.sap/`

Flipper scripting is also possible for less restricted authoring. ADvISER (Li et al., 2020) and ConvLab-2 (Zhu et al., 2020) support user simulations via a GUI to evaluate authored dialogues and diagnose possible issues.

embodiment

Embodiment support is an ECA's capability to perform both verbal and non-verbal behaviors. It is a necessity for developing an ECA and most of the tools support it. IrisTK (Skantze and Al Moubayed, 2012), Visual SceneMaker (Gebhard et al., 2012), the VHToolkit (Hartholt et al., 2013) and ReTiCo/rrSDS (Kennington et al., 2020) include an embodiment. Others, such as RavenClaw (Bohus and Rudnicky, 2009), Disco (Rich and Sidner, 2012), OpenDial (Lison and Kennington, 2016) and Flipper have interfaces available for embodiment. The commercial tools are harder to link to an embodiment, due to the text-only intent-entity marking for agent responses. HALEF (Ramanarayanan et al., 2015) is less suitable for embodiment, due to its focus on telephone-conversation.

design paradigms

Developing dialogues can be done via a bottom-up pragmatic approach, a more theory-driven robust manner or a mixed approach, which are the *design paradigms*. PyDial (Ultes et al., 2017), RASA (Bocklisch et al., 2017), ConvLab-2 (Zhu et al., 2020) and the commercial tools such as DialogFlow are more on the pragmatic side of the design paradigm scale, for quickly developing content with conversational data. Tools such as RavenClaw (Bohus and Rudnicky, 2009) and Disco (Rich and Sidner, 2012) require a theory-driven approach due to their hierarchical way of processing information and are not capable of using conversational data directly for development. The emphasis of Flipper is on using a pragmatic approach when starting to develop dialogues, though for more complex dialogues theory-driven development is also possible, similar to the design paradigm in OpenDial (Lison and Kennington, 2016).

4

dialogue control

The *dialogue control* can be either single or distributed (multiagent) (Cheyer and Martin, 2001). In IrisTK (Skantze and Al Moubayed, 2012) a single component is responsible for the dialogue flow, maintaining transparency of changes in the dialogue state. In complex dialogues a single component for dialogue control can be a bottleneck. RavenClaw (Bohus and Rudnicky, 2009), the VHToolkit (Hartholt et al., 2013), ReTiCo/rrSDS (Michael, 2020) and Flipper are capable of distributed control, using separate components, for example, for backchanneling and deliberate conversation.

interaction management

During a conversation with an ECA, turn-taking and backchanneling are important for a coherent conversation; this is called *interaction management*. The neural network based tools only support rigid turn-by-turn dialogues; there is no managing of other turn behavior like pauses or interruptions. Flipper has a structure similar to RavenClaw (Bohus and Rudnicky, 2009), IrisTK (Skantze and Al Moubayed, 2012) and ReTiCo/rrSDS (Michael, 2020) to support both simple turn-by-turn behavior and more dynamic turn-taking for incremental interactions.

Most dialogue design tools provide a description of their tool and simple

design patterns

examples to run the software. However, an underestimated aspect is how precisely to design the dialogues: which *design patterns* a designer of dialogues could use. IrisTK (Skantze and Al Moubayed, 2012) and Visual SceneMaker (Gebhard et al., 2012) do provide dialogue flow patterns for dialogue designers but design patterns on the higher level dealing with sensory input or behaviors are not provided. Design patterns help dialogue designers with fast decision-making of prototyping their ECA. In this chapter we describe multiple types of design patterns that are helpful in developing dialogues in Flipper.

4.3 Flipper

Flipper, as mentioned in Chapter 3, is a dialogue engine for pragmatic yet robust dialogue management that is applicable in many domains, and has reusable design patterns. Designers of ECAs can use the dialogue engine to quickly create dialogue systems that can be as complex as they like. The software is open-source and available on GitHub.[4]

4.3.1 Architecture

The main concepts in Flipper are the information state and declarative templates written in XML. The information state can be predefined, created at runtime, and/or updated on-the-go. It stores interaction-related information and data in a hierarchical tree-based structure. The information state is represented in JSON format, which is human-readable and easy to integrate with other dialogue components that support working with JSON data structures. Listing 4.1 shows an example information state. Nodes in the information state can be accessed in Flipper by navigating the tree-based data structure using dot notation. For example, the user's name can be accessed through `is.user.name` in Listing 4.1. Information from a dialogue that can be included are, but not limited to, dialogue history, emotional levels and topics. Flipper can be linked to a PostgreSQL database to create a persistent information state. This means that the information state can be restored to a previous valid information state that exists in the database. Such a persistent information state can be used, for example, to track interactions with a user over multiple sessions.

templates

preconditions and effects

The data structure stored in the information state is queried and updated using *templates*. Templates can be grouped and organized in different files according to their related functionality. Each template consists of *preconditions and effects*. Preconditions are sets of rules that describe when a template should be executed. Effects are the associated updates to the information state. Listing 4.2 shows an example template that checks whether a user is present. If so, the user is personally

[4] `github.com/hmi-utwente/flipper-2.0`

Table 4.1: An overview of different dialogue design tools that shows which architecture is used for information processing, the authoring method for dialogues, the embodiment support, the design paradigm, the possibility for interaction management and whether the tool supplies design patterns.

Tool	Information Processing	Authoring Interface	Embodiment	Design Paradigm	Dialogue Control	Interaction Management	Design Patterns
Flipper 2.0	Information state update	Rules, scripting	Yes	Pragmatic or theory-driven	Single or distributed	Yes	Yes
Commercial tools	Neural network	Web-interface for intent-entity mapping	No	Pragmatic	Single	No	No
RavenClaw	Hierarchical plan based	Dialogue task specification	Yes	Theory-driven	Single or distributed	Yes	No
Disco	Hierarchical task based	Hierarchical tree authoring	Yes	Theory-driven	Single	Yes	No
VHToolkit	Information state update	NPC Editor, FLoReS, scripting	Yes	Pragmatic	Single or distributed	Yes	No
IrisTK	Statecharts	Statecharts in IrisFlow	Yes	Pragmatic	Single	Yes	Some
PyDial	Statistical network	Ontologies, user simulation	Yes	Pragmatic	Single	No	No
OpenDial	Probabilistic information state update	Probabilistic rules	Yes	Pragmatic or theory-driven	Single	Yes	Some
HALEF	Any	VoiceXML	No	Pragmatic	Distributed	Yes	No
TrindiKit	Information state update	Formal declaration	Yes	Theory-driven	Single or distributed	Yes	Some
WAMI	Frame-based	Scripting	Yes	Pragmatic	Single	Yes	No
Visual SceneMaker	Statecharts	Sceneflow and Scenescript	Yes	Pragmatic	Single or distributed	Yes	Some
ADvISER	Reinforcement learning	Scripting and user simulation	No	Pragmatic	Single	No	No
ReTiCo/rrSDS	Statecharts	Statecharts and scripting	Yes	Pragmatic	Single or distributed	Yes	No
ConvLab-2	Neural network	Scripting and system evaluation	No	Pragmatic	Single	No	No

```
{ "is" : {
    "user" : {
        "name" : "Alan",
        "speech" : "hello what can you do",
        "emotion" : "happy"},
    "history": {
        "greetByAgent" : false,
        "greetByUser" : false}}}
```


Listing 4.1: An example information state that stores the agent's knowledge of the interaction. The data structure's top-level root node is has a child node user which stores information such as the name of the user, the last recognized user utterance and the current user emotion. Additionally, events, such as greeting intents, are kept track of in the dialogue history.

```
<template id="hello_world">
  <preconditions>
    <condition>is.user.present</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.say">"Hello "+is.user.name+"!
    Nice to meet you!"</assign>
  </effects>
</template>
```

Listing 4.2: Example template where the agent greets a user if they are present.

greeted. Using the information state from Listing 4.1, this template will result in the agent saying the following greeting: "Hello, Alan! Nice to meet you!"

4.3.2 Implementation

Preconditions and effects are evaluated using the GraalJS JavaScript Engine, which supports up to ECMAScript 2020. In Flipper, JavaScript expressions and functions can be used as an imperative addition to the declarative template approach. Finally, Flipper exposes Java objects to be used within templates for further integration with existing (external) software modules. We have created an example project with Java objects to demonstrate how to integrate for example NLU or TTS components and database handling for long-term interaction.

4.3.3 Transaction Model

transaction model

The dialogue engine uses a *transaction model* to ensure reliability. According to Gray and Reuter (1992, p. 6), a transaction is *[...] a collection of operations on the physical and abstract application state*. In Flipper, the check of the preconditions in all templates and execution of their associated effects is considered as one transaction. A transaction is complete when it is successfully committed to a database. A transaction has the following properties (adapted from Gray and Reuter (1992)):

1. Atomicity: information states are atomic, the entire update is applied or nothing changes.
2. Consistency: any update on the information state cannot render the information state invalid.
3. Isolation: though calls for information state updates could in practice occur at the same time, they are executed sequentially, and only one update can happen at the time, to preserve consistency.
4. Durability: once an update is completed successfully, this is reflected in the information state.

In each transaction, the conditions of all templates are checked on a frozen information state. The effects of the templates that are true are executed consecutively. If all effects are executed successfully, the updated information state is committed to the database. If one of the effects fails, all processed effects in the current transaction are rolled back and the information state is restored to the previous state, which is retrieved from the database. This is beneficial for incremental interactions, because it ensures that asynchronous multimodal input and output do not break the entire dialogue flow. Dialogue designers can afterwards see which transactions failed and diagnose problems with their dialogue. Template checking occurs in recurring intervals. A limit can be set on the frequency with which templates are checked. For example, with a frequency of 20 Hz all templates are checked once every 50 ms. Setting a higher frequency may result in a more responsive system, while setting a lower frequency leads to a lower system load.

4.4 Creation of a Dialogue Manager

The first important thing to think about when designing dialogues is the information flow of the dialogue. What type of information is needed from the user and when? What type of information is required for the agent? What should the agent do and when? Which behaviors need to be displayed and when? Here we explain how to create a dialogue system with Flipper and showcase some design patterns using the *sense, think, act* metaphor.

4.4.1 Sensing

sense An interactive ECA needs *sensory input* from the user, for example as in Section 3.3. This information needs to be put into the information state so that concurrent processes can use it. Flipper itself does not contain sensing components, but an example project and external projects that include sensing components are available for download (see Section 4.5).

To receive sensory input from auxiliary devices or software modules we have developed a *middleware* component. This component is a wrapper around existing off-the-shelf messaging and communication services. Currently Flipper supports wrappers for ActiveMQ, ROS, YARP, Apollo/STOMP, TCP/IP, UDP, REST, and USB. Our middleware component listens to messages on a supported communication channel and then places them in the information state. When such messages are received in JSON format they can directly be stored in the information state; otherwise the message has to be preprocessed into a JSON format first.

Once the sensory information has been placed in the information state, it has to be processed to determine the impact on the dialogue flow. To prevent templates from processing the same sensor information twice accidentally, we suggest the following design patterns for dealing with sensory input in Flipper.

As a first simple approach, each template could be required to have an effect that negates its own precondition, such as in Listing 4.3, where the parameter `is.agent.userExpressionEvent` is set from `smile` to `none`. A template could remove the sensor input from the information state once it has processed it. Although this is a pragmatic and quick solution it is not a scalable approach for the long term. Also, it results in verbose templates.

```
<template id="soc_respond_to_smile" conditional="true">
  <preconditions>
    <condition>is.user.events.userExpression === "smile"
    </condition>
  </preconditions>
  <effects>
    <assign is="is.agent.fml.template">"smile_return"</assign>
    <assign is="is.user.events.userExpression">"none"</assign>
  </effects>
</template>
```

Listing 4.3: Example of a template that removes input once it has been processed.

When the impact of a new sensor value should be more multi-faceted, an author could construct a template file with a collection of templates that first dump the raw input in a temporary information state variable and then successively

```
<template id="1">
  <preconditions>
    <condition>is.user.events.userExpression === "smile"</condition>
    <condition>is.user.event.userGesture === "waving"</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.fml.template">"smile_and_wave_return"
    </assign>
  </effects>
</template>
<!-- Many templates could reside here, each triggering on a combi-
nation of is.user.events.userExpression and other preconditions-->
<template id="x">
  <preconditions>
    <condition>is.user.events.userExpression === "smile"</condition>
    <condition>is.weather.current === "sunny"</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.speak">"Beautiful day today!"</assign>
  </effects>
</template>
<template id="last">
  <preconditions>
    <condition>is.user.events.userExpression !== "neutral"</condition>
  </preconditions>
  <effects>
    <assign is="is.user.events.userExpression">"neutral"</assign>
  </effects>
</template>
```


Listing 4.4: Quick design pattern for dealing with a sensory input event. The top templates are triggered by a user's detected facial expression and other sensor information. The last template "cleans up" the sensory input to make sure actions based on such sensory input are only processed once.

process the input. Separating the multiple effects of the new sensor input into multiple templates keeps the templates relatively clean and readable. The execution order of templates is always defined by the order of templates in the template file. A final template can do a cleanup of the raw sensory input once the other templates have finished. See Listing 4.4 for an example of such a template file. However, this solution is useful only when developing small behaviors, because with multiple template files it is hard to know which template is executed last.

Another design pattern for dealing with sensory input is to keep track of a history of sensory input and check against a time or sensor value index whether

```
<template id="add new">
  <preconditions>
    <condition>isNew(is.user.emotion)</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.history.emotions">
     addToArray(is.agent.history.emotions, is.user.emotion)
    </assign>
  </effects>
</template>

<template id="remove old">
  <preconditions>
    <condition>isFull(is.user.emotion)</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.history.emotions">
    removeHeadArray(is.agent.history.emotions)
    </assign>
  </effects>
</template>
```


Listing 4.5: Complex design pattern for dealing with input.

the input has been processed already. This can be done by either keeping track of an index or a timestamp. The downside of this approach is that it creates more overhead (more memory consumption) and is more complex to implement than the other two pragmatic approaches. However, for robust and scalable systems where dialogue designers cannot be sure which other template sets might have access to the same information, this last approach is a necessity. Listing 4.5 shows an example of templates dealing with sensory input in this way. We emphasize that each of these solutions can be applicable in a specific case and that each pattern is a good approach for working with Flipper depending, among other things, on which stage of development the dialogue system is in.

4.4.2 Thinking

think Information from the input can be used by the agent to *think* about it in order to determine an appropriate response in the current dialogue context. This is done in what we call *dialogue behavior templates*. An example of a dialogue behavior template is shown in Listing 4.6.

We encourage designers to make a distinction between high-level and low-level interaction templates. This recommendation follows Lemon et al. (2003), who

```
<template id="soc_sal_returnsalutation" conditional="true">
  <preconditions>
    <condition>!is.agent.history.greetByAgent</condition>
    <condition>is.user.emotion === "happy"</condition>
    <condition>containsKeyword(is.user.speech,
    ["hello","hi"])</condition>
  </preconditions>
  <effects>
    <assign is="is.agent.fml.template">"social_salutation_return"
    </assign>
    <assign is="is.agent.fml.parameters['var.name']">
    is.agent.userName</assign>
    <assign is="is.agent.fml.parameters['emotion.em1']">
    is.agent.userEmotion</assign>
    <assign is="is.history.greetByAgent">true</assign>
  </effects>
</template>
```

Listing 4.6: Example of a template returning a user's happy greeting. This template covers the situation when a user has not previously been greeted by the agent, the user is currently happy, and the user has said "hello" or "hi". In this case the agent should return the greeting with a friendly face, including the user's name.

describe their dialogue design approach as creating high-level dialogue (content) moves, but also handling low-level (management) phenomena like turn-taking, back channelling, and grounding. Turn-taking for example can be done by a state-machine which regulates turns based on current speech activity of the user and agent. By using this conceptual division between content and management templates —a design distinction only; Flipper does not register a formal distinction between the two— some management templates can be reused in different ECAs and different projects. For example, components that contain low-level information state updates for turn-taking and back channelling are applicable in multiple domains and can be used in each agent that requires it, whereas high-level content templates are often not reusable as they contain domain-specific content.

4.4.3 Acting

We have also developed modules for Flipper that use our middleware component to communicate with the behavior realizer of an embodied conversational agent for *acting*, such as described in Section 3.4. These modules can send both Behavioral **act** Markup Language (BML) and Functional Markup Language (FML) (Vilhjálmsson et al., 2007), the latter we mentioned before in Section 3.5.2. Inside Flipper,

```
<template id="behaviour">
  <preconditions>
    <condition>is.agent.behaviours.length !== 0</condition>
  </preconditions>
  <effects>
    <behaviour name="executeBehaviour">
      <object class="behaviorRealiser"
      persistent="behaviorRealiser"/>
      <arguments>
        <value class="String" constant="<bml id='bml1' xmlns='http://
        www.bml-initiative.org/bml/bml-1.0' character='Alice'><gaze
        id='gaze1' target='PERSON1'/></bml>"/>
      </arguments>
    </behaviour>
  </effects>
</template>
```

Listing 4.7: Template sending a String message in BML format with gaze behavior to a behavior realizer.

the parameters for the behaviors need to be determined and set accordingly in a valid BML or FML representation, depending on the behavior realizer. If no embodiment is available, a valid SSML representation is also sufficient for producing speech.

A pragmatic way to deal with agent behaviors is to specify BML or FML strings directly inside templates. Listing 4.7 shows an example of this. The `behaviorRealizer` is a Java module specifically designed for sending BML and FML behaviors of the agent via our middleware component to an external behavior realizer.

An alternative approach is to create a list of BML or FML behaviors. These behaviors can be loaded in the dialogue system from the file system. Dialogue designers can use existing BML behaviors accompanying the Flipper software or create their own. Additionally, these behaviors can be parameterized, and the parameters can be filled using the information from the information state (see Cafaro et al. (2017a)). In Listing 4.6 the assignments of `is.agent.fml.parameters` include setting the name of the user interacting with the system and the emotion of the user. Once the parameters are set, the behavior can be sent through our middleware component to a (BML or FML compliant) behavior realizer, as shown in Listing 4.8.

```
<template id="executeBehaviour">
  <preconditions>
    <condition>is.agent.fml !== ""</condition>
  </preconditions>
  <effects>
    <behaviour name="executeBehaviour">
      <object class="behaviourRealiser"
      persistent="behaviourRealiser"/>
      <arguments>
        <value class="String" is="is.agent.fml"
        is_type="JSONString"/>
      </arguments>
    </behaviour>
    <assign is="is.agent.fml">""</assign>
  </effects>
</template>
```

Listing 4.8: Template that takes an FML request plus its parameters and passes it to the `behaviorRealizer` module for execution.

```
function containsKeyword(utterance, keywords){
  var word;
  list = utterance.split(" ");
  for(word in list){
    var key;
    for(key in keywords){
      if(keywords[key] === list[word]){
      return true;}}}
  return false;}
```

Listing 4.9: Example JavaScript function, one that checks for keywords in a user utterance.

4

4.4.4 Advanced Dialogue Behavior

Dialogue designers might require extra functionality in the design of their dialogues. As the dialogue engine evaluates the templates with JavaScript, it is easy to add existing JavaScript libraries or JavaScript code to perform logic that is cumbersome to express in (declarative) templates. One example applicable to an embodied conversational agent is a function that checks certain keywords in an utterance (shown in Listing 4.9). Other useful JavaScript functionalities are behavioral generators, calculating the appropriate intensity of an emotion of an agent, and timers necessary to know when to perform certain behaviors.

When JavaScript is not expressive enough or when the JavaScript becomes too large to author or maintain, Java classes can be instantiated and integrated in the dialogue engine. Complex functions can be delegated to Java objects that have been created from within the Flipper template collection. This further extends the capabilities and flexibility of Flipper. Examples of useful Java modules are the CoreNLP for natural language understanding (Manning et al., 2014) and BML translators such as ASAP for behavior generation (Van Welbergen et al., 2014).

Connecting with non-Java external components can be done by sending messages across a middleware channel, as mentioned earlier. Information can be exchanged between such external components and dialogues, and retrieved from or stored in the information state. For instance, external reasoners, knowledge bases or natural language generators can interface with Flipper via the supplied middleware and use a separate template file for the handling of their input and output to keep the system modular and reusable. Additionally, this connectivity includes external components that are not "traditional" embodiments for conversational agents, for example a tablet that displays information or an external device (e.g., a coffee machine) that is started automatically when the user requests this in the dialogue with the ECA.

4.5 Applications

In this section we discuss some of the projects that have used Flipper in developing their agent(s). The projects use different types of input and output modalities that are connected to the dialogue engine.

In the previous chapter we already discussed the ARIA as a multimodal information-providing ECA that was developed with the ARIA-VALUSPA (Valstar et al., 2016). Users can ask the agent questions about a specific domain and the agent tells stories to the users. The agent also includes an emotional model that determines whether the agent likes or dislikes the user, based on the user's (non-) verbal responses. For example, turn-management, behavior generation and the emotional model are handled by Flipper templates, whereas external components are used for natural language understanding and (non-) verbal behavior realization. In Chapters 5, 6 and 7 we describe projects in which Flipper serves as a dialogue engine for speech-only oriented dialogue systems.

In another project involving multiagent parties, an external module for synchronization of behaviors was developed and integrated with Flipper templates to create social gaze behavior based on saliency (Kolkmeier et al., 2017a). Saliency indicates what is important during interactions; for example, most gazes will be directed towards the speaker in the current interaction (Ruhland et al., 2015).

In the Council of Coaches project, a platform specifically for multiagent setups has been developed (Op den Akker et al., 2018). This platform is called Agents United (Beinema et al., 2021). The project is oriented towards health coaches, in which users could talk to multiple agents at the same time, each with their expertise in health of food, psychology and exercise. Flipper is used as the conversational intent planner for the agents.

The Snoozle project aimed at helping people sleep using an interactive pillow that lured people to bed. Flipper was used to steer the multimodal behavior of the pillow (Vroon et al., 2017). This is an example where a non-humanoid agent is controlled by Flipper.

Some proposed design patterns result from lessons learned in projects using the first version of Flipper (Ter Maat and Heylen, 2011). In the R3D3 project, Flipper was used for steering the turn-taking and emotive behavior of a receptionist robot combined with a virtual agent (Theune et al., 2017). In the DE-ENIGMA project, involved with child-robot interaction, Flipper was used to control the behavior (speech, facial expressions and gestures) of an emotionally expressive robot. Additionally, modules were developed for a dialogue logger and a dialogue tree within Flipper (Chevalier et al., 2017). Finally, in another child-robot interaction project called EASEL, Flipper was used to control actuated physical learning materials as well as a tablet displaying the GUI of an educational game (Reidsma et al., 2016).

4.6 Discussion and Conclusion

Flipper 2.0 is under active development in the context of several European research projects. We have created a debugging functionality that will give insight into exactly what state the dialogue is in and when certain information state updates will be applied. Additionally, we want to extend our Flipper 2.0 example with many more basic modules for an operational ECA to have a prototype system ready out of the box. For example, we would like to add more possible embodiments, such as social robots as the NAO (Gouaillier et al., 2009) and Pepper (Pandey and Gelin, 2018).

Flipper currently has no graphical user interface for editing the required templates, which the VHToolkit and Visual SceneMaker have. An editor for modifying template files in a tree-like structure would benefit less technically apt dialogue designers. Nevertheless, we see Flipper as a more abstract dialogue management system, connected to external components that each have their own authoring interfaces that help dialogue designers in prototyping a specific component, such as the user simulations in ADvISER (Li et al., 2020).

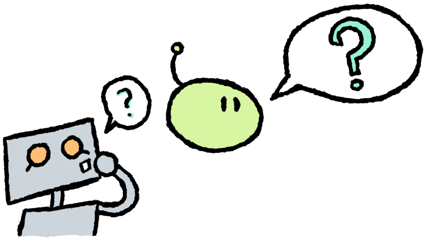
One might think that the rule-based approach used in Flipper is too simple for designing a dialogue system. However, we see machine learning possibilities for the dialogue engine as well. Speech and text oriented approaches using machine learning often require the collection of conversations or the authoring of input/output behavior. Similarly for Flipper, an author could collect information state mappings between sensory input and user behaviors that map to certain agent behaviors, or author information states and behaviors to let a computer learn the most appropriate agent behavior over time. Another option is to connect external machine learning models through our middleware support for specific low-level management tasks such as turn-taking and use this information in high-level templates.

Scalability might be a problem if many update rules need to be integrated. However, we view the reusability of modules as one particular case of scalability for designing dialogues. In the case where an author designs a dialogue system that needs large amounts of data and is open-domain, we suggest to use Flipper for low-level interaction dialogue strategies in combination with, for instance, a cloud-based commercial tool or RASA (Bocklisch et al., 2017) or PyDial (Ultes et al., 2017).

Though between templates the transaction model applies and no dialogue can fail, effects within a template are dependent on order and no full transaction model is applicable. We have a suggestion to deal with this issue, by using two parallel information states, one for writing and one for reading. Still, this leaves the possibility that updates of the information state in an effect block can overwrite each other. A usable tactic for now would be to recommend the user not to create assignments to the same information state in the same effect block but use different templates for that purpose.

The largest advantage cloud-based commercial tools have over Flipper is their scalability. However, these tools lack flexibility for integrating multimodal components for both input and output. In future work we hope to develop a cloud-based Flipper, which could support many users interacting with an ECA at the same time, without being limited to specific smart devices.

We have provided some insight into the development process of dialogues for embodied conversational agents (ECAs) in complex projects, and have presented Flipper 2.0: a tool that makes it easy to quickly and iteratively create dialogues for ECAs, meeting the demands of such projects. We are still developing more design patterns and more features for working with Flipper, such as a supporting more types of embodiment. This tool is particularly useful for people creating dialogues who need to get started quickly, with workable and pragmatic dialogue patterns, yet need to have the possibility to extend their efforts into a complex, multi-faceted, responsive, multimodal dialogue system.



5

BLISS: Question-Asking for Eliciting Self-disclosure in Mental Well-being

This chapter is based on:

- **J. van Waterschoot**, I. Hendrickx, A. Khan, E. Klabbers, M. de Korte, H. Strik, C. Cucchiarini, and M. Theune (2020a). "BLISS: An Agent for Collecting Spoken Dialogue Data about Health and Well-Being". In: *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 449–458

This chapter focuses on building a prototype and testing a social agent in the wild, given the tools and design guidelines introduced in Chapter 3 and Chapter 4. We focus on Dutch spoken dyadic conversations about people's mental well-being. The goal of the social agent is to elicit self-disclosure from end-users and maintain engagement with the user through multiple conversations. The architecture of the social agent uses these modalities from the ARIA from Chapter 3: text and speech processing.

5.1 Introduction

Recent projections show that in the near future the health sector will deal with a growing demand for healthcare, an increasing number of vacancies, and higher expenditures (UHL, 2016; Raad, 2020). Among others, this has led to a paradigm shift in healthcare that emphasizes prevention, citizen empowerment and self-management and in which citizens are increasingly required to assume an independent, self-determining position. Along with these changes, there has been a critical analysis of the current definition of health adopted by the World Health

Organization (WHO) that describes health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”

Huber et al. (2011) discuss the shortcomings of this definition and suggest an alternative definition of health which is defined as “the ability to adapt and to self manage”. The view of *health* adopted in this paper is in line with Huber et al. (2011) and with a Positive Psychology view (Seligman, 2002; Seligman, 2012) in which positive experiences play a central role. In Dutch healthcare systems, this view on positive health and happiness has been widely embraced. Caretakers are trained to focus on the broad definition of health, including physical, mental and social well-being, and more holistic topics such as quality of life and self-management (Ministerie van Volksgezondheid, 2016).

health

This new definition of health requires operationalizations and appropriate instruments for measuring positive health dimensions such as functional status, quality of life and sense of well-being (Huber et al., 2011). Professionals attempt to gain insight into these dimensions through questionnaires and interviews with people who receive long-term or structural healthcare. This leads to insights, assessments and opportunities for positive health for the clients and caretakers. In addition, in-depth qualitative interviews identify opportunities for happiness improvements.

In-depth interviews provide the most insights, but require a serious time investment, both for the actual interview and the analysis and reporting. The contents of this chapter are couched in a larger project, Behaviour-based Language-Interactive Speaking Systems (BLISS), that attempts to offer a solution by developing an intelligent, personalized system that communicates with clients in spoken language to facilitate their self/joint-management of health, wellness and psychological well-being, while measuring them and providing insights about it at the same time. Razavi et al. (2019) developed a similar system and found that communication skills of older adults could be improved through such a system.

In this chapter we report on the first steps undertaken to develop the BLISS social agent for self-management of health, wellness and psychological well-being, in particular the initial phase of data collection. We started with available language resources for the Dutch language to develop a first version of the system that could be used to collect initial data.

5.2 Background

In the early 1990s, the USA’s DARPA launched the Airline Travel Information System (ATIS) project (Price, 1990), which sparked spoken dialogue system (SDS) research. The first SDS for the Dutch language was the Public Transport Information System (OVIS) (Strik et al., 1997), which was a train timetable information system. The spoken language generation part of OVIS consisted of a template-based language generation module linked to a speech synthesis system (Theune et al., 1997; Theune et al., 2001). The OVIS system was developed using a bootstrapping method. As a follow-up to OVIS, the Interactive Multimodal Information eXtraction (IMIX) project (Van den Bosch and Bouma, 2011) developed a multimodal question answering system for Dutch, combining speech and visual modalities. One of its use cases was answering questions about repetitive strain injury; however, this was only for demonstration purposes.

A look at the healthcare applications employing spoken dialogue systems reveals that they have been developed mainly for specific domains such as breast cancer screening (Beveridge and Fox, 2006) or military mental healthcare (Morbini et al., 2012). Persuasive technology has focused mostly on healthcare as well, by supporting people in speedy recovery and taking up exercise or taking medication (Meschtscherjakov et al., 2016). The Council of Coaches (COUCH) project designed a system with multiple virtual agents to provide support for users who have for example diabetes or COPD (Op den Akker et al., 2018).

Chatbots are becoming more popular to offer 24-hour customer support, and we can also see this trend in (Dutch) healthcare. For example, Chantal[1] and Bibi[2] are both virtual general practitioner assistants who can chat in Dutch (written communication) about healthcare issues and practical questions such as making an appointment to speak with the GP. Also personal assistants such as Anne,[3] and robots like Tessa,[4] Alice[5] or Zora,[6] have been put into elderly homes to help older adults (Martinez-Martin and del Pobil, 2018; Burger, 2015; Kardol, 2015). De Graaf et al. (2019) mention that social skills for an agent might not be an effective method to motivate a client to perform a task. In their study, a social robot was put in participants' home to motivate them for undertaking physical activity. The robot's social behavior negatively impacted the users' perception, because it was found to be disruptive to their routines. The authors found that it also takes up to two months for end-users to be accepting of a technology such as a social robot. Many other health applications exist, but these are outside the scope of this chapter. We recommend reading the surveys of Montenegro et al. (2019) and Jaber and McMillan (2020) for the latest developments on conversational agents and SDSs for health.

It is important to prevent the agent from misinforming clients, which is a risk for agents that take spoken input, due to automatic speech recognition errors. Therefore instead of dealing with free speech as input, Bickmore and Picard (2005) suggest using a menu of options or limited text input, to both make the dialogue

[1] https://zaurus.nl/chantal/
[2] http://virtueledoktersassistent.nl/
[3] https://anne4care.nl/
[4] https://www.tinybots.nl/
[5] https://ikbenalice.nl
[6] https://zorarobotics.be/

5

smoother and prevent the system from making crucial errors such as giving users the wrong answer to their questions, because of mishearing the user. Especially in health applications where a high intent accuracy is required, often no free speech is used (Bickmore and Giorgino, 2006). Similarly, the virtual agents from COUCH have speech as output, but the users usually interact with them using input selected from a menu, though there is support for ASR as well (Bosdriesz, 2020).

All the aforementioned social agents are designed to answer domain-specific user questions. These systems can use a structured database for answering questions or natural language processing to extract answers from snippets of text from an unstructured dataset, be information retrieval based or a combination of any (Kolomiyets and Moens, 2011; Calijorne Soares and Parreiras, 2020). In information retrieval, the answers to the user's questions are largely extracted from unstructured documents, which provides more flexibility compared to using a structured database in a knowledge-base approach. More recently, deep learning has become a popular method for a question-answering SDS. In deep learning, the answers the system provides can be generated instead of being directly retrieved from documents (Qu et al., 2019).

Our focus in BLISS is on long-term interaction, asking engaging questions (instead of answering questions) and learning a user happiness model of mental well-being through normal spoken conversation. ELIZA, one of the first chatbots, was rule-based and designed as a therapeutic chatbot that could ask questions to users (Weizenbaum, 1966). Users who talked to ELIZA disclosed personal information and experienced high engagement. Conversations with ELIZA are very different from how people interact with smart devices nowadays. As noted by Radlinski et al. (2019), communication with smart devices is often very command-like in style and does not feel natural to participants. The authors set up a Wizard of Oz (WOz) experiment to collect a dataset of more natural spoken conversations in the context of movie recommendation. They found that these conversations contain far more complex information than what smart devices are capable of understanding now. Retrieving actor names and genres is doable for a virtual assistant to find recommendations, but it is harder to recommend movies if users mention they liked the narrative of a particular movie. Additionally, if users speak disfluently, such as with repeating words, this is harder to interpret for a smart device.

Similarly, with BLISS we aim for a natural conversation between a social agent and users. Specifically for obtaining natural conversation data about personal topics, Zhang et al. (2018) collected PERSONA-CHAT, a dataset containing dyadic text-based chitchat recruited via crowdsourcing. The authors designed personas through crowdsource workers, instructing them to write short descriptive personas, similar to Zhang et al. (2018, p. 2206). Such a description is shown in Example 5.1.

(5.1) I am a vegetarian. I like swimming. My father used to work for Ford. My favorite band is Maroon 5. I got a new job last month, which is about advertising design.

Afterwards, these personas were used for other crowdsource workers to role-play in a dyadic conversation. The authors trained a chatbot on the resulting dataset and found that the chatbot was more engaging to talk to for people than chatbots trained on only resources such as Twitter. More importantly for our research, the conversations contained valuable information about the (fictional) users' personal lives. For example, knowing about someone's favorite food or their family helps an agent to learn about people's well-being.

5.3 Architecture

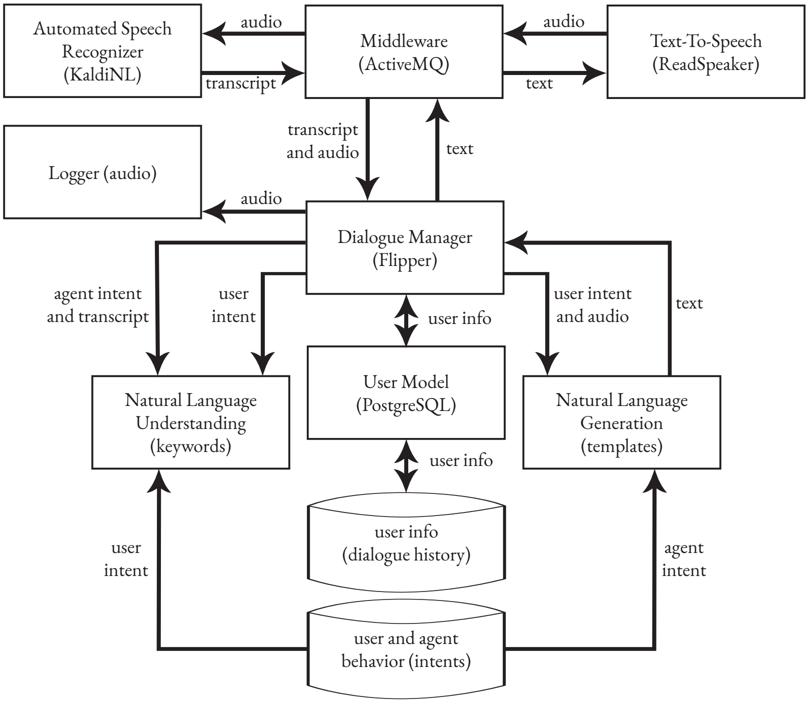
In BLISS, we use the classic spoken dialogue system architecture for our agent, consisting of five main components: ASR, NLU, DM, NLG and TTS. Communication between the components is through the middleware software Apache ActiveMQ,[7] a message broker service. This is a toned-down version of the ARIA-framework, where no visual processing components or embodiment are used. Figure 5.1 shows how these components interact with each other in further detail. In the current implementation of the BLISS agent, the NLU, NLG and DM components run locally on the device, whereas the ASR and TTS components are off-the-shelf products and run as cloud-services. We designed the system in such a way that in the future we will be able to add an embodiment, for example a virtual character or use another speech recognition server for the ASR component, for example with a personalized recognition model.

The whole interaction process can be briefly described as follows. Whenever the ASR receives audio from the microphone connected to the DM, the ASR creates a transcription and sends it to the DM. The DM forwards the transcription to the NLU component, which returns an intent of the user. The DM matches the intent of the user to an intent of the agent and then calls on the NLG component to formulate a behavior of the agent. Once the behavior of the agent has been selected, the TTS receives a message from the DM to realize the agent's behavior by generating the speech. The generated speech is sent to the DM, which plays the audio.

5.3.1 Automatic Speech Recognition

For the ASR, the Corpus Spoken Dutch (CGN)[8] was used to train the acoustic model (AM) and language model (LM) for the speech recognition component.

[7] https://activemq.apache.org/
[8] http://lands.let.ru.nl/cgn/



5

Figure 5.1: Visualization of the BLISS architecture. The BLISS agent consists of a dialogue manager, a natural language understanding and natural language generation component. A database is used to store user information and to retrieve intents for the agent and user. All communication with the TTS and ASR is handled by the ActiveMQ component.

For training the AM, the Kaldi (Povey et al., 2011) framework was used. We implemented a cloud-based speech recognition server and used the neural network based online decoding with iVectors to set up the speech recognition server.[9] The server listens to the audio, which it receives over the internet and sends the decoded transcription.

The user speaks into the microphone of a headset using a laptop and this audio is recorded and sent to the ASR server using a websocket. The ASR can detect the end-of-sentence in the speech signal.

[9] https://github.com/opensource-spraakherkenning-nl/Kaldi_NL

Table 5.1: List of user intents that can be recognized by the prototype with examples (translated to English). The default intent is an inform.

intent	example keyword(s)
question	what do you mean
inform	-
confirm	yes
disconfirm	no
salutation	hello
valediction	goodbye
stalling	ehm
auto-feedback	uh-huh

5.3.2 Natural Language Understanding

The NLU component in our prototype, used to collect the data described in Section 5.4, uses keyword-spotting for intent recognition. We selected relevant intents from the DIT++ taxonomy for our prototype (Bunt et al., 2010). User intents can be classified as question, salutation, inform, valediction, confirm, disconfirm, stalling and auto-feedback (see Table 5.1). Additionally, we use the Dutch Pattern[10] library to extract sentiment from the ASR's output and for retrieving verb phrases (PoS-tagging) from user responses (De Smedt and Daelemans, 2012). The sentiment indicates if a mentioned topic (a verb phrase) is positive or negative for the user. Stop words are filtered with spaCy's[11] default stop word list for Dutch, with some additional stop words that we expect will be said, such as "think" and "find" for expressing opinions and preferences. In our prototype, we made the assumption that the remaining verb phrases represent activities of a particular user. For example, an activity is "hiking".

5.3.3 Dialogue Management

The DM is responsible for responding to the user behavior perceived via the input component (ASR) and for generating the agent behavior that is realized via an output component (TTS). It also controls the NLU and NLG components. The DM of the BLISS agent is based on the dialogue engine Flipper discussed in Chapter 4. In its information state it keeps track of all user transcriptions, topics and intents and of what agent behaviors have been performed. The DM is connected to a PostgreSQL database, where we store all the dialogue information per user.

[10] `https://github.com/clips/pattern`
[11] `https://spacy.io/`

5.3.4 Natural Language Generation

The NLG component is scripted and template-based. In our prototype, the agent follows a script of small talk after which it starts asking the user three pairs of starter and follow-up questions. For the generation of the follow-up questions we use templates, with placeholders for activities users talked about. The placeholders are filled with the verbs extracted from the user utterance by the NLU component, after lemmatizing them to fit in the template. For example, the agent would first ask a starter question, such as "If you could choose one thing you want to do this weekend, what would it be?". After the user answers, a follow-up question is "Sounds good. Why do you like [ACTIVITY]?". So if the user's answer contained "sailing", the full follow-up question would be: "Sounds good. Why do you like sailing?". The agent would repeat this sequence of starter and follow-up questions two more times, and finally it follows a script to close the conversation.

5.3.5 Text-to-Speech Synthesis

The TTS component in our prototype is provided by ReadSpeaker.[12] The current commercially available voices from ReadSpeaker are based on Unit Selection Synthesis (USS). The USS method (Hunt and Black, 1996) relies on a large acoustic database recorded by a professional voice talent, which is searched at synthesis time to find small audio segments which are concatenated to produce a smooth, natural-sounding utterance. This utterance is sent to the dialogue manager for playback.

5.4 Data Collection

For our Dutch spoken prototype, we tried to elicit user information in the health and well-being domain. We required spoken conversational data, specifically about health and well-being. Public corpora such as the JASMIN corpus and CGN unfortunately do not contain this type of data. Therefore we decided to create a prototype of the BLISS agent with mostly scripted capabilities to collect this type of data. This version of the system was tested in several field studies with users, with the following two aims:

1. To find out how people interact with a computer when talking about their daily activities and underlying motivations for these activities.

2. To collect data that could be used for further improvements of the system.

In this section we describe our setup for the data collection, together with an example of the dialogue flow, our preprocessing steps and the meta-data of participants.

[12] https://www.readspeaker.com

5.4.1 Setup

We tested our prototype at three different conferences with a predominantly Dutch-speaking audience. At each of these conferences, we used a stand or a room where users participated in an interaction with the BLISS agent, with different environmental noises. Our setup required an active internet connection for the ASR and TTS cloud-services, a laptop for running the BLISS agent and a headset for speaking and listening to the agent. We asked people passing by our stand to be participants. If they agreed to participate, we gave them an information brochure before participating and asked them to sign a written consent form and provide general demographic information (age range, gender and place of growing up,[13] see Appendix C). If the participant had no further questions, the agent initiated the conversation with the participant. Participants were instructed to repeat themselves if the agent did not respond to their answers. After the interaction, we debriefed participants about how the BLISS agent operates.

5.4.2 Dialogue Flow

In Table 5.2, an example dialogue of a participant with the agent is shown. The agent initiated each conversation with some introductory social dialogue to collect information about the familiarity of the participant with conversational agents, and their sentiment towards them. After this introductory part, the agent asked the participant three times to mention an activity they liked to do, each time with a follow-up question about the motivation behind the activity. Any user utterances following a question by the agent that were classified as an *inform intent* (see Table 5.1) were assumed to be the answer to that question. The agent ignored responses that were only transcribed as "ehm" (*stalling intent*), and waited for another user response instead. If the user spoke at the same time as the agent, the user's speech was recorded, but ignored by the agent. The agent only listened to what the user said after assigning the turn to the user. At the end of the dialogue, the agent asked if participants had changed their opinion about talking to a computer, and finished the conversation.

We implemented two small forms of user adaptation in the BLISS prototype. First, after the agent asked how participants felt about talking with a computer, the sentiment of the user's response was analyzed with Pattern (De Smedt and Daelemans, 2012). The agent then reacted differently to positive (>0.5), negative (<-0.5) or neutral (-0.5<= & <= 0.5) responses. These responses were "You're making me blush!", "I get it, I can sound quite scary." and "I'm looking forward knowing what you think about me at the end." respectively. Second, after it asked the users about the activities they would like to do, the agent used Pattern to

[13] The information about regional differences in speech based on where participants grew up is potentially interesting for other researchers who work with Dutch speech.

5

Table 5.2: Example dialogue between a user and the BLISS agent as interlocutors (I).

#	I	Utterance (translated corrected transcript)
1	Agent	Hello, have you ever talked with a computer before?
2	User	Yes
3	Agent	Okay, which ones have you talked with?
4	User	With Google Home
5	Agent	What do you think about talking with me?
6	User	It is pretty okay
7	Agent	I'm looking forward to knowing what you think about me at the end.
8		Now, I'm curious about you.
9		If you could choose one thing you want to do this weekend, what would it be?
10	User	Reading a bit.
11	Agent	Sounds good. Why do you like that?
12	User	It's very relaxing.
13	Agent	And if you could choose a second activity, what would that be?
14	User	I think that would be cycling.
15	Agent	What would you miss most if you wouldn't do cycling?
16	User	The physical activity, but then I would resort to walking.
17	Agent	Looking back at last week, which activity did you enjoy most?
18	User	Visiting my friends.
19	Agent	Great! What kind of feeling did it give you?
20	User	A lot of fun
21	Agent	Thanks! That's it for now. Do you still think the same about me as before?
22	User	Yes.
23	Agent	Thanks for participating, bye!

extract the verbs from the user's answer (e.g., "cycling" from "I think that would be cycling."). Then the agent filled this in the template of the follow-up question about the user's motivation for doing the activity. If no non-stop word verb is detected in the user utterance, the default placeholder "that" is used.

5.4.3 Preprocessing

In total we recorded 59 sessions. We decided to include a session in our dataset if at least one of the questions by the agent had been answered by the participant. We discarded 4 conversations in which none of the questions had been answered, after which 55 sessions remained. Of the 55 sessions, 9 sessions had latency issues because the reply by the agent was (too) slow. Of these 9 sessions, 4 were incomplete because the participant could not answer all the questions. We decided to keep

Table 5.3: We asked the participants for their region (a), gender (b), age (c), and familiarity with conversational virtual agents (d), where the age is divided into bins for life phases in line with the bins of the CGN. The region represents where participants grew up for most of their life between the ages of 4 and 16.

(a) Region

Region	Users
Dutch	42
Flemish	4
Other	9

(b) Gender

Gender	Users
Female	33
Male	22
Other	0

(c) Age bins

Age	Users
18 –30	23
31 –45	16
46 –60	13
61 –110	3

(d) Familiarity

Exp.	Users
Yes	27
No	23
Unclear	5

these 9 sessions in our dataset, because they contain answers to some questions and the speech itself can still be useful to spoken Dutch researchers, except for learning about response times. Our total dataset thus consists of 55 sessions.

5.4.4 Resource Availability

The collected dialogues, both the transcripts and the audio-files are available for research. Access to this data set is granted after signing a Data Use Agreement for academic research purposes. We refer to the BLISS website[14] for the contact details.

5.4.5 Details

The sessions in our dataset have an average length of 2 minutes and 34 seconds (standard deviation (std) = 60 seconds). If we look at the 46 sessions that did not have any latency issues, the average length is 2 minutes and 18 seconds (std = 24 seconds). The ASR module transcribed 662 utterances in total.

Table 5.3 shows that most of the participants were in the younger age categories (mean = 33,53, std = 14,28). Of the 55 participants, 40% were male. Most of our participants (75%) were from the Netherlands, a few participants were Flemish (7%), while the remainder of the participants had a different country of origin. Around 50% of the users had talked to a spoken dialogue system before, such as Siri, Google Assistant or Amazon Alexa.

[14]BLISS website: `https://bliss.ruhosting.nl/`

5.5 Qualitative Data Analysis

We performed an explorative qualitative analysis of our dataset. We were mainly interested in how people talk to the BLISS agent and which information we could extract from the conversation. We did a preliminary thematic analysis on the dataset to structure the information about what people said (Braun and Clarke, 2006). We also analyzed some dialogue aspects, such as disfluencies (e.g., hesitations and repetitions) and the impressions participants had of the agent.

5.5.1 Activities & Motivations

activities

The BLISS agent wants to learn what makes people happy and healthy. Part of people's happiness and health is determined by the *activities* they undertake, such as hiking, reading and playing games. Therefore the agent needs to learn which activities make people happy and why they choose these particular activities, their motivations. In this section we search for common themes in the user's answers and cluster them.

In Table 5.4b we show the clusters of activities mentioned during the dialogues. For the clustering we extracted the noun and verb phrases from the automatically transcribed answers and grouped them together under the common themes we identified. We excluded all answers that were incomprehensible (incorrect and incomplete transcripts), missing (system error), irrelevant (questions about the system) and answers in which users said that they could not think of another activity. The classes in Table 5.4b are not mutually exclusive, as some users included multiple activities in an answer to an activity question. For example, "walking in nature" can be classified as both a hobby and outdoor activity. After filtering the answers, activities were clustered manually, which resulted in six classes in total.

5

1. *Hobby.* Activities such as watching TV, reading, travelling and doing sports.
2. *Outdoor.* Mentions of an outdoor location, such as the beach, the forest or specific cities.
3. *Resting.* Sleeping, doing nothing or just relaxing.
4. *Social circle.* People described not only the activities, but also with whom they wanted to do this activity, such as with their partner, friends or family.
5. *Social activities.* Activities such as eating out, going to a party or having coffee.
6. *Work.* Work-related activities, such as attending a conference or volunteering.

Table 5.4: The clusters and frequencies (Freq.) of activities people talked about are shown in (a). For example, hobbies includes watching TV, but also walking and sports. A list of motivations for the activities is shown in (b).

(a) Motivations

Motivation Class	Freq.
need for rest	13
wanting to be outside	7
to be in nature	4
to exercise physical activity	4
being at new locations	5
no specific reason	5
desire for interaction	13
being curious and excited	4
just nice	5
relaxation	17
love to do	4
feeling happy and content*	21
focusing on inner self	2
sports	3
interesting	2
need to do fun stuff	2
Total	111

(b) Activities

Activity Class	Freq.
Hobby	57
Social circle	30
Outdoor	22
Social activity	22
Rest	16
Work	14
Total	161

Answers to the third activity question the BLISS agent asked (Table 5.2, turn 17) included more specific activities than the answers to the first and second questions (Table 5.2, turn 9 and 13). For example, activities such as "celebrating a birthday at the office" or "I received my diploma yesterday" were all answers to the third question. To the first and second question, people generally responded with their hobbies like "reading" or "walking".

Table 5.4a shows the *motivations* that were mentioned to the BLISS agent when asked why they liked a certain activity. Again, we used a thematic analysis to extract common themes from the motivations. One remark is that the answer "feeling happy and content" is derived mostly from the answers to the third question about motivation ("What kind of feeling does that give you?"). In response to this question, people often replied with "a good feeling" or "a happy feeling". **motivations**

Table 5.5 shows the categorization of the motivations, which we clustered

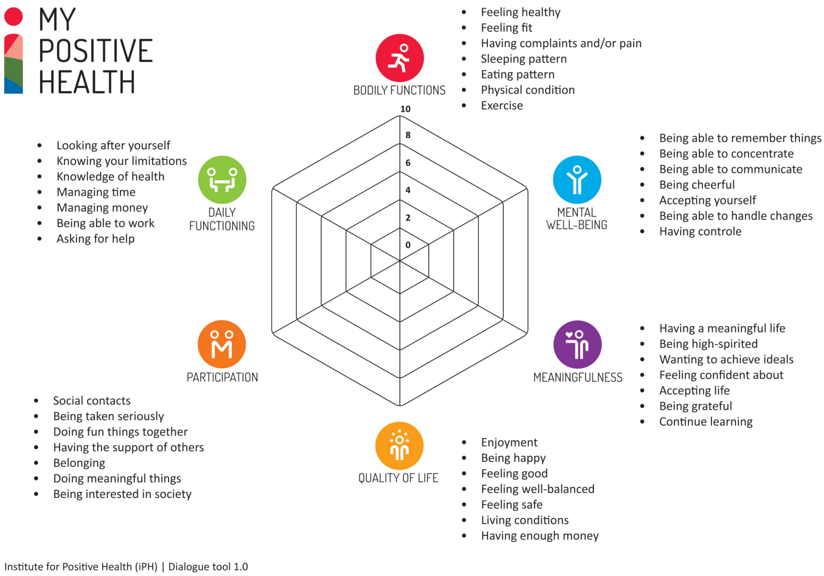


Figure 5.2: Dialogue tool of the Institute of Positive Health (IPH), with the six categories we used for analyzing the motivations in the dialogues with the prototype.

similarly to the clustering of the activities. However, instead of deriving themes from the data, we used the dimensions of the dialogue tool of the Institute of Positive Health (IPH), based on the work of Huber et al. (2011), see Figure 5.2. We filtered the motivations by excluding motivations that were incomprehensible (gibberish transcription), missing (incomplete transcription) or irrelevant (meta-answers) and excluding replies in which the users could not think of a motivation.

Table 5.5 shows that most participants mentioned motivations related to feeling good and wanting to do something, because this makes them feel happy (quality of life).

1. *Quality of life.* Motivations related directly to feeling good and happy and doing things you love.

2. *Daily functioning.* Motivations related to taking time for yourself and knowing what you need.

3. *Participation.* Motivations which include social contacts, such as family or friends and helping out others.

Table 5.5: If we link the motivations to the positive health model of the IPH, based on the work of (Huber et al., 2011), we can see that most users mention motivations related to their general quality of life and daily functioning.

Motivation class	Frequency
Quality of life	43
Daily functioning	31
Participation	18
Physical health	12
Meaningfulness	8
Mental well-being	4
Total	116

4. *Physical health.* Motivations related to wanting to exercise, have a regular sleeping pattern and feeling fit.

5. *Meaningfulness.* Motivations related to finding a purpose, being excited and wanting to learn.

6. *Mental well-being.* Motivations related to mental health and feeling in control of your life.

The first and second motivation question asked by the BLISS agent (Table 5.2, turn 11 and 15) often received responses in the dimension of daily functioning, whereas the third question (Table 5.2, turn 19) mainly received responses in the quality of life dimension. Mental well-being and meaningfulness were not often mentioned as motivations. The second motivation question was less suited from the perspective of positive health, because it asked about what people would miss, instead of asking directly why people liked a certain activity. It would sometimes lead to people repeating the activity they mentioned or saying "I wouldn't really miss anything". The third motivation question was often answered with different variations of "a good feeling".

In Table 5.6 we show the combinations of activities and motivations per question. We combined each of the activity questions with the corresponding follow-up motivation question. For example, if the answer to the activity question was "to go for a walk with friends" (activity classes: *hobby*, *outdoor* and *social circle*) and the reason for this was "it is great to be in nature" (motivation class: *quality of life*), this would add 1 to each of the following combinations: *hobby — quality of life*, *outdoor — quality of life* and *social circle — quality of life*.

Table 5.6: Mentions of activities linked together with the motivations derived from the IPH model. There are six classes for both activities and motivations. All activities and motivations are derived from the transcripts of the ASR.

	Relax-ation	Hobby	Social circle	Social activities	Outdoor	Work	Total
Physical health	3	6	0	1	4	0	14
Daily functioning	9	14	2	1	7	1	34
Mental well-being	1	0	1	1	0	1	4
Participation	1	5	8	5	2	1	22
Meaningfulness	1	2	0	0	0	0	3
Quality of life	1	16	10	6	8	5	46
Total	16	43	21	14	21	8	123

As a result, we see in Table 5.6 that most people who mentioned a hobby, often gave a motivation related to daily functioning or quality of life. Additionally, people could mention multiple motivations for one activity, or have one reason for multiple activities, hence the number totals in Table 5.6 are different from those in Table 5.4. The motivation "feeling happy and content" was often the reason given when BLISS asked: "What kind of feeling does that give you?", which makes this particular question seem more directed than the other motivation questions.

5.5.2 Interaction

Around 10% of all user transcripts contains a word that was explicitly not recognized by the ASR (labeled as "unknown" by the ASR). The ASR recognizes nonverbal utterances like "uh", "uhm" and "mmm". Our dataset contains 55 utterances (on the total of 662) that only consisted of such non-verbal reactions. A common type of ASR error is an incorrect transcription leading to an incomprehensible utterance, like the one shown below, where it can be observed that the ASR wrongly transcribed the user's speech.

ASR	*dat*	*ik*	*mijn*	*ouders*	*dan*	*weer*	*energie*
	that	I	my	parents	than	again	energy

Speech	*dat*	*ik*	*mijn*	*ouders*	*dan*	*weer*	*eens*	*zie*
	that	I	my	parents	then	again	sometime	see

In some cases the dialogue took a very long time to complete due to practical technical issues. This influenced the overall dialogue flow. We show an example of

Table 5.7: Part of a problematic dialogue of a user with the BLISS agent. For example, the user is asking questions about the quality of the system (Turn 2, 7 and 13).

#	Interlocutor	Utterance (translated transcript)
1	Agent	Sounds good. Why do you like that?
2	User	he faltered already with the question.**
3	Agent	And if you could choose a second activity, what would that be?
4	User	uhm then I would go have a drink with my friends in the city
5		yes*
6		yes I do still have a little bit of patience**
7		because he took my text also this time**
8	Agent	What would you miss most if you wouldn't do take?
9	User	if you take, what do you mean with take?**
10	Agent	Looking back at last week, which activity did you enjoy most?
11	User	uh last week i found too active, i liked the best to uh on Sunday go to hockey with the kids
12		mmm.*
13		are we getting any other questions?**
14	Agent	Great! What kind of feeling did it give you?
15	User	uh yes*
16		that makes me happy

a problematic dialogue in Table 5.7, which contains the relevant answers mixed with backchannels (indicated with *) and meta-level (indicated with **) utterances where the participant gives explicit feedback on the system quality. Note that the current version of the BLISS agent does not contain mechanisms to cope with explicit problem signals from users. For example, as seen in Table 5.7, the agent does not "understand" that the answer of the user (Turn 2), is not an answer to the question it asked and continues asking the next question (Turn 3).

We asked users for explicit feedback on their conversation with the agent. We started with a direct question to establish their familiarity with conversational agents, followed by an open question to determine their stance towards dialogue systems. Around 50% of the users had talked to a dialogue system before such as Siri, Google Assistant, Google Home or Amazon Alexa. Most users (43.6%) were positive about engaging in a conversation with the agent ("amusing", "nice", "interesting"), 20% had a more negative, cautious attitude ("weird", "ill at ease") and in 36.4% of the answers we had a neutral answer ("it depends") or we could not determine the sentiment of the user answer.

At the end of the conversation we asked whether their stance had changed after speaking to the system. This was the case in 16% of the cases. Moreover, at the end of the dialogue, the vast majority of people were positive about the system ("yes, i still like you"); and some users gave constructive feedback about the voice quality ("well to be honest, I find your voice a bit forced"), dialogue flow ("yes I rather like you but you are a bit slow"), and level of comprehension ("you still need to learn to have a conversation").

5.6 Discussion

The BLISS agent is a work in progress, and a new version is planned for the near future that is more scalable and easier to test with end-users. While using the prototype for data collection, we noticed several issues with different components of the system. Some of these issues are related to technical implementation (hardware and software), while others are more related to the usage of the system. One of the important components is the NLG, which generates the questions for the user based on the initial answer. We observed that sometimes the quality of the generated follow-up motivation question (second or third question) was not good and sometimes the question did not even make sense. For example, a follow-up question of the agent was "Why do you like going?", as it used the verb "going" from the previous user sentence: "I like going to the cinema or going out for dinner." In this case the NLU component did not extract the complete activity for the NLG component. In such cases users often responded with a meta-question about the system which broke the flow of the dialogue, for example "What are you saying?".

code-switching

Although most of the participants were native speakers of Dutch, sometimes they would use *code-switching*, which means that they would use English words while talking to the system in Dutch. Because the ASR did not contain those words in the acoustic and language model, it produced recognition errors for such words. Our conversations did not contain any dialect words and no mistakes were made because of that, but should be included in a newer ASR model for situations in which these words occur more frequently, which we do expect with older adults for example. Besides the words themselves, different pronunciations were not found to be a difficult problem, and we were positively surprised by the ASR picking up most words with an accent. At the conferences we also had latency issues with the ASR and TTS. Since these components are both cloud-based and dependent on a stable internet connection, the delays would produce some discomfort to the participants, and they would try to repeat their answer, or were too quick in answering. If a stable connection exists, the ASR has about a second of delay and the TTS works almost immediately. Also, for short answers like the "yes" or "no" answers of the participants, recognition was poor, and manual interference (transcribing the user speech for the BLISS agent) was required to resume the dialogue.

Our prototype typically does not yet have the ability to deal appropriately with situations in which the user doesn't respond with an answer to the question. Sometimes users repeated the question before answering, hesitated or requested some elaboration. Often the agent interpreted the user's response as the answer to the question it had just asked. For example, the agent could ask: "What would you like to do this weekend?" and the user would respond "This weekend. Let me think", after which the agent could ask "What kind of feeling did thinking give you?". Hayano (2013) investigated how often people do not respond to questions with an answer in human-human conversation. This happens about 15%-24% of the time. However, such cases need to be covered as well, so implementing a fall-back strategy for these events would help resolve this issue.

Even though the ASR did not recognize all the utterances by the users correctly, we found that even with errors and a noisy environment, relevant information can be extracted and for each of the users we did find at least one activity. This means that even though the speech recognition is not perfect, it is usable for retrieving this type of user information. This is an important finding because it indicates that in real world conditions, the system could be usable. Additionally, we aim for a long-term interaction solution where only activities and motivations that are repeated during multiple conversations are trustworthy pieces of information for the user model.

5.7 Future Work

One of the goals of BLISS is to provide users with personalized dialogues. For starters, the users' speech is recorded together with the transcripts obtained through ASR. These speech recordings can be used to adapt the ASR's language and acoustic model so that it can better recognize the users' speech and to improve and personalize the dialogue. Additionally, we will use the data for creating a more personalized happiness model. This means that the agent should be able to detect full user answers more appropriately, such that it waits until the user is done answering. We will also improve the agent's activity and motivation extraction to create a correct user model. With the collected data we can create a method for this purpose.

However, if we want to apply machine learning methods, much more data is required. This could be accomplished by extracting information from health records or by training a model on generic Dutch spoken dialogues between humans and fine-tuning it to smaller datasets about positive health (Vaswani et al., 2017). At the moment of writing we are contacting several health organizations and companies to see whether such data can be made available for specific case studies. An important aspect is of course that this is done under conditions of security and privacy, with approval from the ethics committee and in agreement with GDPR regulation. Thus far people have only interacted once with the BLISS agent. In the next two chapters we propose a design of and pilot a long-term real world study with a similar architecture as the BLISS agent.

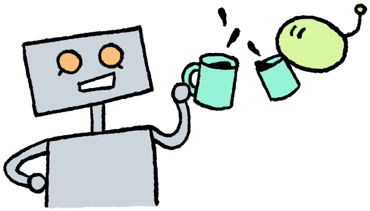
With regard to the TTS, we note that one of the drawbacks of the current TTS module is that the USS method is not very flexible, which limits the extent to which personalization and expressiveness can be accomplished. To tackle this issue, a neural TTS speech synthesis system is currently being developed by ReadSpeaker. Such a neural speech synthesis system can provide high-quality speech and add much more flexibility than is possible with USS systems (Habib et al., 2019). In particular, the aim is to personalize speech output to the users by being able to flexibly modify the input to the model, through pronunciation adaptation or changing emphasis for certain words (Shechtman and Sorin, 2019). That way, new voices are likely to sound more appropriate for the conversational setting of a dialogue and will sound more empathetic for the users (James et al., 2020). Additionally we have not yet exhausted using SSML for changing prosody or pitch to increase speech quality in a more controlled way.

One of the issues that we derived from this data collection is the impact of phrasing questions on users' answers. Additionally, we found that people might not always respond with an answer. In the next chapter, we create a model for addressing different types of questions and dealing with non-answers. Our focus of the questions is again on extracting activities and interests (i.e., topics) from the user. In Chapter 7 we evaluate this model in the real world and gain more insights into how people respond to these questions.

In this chapter, we have shown the results of our first data collection with a prototype of the BLISS agent, which already gives some useful insights into aspects of people's well-being, such as how people tend to describe their activities and how the agent should cope with a variety of responses given during a spoken dialogue. For the future, we intend to extend the scope of BLISS by incorporating more specific health contexts. Moreover, since self-management is a crucial element in the definition of health adopted in our research (Huber et al., 2011), our future work will investigate how the knowledge obtained through the dialogues can best be employed in the context of BLISS to realize a system that is capable of supporting users to self-manage their health and well-being.

Part III

Personalizing Dialogue in Long-Term Interaction



6 Personalized Question-Asking in Casual Conversations

6.1 Introduction

A social agent that can ask questions is no novel thing, with ELIZA being the first social agent capable of holding interesting conversations without much knowledge (Weizenbaum, 1966). More recently, a social chatbot capable of asking questions was compared to filling in surveys to see which method could provide better qualitative information (Xiao et al., 2020). Xiao et al. (2020) found in their study that people were more engaged with the chatbot and responses given to the chatbot were also more informative compared to the responses of the people who filled in the survey. Machine learning methods exist to generate questions as well, but these methods in general require fixed pairs of question-answers with ground truths (Su et al., 2018; Isonishi et al., 2021). However, for casual conversation, this ground truth might not exist.

CoffeeBot

In this chapter we discuss the design of a social conversational agent for casual conversations: CoffeeBot. *CoffeeBot* is a social agent capable of personalized spoken casual conversation in long-term interactions. The name of the CoffeeBot stems from a location where colleagues have casual conversations: near the coffee machine. The CoffeeBot's goal is to "learn" during these conversations about people's interests and to personalize the long-term interactions based on these users' interests by asking questions. We start this chapter by explaining a theory of casual conversation and how to apply it in the real world. Next, we model the CoffeeBot and prototype it based on the architecture in Chapter 3.

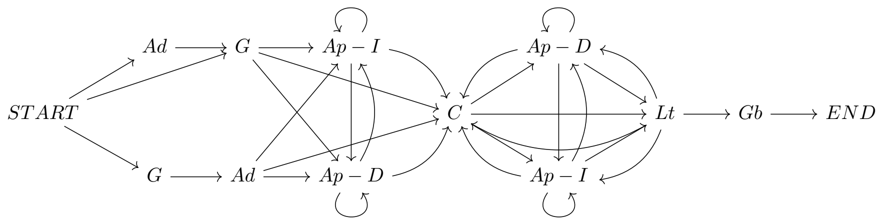
casual conversations

Casual conversations are conversations such as talking at the coffee machine with colleagues, pub conversations or dinner table talk (Gilmartin et al., 2018). These types of conversations can reveal interesting personal information about the interlocutors, as they occur frequently and contribute to the social relationship between people (Eggins and Slade, 2001). Interestingly enough, most people feel like they talk about nothing or nothing happens during these types of conversations. However, Eggins and Slade (2001, p. 16) point out that "[...] casual conversation disguises the significant interpersonal work it achieves as interactants enact and confirm social identities and relations". They call it the *paradox* of casual conversation, because most people feel like they are talking about nothing, though these conversations show how people view the world around them. It is during these types of conversations that people feel most at ease and relaxed, thus revealing more personal information about themselves.

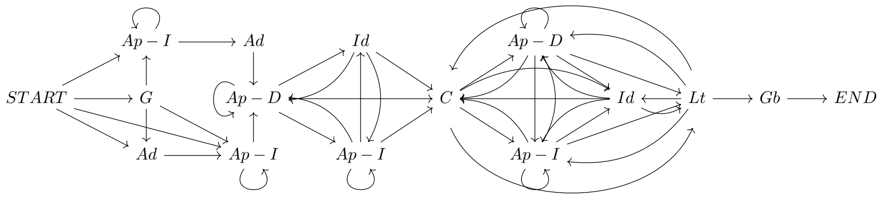
Ventola (1979) states that casual conversations are "verbal interactions in casual encounters" and these interactions are spontaneous and important for establishing and maintaining social relationships. These conversations happen in spoken face-to-face situations, and differ between strangers (maximal social distance) and close friends (minimal social distance). Additionally, these conversations can be minimal or non-minimal. In minimal conversations, only phatic (verbal) expressions (greeting, nodding) are used, such as two friends passing by and greeting or calling each other. Non-minimal conversations involve more than phatic verbal expressions to maintain good social relationships, and must have a deeper content component. However, non-minimal conversation with minimal social distance seem to only apply for human-human interactions and not human-agent, because people do not expect or want to maintain close social relationships with social agents (Clark et al., 2019). Regardless, in their study on how people view social relationships with both people and agents, Clark et al. (2019) state that the majority of people not wanting or expecting social relationships with agents are largely based on goal-oriented interactions of people with social agents, without considering long-term interactions. We therefore believe that implementing a casual conversation strategy for social agents for long-term interaction can still be beneficial, because people possibly change their expectations after longer exposure to an agent. For example, a social agent could learn over time when and with whom certain topics are more relevant to talk about or not, or for example an agent that can adapt to when a user wants to talk, which differs per individual (Heylen et al., 2012).

6.2 Structure of Casual Conversations

A good way to start the dialogue design is to learn about overall dialogue structure and which elements are required for non-minimal casual conversations. We use Ventola (1979)'s model of casual conversation as a starting point. Based on the formalization of Ventola, we created state-based dialogue structures of non-minimal conversations with friends and strangers (see Figure 6.1). The elements of a casual conversation are: i) greeting (`G`), ii) addressing (`Ad`), iii) indirect approach (`Ap-I`),



(a) A visualization of the structure of non-minimal casual conversations with friends.



(b) A visualization of the structure of non-minimal casual conversations with strangers.

Figure 6.1: Our state-based visualizations of non-minimal casual conversations, based on the work of Ventola (1979)

iv) direct approach (`Ap-D`), v) identification (`Id`), vi) centering (`C`), vii) leave-taking (`Lt`) and viii) goodbye (`Gb`). A greeting consists of nodding or a verbal "hi" or "hello". An example of an address is "sir" or "Alice". *Indirect approaches* relate to immediate situations, indirectly related to a person, such as asking about the weather or bus times. *Direct approaches* are about a person themselves, usually by knowing some context about the person, such as asking how somebody's husband is doing or how they liked a certain conference. Identification only occurs in conversations with strangers, and possibly multiple times, because when meeting new people, a name mentioned once is easily forgotten. *Centering* is about diving into depth about a topic, in which there are usually follow-up questions, opinions and statements shared about the same topic. Announcing the end of the conversation is part of leave-taking. The leave-taking element is often used as a form of politeness, but also provides an opportunity for the other interlocutor to say some last things or perhaps make the conversation more interesting and prevent the other interlocutor from leaving. A goodbye happens at the end of the conversation after leave-taking when the interlocutors each go their separate way.

indirect and direct approach

centering

In non-minimal casual conversations with friends or colleagues, only a greeting, centering, leave-taking and goodbye are required as the bare minimum, such as shown in Table 6.1. An example of a meeting between two strangers is shown in Table 6.2.

Table 6.1: Example of a non-minimal conversation between interlocutors (I) Alice and Bob who are close colleagues.

I	Action	Structure
Alice	Bob, hi!	Greeting, addressing
	I read this great paper about embodied conversational agents, should I send it to you?	Centering
Bob	Sure	Centering
Alice	Alright, gotta run. See you around.	Leave-taking, goodbye
Bob	Bye	Goodbye

Table 6.2: Example non-minimal conversation between two interlocutors Cindy and Dave, meeting at a conference as strangers.

I	Action	Structure
Carol	Can I ask a question about your talk, sir?	Indirect approach, addressing
Dave	Sure, what is on your mind?	Indirect approach
Carol	Thanks! My name is Carol. Have you looked at other contextual factors in your study?	Identification, direct approach
Dave	Yes we did look at some additional factors, like the location, age and time of day, but we could not find a significant effect either.	Centering
...	...	Centering, direct approach, indirect approach, identification
Carol	Thanks so much again for taking the time. Bye.	Leave-taking, goodbye

6

An important distinction is that casual conversation $\neq$ small talk. Small talk can be a part of casual encounters and is usually related to the indirect and direct approach elements of the casual conversation structure (Ventola, 1979). Gilmartin et al. (2018) simplified Ventola's casual conversational model, in which the greeting and address are merged into `G`, identification, indirect approach and direct approach are merged into `A` and goodbye and leave-taking are merged into `L`. We modified this model to fit non-minimal conversations with the CoffeeBot in which multiple leave-takings can occur (Figure 6.2).

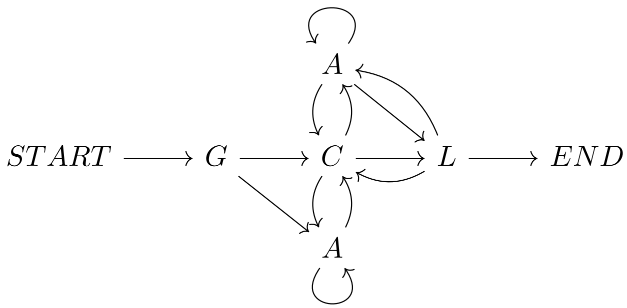


Figure 6.2: A simplified version of Ventola (1979)'s casual conversation structure adapted from Gilmartin et al. (2018), with a specific component for leave-taking (L).

6.3 Remembering in Real World Casual Conversations

Casual conversations are spontaneous conversations that happen frequently in the real world and the CoffeeBot is designed for just these types of long-term interactions with people. Initially, all people are strangers to the CoffeeBot (maximum social distance), though over time, the CoffeeBot tries to minimize social distance between the user and itself. Long-term interactions can have a duration of at least a week, up until months or even years, for both social agents (Bickmore et al., 2010) and robots (Leite et al., 2013). Leite et al. (2013) emphasize the importance of conducting long-term studies, for example to prevent the impact of the novelty effect on the interaction evaluation. The large majority of long-term studies discussed by Leite et al. (2013) involve task-oriented interactions, such as for health and educational purposes. Instead, the CoffeeBot has no explicit task and only learns about users by frequently making casual conversation. To develop a social relationship with its users, in other words, to decrease social distance over time, the CoffeeBot requires a memory component that contains personal information about the user.

Personal information can be as simple as remembering someone's name (Glas et al., 2017) or recall information from previous interactions. For example, for a real-estate selling agent, the style preferences and job type of a user are personal, and the agent should adapt to this specific information to achieve its goal of selling (Richards and Bransky, 2014). This type of information is usually gathered in the direct (and occasionally indirect) approach element of the casual conversation model. Campos and Paiva (2010)'s chatbot *MAY* chit-chatted with teenagers and recalled details about past conversations such as earlier mentioned names and activities. Third-party raters were shown two conversations, one with a version of MAY with memory and without. The observers rated the intimacy and companionship higher of the chatbot that remembered details. In a follow-up study, Campos et al. (2018) tested a method using semantic similarities of past conversations to personalize agent utterances. Utterances that were semantically similar to topics mentioned in past conversations between the agent and user were used more often. However, they did not find a significant difference between asking questions from memory or new questions, possibly due to too implicit memory structure or lack of reasoning about the topics in the memory. The CoffeeBot should therefore be explicit in talking about its memories, for example referring to previous conversations and performing some basic reasoning about the topics to show a shallow sense of understanding.

The topics the CoffeeBot could talk about were similar to Kennedy et al. (2017)'s social agent *Kevin*, which was meant for having dyadic casual conversations with office workers. Topics that people can talk to Kevin about vary from talking about their weekends, upcoming or past work meetings, ask feedback from peers or discuss new ideas with peers. In Kennedy et al.'s experiment, the authors placed the robot, a Furhat (Al Moubayed et al., 2012), at the workplace near the coffee machine, so that it could easily interact with the coworkers over the course of three weeks. *Kevin*'s dialogue structure was based on a probabilistic dialogue graph authored beforehand. When Kevin could not answer a question, this was marked as a "failed response" and sent to crowdsource workers, who formulated a response given an unanswered question. This response was added to the dialogue graph, such that Kevin could use this response the next time a similar response should be given and Kevin could also learn from previous user responses in interactions and added those responses to its dialogue graph to use in upcoming interactions. The CoffeeBot has a different model than Kevin, because the focus of the CoffeeBot is on personalized information elicitation from the user. The CoffeeBot is only capable of answering a few questions and is focused on discovering user interests such as hobbies and plans by asking personalized questions. Another major difference is that the CoffeeBot learns to converse on an individual level specific to a user, in contrast to Kevin who learns a better global model from the interactions with all users.

Another similar approach to casual conversation was used by Gockley et al. (2005), who conducted a 9-month-long study with a receptionist robot called *Valerie*. Valerie was a robot with a monitor on top to show her virtual human head and was placed at a receptionist desk at the workplace. She was capable of understanding and producing speech as well as receiving typed input. Her goal was to interact with office workers on a day-to-day basis, and she could provide information about meetings and offices as well as talk about her personal life. She was programmed with an elaborate backstory which she shared with people over time. Gockley et al. found that the novelty effect was present in the first week of the experiment with many interactions, but after this week the number of interactions flattened to a steady number for the whole duration of the experiment. Many users however did not interact with Valerie more than half a minute, because she often had long monologues about her personal life, which was not pleasant to listen to for long, as she had no emotional speech capabilities. Gockley et al. suggested asking more about the user interests to make the interactions more enjoyable for users, which is exactly what we have designed for in the CoffeeBot.

Heylen et al. (2012) discussed their experience with deploying a rabbit-like companion robot in two people's homes. The authors concluded that just deploying a social robot in the real world does not seem to have much benefit compared to lab studies, because the two participants felt more obligated to interact with the robot, where one participant was even quite negative about it. The dialogues and especially the timings of the dialogues were preventing a good experience for the users. The dialogues felt repetitive, and the robot started conversations at inconvenient times, which participants dutifully responded to. As Heylen et al. (2012) concludes, putting a robot in people's homes does not make it necessarily an ecologically valid experiment.

Irfan et al. (2020) deployed a barista Pepper robot in the wild with better dialogue capabilities, but ran into practical issues as well in a pilot study. These were issues with speech recognition and interpreting complicated user requests. However, the conversations were perceived as more engaging by participants when the robot used a personalization strategy than if it did not. Its setup is similar to the deployment of the CoffeeBot in the real world, because the start of the dialogue was only initialized by the user's convenience. However, the CoffeeBot did not rely on performing a certain task, such as supporting healthy eating or serving coffee.[1]

6.4 Topic-based User Model

According to De Carolis et al. (2013, p. 92), "user modeling allows socially intelligent systems to adapt to the users' behavior by constantly monitoring it and by continuously collecting their direct and indirect feedback". A user model can consist of information such as, but not limited to, the background of the user, the environment of the interaction, the moods of the user and data about the system interacting with the user. The CoffeeBot has implicitly built a user model over a longer period of time. This *user model* consists of the topics mentioned by the user and recognized by the CoffeeBot, the frequency of the topic, the last date the user mentioned the topic and the user's sentiment about the topic. A similar study combining topics and their sentiment was done by Langlet and Clavel (2016). Langlet and Clavel analyzed conversations from the multimodal SEMAINE database

user model

[1]Perhaps in a future version, as some participants and colleagues of ours were very interested in the CoffeeBot serving coffee itself.

(McKeown et al., 2012) and could successfully recognize topics and classify them as user likes or dislikes. In addition to the user model, the CoffeeBot stores meta-information about the interaction: date, number and length of interactions and number and length of turns.

topic detection

In Section 2.3.2, we discussed possible definitions of topics and how topics are detected in texts. The importance of *topic detection* for the CoffeeBot is to "understand" what the user is talking about during the coffee machine conversations. These topics serve as the main source for personalizing the conversation. For the CoffeeBot, a simplified version of the model of Grosz and Sidner (1986) is used for detecting topics in a sentence, a method also used by Elvir et al. (2017). Rats (1996) based her definition of topics on Grosz and Sidner's model of focus terms. Focus terms are the focus of a sentence and are often placed in the subject position of sentences. Topics in (direct) object position can be used for topic switches, by placing them in subject position later. For example, if one person in a dyadic conversation states "Tomorrow I have to give a lecture to students.", the possible next topics are "a lecture" and "students". A follow-up turn of the other interlocutor could then be: "A lecture can be fun to give." or "Hopefully the students won't give you a hard time." A similar sentiment approach to Langlet and Clavel (2016) has been added to the CoffeeBot's model for detecting if a user likes a topic. The sentiment for the CoffeeBot is not calculated on a word (phrase) level, but calculated over a whole turn in the conversation, which could be multiple sentences. Only the lexical aspects are taken into account in the sentiment analysis, using Pattern (De Smedt and Daelemans, 2012). Additional aspects that would make the sentiment method more robust would be detecting non-verbal valence and arousal using audio or video. The sentiment consists of a numeric value between -1 and 1 for polarity (valence) and between 0 and 1 for intensity, where a low score for polarity means negative, and a high score means positive.

A sentence can contain multiple topics, which we divided into two categories for the CoffeeBot: topic terms and topic events. The CoffeeBot determines the topic term of a sentence by its phrases, their syntactic role in relation to the verb (subject, object or direct object) and the words' PoS tag (noun or verb phrase). The importance of the topic event depends primarily on the syntactic role, where SUBJECT > OBJECT > INDIRECT_OBJECT and secondarily on the PoS, where NP > VP. A stop word list is used for preventing pronouns and auxiliary verbs from being extracted as topic terms. For extracting verb phrases, a complement is extracted from the topic. For example, in the sentence "Next week I will visit my grandparents.", the extracted NP-based topics are "week" and "my grandparents". The extracted VP-based topic term is "visit", with the complement "my grandparents". Topic events consist of predicates that can include information such as locations and timestamps, extracted with semantic role labeling (SRL). Using the previous example sentence, an extracted event consists of the main verb (`V`) "visit", a temporal phrase (`ARG-TMP`) "next week" and verb arguments `A0` "I" and `A1` "my grandparents". The difference between topic terms and events is that terms do not contain any context of the topic and events contains context that is extracted with SRL, such as locations. We made this distinction for more easily grouping mentions of a topic together (terms) and have a list of more context-sensitive topic can be chosen (events) for the CoffeeBot.

Finally, the user model has a forgetting component for the topics. If topics are not mentioned for a while, these topics become less likely to be mentioned in a subsequent interaction. A forgetting mechanism is convenient for efficiency, scalability and adaptability (Lim et al., 2009).

6.5 Prototyping CoffeeBot

To provide the CoffeeBot with spoken dialogue and topic detection, Flipper serves as a basis for the dialogue engine (see Chapter 4). The components that are required for the CoffeeBot are authentication recognition, an NLU and NLG component, an ASR service and a TTS service. For all the technologies for the implementation of the CoffeeBot, we prefer to use open-source or easy-accessible software packages. The architecture of the CoffeeBot[2] is shown in Figure 6.3.

6.5.1 Dialogue Management

Dialogue management is handled by Flipper templates in the move planner. All CoffeeBot components and their information processing are shown in Figure 6.3. For each of the components, except for the user model, the input and output is stored in the information state of Flipper. The topic-based user model is stored in Flipper's database. Flipper has a middleware wrapper that the CoffeeBot uses for components such as the ASR and TTS services of Google and ReadSpeaker[3] respectively. We implemented rules in Flipper templates for information processing, such as sending ASR output to the Affective & Natural Language Understanding component of the CoffeeBot or retrieving relevant topics for Move Planning from the Topic-based User Modeling. Affective & Natural Language Understanding determines the intent of the user, Authentication is for access to a personalized interaction for the user and retrieve the correct topic model and Natural Language Generator generates agent language from behavior specifications by the Move Planner.

[2] `https://gitlab.com/bliss-nl/babbelbot`
[3] `https://www.readspeaker.com`

6

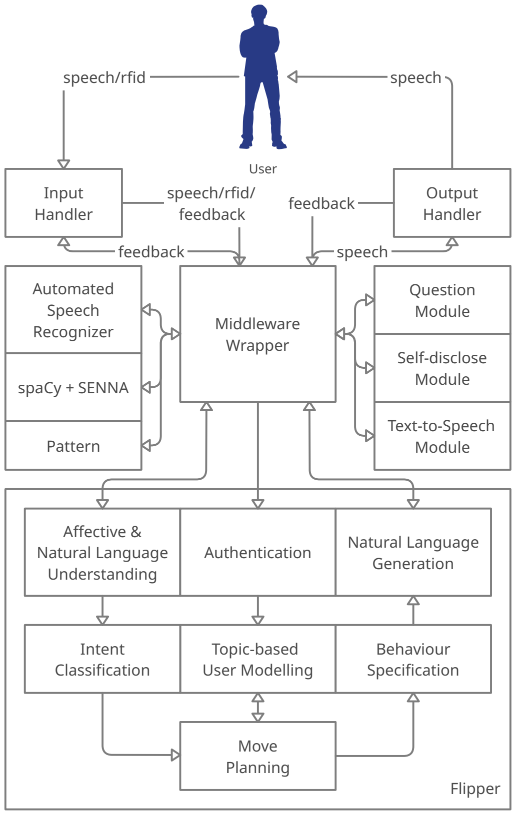


Figure 6.3: Architecture of the CoffeeBot.

6.5.2 Sensors

Two sensor components were the minimum requirements for the CoffeeBot to work.

- A scanning card mechanism for authentication of a user when they have a conversation with the CoffeeBot.
- A speech recognizer that can transcribe distant audio (about 1.5 meter). The CoffeeBot requires the transcripts for detecting topics and performing sentiment analysis.

Good authentication is an important feature for consistent interactions over a longer period of time, accessibility and protecting personal data. Authentication of users is commonly done with usernames and passwords on websites, but these are cumbersome to use for public social robots. Biometric measures that improve accessibility are face, voice or fingerprint recognition, however, this data is very personal and does not allow for direct anonymization. Gockley et al. (2005) used a card swiping mechanism for authenticating users and found that this was cumbersome. They suggested to use RFID instead, which has been a common way of authentication in long-term interactions (Kanda et al., 2010; Heylen et al., 2012; Campos et al., 2018; Davison et al., 2020). The scanning card mechanism of the CoffeeBot has been implemented by attaching an Arduino Uno with an MFRC-522 for reading RFIDs. All IDs are stored with 256 AES encryption. Students and employees are expected to have a card with an RFID, also commonly used for getting coffee, and can use this to communicate with the CoffeeBot. However, if participants would like more privacy, an anonymous card not directly linked to them can be used for all interactions.

We implemented two options for ASR, either our open-source Kaldi speech recognizer used before in Chapters 3 and 4 or Google's Speech-to-Text implementation. The former is more privacy-friendly and customizable and the latter is more robust against noise and has lower WER. Either option sends the transcriptions over the middleware to Flipper for affective and natural language understanding. The ASR provides Flipper with end-of-sentence detection, (intermediate) transcriptions and transcription confidence levels.

6.5.3 Natural Language Understanding

During the conversations with the CoffeeBot, the intents of users should be recognized. Intents are classified based on previous user and agent intents and keywords, similar to the *FLoReS* dialogue manager (Morbini et al., 2012). Certain patterns of keywords are matched to one of eight possible user intents based on the DIT++ taxonomy (Bunt et al., 2010): question, inform, confirm, disconfirm, repeat, salutation, valediction or backchannel. If no intent can be determined, its default class is inform. The spaCy package provides numerous tools for natural language processing.[4] spaCy's PoS tagger and dependency parser (DP) are used for extracting topic terms. Extracting the semantic role (SR) of user utterances for topic events is implemented with SENNA (Collobert et al., 2011). Pattern's sentiment analysis tool measures the polarity and intensity of user turns based on the words in the ASR transcription. A positive polarity is interpreted as a sign of interesting topics and a negative polarity is a sign of disinterest in the topics of the sentence (De Smedt and Daelemans, 2012).[5] Finally the timestamp is saved together with the topic. If a temporal marker such as "yesterday" or "next week" is mentioned by the user, this term is translated to UTC and saved as a timestamp of the topic, oth-

[4] `https://spacy.io/`

[5] `https://github.com/clips/pattern`

6

erwise a timestamp of mention of the topic is stored. This temporal information is necessary to resume topics from previous conversations with context.

All sensory information and information processed by the NLU component is stored in the information state during the conversation. Long-term necessary information, such as topics of interest and their frequency is stored inside the database connected to Flipper.

Answering questions is not the primary goal of the CoffeeBot, though for convenience, it has a limited hand-crafted capability for answering questions. The CoffeeBot can tell about its purpose and answer some social chit-chat, but is not designed as a virtual assistant for booking meetings or answering factoid questions. It uses keyword spotting for answering questions.

6.5.4 Natural Language Generation

The natural language generation (NLG) component is the core component of the CoffeeBot for making casual conversation. Before any natural language generation takes place, the intent for the agent has to be specified. Below are the included higher level categories with agent intents between brackets. The intents are similarly to the user intents, based on DIT++ (Bunt et al., 2010). Once the CoffeeBot has selected an intent, it proceeds to specify an utterance to be generated.

- Self-disclosure (inform, question)
- Social talk (salutation, valediction, introduction)
- Question-Answering (inform)
- Question-Asking (question)
- Interaction Management (repeat, contact)

Bickmore et al. (2010) emphasize how dialogue designers can create agents for use over a longer period of time by ensuring a high level of engagement. The authors state that more variable behavior and a more personal background for the agent will help with improving the engagement of the user, though it could negatively impact task performance for more task-oriented dialogues. The CoffeeBot is not designed for a task and therefore its self-disclosure and question generation for the NLG component should be personal and diverse enough to retain engagement of users respectively.

The CoffeeBot has two ways of generating utterances: canned text or via templates. Canned text is used for most agent intents as shown in Figure 6.4. Canned texts are sentences authored by dialogue designers which can be chosen from a list and can directly be used for the agent's utterance. The question-asking part is the main component of the CoffeeBot. According to the casual conversation

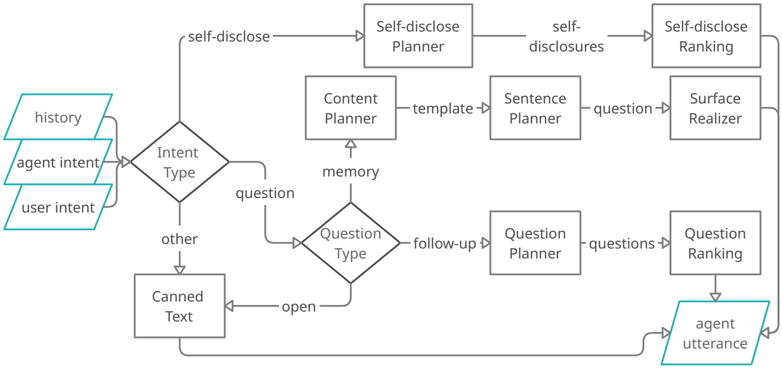


Figure 6.4: The natural language generator (NLG) component in more detail, showing the pipelines for generating agent utterances.

model (Figure 6.2), questions are asked in a conversation in either one of two phases: approach (A) and centering (C).

Templates are used for generation of the questions of the CoffeeBot. A good question generation model is vital to get high-quality responses from users and maintain high engagement in the long term with the CoffeeBot. The question model consists of three types of questions: starter, follow-up and memory. These types map directly to three elements of the casual conversation model: indirect approach, centering and direct approach. During indirect-approach, starter questions are asked. An example of a starter question is “What is the best restaurant in your area?”. In the centering element, follow-up questions are asked. Follow-up questions are constructed with templates from Mandasari (2019) based on the speed date dataset of Huang et al. (2017). An example follow-up question is if the agent asked before “Where did you go to a conference?” and the user responds with “France”, the agent would ask a follow-up question such as “Where in France did you attend a conference?”. During direct approach, the CoffeeBot asks questions related to topics mentioned by a person in previous conversations and retrieves them from its memory. These are questions such as “Last week you talked about attending a conference, how did it go?”. The CoffeeBot uses canned text and two different types of templates for generating these questions. Canned text is used for starter questions, which is directly taken from the list of conversation starters. The first type of templates are follow-up questions generated by a question planner with templates based on semantic role labeling (SRL). The second type, memory questions, are based on a classical NLG pipeline with a content and sentence

6

planner as well a surface realizer to generate the agent utterance.

Additionally, the CoffeeBot has templates for answering questions about itself for self-disclosure (Lundell Vinkler and Yu, 2020). These answers are part of the self-disclosure component, which contains preferences of the CoffeeBot. For example, the CoffeeBot could say “I’m not really into football, but I really like basketball.”, if the user would mention something about sports.

6.5.4.1 Starter Questions

starter questions *Starter questions* are questions that start a new topic or are introductory and are not related to the direct context of a conversation. To the best of our abilities we looked for empirically validated starter questions that people use, but found no such studies related to casual conversation, only for domain-specific instances (Kellermann, 2007; Rothe et al., 2018). Consequently we used an online set of 250 conversation starters as the available starter questions the CoffeeBot could ask.[6] These questions go beyond the standard questions of occupation and hobbies to trigger interesting and entertaining responses, such as “What piece of technology would look like magic or a miracle to people in medieval Europe?”. The CoffeeBot asks these questions mainly in the initial interactions to get to know the user interests, because it does not have a populated user model yet.

6.5.4.2 Follow-up Questions

follow-up questions Huang et al. (2017) conducted a study to investigate the effect of asking *follow-up questions*. Huang et al. (2017, p. 432) define follow-up questions as “questions that encourage the partner to elaborate on the content of their prior conversational turn”. In their study they paired up people online for speed dates via a text-based interface. One of the speed daters was instructed to ask few/average/many questions to the other after which both speed daters rated the quality of their date by stating if they want to go on a second date. Huang et al. found that the people asking the most follow-up questions were liked better overall. They released a dataset which contains annotated types of questions: starter and follow-up questions.

6

Mandasari (2019) has worked on a template-based framework for follow-up question generation based on the annotated follow-up questions in Huang et al.’s dataset. The templates of the framework are based on combining an interrogative type (e.g., why, how, when, where), auxiliary words (e.g., do, not) and semantic roles, by using a natural language parser, SENNA, which can extract semantic role (SR) labels (Collobert et al., 2011) from the follow-up questions in the speed date dataset, such as locations and verb arguments. An example of a template is `WHY + aux + n't + A0 + V + A1?` that can generate the question “Why don’t you have a car?”. In total 514 different types of follow-up questions were

[6] `https://conversationstartersworld.com/250-conversation-starters/`

Table 6.3: Example of a follow-up question generated based on a detected semantic role (Mandasari, 2019).

Interlocutor	Utterance	Semantic Role	Question Type
Agent	Where are you from?	-	starter
User	I'm from the Netherlands.	ARG-LOC	-
Agent	How's the weather in the Netherlands?	-	follow-up

generated from 295 questions in the dataset. After evaluating the questions with external raters based on grammaticality and coherence, 60 templates remained. These templates are used by the CoffeeBot for follow-up question generation if it finds specific patterns of topics in the most recent user response. An example is shown in Table 6.3. Here the agent starts with a starter question, after which the user gives a reply with a recognized location using SRL. This location is part of a specific template in Mandasari (2019)'s model, `ARG-LOC`, for which it has a template that asks about the weather in a specific location.

6.5.4.3 Memory Questions

The third type of questions, *memory questions*, are generated with a similar approach as the follow-up question generation, by writing templates with placeholders for topics. Olafsson et al. (2016) introduced an annotation schema for topic development over multiple conversations. In this annotation schema, there is a distinction between two types of topic changes based on memory: changing a topic based on a previous utterance in the current conversation (reintroducing) or on a previous utterance in a previous conversation (reminding). The memory questions have the same purpose, reintroducing or reminding the user of a topic. In contrast to the follow-up questions, which use only the previous user utterance, the memory questions use the topics in the user model and conversation history as input. The former contains the topics to fill in the question templates. The latter prevents repetitiveness by checking the current conversation's history if the question has not already been asked. Memory questions that have a high semantic similarity with the previous agent and user utterance are more likely to be selected as the next memory question. Memory questions are only asked if the CoffeeBot knows sufficient topics about the user. Additionally, the CoffeeBot indicates in the question why it introduces certain topics by using time markers (Burkert et al., 2010), such as a prefix of "Yesterday" in the question "Yesterday we talked about your conference, how did it go?". The memory question generation uses a classical

memory questions

6

natural language generation approach for questions, consisting of a pipeline of three components (Gatt and Reiter, 2009):

- Content Planner: Select appropriate question template
- Sentence Planner: Select an appropriate question based on the template
- Surface Realization: Perform the surface realization of the question

content planner In the *content planner*, the CoffeeBot determines what template it should select for a memory question. A number of templates was specifically created for the CoffeeBot to ask memory questions to the users. Each template includes a placeholder for any NPs, VPs and SR labels that are related to the interests of the user. Highly frequent topics that have not been asked about (recently) in the conversations with the user are added as possible candidates for the template and one is selected based on frequency, semantic similarity, last time mentioned and sentiment. Additionally, the content planner ranks the available templates based on how well the topic fits the PoS and SR of the template's placeholder. The highest ranking template is then passed on to the sentence planner.

sentence planner In the *sentence planner*, the CoffeeBot determines if it will ask a topic event or topic term question. For a topic event, all syntactic roles (e.g. subject, object, verb) and semantic roles (e.g., locations, adjective modifiers) are generated. In the case of a topic term, the appropriate form of the topic is either an NP or a VP to generate. The template is filled in this stage with candidate topics. Based on the template, the sentence planner generates candidate questions and selects a question from these candidates. The question contains all components to form the question, e.g., subject, object, topic and other sentence components, but does not yet have the proper surface form (all words are in lemma form). In addition, the sentence planner adds a temporal component such as "Last week", if the topic selected by the content planner was mentioned last week. These temporal elements are canned text and selected based on when the last conversation took place with the selected topic. According to Burkert et al. (2010), it is good to have temporal components as long as they are not too precise. Therefore the most precise temporal marker is "today" or "last night", without mentioning specific hours or minutes.

surface realizer A *surface realizer* takes a question from the sentence planner and generates an utterance. SimpleNLG is a surface realizer designed for text generation (Gatt and Reiter, 2009). It has been extended in multiple languages, including German, French and Dutch (De Jong and Theune, 2018). The SimpleNLG software library was developed to perform surface realization of sentences, and it takes care of the structure of the sentence and the morphology of the words. SimpleNLG can generate grammatical sentences from the questions of the sentence planner, because the sentence planner contains information about the appropriate syntactic roles for the structure in SimpleNLG. SimpleNLG can match the correct PoS and

SR of the topics to the placeholders in the templates and takes care of the surface grammar. For example, if the sentence planner passes on the sentence temporal marker "yesterday", interrogative "how", subject "you", verb "like" and topic term "visit the conference", the surface realizer generates the question "How did you like visiting the conference yesterday?".

Unfortunately, SimpleNLG does not have great support for the generation of questions, *interrogatives*, because it was not designed for generating questions in a dialogue. However, we modified SimpleNLG's approach to generating questions such that it can generate questions that are usable by the CoffeeBot. The CoffeeBot's language is English, and therefore we only focused on the English question generation by SimpleNLG. The first problem was that it did not support all the question words, such as *when* and *which*. The second problem was that for structuring the sentences, subjects and objects were added in the wrong order. For example, SimpleNLG would generate questions such as *How did you the letter give?*. The first problem was relatively easy to solve by just adding the question types to the interrogative types of SimpleNLG. The second problem required the implementation of a new structuring algorithm for interrogative types and add it to SimpleNLG. These additions are available online for other researchers interested in using interrogatives in SimpleNLG for their dialogue system.[7] A Dutch version for generating interrogatives in SimpleNLG was developed for supporting the BLISS project as well.

interrogatives

An example of a memory question is shown in Table 6.4. Turn 1 shows the user mentioning a temporal argument (`ARG-TEMPORAL`), and a candidate topic mention, "a conference about human-robot interaction", which are saved in the user model. A week later the agent remembers the topic from the last conversation and uses a template "How was your `ARG-TOPIC` `ARG-TEMPORAL?`". Note that because the user mentioned a time marker, this can be retrieved from the user model and the timestamp gets translated to the correct surface format. Because a week has passed since the topic mention, the selected phrase for the time marker is "a week ago".

6.5.4.4 Question Model

Our question model is loosely based on the topic-schema developed by Olafsson et al. (2016). Their schema includes topic shifts and introduction as in the analysis of Rats (1994), but extends to long-term interaction with reintroducing topics or reminding users of topics. For determining the question type in the NLG component, the CoffeeBot connects with a remote question server to retrieve one of the three types of questions. The algorithm for selecting the question is shown in Algorithm 1. The questions always starts with an introductory starter question

[7] `https://github.com/Barachia/SimpleNLG-NL`

Table 6.4: Example of a memory question generated based on a *time marker* and **topic** in the user model.

Interlocutor	Utterance	Semantic Role	Topic
User	*Tomorrow* I will attend **a conference about human-robot interaction**.	ARG-TEMPORAL	conference
...	(One week later)		
Agent	How was your conference about human-robot interaction last week?	-	conference

for both new users and familiar users. It asks a number of follow-up questions after a starter or memory question, either until a random number between 3 and 5 is reached or if there are no possible follow-up questions to generate. It then randomly asks a starter question or a memory question. If it chooses a starter question, it does not repeat recent questions. If it chooses a memory question, it proceeds with an exhaustive vs. explorative approach, where in some cases it chooses a topic that has been discussed many times before and is sure to be a user interest (exhaustive) and other times it picks a topic that has only been mentioned at most a couple of times (explorative), as long as the sentiment is positive. A correct balance between the two approaches is necessary to show personalization based on frequent topics that are definitely a user interest and learning more about the user by asking about infrequent topics (Campos et al., 2018). At first, the CoffeeBot picks explorative topics, because the user model is not populated with enough topics yet. After enough topics have been learned, the odds of asking exhaustive or explorative are equal (fifty-fifty).

6.5.5 Behavior Generation

Once the NLG module has generated a sentence, this sentence is sent from Flipper to a RESTful API in SSML format to a supported TTS service. Once Flipper retrieves the sentence audio file from the TTS service, this is played back to the user. The CoffeeBot supports Google Cloud Text-To-Speech (Weiss et al., 2021), MaryTTS (Schröder et al., 2011) and ReadSpeaker voices. As well as for the ASR, the choice to pick one is a trade-off between naturalness, customizability of the voice and data privacy. The Google voices provide the best experience with regard to naturalness with a Wavenet voice. The MaryTTS voices are fully customizable and are more privacy-friendly, because no data is stored at a third party, as is the case

Algorithm 1 The CoffeeBot's algorithm for selecting a question.

$followups \leftarrow 0$
$followuplimit \leftarrow rand(3,5)$
$timeout \leftarrow 360$
$goodbye \leftarrow FALSE$
$question \leftarrow starter$
$askQuestion \leftarrow question$
$answer \leftarrow usertext$
while $\neg goodbye$ **AND** $\neg timeout$ **do**
    if $answer$ suitable for follow-up **AND** $followups < followuplimit$ **then**
        $followups \leftarrow followups + 1$
        $question \leftarrow followup$
    else if memory has topic **AND** topic freq. $\geq$ 5 **AND** topic sent. $\geq$ -0.2 **then**
        $question \leftarrow memory$
        $followuplimit \leftarrow rand(3,5)$
    else
        $question \leftarrow starter$
        $followuplimit \leftarrow rand(3,5)$
    end if
    $askQuestion \leftarrow question$
    $answer \leftarrow usertext$
end while

with Google. The ReadSpeaker voices are somewhere in between, which offers a good balance between customizability, privacy protection and voice quality. With regard to privacy, data is shared with ReadSpeaker servers, but no data is stored.

6.6 Conversation Flow

Whenever the user approaches the CoffeeBot and let it scan their RFID card, there are two ways the conversation can start (Figure 6.5). When a card has been scanned, the CoffeeBot has to determine whether it has met this person before. If that is not the case, the agent briefs the user on its goal and type of data collection. Participants can also stop the interaction altogether by saying "goodbye". If the user is known to the CoffeeBot, or it has introduced itself to a new user, the CoffeeBot asks a starter question. The user can either answer the question, after which the CoffeeBot asks another question. If the user does not give an answer, but instead asks a question, the CoffeeBot answers the question if it knows the answer. If it does not know the answer, it states it does not know and continues with asking the next question. If the user or CoffeeBot is done with the conversation, either can end the conversation. The user can also refrain from responding to the CoffeeBot. The

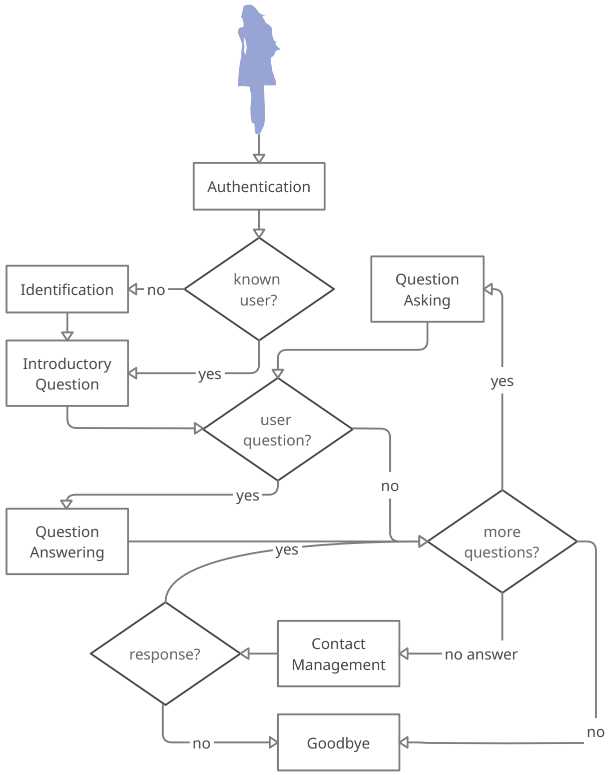


Figure 6.5: Dialogue flow of a conversation. The CoffeeBot checks if it knows the user, after which a series of questions start until time runs out or the user ends the conversation.

CoffeeBot then tries to search for contact again for two tries. After that, it closes the conversation. An ideal example of a possible conversation between a user and the CoffeeBot is shown in Table 6.5.

6.7 Discussion

We reflect on some features of the CoffeeBot and how it can be extended to be better at casual conversation. For example, the tracking of topics in a dialogue technique is relatively simplistic in the CoffeeBot. Though the tracking by means of phrases and frequencies makes it robust to ASR errors, more implicit topic information is lost in the recognition. Online resources can provide a larger range

Table 6.5: Possible interaction between a user and the CoffeeBot.

Interlocutor	Action	Utterances
User	Walks to the coffee machine and scans card	
CoffeeBot	Identifies a person	
CoffeeBot	Greets the user and asks them to participate	Hi! Are you up for a chat?
User	Greets and confirms to participate	Hi, sure.
CoffeeBot	Ask question	What did you do yesterday?
User	Answers question	I played squash.
CoffeeBot	Asks follow-up question	Do you play squash regularly?
User	Answers question	Yes
CoffeeBot	Asks starter question	What animal do you really hate?
User	Answers question	I really don't like mosquitoes.
CoffeeBot	Preclose	Let's talk more later!
User	Responds + greets back	Ok. Bye
CoffeeBot	Greets back	Bye bye
User	Walks away	
CoffeeBot	Store conversation and prepare for new conversation	

of possible topics. VerbAtlas is focused on creating a richer way than SENNA to represent semantic relations of verbs and universal dependencies, including more specific types of arguments and semantic relatedness between verbs (Di Fabio et al., 2019). Question templates based on VerbAtlas can be more diverse, because the arguments in VerbAtlas have more specific properties to deal with context. With VerbAtlas, verbs in question templates can be replaced with semantically similar verbs with the same type of arguments and context, which is not possible with SENNA. For example, the verb distinguish has multiple synonyms such as differ, perceive and recognize. In the sentence, "I distinguish four different topics", distinguish can really only be replaced by differ to have the same semantic meaning. Another source for a larger list of possible topics is Wikipedia. Wikipedia contains a vast amount of information and each page can serve as a possible topic (Breuing et al., 2011). A feasible extension of the CoffeeBot's generation module for language is a sequence-to-sequence method with self-attention for explicitly modeled personality traits, such as in the studies of Zheng et al. (2020) and Liu et al. (2020). However, the studies were evaluated with a single response to a prompt, and it is difficult to estimate how well this method would work in longer conversations.

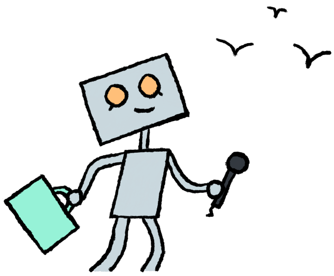
Topics and their sentiment are based on the valence scores of the Pattern library (De Smedt and Daelemans, 2012). Sentiment detection with Pattern is rather context-insensitive and not robust, therefore it can be beneficial to add another framework such as VADER (Hutto and Gilbert, 2014) or LIWC (Kahn et al., 2007) for more robust sentiment detection. Moreover, not only the textual sentiment of the user utterance should play a role in determining whether the user likes a topic or not. Using sentiment recognition such as with openSMILE in the SSI toolkit (Chapter 3) could provide more robust feedback indicating whether a user actually likes a topic that is being discussed or not (Wagner et al., 2013).

In the current question model and conversation flow, the agent asks many questions. Though the CoffeeBot can answer some user questions and talk about itself with self-disclosure, the conversation can get a bit one-sided and less engaging. Paranjape et al. (2020) developed an Alexa voicebot with an opinion-module that leads to a more engaging agent to talk to for users and includes a larger database of knowledge for answering user questions. Their voicebot was oriented towards more mixed initiatives, though they found that there was big cognitive load on the users who had to take the initiative. It might therefore be worthwhile to research how much a social agent should take or give the initiative as well. Adding a story telling capability for the CoffeeBot such as in the work of Gockley et al. (2005) could have two advantages: increasing user engagement by making them curious about the story and balancing the type of conversation between user and agent.

The current forgetting component in the CoffeeBot simply forgets a topic if that topic is not mentioned anymore after one month. A more complex memory component should remember static information, such as a name, that the CoffeeBot should never forget, but also it should be possible for users to ask if the agent would remove topics from its memory (Lim, 2012). For example if a topic is too embarrassing or if it is incorrect. However, this mechanism of forgetting should be built with caution and preferably includes an accessible option for the user to check and correct their data that is known to or about to be forgotten by the CoffeeBot.

The CoffeeBot is a purely spoken dialogue system with a speaker, but does support animated behavior such as facial expressions or gestures. The TTS component can be extended to also include support for BML to support animated behavior in some robots and virtual agents. BML can be mapped directly onto virtual humans, such as SmartBody, Unity3D or Greta (Thiebaux et al., 2008; Kolkmeier et al., 2017a; Poggi et al., 2005) and robots such as the Zeno (Davison et al., 2020) and NAO (Le and Pelachaud, 2012). An embodiment with more features would increase engagement of the user interacting with the CoffeeBot.

To conclude, we used the tools of dialogue design and prototyping from the previous chapters to develop a dialogue structure for long-term casual conversation. We view our approach to casual conversation modeling with the CoffeeBot as only a first step in developing interactions with casual conversation. We focused on the aspect of question-asking, but much more research is necessary with this type of conversation and how we can use it to enrich interactions with social agents (Gilmartin et al., 2018). The use of non-verbal cues and how to handle multi-party casual conversations are other interesting questions in the field of casual conversation. In the next chapter we discuss the results of how the casual conversation model worked in the real world.



7 Evaluating Question-Asking in Casual Conversations in the Real World

This chapter is based on this paper:

- **J. van Waterschoot** and M. Theune (2021). “Evaluating Conversational Question Generation: CoffeeBot”. In: *Proceedings of Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI 2021)*. ACM, p. 6

7.1 Introduction

Evaluating the performance of interactions with social agents is no easy task. Often to evaluate these interactions, researchers perform lab studies in controlled conditions. However, these lab conditions might not necessarily apply to the real world. The real world has often more variability than researchers can account for in controlled lab studies, making it harder for researchers to validate their intervention. In a recent survey of the Intelligent Virtual Agent conference proceedings, only 20 out of 276 studies were field studies (Norouzi et al., 2018). One of the major issues with many social agents is their deployment in the real world. More often than not, after a research project with social agents has ended, the robot or agent is locked away and never to be seen again. Though open platforms exist such as the ones discussed in Section 2.2.1, 4.2 and the ARIA framework in Chapter 3, many platforms are proprietary, poorly documented or not accessible. As a result, social conversational agents for real world deployment are built from scratch. Recently, more platforms are published with open-source code together with tutorials, which increases the accessibility for other users (Ultes et al., 2017; Bohus et al., 2017; Li et al., 2020). The HRI community has made an effort to move towards more deployment of robots and agents in the real world (Rosenthal-von der Pütten et al., 2016). The theme of the Human Robot Interaction conference in 2020 was *Real World Human-Robot Interaction*, focusing on real world application of agents. However, real world deployment does not always meet the ecological validity one can expect outside the laboratory. For example, in Heylen et al. (2012), the deployment of a social robot at people's home which did not meet all expectations of participants lowered their opinion of the robot. Additionally, people still felt like they were taking part in an experiment and dutifully interacted with the robot, which is not representative of a natural setting.

In this chapter we review existing real world evaluation methods. Our goal is to find measures that are usable in real world studies and focus on evaluating the user experience and personalization. In Chapter 2.3 we mentioned Fan and Poole (2006, p. 183)'s definition on personalization, which is "a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its personal relevance to an individual or a category of individuals". In the case of the CoffeeBot, personalization is the process of asking questions that are personally relevant for each individual. The CoffeeBot uses similar language as the user, or talks about topics the user is interested in. After highlighting different measures for user experience and personalization, we discuss the setup of a pilot for a real world study with the CoffeeBot introduced in Chapter 6. Finally, we discuss the results of the pilot for the CoffeeBot study and its implications for a design of a full real world study.

7.2 Background

In this section we discuss i) available frameworks and methods for the evaluation of social agents, ii) questionnaires for measuring acceptance and usability and iii) measures for long-term real world evaluation.

7.2.1 Frameworks for Social Agent Evaluation

evaluation framework

Several evaluation frameworks for social agent evaluation are shown in Table 7.1. Choosing a framework helps focus on *what* we want to measure without considering the *how* in too much detail. PARADISE (Walker et al., 1997) is seen as one of the first commonly used *evaluation frameworks*. It focuses on two aspects of dialogue: task success and minimization of costs. The former is not relevant for the CoffeeBot: there is no task to succeed in. The latter is divided into two categories: efficiency and quality measures. Examples of how to measure efficiency are number of utterances and total dialogue time, and examples of quality are repair ratio and inappropriateness of agent responses. However, PARADISE does not capture social aspects of the interaction.

Table 7.1: Frameworks for evaluating social agents

Framework	Focus
PARADISE (Walker et al., 1997)	task success, minimize costs
USUS (Weiss et al., 2011)	usability, acceptance and societal impact
Social Acceptance Model (De Graaf et al., 2019)	usefulness, use intention, norms and beliefs
USUSGoals (Wallström and Lindblom, 2020)	usability, acceptance, societal impact and user experience

Weiss et al. (2011) captured usability, acceptance and societal impact together in their theoretical framework, Usability, Social acceptance, User experience and Societal impact (USUS). Similarly to the work of Heerink et al. (2010), USUS is focused on robots from a utilitarian perspective (De Graaf et al., 2019). De Graaf et al. (2019) propose a "social acceptance model", which also includes users' norms and beliefs about social interaction and a focus on whether users would use the social agent again in the future. Wallström and Lindblom (2020) expanded on Weiss et al.'s framework, calling it USUS Goals and added user experience as an important part of evaluation, especially for long-term evaluation. User experience factors include the embodiment of the robot, the perception of the robot by the user, the feeling of security and trust, the affect of the robot and the co-experience of the user and robot during interaction.

7.2.2 Questionnaires for Acceptance and Usability of Social Agents

The common way of evaluating social agents with users is by providing a questionnaire that can consist of a list of statements and/or questions with respect to users' experience with a system. Questionnaires are widely used within HRI as a self-report method to measure different phenomena (Rueben et al., 2020). Using a validated questionnaire is beneficial for researchers, because the results of the study can be comparable and researchers do not have to develop and validate their own method. However, researchers should be aware of the limitations that come with reusing or adapting *questionnaires*. In this section about evaluation, we discuss the development of questionnaires for users and technology in HRI research. An overview of all considered questionnaires for acceptance and usability in human-agent interaction is listed in Table 7.2, with an example question from each of the questionnaires next to it.

questionnaire

7

In the early stages of HRI research, questionnaires that were used in general

Table 7.2: Questionnaires used for evaluating acceptance and usability of social agents.

Method	Focus	Example
Systems Usability Scale (Brooke, 1996)	Effectiveness, efficiency and satisfaction	I thought the system was easy to use.
Software Usability Measurement Inventory (SUMI) (Kirakowski and Corbett, 1993)	Usability	I would not like to use this software every day
Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988)	Mood	Likert rating of *Interested*
Technology Acceptance Model (TAM) (Davis, 1989)	Acceptance	I find it cumbersome to use [TECHNOLOGY]
Negative Attitude toward Robot Scale (NARS) (Nomura et al., 2006)	Negative mood	I would feel uneasy if robots really had emotions.
Social presence (Jung and Lee, 2004)	Engagement	How much did you feel as if [TECHNOLOGY] was communicating with you?
Godspeed (Bartneck et al., 2009)	Perception of robot	Likert rating of Fake —– Natural
Almere model (Heerink et al., 2010)	Acceptance with context	I'm certain to use the robot during the next few days
Robotic Social Attribute Scale (RoSAS) (Carpinella et al., 2017)	Perception of robot and its social membership	Likert rating of *Competent*

for technology were applied for social agents as well. One of the first general purpose technology scales includes the Systems Usability Scale (SUS) (Brooke, 1996). Brooke (1996) developed a small quick and dirty scale for testing technology with end-users, with only 10 statements on a 5-point Likert (Likert, 1932) scale, including statement such as "I thought the system was easy to use". The focus of the scale was measuring effectiveness, efficiency and satisfaction of the technology. Kirakowski and Corbett (1993) focused on usability of technology using the Software Usability Measurement Inventory (SUMI) scale, which has 50 statements such as "I would not like to use this software every day." Another model focusing on the acceptance of technology is the Technology Acceptance Model (TAM) (Davis, 1989). Statements according to this model include "I find it cumbersome, to use [TECHNOLOGY NAME]" about perceived ease of use and usefulness. Hone and Graham (2001) developed the Subjective Assessment of Speech-System Interface (SASSI) questionnaire for evaluating usability of speech-based system, with statements like "It is clear how to speak to the system" and "I was able to recover easily from errors".

One of the most commonly used questionnaires in the HRI community is

Bartneck et al. (2009)'s Godspeed, an instrument designed for measuring users' perception of robots based on five concepts of the constructs anthropomorphism, animacy, likeability, perceived intelligence and perceived safety. The questionnaire has been translated into multiple languages. Carpinella et al. (2017) state that the Godspeed questionnaire is incomplete for HRI research, because it omits factors such as robots' perceived social category membership and focuses on specific designs of robots. Starting with the work of the Godspeed questionnaire and literature on social perception and categories in psychology, the Robotic Social Attribute Scale (RoSAS) measure has been developed. The three RoSAS factors are warmth, competence and discomfort. The measure has been validated for HRI by Pan et al. (2017), though with only a small sample.

The authors of the Almere model for acceptance considered the TAM too limited for social robots and included more factors for acceptance, mostly related to the context (Heerink et al., 2010). The Almere model includes questions about long-term use, though these statements are limited to the near future, such as "I'm certain to use the robot during the next few days". De Graaf et al. (2019) consider Heerink et al. (2010)'s model to be too limited for rating the acceptability of social robots. One of their arguments is that the Almere model is not theoretically well-grounded with clear indicators why certain factors of acceptability are included or excluded. Another argument De Graaf et al. make is that the Almere model has only been tested on specific user groups, and the dataset it was tested on was too heterogeneous to generalize.

Acceptance can also be measured through the positive and negative affect people have for social agents. The Positive and Negative Affect Schedule (PANAS) has been used to measure people's mood through self-reporting, also in the long-term (Watson et al., 1988; Rosenthal-von der Pütten et al., 2013). Similarly, the Negative Attitude toward Robot Scale (NARS) measures social anxiety of people wanting to use social robots (Nomura et al., 2006). Jung and Lee (2004) found that an agent being physically co-located with a user can have a positive effect on engagement and thus impact willingness to interact with a social agent again.

Werner (2020) looked at real world evaluation methods of social agents in 24 European projects. Most of the evaluation methods discussed in the survey are aimed at evaluating an assistive companion for older adults, a participant's quality of life with the technology or are developed to push the field of HRI forward. Werner (2020) found that many of the evaluation methods have not been properly tested due to technical issues of a prototype, and thus the technology has to first catch up with the method before the method can be validated (Pripfl et al., 2016). Moreover, people participating in real world studies still notice that they are being observed and exhibit the same socially acceptable behavior as in a lab study (Heylen et al., 2012). Werner (2020) also states that some methods have only been tested in studies where the participants are too heterogeneous or too small to draw conclusions from (Pripfl et al., 2016; Östlund et al., 2015). Finally, there are few standardized methods in HRI research, and those that exist, such as the Almere model, are oriented towards *only* acceptance of social agents and not future use (Heerink et al., 2010).

Criticisms of questionnaires should be taken into consideration. A critical review by Krägeloh et al. (2019) discussed six questionnaires for acceptance of social robots that have been validated, which includes NARS and RoSAS, and found that all of these scales have limited or inconsistent evidence thus far. Questionnaires that are too long will become a nuisance to participants (Bickmore et al., 2010).

7.2.3 Long-term Evaluation

Whereas the above methods are commonly used for evaluating user perception of social agents, they have not been specifically designed towards long-term interaction in the real world . However, for effective evaluation of the CoffeeBot long-term interaction is a requirement. In this section we will highlight measures that have been used in previous studies that conducted field studies.

Questionnaires are the most commonly used metric in long-term studies. They are used for measuring friendship, social support and social presence (Leite et al., 2014), user attitude (Gunson et al., 2020), perceived safety (Pripfl et al., 2016), repetitiveness (Bickmore et al., 2010), interpersonal relationship (Schulman, 2013; Campos et al., 2018), immediacy (Schulman, 2013), (communication) satisfaction (Mattar and Wachsmuth, 2014; Coronado et al., 2018; Campos et al., 2018; Bajones et al., 2019; Gunson et al., 2020; Irfan et al., 2020), language style (Gunson et al., 2020), enjoyment (Bickmore et al., 2010; Gunson et al., 2020), engagement (Bickmore et al., 2010; Leite et al., 2014), intelligence (Campos et al., 2018), usefulness (Kanda et al., 2010), flexibility and usability (Bajones et al., 2019) and continuation of conversation (Bickmore et al., 2010). In these studies, it is mostly users themselves or observers evaluating the interaction, but sometimes crowdsource workers are used for large-scale evaluation with questionnaires, to retrieve data with more power to draw conclusions from, though these crowdsourcing studies have their limitations (Santhanam et al., 2020). A downside of questionnaires is that they often measure attitude and not actual behavior. Questionnaires are susceptible to socially acceptable answers, and people are not always the best in self-reflecting on their beliefs and actions. Therefore besides questionnaires other measurements are necessary.

Observations of human behavior can tell you much about how the technology has been accepted, and especially in the long-term, about how users form relationships with social agents (Huttenrauch and Eklundh, 2002; Grandgeorge, 2020).

observations

Behavior *observations* can be used for measuring failures, moods, motivation, reaction and communication quality in multi-user robot interactions (Huttenrauch and Eklundh, 2002; Hebesberger et al., 2016), analyzing micro-behaviors and sequential analysis (Sabanovic et al., 2006; Irfan et al., 2020) and measuring engagement (Sabanovic et al., 2006; Kanda et al., 2010). Additionally, familiarity, intelligence and perceived familiarization are measured by annotating recordings of behaviors (Kanda et al., 2010).

interviews

Interviews are great for getting results that are not limited by a questionnaire's statements and questions, though they take more time and should be carefully constructed (Galvão Gomes da Silva et al., 2018; Bethel et al., 2020). Leite et al. (2014) used semi-structured interviews for evaluating user attitude towards a robot and Weiss and Hannibal (2018) measured user reaction and satisfaction with interviews during house visits. Interviews can reveal much information about usefulness (Huttenrauch and Eklundh, 2002; Hebesberger et al., 2016; Pripfl et al., 2016; Weiss and Hannibal, 2018) and usability (Pripfl et al., 2016; Bajones et al., 2019) of social agents as well. For both observations and interviews the analysis costs are higher than with quantitative methods (e.g., questionnaires), even with more automated analysis being possible with today's technology, such as speech transcriptions and image recognition. One other qualitative method that helps a project in the pilot stage are focus groups, mentioned by Davison et al. (2020). Evaluating a system with domain experts, or focus groups, helps to prevent common pitfalls in social agent design and can be used to make sure the target group is open to trying out the technology, as was done in Clark et al. (2019).

thematic analysis

Another evaluation method in long-term studies is that of *thematic analysis*, where themes are systematically determined from recorded data, usually by annotating (Braun and Clarke, 2006). For example, Campos et al. (2018) looked at the diversity of topics as a measure for analyzing the interpersonal relationship between user and agent. For the domain of chatbots, Venkatesh et al. (2018) and Paranjape et al. (2020) looked at topical diversity across the conversations. Weiss and Hannibal (2018) also asked participants to keep a diary of activities as a reference to the human-agent interaction data. Additionally, the logs of systems can give insight into certain structures of conversations (Bajones et al., 2019).

ratio data

Finally there are methods that fit into the PARADISE (Walker et al., 1997) framework component of cost minimization. These do not directly measure user attitudes, but can provide an indication of the overall user experience through quantitative analysis of *ratio data* from logs. Generic methods include the duration of words (Bickmore and Cassell, 1999; Schulman, 2013), number of prior conversations (Schulman, 2013; Campos et al., 2018), completed interactions (Schulman, 2013; Kennedy et al., 2017), number of interactions per person, (Gockley et al., 2005; Kanda et al., 2010; Kennedy et al., 2017; Trinh et al., 2018; Coronado et al., 2018), conversation duration (Gockley et al., 2005; Venkatesh et al., 2018; Gunson et al., 2020), errors made by the system (Venkatesh et al., 2018), number of turns (Campos et al., 2018; Gunson et al., 2020; Paranjape et al., 2020), user and system utterance length (Paranjape et al., 2020), word count (Gratch et al., 2007), number of user utterances (Trinh et al., 2018), total number of interactions (Gockley et al., 2005), number of participants (Gockley et al., 2005) and user ratings (Gockley et al., 2005; Leite et al., 2014; Hebesberger et al., 2016; Venkatesh et al., 2018; Paranjape et al., 2020).

Most of these data can indicate forms of engagement or system performance. However, these measures are highly dependent on context. For example, neither many nor few turns are indications of good engagement or performance per se. In task-oriented dialogues, fewer turns could be more efficient, but some users might want to express a single intent per turn, as opposed to multiple intents in a longer turn. For casual conversation there is no direct relation between turns and engagement. Some users prefer to have frequent, but very brief conversations, whereas others might enjoy having lengthy discussions. More specific cost measures include percentage of user questions and repeat requests for clarity of the dialogue (Trinh et al., 2018), tasks completed in task-oriented dialogues (Trinh et al., 2018; Gunson et al., 2020; Irfan et al., 2019), suggestions of the system followed (Campos et al., 2018; Coronado et al., 2018) and acknowledged by the user (Coronado et al., 2018) for usefulness, and recurring topics (idiosyncrasy) for measuring interest (Kennedy et al., 2017).

Ganster et al. (2010) emphasize the difficulty in comparing HRI studies, since many measures are used. For questionnaires, Godspeed (Bartneck et al., 2009) was a first attempt to standardize measures for HRI studies and later RoSAS (Carpinella et al., 2017) was introduced to include more discomfort and warmth aspects of HRI. Ganster et al. (2010) claim that many studies do not utilize temporal aspects, such as memory in agents, or do not measure regularly over a longer period of time.

There does not seem to be any direct metric for personalization in longitudinal studies. Measures for engagement and empathy come close and probably are the most usable for evaluating personalized interactions (Leite et al., 2014; Bickmore et al., 2010). Most studies do not contain a memory component that makes effective use of its repeated interactions with participants (Leite et al., 2013). Measures for assessing the interpersonal relationship would be useful, but they need to be generalized to be applicable to any type of longitudinal study. For example, the Working Alliance Inventory (Horvath and Greenberg, 1989) measures the interpersonal relationship between therapist and client, but it remains to be seen if these types of domain specific questionnaires can be generalized. For measuring closeness in dyadic conversations, Hecht (1978)'s Interpersonal Communication Satisfaction Inventory (ICSI) questionnaire has been used before in personalization and long-term studies with social conversational agents (Mattar and Wachsmuth, 2014; Campos et al., 2018; Skjuve and Brandzaeg, 2019).

7.3 Aim

The aim of the CoffeeBot is to see whether question-asking from memory in casual conversation can lead to more personalized interactions. We deployed the CoffeeBot for a pilot study in which we focused on usability of the system. By first piloting the real world study, we hope to prevent many of the issues experienced in other real world studies (Werner, 2020). The main interests in this pilot are i) the interaction design of question asking, ii) detecting issues in the setup and iii) usability of the measures. The most important aspect of the setup is deploying the CoffeeBot in the real world for a longer period of time, at least for a month, with a weekly casual conversation with each participant.

7.4 Method

Instead of using a virtual agent for embodiment (Valstar et al., 2016), we opted for a physical robot, which would draw more attention in a public space (Segura et al., 2012). Additionally, a virtual human might set expectations higher than a robot with less humanoid characteristics. We also wanted to emphasize the speech capabilities rather than other behavior. Therefore we designed a low-cost low-fidelity prototype that could be deployed for a longer period of time in a real world public place, such as a coffee shop or near a coffee machine.

7.4.1 System setup

The CoffeeBot was embodied as a silver-colored cardboard cut-out robot, shown in Figure 7.1. The head of the robot contained a Bluetooth speaker, used for recognition of the user speech and the production of speech synthesis for the CoffeeBot. The lower body contained an Arduino Uno, with an MFRC522 connected for scanning RFID (radio-frequency identification) cards, similar to the identification method used by Davison et al. (2020). Both components were connected to a laptop, hidden out of sight, that ran the CoffeeBot's autonomous core system. A remote server ran NLU and NLG components to reduce the strain on the laptop. The agent used Google Cloud Speech for ASR and ReadSpeaker's British English voice James for TTS.

7.4.2 Participants

There was no specific target group for the CoffeeBot, except that participants had to be i) 18 years or older and ii) 2) had to be relatively fluent in English. We deployed the CoffeeBot at two different locations in late 2020, one at a university of applied sciences (L1) and one at a university college (L2). All participants in the study were students attending either of the universities. Before interacting

(a) The CoffeeBot located in a coffee shop at location L1.

(b) The CoffeeBot located near the coffee machine at location L2.

Figure 7.1: The two locations of the CoffeeBot evaluation.

with the CoffeeBot located at L1, students filled in an informed consent form. At L2 the informed consent could only be given by scanning a QR code that had a digitalized version of the informed consent form, which was stored on our university's server. The Ethics Committee of our faculty approved the forms and checked that all data collection was in compliance with university policy and General Data Protection Regulation (GDPR). At both locations, participants were either recruited personally by us or signed up through a link/QR code near the CoffeeBot that lead to an online consent form.

Recruitment was less than ideal with many of the staff and students working from home (most of the time) during the COVID-19 pandemic. During recruitment and deployment on-site, we wore masks and kept 1,5m distance. We put up a note near the CoffeeBot that only one person at a time could talk to it. We asked participants to talk to the CoffeeBot at least three times, preferably on different days, but it was up to participants when and if they would interact with the CoffeeBot.

7.4.3 Procedure

At L1, the CoffeeBot was placed on the side of the room, where a participant could sit down. For L2, the CoffeeBot was placed next to an automated coffee machine in a central location. Participants initialized the conversation by holding their card or key with RFID in front of the CoffeeBot's body.

In the first interaction, the CoffeeBot would introduce itself and its goal of wanting to get to know the participant. It also gave the instruction that the participant could end the conversation by saying "goodbye". After the introduction, the CoffeeBot started a round of questions. The user might answer these or not, but at the end of the user's turn, the CoffeeBot asked another question, either as a follow-up on what the user said or introducing a new topic. The conversation could go on for up to 6 minutes, after which the CoffeeBot ended the conversation with a message to hopefully see the user again soon, or the user could end the conversation earlier by saying "goodbye". In subsequent conversations with the same participant, the CoffeeBot would start by saying it would ask questions and continue to do so until either the CoffeeBot or participant ended the conversation.

7.4.4 Data collection & Measurements

In the pilot experiment we collected data from three different sources: i) from the CoffeeBot itself, which made speech recordings, transcriptions and logs from its interactions with the user, ii) two questionnaires filled in by users and iii) semi-structured interviews held at the end of the experiment by the researcher.

7.4.4.1 Recordings, transcriptions and logs

The speech recordings were saved in 16-bit PCM wav format named by date and user ID. Each recording started when a user scanned their RFID card and stopped when the user or CoffeeBot ended the conversation. Additionally, the CoffeeBot saved the transcriptions of the speech and logs of the interaction in a database in JSON format. The logs contain the meta-data of the conversations, such as the user ID, the number of interactions, the sentiment levels of user utterances and the dialogue history with timestamps. We also calculated how many topics mentioned by the user were not recalled by the CoffeeBot and how many follow-up questions and self-disclosures were incorrectly used.

7.4.4.2 Questionnaires

We compiled two questionnaires, which took about 4-5 minutes each to complete. One was sent to participants after their first interaction, and the second one was sent after the experiment ended. The second questionnaire contained additional questions about how often people had been at the university during the run of the experiment and how often they had left their working place for a break (see Appendix A.2). We asked these additional questions to compare the number of opportunities there were for interacting with the CoffeeBot with the number of actual interactions. Both questionnaires contained the same items, which were taken from four different questionnaires, as described below. All questionnaires were standardized to a 7-point Likert scale.

- ICSI (Hecht, 1978)
- McGill Friendship Questionnaire (Mendelson and Aboud, 1999)
- Social presence (Jung and Lee, 2004)

- RoSAS questionnaire (Carpinella et al., 2017)

The first two questionnaire components focused on the *personalization* aspect of the CoffeeBot in the open-domain casual conversation. The first was based on the ICSI measure by Hecht (1978), which has been specifically designed for dyadic conversations between either friends, strangers or acquaintances and can measure closeness in computer-mediated communication (Tidwell and Walther, 2002) (Appendix A.2.1). ICSI has been applied before in HRI research, in a study about conversational memory (Campos et al., 2018) and for measuring user experience (Skjuve and Brandzaeg, 2019). The second component was based on the McGill Friendship questionnaire (Mendelson and Aboud, 1999), which was used by Leite et al. (2014) to evaluate how helpful and encouraging the participants found their social robot. Four out of the six categories of the McGill Friendship questionnaire did not apply to the CoffeeBot's type of casual conversation: help, reliable alliance, self-validation and emotional security, so we did not include these in our questionnaire (Appendix A.2.2). The two categories we did include were intimacy and stimulating companionship. Intimacy is about being honest, expressing yourself and how comfortable you are with sharing personal information. The items for stimulating companionship measure how enjoyable the conversation is. They served as an indicator for engagement and if people wanted to talk again to the CoffeeBot.

The purpose of the other two questionnaire components was to measure the quality of the CoffeeBot in general: RoSAS for the perception users had of the CoffeeBot (Carpinella et al., 2017) and a social presence questionnaire for measuring their engagement (Jung and Lee, 2004). We used the RoSAS questionnaire because it has been used commonly as an evaluation tool for social robots and agents and is well-known to HRI researchers (Krägeloh et al., 2019) (Appendix A.2.4). We are aware of Werner (2020)'s statement that the validation evidence for RoSAS has been rather limited (Krägeloh et al., 2019; Pan et al., 2017), but we believe it does strike a good balance in completeness (i.e., measuring the discomfort) and as a standard measure in HRI.

There are different interpretations of social presence in a human-human context (Short et al., 1976), though we used the perceived social presence of the robot as a measure of engagement, similarly to Jung and Lee (2004). We measured the social presence of the CoffeeBot with Jung and Lee (2004)'s questionnaire to gauge the feeling of users being socially present with the CoffeeBot (Appendix A.2.3).

7.4.4.3 Interview

At the end of the experiment, together with the second questionnaire, an email invitation for a semi-structured interview was sent to all the participants to informally talk about their experience with the CoffeeBot. See Appendix A.3 for the leading interview questions.

7.5 CoffeeBot Pilot Results

Thirteen people gave informed consent to participate in the experiment. Of these participants, 2 never talked to the CoffeeBot. Out of the 11 people who did interact with the CoffeeBot, 6 people interacted with it more than once and 5 interacted with it only once. Of these 5 participants, only one interacted with the CoffeeBot with more than 2 (user) turns. If we filter out participants who had no more than one turn per interaction, 7 participants (3 male, 2 female, 2 non-binary) remain. The results we describe relate to these final 7 participants. An anonymized example of the conversation is shown in Table 7.3.

The CoffeeBot was deployed for two weeks at location L1 and three weeks at location L2, with one week in between. In the last week at location L2, a strict COVID-19 lockdown was enforced. At location L1, only one participant interacted with the CoffeeBot more than once, but this participant did not have more than one turn per interaction, and no questionnaires were filled in by any of the participants. Therefore the results of the questionnaire and interviews are only from participants who talked to the CoffeeBot at location L2.

Unfortunately, the low number of interactions means that not enough topics are recalled in any of the conversations to trigger memory-based questions for personalization. Another issue was a bug in the ASR component, which crashed in the first interaction with a participant at L1. Though in subsequent interactions this bug did not occur, it is likely the bot left a bad impression in the first interaction with users.

7.5.1 Interaction ratio data

We are interested if the CoffeeBot can mimic short coffee talk conversations, for multiple interactions and for around 5 minutes of conversation. The average number of turns per interaction was 10,6 (σ=6,68), the average duration was 3 minutes and 7 seconds. The average number of interactions was 2,42 (σ=0,98). The average topic recall was 0.607: around 3 out of 5 topics were recalled. About half the follow-up questions were inappropriate (0.46). Self-disclosure error rate was similar to the reverse topic recall rate with 0.4. See Table 7.4 for the results per participant. The topic recall was calculated by taking the number of recalled topics (topic terms and events) divided by the total number of topics mentioned. For the self-disclosure and follow-up errors, the error was calculated by manually annotating coherence and dividing incoherent self-disclosure and follow-up questions by the total number of self-disclosure and follow-up questions by the agent, respectively. Incoherent self-disclosure is shown in Example 7.1, in which the agent interprets the user's response to a starter question as a type of game. Example 7.2 shows an incoherent follow-up to a self-disclosure, in which *rap music* is interpreted as a location.

Table 7.3: Part of the first conversation between Tessa (a pseudonym for L2U7) and the CoffeeBot. I = Interlocutor, CB = CoffeeBot and U = User. The intent column contains the recognized and generated intent for the user and agent respectively. The topic terms and topic events are the types of topics recalled by the agent.

#	I	Utterance	Intent	Topic terms	Topic events
1	CB	Hello! I don't believe we have met, what is your name?	greeting		
2	CB	Hello! I don't believe we have met, what is your name?	repeat		
3	U	my name is Tessa	inform	Tessa	
4	CB	It's a pleasure to meet you. I am CoffeeBot.	inform		
5	CB	My goal is to get to know you by asking you questions. Every time we meet, I will ask you a couple of questions.	inform		
6	CB	Do you have some time available next weeks to talk to me?	question		
7	U	yes	confirm		
8	CB	Okay. I have some questions. You can tell me to stop by saying 'goodbye'.	inform		
9	CB	Where is the best place to take a date?	question		
10	CB	Where is the best place to take a date?	repeat		
11	U	coffee shop	inform	coffee shop	
12	CB	Which sports do you like?	question		
13	CB	Which sports do you like?	repeat		
14	U	I like badminton	inform	badminton, like badminton	
15	CB	I see. Why do you like badminton?	question		
16	U	because I get to play with my friend	inform	my friend, play with my friend	I get to play with my friend
17	A	Would you like to be famous? Why?	question		
18	U	I don't want to be famous	inform		I not want to be famous
19	A	Alright. Why don't you want to be famous?	question		
20	U	because I think capitalism is bad	inform	capitalism	I think capitalism

Table 7.4: Ratio results of the CoffeeBot pilot. The user ID (UID) indicates the location (L) and participant (U) number. The interactions indicate total (different days) interactions.

UID	total turns	inter-actions	mean length	topic recall	self-disclosure error	follow up error
L1U4	8	1	3:53	1/2	0/0	0/0
L2U1	30	3(2)	2:39	11/21	0/0	3/7
L2U2	12	2(1)	2:06	7/9	1/1	0/2
L2U4	24	4(1)	2:08	12/17	1/4	3/6
L2U5	35	3(2)	2:48	9/19	0/0	4/6
L2U6	15	2(1)	2:17	5/12	0/0	3/4
L2U7	50	2(2)	5:59	32/47	0/0	6/16

(7.1) A: Who is most influential in your life?
U: my family
A: I like to play games the most out of all interests. What do you think about interests, do you have a favorite?

(7.2) A: What is your favorite music? I think mine is dance music.
U: rap music
A: I see. Where is rap located?

7.5.2 Memory

The CoffeeBot has a user model with topics for each participant in its memory. A topic contains the name, the frequency of how often it has been mentioned by either the user or the agent, the first time a topic is introduced as well as the most recent occurrence, the phrase type and the sentiment. Topic events also contain the arguments for generating a question about the event. Listing 7.1 shows a good topic term, “japanese food sushi”, that reveals an interest of the user. An example of a good topic event, “I wish I was super smart”, is shown in Listing 7.2. Unfortunately, not all topics were relevant and interesting, such as the examples in Listings 7.3 and 7.4. Mostly these types of topics were extracted from user intents that were too general, or they were not answers to questions, but meta-interaction such as asking for a repetition or stating they did not understand the CoffeeBot.

```
{ "japanese food sushi":
  { "name": "japanese food sushi",
    "frequency": { "user": 1, "agent": 0 },
    "firstTime": "2020-11-05-14-40-40-CET",
    "lastTime": "2020-11-05-14-40-40-CET",
    "type": "NP",
    "sentiment": { "polarity": 0, "intensity": 0 }}}
```


Listing 7.1: Example of a good topic term in the user model of one participant.


```
{ "text": "I wish I was super smart",
  "frequency": { "user": 1, "agent": 0 },
  "firstTime": "2020-11-05-14-38-45-CET",
  "lastTime": "2020-11-05-14-38-45-CET",
  "sentiment": { "polarity": 0.273, "intensity": 0.655},
  "args": { "A0": "I", "V": "wish", "A1": "I was super smart" }}
```


Listing 7.2: Example of a good topic event in the user model of one participant.

7.5.3 Questionnaire

In total, 6 participants filled in the first questionnaire and 2 participants filled in the second questionnaire. Only one of the participants filling in the second questionnaire talked to the CoffeeBot on separate days. Though this is limited data, we do want to mention the results of the first questionnaire. For the ICSI part, most participants were favorable about having another conversation (median=5), enjoyed the conversation (5) and were satisfied (5.5). However, they did not feel the conversation went smoothly (2.5). Additionally, the agent performed poorly in terms of invoking user laughter (3.5), letting the user know they were

7


```
{ "all kind":
  { "name": "all kind",
    "frequency": { "user": 1, "agent": 0 },
    "firstTime": "2020-11-05-14-39-09-CET",
    "lastTime": "2020-11-05-14-39-09-CET",
    "type": "NP",
    "sentiment": { "polarity": 0, "intensity": 0 }}}
```


Listing 7.3: Example of a bad topic term in the user model of one participant.

```
{ "text": "stop what did you say",
  "frequency": { "user": 1, "agent": 0 },
  "firstTime": "2020-11-05-14-38-31-CET",
  "lastTime": "2020-11-05-14-38-31-CET",
  "sentiment": { "polarity": 0, "intensity": 0 },
  "args": { "V": "stop", "A1": "what", "A0": "you" }}
```

Listing 7.4: Example of a bad topic event in the user model of one participant.

communicating effectively (3.5) and talking about interesting things (3). On all points of the intimacy part of the McGill friendship questionnaire, the CoffeeBot scored extremely poor to poor ($\leq$3). Only for the statement "CoffeeBot would listen if I talked about my problems" the ratings were average (4). The CoffeeBot scored much better on all items of the stimulating companionship scales ($\geq$4), where only the statement "CoffeeBot has good ideas about entertaining things" was rated as poor (3). Social presence scored mostly mediocre, with the CoffeeBot not being very life-like (3), but it was considered quite sociable (5) and people felt it communicated with them (4). On the RoSAS scale, the interaction with the CoffeeBot was found to be interactive (5), strange (4.5), competent (4) and responsive (4.5), whereas it was not found to be emotional (2), compassionate (2), aggressive (1), dangerous (1) or feeling (2.5).

7.5.4 Interview

Of the 7 participants, 2 consented to a separate short interview. The other participants were sent reminders via email, but did not respond. The length of each interview was about 15 minutes. Both interviewees said that it was easy to interact with the bot, though the signing up process was a bit difficult to understand, and they had to repeat themselves. One of the interviewees said that the robot looked cute, whereas the other interviewee would have liked to see something like movement of the arms to make it more engaging. Neither of them found the questions the CoffeeBot asked to be too personal, and both thought the topics were appropriate for the setting. Annoyances mentioned by the interviewees were that the CoffeeBot repeated itself often, made some weird topic switches and the voice was sometimes hard to understand. When asked about other practical uses for a CoffeeBot-like robot, one interviewee mentioned it could help train interviewing skills or could be positioned in waiting rooms.

7.5.5 Exploratory Analysis

Despite the limited amount of data of our pilot, we believe some conversations with the CoffeeBot are worth looking into. Besides the results reported above, we conducted an exploratory analysis of all the conversations. All the examples included are directly retrieved from the conversations without changing the ASR and TTS transcription, except for anonymization, and capitalization and punctuation for readability.

Example 7.3 from the conversation in Table 7.3 shows that the agent frequently repeated itself after the user did not respond quickly enough or the ASR did not pick up an utterance. In total this happened 23 times across all the conversations.

(7.3) A: Which sports do you like?
A: Which sports do you like?
U: I like badminton.

The follow-up questions asked by the CoffeeBot were sometimes incoherent. Sometimes this was due to ASR issues, such that in Example 7.4 a TV show's name is transcribed as a verb. Other times the lack of common sense or deeper understanding led to weird questions, such as in Example 7.5, in which the CoffeeBot does not understand that it asked the starter question about the friend and therefore the follow-up question is not coherent.

(7.4) A: If you could bring back one TV show that was canceled, which one would you bring back?
U: You want to Brody.
A: When do you usually want to Brody?

(7.5) A: What was your best friend's name growing up?
U: My best friend's name growing up was [NAME].
A: Alright. Who is [NAME]?

The questions the CoffeeBot asked were always open-ended questions, to elicit a more extensive response from the user. We were curious to see how informative people's responses were, so we measured the amount, type and sentiment of self-disclosure. We considered also annotating the sensitivity of the disclosed information and categorizing more types of self-disclosure (Jourard, 1961; Knijnenburg et al., 2013; Pickard et al., 2016), but the types are heavily dependent on the starter questions, and we found no user responses that contained a different type than elicited by the question. We first annotated the questions if they contained a user response. Questions asked by the CoffeeBot after which the user said goodbye are omitted, as well as repetitions of questions. In total there were 118 questions, of which 76 were starter questions, 42 were follow-up questions and none were memory questions. Of the 118, 24 questions were responded to with a counter question of the user or an answer that the ASR could not transcribe correctly. After removing these responses from the set, 94 questions remain that have an answer as a user response, which we call the question-answer set.

For the amount of disclosure, we took Ravichander and Black (2019)'s definition of voluntary and involuntary self-disclosure. Voluntary means that a user self-discloses something that is not directly addressed, and involuntary is self-disclosing information directly addressed by the question. We only found one case in the question-answer set that showed voluntarily sharing of information, shown in Example 7.6. In this example, the user discloses their best friend's name (involuntary disclosure) and informs the CoffeeBot about the meaning of the name (voluntary disclosure). Sometimes a question included two options in one and a user gave a response with both answers, such as in Example 7.7, this still counts as involuntarily.

(7.6) A: What was your best friend's name growing up?
U: [NAME], which means kitchen.

(7.7) A: Would you like to be famous? Why?
U: No I would not like to be famous, because everybody is famous dances.

As categories of self-disclosure in user answers, we take the three categories from Barak and Gluck-Ofri (2007): information, thoughts and feelings. Though these categories were formulated for annotating internet forum data, the categories do apply for casual conversation as well. Information relates to general information disclosed by the user ("One night ago I went for dinner.") Thoughts are related to opinions about general concepts, plans or related to one self. If in a user utterance the phrase "I like" or "I think" exists, the utterance is classified as a thought. Feelings are any answers expressed as for example anger, inconvenience and frustration. An emotion ("angry") or the phrase "I feel" was an indicator to classify an utterance as feelings. In our case, we also count positive thoughts and feelings, as opposed to the definition in Barak and Gluck-Ofri (2007). Of the 93 question-answer set, 16 items were information, 73 were thoughts and 5 were about feelings. However, because almost all answers are due to involuntarily sharing of information, these categories are almost directly related to the questions rather than the answers.

Finally we annotated the sentiment values, or polarity, of the question-answer set, partially based on the model of Bak et al. (2014). Bak et al. (2014) considers three levels of self-disclosure: general, medium and high sensitivity. General self-disclosure items are factual answers to questions, such as who the coach is of a

Table 7.5: This table shows the combinations of disclosure type and sentiment for each of the 94 question-answer items.

Type / Sentiment	Positive	Negative	Neutral	Total
Information	0	0	16	16
Thoughts	19	5	49	73
Feelings	0	2	3	5
Total	19	7	68	94

football team. The medium level occurs when for example, a user would say that they love a certain football coach, which can be detected through a combination of a personal pronoun ("I") and an opinion verb ("like"). Finally there is high sensitivity, which is secretive and negative information, such as negative remarks about physical appearance and mental and physical conditions. We took a slightly different categorization from Bak et al. in which we make a separation between neutral, positive and negative instead of the three levels of sensitivity. The neutral level contains the same items as general self-disclosure, or medium with no clear positive or negative words ("surprise" or "nervous"). The positive level contains positive items of a medium sensitivity, and the negative level contains negative items of medium and items of high sensitivity. We labeled any utterance with "I like" or keywords like "good" as positive, utterances with "I didn't like" or "bad" as negative and all others as neutral. In total, 19 were positive, 7 were negative and the remaining 68 were neutral. The user utterance in Example 7.3 is classified as positive, whereas the user utterance in Example 7.7 would be classified as negative.

We show the combination of type and sentiment in Table 7.5. Overall, types of questions-answer items that were about feelings are neutral or negative. For the information type, the sentiment was always neutral. Thoughts were mostly neutral, with only a couple of items being negative. All positive sentiment was found only in the category of thoughts.

7.6 Discussion

Unfortunately, we did not collect enough data to evaluate our methodology and interaction design properly. The data that we have is limited with 7 participants and at most interactions on two different days, and no memory questions were asked. Therefore we will discuss the results of the CoffeeBot evaluation on an exploratory level.

The foremost limitation of this research was the small number of participants and interactions for each participant. The COVID-19 pandemic and a lockdown severely restricted the number of interactions possible over the course of 2 or 3 weeks. Our analysis is therefore impacted by the novelty effect, because users did not get familiar with the CoffeeBot.

Even though the Google ASR worked relatively well in a not-so crowded coffee place, the ASR had some issues with speaker diarization; separating background voices from the participant's. About 40% of the topics were not recalled, mostly due to ASR errors. Irfan et al. (2020) found the same issues with their social robot as barista in a coffee shop. They recommend constraining the grammar of the ASR model and adapting it to non-native English as well. Follow-up questions were misclassified or inappropriate due to ungrammatical sentences that could not be correctly parsed with the current SRL implementation. Inoue et al. (2020)'s work on follow-up questions might be more robust to ungrammaticalities and can be adapted for open-domain chat. Additionally, some sentences spoken by the TTS were not comprehensible for participants, as also mentioned by one of our interviewees.

Despite these errors, participants did still enjoy the conversations and company of the CoffeeBot, though they felt the flow of the interaction was insufficient. Unfortunately, due to the limited interactions and no topic being discussed at least 5 times as per the question model in Section 6.5.4.4, none of the memory-based questions that should trigger a sense of personalization were asked in any of the interactions.

We did perform an exploratory analysis on our data to look for possible patterns and measure the quality of self-disclosure. In total 94 topics were disclosed across all participants. We found that only on one occasion voluntary information was shared, and the starter and follow-up questions did not elicit more information from the user than the involuntary information. Of the in total 94 topics, most were a positive or neutral response and were opinions of people. However, this is likely due to the phrasing of the questions, which primes users to answer with certain phrases. Negative responses hardly occurred, which could be due to politeness of people or people preferring not to talk about negative things with a robot they hardly know.

The design of the CoffeeBot's embodiment is simple, to prevent high expectations from participants. However, as mentioned by one of the interviewees, the CoffeeBot could have benefited from engaging behavior such as movement without raising expectations too much. Another issue was turn-taking in the interaction, where both the robot and user's speech overlapped occasionally, and the robot repeated itself sometimes. For the robot this could be solved by having better end of turn detection or adding a component to see if the user is still thinking about a question. From the user's perspective, turn-taking can be improved by adding behaviors in the robot so that the user is signaled about when to speak.

7.7 Future Work

In future work, we want to deploy the CoffeeBot for at least two months at the same location in a public space. Two months should be sufficient for the robot to be socially accepted and reduce the novelty effect (Leite et al., 2013; de Graaf et al., 2016). We want to conduct a study to see if the question model with memory questions would have a benefit in the long-term on engagement and interaction quality. We plan to set up a between-subject study where the experiment group gets all three question types, including memory-based questions, and the control group only gets starter and follow-up questions. We will also finalize the prototype based on the feedback from the participants, for example by either improving the current prototype (e.g., moving arms, adding lights to guide turn-taking) or using an existing social robot that has these features.

Additionally, we want to check if other methods would be applicable for the CoffeeBot evaluation. For example, a social relationship could be measured with Brennan et al. (1998)'s self-report measure for adult attachment. However, as Clark et al. (2019) note, people are not yet expecting to create (equal) social relationships with social agents. Furthermore, the benefit for people interacting with CoffeeBot could be too implicit, because the conversations in theory only help the CoffeeBot to build a better user model. We would like to see how we can more explicitly show the benefit of the personalization through question asking to participants for a better balance of reciprocity.

With the changes based on the pilot study in mind, we have formulated the following hypotheses for a full real world user study with the CoffeeBot with respect to the question asking model.

H_1: Participants feel more close with the agent that personalizes based on memory questions.

a) Participants will rate the quality of interpersonal communication higher for the agent asking memory questions.

b) Participants will have a more positive attitude towards the agent asking memory questions.

H_2: Participants are more engaged with the agent that personalizes based on memory questions.

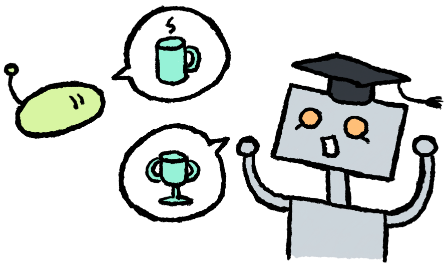
a) The conversations between participants and the agent asking memory questions will have more topics and more depth with respect to the topics.

b) The conversations between participants and the agent asking memory questions will have more and longer conversations.

Hypothesis 1 should be evaluated through self-report questionnaires and hypothesis 2 should be evaluated through analyzing the ratio data of the interaction as per the method of the pilot study.

Part IV

Discussion and Conclusion



8 Discussion and Conclusion

In this final chapter of the thesis we discuss the answers to our research questions and reflect on our research output in Section 8.1. In Section 8.2 we discuss what we believe are important steps forward for developing personalized social conversational agents. Finally, we provide a take-home message in Section 8.3.

8.1 Findings

We highlight our findings with respect to the three research questions introduced in Chapter 1 related to: i) designing and prototyping a multimodal agent, ii) personalizing a conversation and iii) measures in long-term real world evaluation.

8.1.1 Designing and Prototyping Social Conversational Agents

- RQ1: How can dialogue designers effectively and iteratively prototype a social conversational agent?

In Chapter 2 we discussed modalities that can be used by the user and the agent: text, speech, video, touch and physiology. Each of them have their merits with respect to reliability, accessibility and informativeness. We highlighted available tools for designing multimodal (prototype) conversational agents, such as Visual SceneMaker (Gebhard et al., 2012) and the Virtual Human Toolkit (Hartholt et al., 2013). After considering the different modalities and tools for social behavior interpretation and generation in Chapter 2, we presented the ARIA framework in Chapter 3 with text, speech and video as available modalities. We showcased social conversational agents that can be developed within the ARIA framework. Dialogue design ranges from prototyping scripted chat dialogues to more multimodal adaptive conversations based on socio-emotional recognition of users with audio and video modalities. In Chapter 4 we presented the dialogue engine *Flipper*, and how it integrates with multimodality, natural language processing and dialogue management. Additionally, we developed design patterns for dialogue development illustrated with examples, which can be applied with different dialogue design tools, such as in the ARIA framework, as well as in other existing frameworks. We emphasize the need for reusability of tools, which we achieved through building a modular agent around Flipper as well as the sharing of example conversations as a quick starting point for dialogue designers. In Chapters 5 and 6 we designed two prototypes of social conversational agents based on Flipper and the design patterns. In Chapter 5 this was the BLISS agent, which was iteratively improved to ask questions about people's happiness. In Chapter 6 we focused on prototyping long-term real world interactions for casual conversations. Though not the only way, we hope that dialogue designers can more effectively develop a social conversational agent, with our comparison of strengths of different modalities and tools in an application domain, and by providing our own modular tool with design patterns.

8.1.2 Personalizing Conversations with Social Conversational Agents

- RQ2: How can dialogue designers personalize the interaction between a user and a social conversational agent?

In Chapter 2 we stated that personalization is "a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its personal relevance to an individual or a category of individuals" (Fan and Poole, 2006, p. 183). Personalization can be achieved through using socio-emotional modalities as described with the ARIA framework in Chapter 3. With the ARIA system, the virtual human Alice adapted for example to a user's mood, both verbally and with facial expressions and gestures. If the user was smiling, the agent would look more happy and sound friendlier. Another method of personalization is through adaptive question-asking such as with the BLISS agent in Chapter 5 and the CoffeeBot in Chapter 6. With these two prototypes we aimed at eliciting self-disclosure of topics the user is interested in to build a user model. The self-disclosure of these topics in turn led to more personalized questions and thus more topics for the user model, which was an essential part in the design of the CoffeeBot, which is meant for long-term real world interaction. We found that personalization does not need to be based on a complicated model. Some people appreciated some form of personalization through recognition of their words with only superficial natural language processing, which was the case during data collection described in Chapter 5 with the BLISS agent. However, the BLISS agent failed at showing understanding non-standard user responses with its mostly scripted dialogue. By keeping the conversations mostly agent-initiative, we prevented breakdowns of conversations in which the agent did not know how to respond to the user. In Chapter 5 we also conducted a thematic analysis and found that despite the ASR or TTS not optimally working, a user model was constructed with at least one correct topic of interest. Though this is limited in possibilities for personalization, such a user model will at least contain some user interests, even from relatively short conversations. The CoffeeBot described in Chapter 6 had a more flexible dialogue in a casual conversation structure, though the question generation did formulate some weird questions occasionally during the pilot in Chapter 7. However, participants said they did not mind the CoffeeBot making mistakes and it did not break down the interaction for them. The personalized question-asking with a user model of topics was developed for more engaging casual conversations with the CoffeeBot, but can be used for other purposes. For example, building a recommender system fine-tuned with the user model or providing insight for healthcare workers to know what is important for a person's quality of life. We believe our personalization strategy of capturing user topics and asking questions can be applied in task-oriented dialogues, for example by adding personalized casual conversation or small-talk to the task-related dialogue or by using user topics in sentences of a conversational agent to personalize towards that particular user.

8.1.3 Evaluation of Social Agents in the Long-term in the Real World

- RQ3: How can dialogue designers measure the effect of personalization on acceptance in long-term real-world interactions with a social conversational agent?

In Chapter 7 we compared different ways of evaluating human-agent long-term real world interactions, based on mostly Human-Robot Interaction (HRI) research in longitudinal studies (Leite et al., 2013). A long-term study can be considered real world when it lasts more than two months with at least weekly interactions and when it is deployed in an unsupervised environment without intervention of experimenters. A common way to measure a social non-task-oriented conversational agent's performance in real world interactions is that of acceptance, thus an evaluation should include a measure for acceptance. Most studies involving real world interactions use questionnaires, meta-data of interactions (such as number of turns) or interviews. Existing questionnaires for long-term studies about acceptance are Heerink et al. (2010), Weiss et al. (2011), and de Graaf et al. (2016). Indirect ways of measuring acceptance are by tracking the frequency and number of turns in interactions over time, or as in interviews, asking users about current and future use. In our study with the BLISS agent in Chapter 5, we evaluated a prototype in the real world with a relatively diverse group of adults and we asked participants afterwards if they would like to interact with the agent in the future. Users' feelings were mostly positive and users indicated a willingness to talk to BLISS agent again, but because the study only involved a single interaction with users, we cannot say for certain they would actually talk to the agent again and for how long. In our pilot with the CoffeeBot in Chapter 7 we used questionnaires for measuring the effect of personalization on the rapport between a user and the CoffeeBot (Hecht, 1978; Mendelson and Aboud, 1999) and the quality of the CoffeeBot in general (Jung and Lee, 2004; Carpinella et al., 2017), the meta-data, a thematic analysis and an interview to see how usable the prototype was. Critical comments in the interview were about the sound quality and the repetitiveness, whereas the positive comments were about interesting questions to ponder about. The questionnaires did not seem to be too much trouble for participants, especially when they are spread out over time. We think that focusing on long-term use with acceptance is important, regardless of the goal of the long-term real world interaction design. Measuring the effect of personalization through the evaluation setup of the CoffeeBot questionnaires, interviews and meta-data is a good way to start, but we have yet to conduct a full study to reliably make conclusions about the methodology.

8.2 Future Work

We face many challenges ahead in the world of designing personalized conversational social agents. Here we highlight the ones we believe deserve attention.

8.2.1 Software & Privacy

The software that was developed during this thesis is open-source and available to use for other researchers. An emphasis on developing good open-source tools and readily available software is necessary for conducting research without too much focus on software development. There is always a trade-off between using commercial software that is more state-of-the art and more accessible open-source software. However, commercial software often has its own ecosystem and does not allow for much flexibility in terms of dialogue design. Only until a decade ago companies offered mostly closed-source software to researchers, but business models have shifted to more open-source and collaborative efforts with research departments. Business models are more often based on offering service instead of software such as with spaCy and RASA (Bocklisch et al., 2017). More research groups make their software publicly available to stimulate use by other parties. However, good documentation, such as design patterns and examples, is often lacking from academic software and is a key feature to make the software usable for others.

An important aspect with personalization is the privacy of users. Especially with virtual assistants making their way into our lives and more smart devices collecting personal data in the cloud, researchers and users should be aware of keeping the data safe. The ARIA, BLISS and CoffeeBot systems we developed can be fully independent of cloud-services. Though with these systems you can be sure no data will be shared with third parties, there is a greater burden on the user and dialogue designer to keep the local software and hardware safe for use. Usually these systems are left without maintenance after completion and run the risk of security issues. Additionally, using more cloud-services in a conversational agent provides easier access to features that otherwise need to run locally and this could result in compatibility issues with users' own hardware.

We believe that a combined approach with a focus on accessible user control of data is the step forward, regardless of platform. User control of data can be provided through apps that show collected data, past conversations or summaries, in which the user can modify information (Hendrickx et al., 2021). Another approach of control is working with a data daemon, which is an assistant that helps the user in managing their personal data (Toussaint et al., 2021).

8.2.2 Long-term Real World User Studies

We have proposed a framework, design guidelines and evaluation metrics for designing long-term social agents, but all the user studies and data collections described in this thesis have been exploratory or short-term in nature. We therefore see the following challenges that future work on these aspects could tackle.

Specifically for spoken dialogue systems, incremental processing is essential, compared to chatbots. In long-term real world studies, a lack of incremental processing can become a nuisance for users to wait until an agent is finished talking or to be unable to interrupt the agent, because a user might already know what to say. The framework with moves we proposed in Chapter 3 is fit for incremental processing with support for interruptions, but not many other existing platforms are available (Michael, 2020). Schlangen and Skantze (2009) proposed a model for incremental dialogue processing, with a focus on turn management. A recent overview by Skantze (2021) shows that state-of-the-art systems use either chatbots or explicit turn-taking with wake-words. Additionally, neural models of turn-taking are often trained on data that might not be representative for the goal of a social agent. With more datasets of spoken and multimodal data becoming available, we hope that better models for turn-taking will become available or will be easier to train for dialogue designers themselves, with hopefully higher engagement in long-term interaction (Oertel et al., 2013; Buschmeier et al., 2014; Cafaro et al., 2017b; Nazareth et al., 2019b).

Evaluating long-term real world studies remains challenging. In Chapter 7 during the evaluation of the CoffeeBot we carefully deliberated about measurements for long-term real world evaluation, in which we tried to balance relevance and depth of the measurements with low cognitive workload for users. In the case of the method for the CoffeeBot, questionnaires about acceptance and usability, interviews and the meta interaction data were used for measuring the effect of its personalization strategy on self-disclosure and social relationship in a real world setting. Questionnaires are accessible, good to understand for end-users and usually more likely to be validated than other measures, but newer (automated) measures for long-term evaluation are not widely available yet (Werner, 2020). We cannot confirm if our approach for long-term real world evaluation is correct due to our small sample. More attention should be directed towards validating real-world long-term evaluation measurements that can be used in multiple contexts. However, there will probably always be confounding factors in a real world study, which makes complete validation improbable.

During our data collections for both the BLISS agent and the CoffeeBot, we found that most end-users commented on the quality of the voice of the agent, the text-to-speech synthesis (TTS). TTS for conversation has been a field that deserves more attention, especially with the rise of virtual assistants. Most of these systems have started with voices based on unit-selection for reading aloud written text. Now voice systems have evolved to using neural networks such as with Tacotron (Wang et al., 2017). These voices sound more fluent, but still lack empathy and control for affect in the voice, which was achieved with previous unit-selection systems such as the Sensitive Artificial Listener (Schröder et al., 2009). Depending on the level of control a dialogue designer would need, we recommend two approaches. One is integrating SSML (Shuang and Burnett, 2010) for adding affect to speech which leaves more control of the affect of the voice to the designer. The other approach is an automatic method for end-2-end affective speech synthesis (Wu et al., 2019), which could be more useful if control is not that important to the dialogue designer.

8.3 Concluding Remarks

Our main research question was "How can we design a social conversational agent capable of personalizing interactions with users in the real world?". We have provided an overview of different frameworks and tools and developed our own framework with which dialogue designers can start prototyping a social conversational agent, with multimodal capabilities and flexible dialogue management.

Furthermore, acceptance and personalization do not necessarily require complex machine learning models if dialogues are designed correctly. Personalization through machine learning and implicitly learning large quantities of data about a person is yet unsuccessful in creating significantly better user experiences in long-term interaction with conversational agents, regardless of the ethical implications.

We emphasize that personalization is more about making the user feel understood and focusing on smaller components of a dialogue that contribute to a feeling of personalization. By no means is our method of personalization an exhaustive one. Multimodality and personalized question generation are but a few of important components of designing personalized social conversational agents. The latter is a highly underresearched topic, especially in non-task-oriented domains such as in casual conversation.

Finally we argued for the importance of testing social conversational agents in the wild. In the last decade, the number of conducted long-term field studies has been increasing (Leite et al., 2013). However, in the wild does not necessarily equate with ecological validity (Heylen et al., 2012). A step forward would be to embrace that mistakes will happen in conversations in the wild and to take these into account for a dialogue design to make it more robust. This is not limited to mistakes in (non-) verbal natural language understanding and generation, but also understanding the social context.

With this thesis we aim to give future social conversational agent designers a starting kit with tools and knowledge to keep innovating and to design agents which can provide a better user experience in the long-term in the real world.

Part V

Appendices

A
Questionnaires and Interview Questions

A.1 Book-ARIA Questionnaire

ARIA-VALUSPA: QUESTIONNAIRE

Thank you for participating in the ARIA-VALUSPA study! We hope that you had a pleasant experience and that you enjoyed your conversation with Alice. Were you able to find the answer to the question?

As a participant of this study, you will be asked to fill this questionnaire after every interaction. It has been designed to help us measure your experience in terms of usability, flexibility, satisfaction and enjoyment.

If you have any questions, do not hesitate in letting us know.

A. ABOUT YOU:

Date of the experiment:

Age:

Gender:

Mother tongue:

English proficiency: Low / Intermediate / High / Native

Scenario: Please, select the objective you were assigned in this experiment.

<table>
<tr><th colspan="2">Were you able to get the correct answer? Please tick relevant box below.</th><th rowspan="2">Objective / Question</th></tr>
<tr><th>Yes</th><th>No</th></tr>
<tr><td></td><td></td><td>Find out about Alice’s real name</td></tr>
<tr><td></td><td></td><td>Find out how many siblings Alice actually has</td></tr>
<tr><td></td><td></td><td>Find out about the role of Oxford</td></tr>
<tr><td></td><td></td><td>Other objective (please specify):</td></tr>
</table>

B. ABOUT THE EXPERIENCE

Please, rate your degree of agreement with the following statements from 1 (strongly agree) to 5 (strongly disagree).

		Strongly Agree 1	2	3	4	Strongly disagree 5
1	I thought the system was easy to use.					
2	I think that I would like to engage in conversations with Alice more often using the system.					
3	I found the system more complex than necessary.					
4	I think that I would need the support of a technical person to be able to use this system.					
6	I found the various functions in this system were well integrated.					
7	I would imagine that most people would learn to use this system very quickly.					
8	I found the system very cumbersome to use.					
9	I felt very confident using the system.					
10	I needed to learn a lot of things before I could use this system.					
11	I found the system simple to use.					
12	I think that the system is user friendly.					
13	I thought that the system requires the fewest steps possible to accomplish what I want to do with it.					
14	I found that using the system is effortless.					
15	I can use it without written instructions.					
16	I don't notice any inconsistencies as I use it.					
17	I think both occasional and regular users would like it.					
18	I can use it successfully every time to have a meaningful conversation.					
19	I would recommend it to a friend.					

20 It is fun to use.					
21 It works the way I want it to work.					
22 It is wonderful.					
23 I feel that it is a great way to spend my leisure time.					
24 It is pleasant to use.					
25 The conversations were engaging					
26 The plot was engaging					
27 I think Alice was listening to what I said					
28 I think Alice could understand what I said					
29 I think Alice responded appropriately to what I said					
30 I think Alice understood my feelings					
31 I think Alice was responsive to my feelings					
32 I think Alice could express her emotions in a way I could understand.					
33 I am satisfied with the system.					

C. OTHER COMMENTS:
Do you have any other comments about your experience? Let us know!

A.2 CoffeeBot Questionnaires

Here we list the questions of the questionnaire for each item we were interested in. All items are rated on a 7-point Likert scale.

A.2.1 Interpersonal Communication Satisfaction Inventory

- CoffeeBot let me know I was communicating effectively.
- I would like to have another conversation like this one.
- CoffeeBot genuinely wanted to get to know me.
- I was NOT satisfied with the conversation.
- I actually had something else to do.
- I felt that during the conversation I was able to present myself as I wanted CoffeeBot to view me.
- CoffeeBot understood what I said.
- I was very satisfied with the conversation.
- CoffeeBot expressed a lot of interest in what I had to say.
- I did NOT enjoy the conversation.
- CoffeeBot did NOT provide support for what he was saying.
- I felt I could talk about anything with CoffeeBot.
- We each got to say what we wanted.
- I felt that we could laugh easily together.
- The conversation flowed smoothly.
- CoffeeBot changed the topic when his feelings were brought in the conversation.
- CoffeeBot frequently said things which added little to the conversation.
- We talked about something I was NOT interested in.

A.2.2 McGill Friendship Questionnaire

- Intimacy:
  - CoffeeBot is someone I can tell private things to.
  - CoffeeBot knows when I'm upset.
  - CoffeeBot is someone I can tell secrets to.
  - CoffeeBot knows when something bothers me.
  - CoffeeBot would listen if I talked about my problems.
  - CoffeeBot would understand me if I told him my problems.
  - CoffeeBot is easy to talk to about private things.
  - CoffeeBot understands my feelings.
- Stimulating companionship:
  - CoffeeBot is fun to do things with.
  - CoffeeBot tells me interesting things.
  - CoffeeBot has good ideas about entertaining things to do.
  - CoffeeBot makes me laugh.
  - CoffeeBot is exciting to talk to.
  - CoffeeBot is enjoyable to be with.
  - CoffeeBot is exciting to be with.
  - CoffeeBot is fun to stand and talk with.

A.2.3 Social Presence

- How sociable was the CoffeeBot?
- How personal was the CoffeeBot?
- How life-like was the CoffeeBot?
- How sensitive was the CoffeeBot?
- While you were interacting with the CoffeeBot, how much did you feel as if he was a social being?
- While you were interacting with the CoffeeBot, how much did you feel as if he was communicating with you?

Scary	Emotional	Dangerous
Knowledgeable	Compassionate	Awkward
Reliable	Organic	Aggressive
Interactive	Social	Awful
Responsive	Feeling	Strange
Capable	Happy	Competent

A.2.4 RoSAS

Instead of statements, for RoSAS participants had to rate whether the word described the CoffeeBot very well (7) or not at all (1).

A.2.5 CoffeeBot First Questionnaire

In addition to the four questionnaires, we asked participants the following questions:

- What is your birth year?
  - (Any number between 1900 and 2002)
- What is your gender?
  - Female
  - Male
  - Non-binary
  - Prefer not to say

A.2.6 CoffeeBot Second Questionnaire

We asked these two questions to put into context how often people had spontaneous opportunities to talk to the CoffeeBot.

- How often were you in your own standard workplace over the course of two (three) weeks? Half-days count as days, as long as you were in and around your own workplace for more than 3 hours. If you spent a full or half-day at another workplace than your own standard workplace, do NOT count those days.
  - Less than 1 day
  - 1-3 days
  - 4-7 days

  - 8-10 days

- How often did you walk away from your desk on a full work day at your standard workplace over the last two (three) weeks? Don't include coming into the workplace at the start of your workday or going home. Include walks for going to the toilet, printing, getting a snack or lunch or talking to another person.
  - Over 20 times a day
  - Around 10 to 20 times a day
  - Around 5 to 10 times a day
  - Around 0 to 5 times a day
  - Never

- Did you talk to the CoffeeBot more than once?
  - Yes
  - No

A.3 CoffeeBot Interview Questions

The semi-structured interview contained five questions:

- How easy was it for you to interact with the CoffeeBot?
- What would you change about the appearance of the CoffeeBot?
- What kind of topics would you like to discuss with the CoffeeBot?
- What were some annoying things about the CoffeeBot?
- What practical uses do you see for a CoffeeBot?
- What was your most memorable conversation with the CoffeeBot?

B
Experiment and Ethical Forms

B.1 Book-ARIA: Alice's Adventures in Wonderland

Thank you for participating in the ARIA-VALUSPA study! We hope that you had a pleasant experience and that you enjoyed your conversation with Alice. Were you able to find the answer to the question?

As a participant of this study, you will be asked to fill this questionnaire after every interaction. It has been designed to help us measure your experience in terms of usability, flexibility, satisfaction and enjoyment.

If you have any questions, do not hesitate in letting us know.

B.2 Coffeebot Informed Consent

Participant #:

ID #:

INFORMED CONSENT

Introduction:
The Human Media Interaction group at the University of Twente conducts research on conversations with social agents and robots. Your informed consent is given with regards to the 'Small-talk with the Coffeebot 2020' experiment. You can find more information about this specific project in the accompanied information brochure.

Principal researchers:
Jelte van Waterschoot[1], Mariët Theune[1], Dirk Heylen[1]

Contact information:
If you have any questions regarding this research, you can contact Jelte van Waterschoot (j.b.vanwaterschoot@utwente.nl), or another principal researcher or Petri de Willigen, secretary of the Ethical Committee (address: P.O. Box 217, 7500 AE Enschede (NL), tel. 053-489 2085, e-mail: ethics-comm-ewi@utwente.nl). The committee consists of independent experts of the university and are available for answering questions regarding this research.

Research: Small-talk with the Coffeebot
I understand that for the duration of this experiment:

- I am clearly informed about the research. The research goals and methods have been sufficiently explained and I had ample opportunity to ask questions.
- I understand that I can withdraw my consent for the experiment at any given moment, without giving a reason and without any further consequences.
- I give consent for the collection of audio recordings and anonymized transcriptions as described in the information brochure for this experiment research purposes.

Audio recordings will only be analyzed by the researchers and will never be disclosed to third parties for demonstration or reporting purposes. You can ask for deletion of your data until up to one month after the experiment has ended. All research material will be processed and stored in accordance with the General Data Protection Regulation (GDPR). All data will be stored for a minimum of 10 years, in accordance with the UT data policy.

Date: Place:

Signed in duplicate
Name participant: Signature participant:

E-mail participant: ID:

I have provided explanatory notes about the research. I declare myself willing to answer to the best of my ability any questions which may still arise about the research.

Name researcher: Signature researcher:

The extra copy of this consent form is for you to keep.

B.3 Coffeebot Information Brochure

Information Brochure Pilot: Coffeebot, long-term interaction through question generation

Dear participant,

we would like to inform you about the research you have applied to participate in.

Introduction

You will participate in a small talk conversation with a ‘coffeebot’, a social agent. This agent is unfortunately not able to make coffee for you (yet), but is designed to mimic coffee-machine talk. Think about meeting someone at the coffee machine and telling them about a conference or a paper, or maybe about a hobby. During the experiment, you will occasionally interact with the coffeebot through spoken interaction. The coffeebot will ask you a number of questions every time. The coffeebot **is very limited** in questions you ask. The questions are small-talk related, such as ‘What is your favorite TV show?’ or “Who’s your favorite actor?” to “Why do you like playing football?”.

The experiment will take place in November 2020 in a public spot, such as a coffee corner. In the proposed research, entitled "Coffeebot: long-term interaction through question generation", speech recordings are made during the experiment and a computer will talk with you. You will be asked to fill in a questionnaire at the start of the experiment and at the end of the experiment (15 min each), as well as a post interview being held at the end of the experiment (15 min). The aim of the research is to establish whether a certain model of questioning by a social agent has an impact on the quality and engagement of a conversation. The research could contribute valuable insight into how people communicate with social agents. For this research, there are two important aspects which you should be aware of.

Data

Firstly, all conversations with the social agent will be recorded, that means that your speech will be recorded as well. If you object to this, you may not participate in the research. The speech data will be used for analysis of the conversations for this research, for example pitch changes. Additionally, we analyze for example the length of the conversation and utterances as well as identify interesting topics based on the transcriptions. None of your personal data (audio recordings, raw transcriptions) shall be shared for purposes outside this research study and only stored securely in accordance with the UT’s data regulations for 10 years. Your personal speech data will not be published in any shape or form. Questionnaire results, anonymized transcriptions and anonymized meta-data such as pitch, volume and voice activity will be used for publication in papers and presentations, but not for sharing.

If you wish to withdraw from this research, you can do this for up until one month after the experiment has ended. After this, the original data will be anonymized (e.g. in transcriptions, we replace names by fictional names) and not be traceable back to one person.

Questions

Secondly, the social agent might ask inappropriate questions, because of its autonomous behavior. For example, a question could occur as “Do you like visiting your grandma?”, whereas your grandma might have passed away very recently.
Even though we ask you to be benevolent to the social agent and truthful as much as possible during the interaction, you can stop the interaction at any time or ignore statements by the agent if you want to by saying ‘I have to go’, ‘goodbye’ or ‘stop talking’ or walking away from the agent (it will time out after ~20 seconds).

Procedure

The experiment requires you to participate at least 3 times for about 5 minutes over the course of two weeks. It requires you to use an RFID card to initiate the interaction by **swiping the robot**, which will be an anonymized identifier for you during the experiment. After the experiment has ended you will be briefed and open-ended interviewed about the results. If you wish to be informed about the results of the study, you can ask the experimenter.

Interview

During a post-interview after the experiment, we will have an open-ended discussion about your experience with the Coffeebot. Questions will be:

- What was your most memorable conversation with the coffeebot?
- What were some annoying things about the coffeebot?
- What would be a practical use for such a coffeebot?
- How could we make the interaction more engaging with the coffeebot?

Yours sincerely,

Coordinator: Jelte van Waterschoot
Department of Human Media Interaction
Faculty of EEMCS
University of Twente
Tel: +31 (0) 53 489 3100
Email: j.b.vanwaterschoot@utwente.nl

Research leader: Mariët Theune
Tel: +31 (0) 53 489 3817
Email: m.theune@utwente.nl

C
CGN Annotation Schema

Table C.1: The genders corresponding to the sex of the Corpus Spoken Dutch (CGN)

Code	Sex
sex1	Male
sex2	Female
sexX	Unknown
null	Other

Table C.2: The regions corresponding to the `resRegion` of the Corpus Spoken Dutch (CGN)

Code	Region
regN2b	Oost Utrecht, excl. de stad Utrecht
regN2a	Zeeland, incl. Goeree Overflakkee
regN2c	Gelders rivierengebied, incl. Arnhem en Nijmegen
regN2d	Veluwe tot aan de IJssel
regN2e	West Friesland
regN2f	Polders
regN1a	Zuid-Holland, excl. Goeree Overflakkee
regN1b	Noord-Holland, excl. West Friesland
regN1c	West Utrecht, incl. de stad Utrecht
regN3a	Achterhoek
regN3b	Overijssel
regN3c	Drenthe
regN3d	Groningen
regN3e	Friesland
regN4a	Noord-Brabant
regN4b	Limburg (Nederland)
regNx	Nederland -overig
regV1	Antwerpen en Vlaams-Brabant
regV2	Oost-Vlaanderen
regV3	West-Vlaanderen
regV4	Limburg (Vlaanderen)
regW	Wallonië
regVx	Vlaanderen -overig
regZ	Regio buiten Nederland en Vlaanderen
regX	Regio onbekend
null	Unknown

Publications

Conferences

- M. Valstar, S. Dermouche, C. Pelachaud, E. Coutinho, B. Schuller, Y. Zhang, D. Heylen, M. Theune, **J. van Waterschoot**, T. Baur, A. Cafaro, A. Ghitulescu, B. Potard, J. Wagner, E. André, L. Durieu, and M. Aylett (2016). "Ask Alice: An Artificial Retrieval of Information Agent". In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*. ACM Press, pp. 419–420. DOI: 10.1145/2993148.2998535

- A. Cafaro, M. Bruijnes, **J. van Waterschoot**, C. Pelachaud, M. Theune, and D. K. J. Heylen (2017a). "Selecting and Expressing Communicative Functions in a SAIBA-Compliant Agent Framework". In: *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings*. Springer, pp. 73–82. DOI: 10.1007/978-3-319-67401-8_8

- **J. van Waterschoot**, M. Bruijnes, J. Flokstra, D. Reidsma, D. Davison, M. Theune, and D. Heylen (2018a). "Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents". In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 43–50. DOI: 10.1145/3267851.3267882

- **J. van Waterschoot**, G. Dubuisson Duplessis, L. Gatti, M. Bruijnes, and D. Heylen (2018b). "An Information-Providing Closed-Domain Human-Agent Interaction Corpus". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), pp. 2746–2752

- D. S. Nazareth, C. Burghardt, A. Capra, P. Cristoforetti, W. Lam, **J. B. van Waterschoot**, G. J. Westerhof, and K. P. Truong (2019a). "Babbelbord: A Personalized Conversational Game for People with Dementia". In: *Dementia Lab 2019. Making Design Work: Engaging with Dementia in Context*.

Springer International Publishing, pp. 169–173. DOI: 10.1007/978-3-030-33540-3_18

- **J. van Waterschoot**, I. Hendrickx, A. Khan, E. Klabbers, M. de Korte, H. Strik, C. Cucchiarini, and M. Theune (2020a). "BLISS: An Agent for Collecting Spoken Dialogue Data about Health and Well-Being". In: *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 449–458

Workshops

- K. Bowden, T. Nilsson, C. Spencer, K. Cengiz, A. Ghitulescu, and **J. B. van Waterschoot** (2017). "I Probe, Therefore I Am: Designing a Virtual Journalist with Human Emotions". In: *Proceedings of eNTERFACE'16 - The 12th Summer Workshop on Multimodal Interfaces*. Centre for Telematics and Information Technology (CTIT), pp. 47–53

- **J. van Waterschoot** and M. Theune (2017). "Topic-Based Personalization of Dialogues with a Virtual Coach". In: *Persuasive Embodied Agents for Behavior Change Workshop at the 2017 Intelligent Virtual Agents Conference*, p. 5

- **J. van Waterschoot** (2017). "YRRSDS Position Paper". In: *Proceedings of the 13th Workshop on Spoken Dialogue Systems for PhDs, PostDocs & New Researchers*. SIGDial, pp. 51–52

- **J. van Waterschoot** and M. Theune (2018). "Challenges in Long-Term Evaluations of ECAs in the Real World". In: *First Workshop on Methodology and the Evaluation of Intelligent Virtual Agents (ME-IVA) at the Intelligent Virtual Agent Conference*, p. 2

- **J. van Waterschoot** and M. Theune (2019b). "Question-Answering Using Phonetic Information for Offthe Shelf Speech Recognition in Interactive Systems" (Groningen, Netherlands)

- **J. van Waterschoot** and M. Theune (2019a). "Measures in Longitudinal Interactions with Embodied Conversational Agents". In: *2nd Workshop on Methodology at Intelligent Virtual Agent Conference (ME-IVA 2019)*, p. 3

- **J. van Waterschoot**, I. Hendrickx, and M. Korte de (2020b). "BLISS: A Collection of Dutch Spoken Dialogue about What Makes People Happy". Poster (Utrecht, The Netherlands)

- **J. van Waterschoot** and M. Theune (2021). "Evaluating Conversational Question Generation: CoffeeBot". In: *Proceedings of Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI 2021).* ACM, p. 6

- L. van Uffelen, **J. van Waterschoot**, and M. Theune (2021). "Show Me Yours If I Show You Mine: Self-Disclosure in Conversational Agents". In: *HRI 2021 Workshop Robo-Identity: Artificial Identity and Multi-Embodiment*. ACM, p. 4

- **J. van Waterschoot**, I. Hendrickx, A. Khan, C. Cucchiarini, H. Strik, L. ten Bosch, and R. Tieben (2021). "Spoken Conversational Agents for Older Adults: Who Are the Stakeholders and What Do They Expect?" In: *Proceedings of the ACM IUI 2021 Workshops*. CEUR-WS, p. 6

- I. Hendrickx, **J. van Waterschoot**, A. Khan, L. ten Bosch, C. Cucchiarini, and H. Strik (2021). "Take Back Control: User Privacy and Transparency Concerns in Personalized Conversational Agents". In: *Proceedings of the ACM IUI 2021 Workshops*. CEUR-WS, p. 6

Other

- S. Koenderink, K. Truong, D. Nazareth, **J. van Waterschoot**, and R. de Jong (2019). "Twentse toepassingen van TST in de zorg". In: *DIXIT: Taal en Spraaktechnologie en Gezondheidszorg* 16 (December 2019), pp. 12–13

- C. Cucchiarini, I. Hendrickx, H. Strik, L. ten Bosch, M. Theune, **J. van Waterschoot**, and R. Tieben (2020). "BLISS dialoogsysteem". In: *DIXIT: Taal en Spraaktechnologie van betekenis* 17, pp. 18–19

Bibliography

Adiwardana, D., M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le (2020). *Towards a Human-like Open-Domain Chatbot*. arXiv: 2001.09977.

Akker, H. op den, R. op den Akker, T. Beinema, O. Banos, D. Heylen, B. Bedsted, A. Pease, C. Pelachaud, V. T. Salcedo, S. Kyriazakos, and H. Hermens (2018). "Council of Coaches - A Novel Holistic Behavior Change Coaching Approach". In: *Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and E-Health (ICT4AWE)*. SciTePress, pp. 219–226. DOI: 10.5220/0006787702190226.

Al Moubayed, S., J. Beskow, G. Skantze, and B. Granström (2012). "Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction". In: *Cognitive Behavioural Systems*. Springer, pp. 114–130. DOI: 10.1007/978-3-642-34584-5_9.

Allan, J. (2002). "Introduction to Topic Detection and Tracking". In: *Topic Detection and Tracking: Event-Based Information Organization*. Springer US, pp. 1–16. DOI: 10.1007/978-1-4615-0933-2_1.

Allan, J., J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang (1998). "Topic Detection and Tracking Pilot Study Final Report". In: *Proceedings of the DARPA Boradcast News Transcription and Understanding Workshop*. Morgan Kaufmann, pp. 194–218. DOI: 10.1184/R1/6626252.v1.

Allen, J., G. Ferguson, and A. Stent (2001). "An Architecture for More Realistic Conversational Systems". In: *Proceedings of the 6th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, pp. 1–8. DOI: 10.1145/359784.359822.

Almaev, T. R. and M. F. Valstar (2013). "Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition". In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, pp. 356–361. DOI: 10.1109/ACII.2013.65.

Aminuddin, R., A. Sharkey, and L. Levita (2016). "Interaction with the Paro Robot May Reduce Psychophysiological Stress Responses". In: *2016 11th*

ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 593–594. DOI: 10.1109/HRI.2016.7451872.

Aranyi, G., F. Pecune, F. Charles, C. Pelachaud, and M. Cavazza (2016). "Affective Interaction with a Virtual Character Through an fNIRS Brain-Computer Interface". In: *Frontiers in Computational Neuroscience* 10. DOI: 10.3389/fncom.2016.00070.

Aylett, M. P. and C. J. Pidcock (2007). "The CereVoice Characterful Speech Synthesiser SDK". In: *Proceedings of the 7th International Conference on Intelligent Virtual Agents*. Springer-Verlag, pp. 413–414. DOI: 10.1007/978-3-540-74997-4_65.

Bajones, M., D. Fischinger, A. Weiss, P. D. L. Puente, D. Wolf, M. Vincze, T. Körtner, M. Weninger, K. Papoutsakis, D. Michel, A. Qammaz, P. Panteleris, M. Foukarakis, I. Adami, D. Ioannidi, A. Leonidis, M. Antona, A. Argyros, P. Mayer, P. Panek, H. Eftring, and S. Frennert (2019). "Results of Field Trials with a Mobile Service Robot for Older Adults in 16 Private Households". In: *ACM Transactions on Human-Robot Interaction* 9.2, 10:1–10:27. DOI: 10.1145/3368554.

Bak, J., C.-Y. Lin, and A. Oh (2014). "Self-Disclosure Topic Model for Classifying and Analyzing Twitter Conversations". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pp. 1986–1996. DOI: 10.3115/v1/D14-1213.

Baltrusaitis, T., A. Zadeh, Y. C. Lim, and L.-P. Morency (2018). "OpenFace 2.0: Facial Behavior Analysis Toolkit". In: *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 59–66. DOI: 10.1109/FG.2018.00019.

Banchs, R. E. (2012). "Movie-DiC: A Movie Dialogue Corpus for Research and Development". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*. Association for Computational Linguistics, pp. 203–207.

Barak, A. and O. Gluck-Ofri (2007). "Degree and Reciprocity of Self-Disclosure in Online Forums". In: *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 10.3, pp. 407–417. DOI: 10.1089/cpb.2006.9938. pmid: 17594265.

Bartneck, C., D. Kulić, E. Croft, and S. Zoghbi (2009). "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots". In: *International Journal of Social Robotics* 1.1, pp. 71–81. DOI: 10.1007/s12369-008-0001-3.

Beinema, T., D. Davison, D. Reidsma, O. Banos, M. Bruijnes, B. Donval, Á. F. Valero, D. Heylen, D. Hofs, G. Huizing, R. B. Kantharaju, R. Klaassen, J. Kolkmeier, K. Konsolakis, A. Pease, C. Pelachaud, D. Simonetti, M. Snaith,

V. Traver, J. van Loon, J. Visser, M. Weusthof, F. Yunus, H. Hermens, and H. op den Akker (2021). "Agents United: An Open Platform for Multi-Agent Conversational Systems". In: *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, pp. 17–24. DOI: 10.1145/3472306.3478352.

Berg, M. M. (2015). "NADIA: A Simplified Approach Towards the Development of Natural Dialogue Systems". In: *Natural Language Processing and Information Systems*. Springer International Publishing, pp. 144–150. DOI: 10.1007/978-3-319-19581-0_12.

Bethel, C. L., J. E. Cossitt, Z. Henkel, and K. Baugus (2020). "Qualitative Interview Techniques for Human-Robot Interactions". In: *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer International Publishing, pp. 145–174. DOI: 10.1007/978-3-030-42307-0_6.

Beveridge, M. and J. Fox (2006). "Automatic Generation of Spoken Dialogue from Medical Plans and Ontologies". In: *Journal of Biomedical Informatics* 39.5, pp. 482–499. DOI: 10.1016/j.jbi.2005.12.008.

Bickmore, T. W. and J. Cassell (1999). "Small Talk and Conversational Storytelling in Embodied Interface Agents". In: *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. AAAI, pp. 87–92.

Bickmore, T. W. and T. Giorgino (2006). "Health Dialog Systems for Patients and Consumers". In: *Journal of Biomedical Informatics* 39.5, pp. 556–571. DOI: 10.1016/j.jbi.2005.12.004.

Bickmore, T. W. and R. W. Picard (2005). "Establishing and Maintaining Long-Term Human-Computer Relationships". In: *ACM Transactions on Computer-Human Interaction* 12.2, pp. 293–327. DOI: 10.1145/1067860.1067867.

Bickmore, T. W., D. Schulman, and L. Yin (2010). "Maintaining Engagement in Long-Term Interventions with Relational Agents". In: *Applied Artificial Intelligence* 24.6, pp. 648–666. DOI: 10.1080/08839514.2010.492259.

Bickmore, T. W., H. Trinh, S. Olafsson, T. K. O'Leary, R. Asadi, N. M. Rickles, and R. Cruz (2018). "Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant". In: *Journal of Medical Internet Research* 20.9, e11510. DOI: 10.2196/11510.

Bird, S., E. Klein, and E. Loper (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. "O'Reilly Media, Inc." 506 pp.

Blei, D. M., A. Y. Ng, M. I. Jordan, A. S. Edu, M. I. Jordan, and J. B. Edu (2003). "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3, pp. 993–1022.

Bocklisch, T., J. Faulkner, N. Pawlowski, and A. Nichol (2017). "Rasa: Open Source Language Understanding and Dialogue Management". In: *Conversa-*

tional AI Workshop at the 31st Conference on Neural Information Processing Systems (NIPS 2017). NIPS, pp. 1–9.

Bohus, D., S. Andrist, and M. Jalobeanu (2017). "Rapid Development of Multimodal Interactive Systems: A Demonstration of Platform for Situated Intelligence". In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, pp. 493–494. DOI: 10.1145/3136755.3143021.

Bohus, D., A. Raux, T. Harris, M. Eskenazi, and A. Rudnicky (2007). "Olympus: An Open-Source Framework for Conversational Spoken Language Interface Research". In: *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*. Association for Computational Linguistics, pp. 32–39.

Bohus, D. and A. I. Rudnicky (2009). "The RavenClaw Dialog Management Framework: Architecture and Systems". In: *Computer Speech and Language* 23.3, pp. 332–361. DOI: 10.1016/j.csl.2008.10.001.

Bos, J., E. Klein, O. Lemon, and T. Oka (2003). "DIPPER: Description and Formalisation of an Information-State Update Dialogue System Architecture". In: *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*. Association for Computational Linguistics, pp. 115–124.

Bosch, A. van den and G. Bouma, eds. (2011). *Interactive Multi-Modal Question-Answering*. Springer Berlin Heidelberg. 1-279. DOI: 10.1007/978-3-642-17525-1.

Bosdriesz, L. (2020). "Adding Speech to Dialogues with a Council of Coaches". MA thesis. University of Twente. 120 pp.

Bowden, K., T. Nilsson, C. Spencer, K. Cengiz, A. Ghitulescu, and J. B. van Waterschoot (2017). "I Probe, Therefore I Am: Designing a Virtual Journalist with Human Emotions". In: *Proceedings of eNTERFACE'16 - The 12th Summer Workshop on Multimodal Interfaces*. Centre for Telematics and Information Technology (CTIT), pp. 47–53.

Braun, D., A. Hernandez Mendez, F. Matthes, and M. Langen (2017). "Evaluating Natural Language Understanding Services for Conversational Question Answering Systems". In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 174–185. DOI: 10.18653/v1/W17-5522.

Braun, V. and V. Clarke (2006). "Using Thematic Analysis in Psychology". In: *Qualitative Research in Psychology* 3.2, pp. 77–101. DOI: 10.1191/1478088706qp063oa.

Brennan, K. A., C. L. Clark, and P. R. Shaver (1998). "Self-Report Measurement of Adult Attachment: An Integrative Overview". In: *Attachment Theory and Close Relationships*. The Guilford Press, pp. 46–76.

Breuing, A. and I. Wachsmuth (2012). "Let's Talk Topically with Artificial Agents! Providing Agents with Humanlike Topic Awareness in Everyday Dialog Situations." In: *Proceedings of the 4th International Conference on Agents and Artificial Intelligence*, pp. 62–71.

Breuing, A. and I. Wachsmuth (2013). "Talking Topically to Artificial Dialog Partners: Emulating Humanlike Topic Awareness in a Virtual Agent". In: *Communications in Computer and Information Science* 358.980, pp. 392–406. DOI: 10.1007/978-3-642-36907-0_26.

Breuing, A., U. Waltinger, and I. Wachsmuth (2011). "Harvesting Wikipedia Knowledge to Identify Topics in Ongoing Natural Language Dialogs". In: *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. Vol. 1. IEEE, pp. 445–450. DOI: 10.1109/WI-IAT.2011.158.

Broekens, J., M. Heerink, and H. Rosendal (2009). "Assistive Social Robots in Elderly Care: A Review". In: *Gerontechnology*, pp. 94–103.

Brooke, J. (1996). "SUS - A Quick and Dirty Usability Scale". In: *Usability Evaluation in Industry*. 1st. Taylor and Francis, pp. 189–195.

Brown, G. and G. Yule (1983). *Discourse Analysis*. 1st. Cambridge University Press. 302 pp. DOI: 10.1017/CBO9780511805226.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei (2020). "Language Models Are Few-Shot Learners". In: *34th Conference on Neural Information Processing Systems*. NIPS.

Bruijnes, M., R. op den Akker, A. Hartholt, and D. Heylen (2015). "Virtual Suspect William". In: *Intelligent Virtual Agents*. Springer International Publishing, pp. 67–76. DOI: 10.1007/978-3-319-21996-7_7.

Bunt, H., 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 (2010). "Towards an ISO Standard for Dialogue Act Annotation". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), pp. 2548–2555.

Burger, S. director. (2015). *Ik Ben Alice / Alice Cares*. KRO-NCRV.

Burkert, O., C. Brom, R. Kadlec, and J. Lukavský (2010). "Timing in Episodic Memory: Virtual Characters in Action". In: *Proceedings of the Remembering Who We Are - Human Memory for Artificial Agents Symposium at the AISB 2010 Convention*. AISB, p. 9.

Burtsev, M., V. Logacheva, V. Malykh, I. V. Serban, R. Lowe, S. Prabhumoye, A. W. Black, A. Rudnicky, and Y. Bengio (2018). “The First Conversational Intelligence Challenge”. In: *The NIPS ’17 Competition: Building Intelligent Systems*. Springer International Publishing, pp. 25–46. DOI: 10.1007/978-3-319-94042-7_2.

Buschmeier, H., Z. Malisz, J. Skubisz, M. Wlodarczak, I. Wachsmuth, S. Kopp, and P. Wagner (2014). “ALICO: A Multimodal Corpus for the Study of Active Listening”. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*. European Language Resources Association (ELRA), pp. 3638–3643.

Cafaro, A., M. Bruijnes, J. van Waterschoot, C. Pelachaud, M. Theune, and D. K. J. Heylen (2017a). “Selecting and Expressing Communicative Functions in a SAIBA-Compliant Agent Framework”. In: *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings*. Springer, pp. 73–82. DOI: 10.1007/978-3-319-67401-8_8.

Cafaro, A., J. Wagner, T. Baur, S. Dermouche, M. Torres Torres, C. Pelachaud, E. André, and M. Valstar (2017b). “The NoXi Database: Multimodal Recordings of Mediated Novice-Expert Interactions”. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, pp. 350–359. DOI: 10.1145/3136755.3136780.

Calijorne Soares, M. A. and F. S. Parreiras (2020). “A Literature Review on Question Answering Techniques, Paradigms and Systems”. In: *Journal of King Saud University - Computer and Information Sciences* 32.6, pp. 635–646. DOI: 10.1016/j.jksuci.2018.08.005.

Calix, R. A., S. A. Mallepudi, B. Chen, and G. M. Knapp (2010). “Emotion Recognition in Text for 3-D Facial Expression Rendering”. In: *IEEE Transactions on Multimedia* 12.6, pp. 544–551. DOI: 10.1109/TMM.2010.2052026.

Calvo, R. A. and S. D’Mello (2010). “Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications”. In: *IEEE Transactions on Affective Computing* 1.1, pp. 18–37. DOI: 10.1109/T-AFFC.2010.1.

Campos, J., J. Kennedy, and J. F. Lehman (2018). “Challenges in Exploiting Conversational Memory in Human-Agent Interaction”. In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*. IFAAMAS, pp. 1649–1657.

Campos, J. and A. Paiva (2010). “MAY: My Memories Are Yours”. In: *Intelligent Virtual Agents*. Springer, pp. 406–412. DOI: 10.1007/978-3-642-15892-6_44.

Cang, X. L., P. Bucci, A. Strang, J. Allen, K. MacLean, and H. S. Liu (2015). “Different Strokes and Different Folks: Economical Dynamic Surface Sensing and Affect-Related Touch Recognition”. In: *Proceedings of the 2015 ACM on Inter-*

national Conference on Multimodal Interaction. Association for Computing Machinery, pp. 147–154. DOI: 10.1145/2818346.2820756.

Canonico, M. and L. D. Russis (2018). "A Comparison and Critique of Natural Language Understanding Tools". In: *Proceedings of the Ninth International Conference on Cloud Computing, GRIDs, and Virtualization*. IARIA, p. 6.

Carletta, J., S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner (2005). "The AMI Meeting Corpus: A Pre-Announcement". In: *Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction*. Springer-Verlag, pp. 28–39. DOI: 10.1007/11677482_3.

Carletta, J., A. Isard, S. Isard, J. C. Kowtko, G. Doherty-Sneddon, and A. H. Anderson (1997). "The Reliability of a Dialogue Structure Coding Scheme". In: *Journal Computational Linguistics* 23.1, pp. 13–31. DOI: 10.1024/0036-7281.144.5.213.

Carpinella, C. M., A. B. Wyman, M. A. Perez, and S. J. Stroessner (2017). "The Robotic Social Attributes Scale (RoSAS)". In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Vol. Part F1271. Association for Computing Machinery, pp. 254–262. DOI: 10.1145/2909824.3020208.

Carroll, L. (1865). *Alice's Adventures in Wonderland*. Ed. by D. Widger. 1st. Vol. 1. Project Gutenberg. 1-21. DOI: 10.1017/CBO9781139568555. pmid: 11615418.

Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone (1994). "Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents". In: *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. Association for Computing Machinery, pp. 413–420. DOI: 10.1145/192161.192272.

Cassell, J., H. H. Vilhjálmsson, and T. Bickmore (2004). "BEAT: The Behavior Expression Animation Toolkit". In: *Life-Like Characters: Tools, Affective Functions, and Applications*. Springer, pp. 163–185. DOI: 10.1007/978-3-662-08373-4_8.

Chali, Y. and S. A. Hasan (2015). "Towards Topic-to-Question Generation". In: *Computational Linguistics* 41.1, pp. 1–20. DOI: 10.1162/COLI_a_00206.

Chang, J. P., J. Boyd-Graber, S. Gerrish, C. Wang, and D. M. Blei (2009). "Reading Tea Leaves: How Humans Interpret Topic Models". In: *Proceedings of the 22nd International Conference on Neural Information Processing Systems*. Curran Associates Inc., pp. 288–296.

Chaoua, I., D. R. Recupero, S. Consoli, A. Härmä, and R. Helaoui (2018). "Detecting and Tracking Ongoing Topics in Psychotherapeutic Conversations". In:

Proceedings of the First Joint Workshop on AI in Health. Vol. 2142. CEUR-WS, pp. 97–108.

Chauhan, D. S., D. S R, A. Ekbal, and P. Bhattacharyya (2020). "Sentiment and Emotion Help Sarcasm? A Multi-Task Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 4351–4360.

Chevalier, P., J. J. Li, E. Ainger, A. M. Alcorn, S. Babovic, V. Charisi, S. Petrovic, B. R. Schadenberg, E. Pellicano, and V. Evers (2017). "Dialogue Design for a Robot-Based Face-Mirroring Game to Engage Autistic Children with Emotional Expressions". In: *Social Robotics*. Springer International Publishing, pp. 546–555. DOI: 10.1007/978-3-319-70022-9_54.

Cheyer, A. and D. Martin (2001). "The Open Agent Architecture". In: *Autonomous Agents and Multi-Agent Systems* 4.1, pp. 143–148. DOI: 10.1023/A:1010091302035.

Chowanda, A., P. Blanchfield, M. Flintham, and M. Valstar (2014). "ERiSA: Building Emotionally Realistic Social Game-Agents Companions". In: *Intelligent Virtual Agents*. Springer International Publishing, pp. 134–143. DOI: 10.1007/978-3-319-09767-1_16.

Clark, H. H. (1996). *Using Language*. Cambridge University Press. 446 pp.

Clark, L., N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan (2019). "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, pp. 1–12. DOI: 10.1145/3290605.3300705.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa (2011). "Natural Language Processing (Almost) from Scratch". In: *The Journal of Machine Learning Research* 12 (null), pp. 2493–2537.

Coronado, M., C. A. Iglesias, Á. Carrera, and A. Mardomingo (2018). "A Cognitive Assistant for Learning Java Featuring Social Dialogue". In: *International Journal of Human-Computer Studies* 117, pp. 55–67. DOI: 10.1016/j.ijhcs.2018.02.004.

Cowan, B. R., N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira (2017). "What Can I Help You With? Infrequent Users' Experiences of Intelligent Personal Assistants". In: *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, pp. 1–12. DOI: 10.1145/3098279.3098539.

Cucchiarini, C., I. Hendrickx, H. Strik, L. ten Bosch, M. Theune, J. van Waterschoot, and R. Tieben (2020). "BLISS dialoogsysteem". In: *DIXIT: Taal en Spraaktechnologie van betekenis* 17, pp. 18–19.

Davis, F. D. (1989). "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". In: *MIS Quarterly* 13.3, pp. 319–340. DOI: 10.2307/249008.

Davison, D. P., F. M. Wijnen, V. Charisi, J. van der Meij, V. Evers, and D. Reidsma (2020). "Working with a Social Robot in School: A Long-Term Real-World Unsupervised Deployment". In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, pp. 63–72. DOI: 10.1145/3319502.3374803.

De Carolis, B., I. Mazzotta, N. Novielli, and S. Pizzutilo (2013). "User Modeling in Social Interaction with a Caring Agent". In: *User Modeling and Adaptation for Daily Routines: Providing Assistance to People with Special Needs*. Springer, pp. 89–116. DOI: 10.1007/978-1-4471-4778-7_4.

De Smedt, T. and W. Daelemans (2012). "Pattern for Python". In: *Journal of Machine Learning Research* 13.66, pp. 2063–2067.

Degottex, G., J. Kane, T. Drugman, T. Raitio, and S. Scherer (2014). "COVAREP — A Collaborative Voice Analysis Repository for Speech Technologies". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 960–964. DOI: 10.1109/ICASSP.2014.6853739.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

Di Fabio, A., S. Conia, and R. Navigli (2019). "VerbAtlas: A Novel Large-Scale Verbal Semantic Resource and Its Application to Semantic Role Labeling". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, pp. 627–637. DOI: 10.18653/v1/D19-1058.

Eggins, S. and D. Slade (2001). *Analysing Casual Conversation*. 1st. Continuum. 333 pp.

Ekman, P. (1999). "Basic Emotions". In: *Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd, pp. 45–60. DOI: 10.1002/0470013494.ch3.

Ekman, P., W. V. Friesen, and J. C. Hager (2002). *Facial Action Coding System: The Manual*. Research Nexus. 512 pp.

El Asri, L., H. Schulz, S. Sharma, J. Zumer, J. Harris, E. Fine, R. Mehrotra, and K. Suleman (2017). "Frames: A Corpus for Adding Memory to Goal-Oriented

Dialogue Systems". In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 207–219. DOI: 10.18653/v1/W17-5526.

Elvir, M. (2010). "Episodic Memory Model For Embodied Conversational Agents". MA thesis. University of Central Florida. 199 pp.

Elvir, M., A. J. Gonzalez, C. Walls, and B. Wilder (2017). "Remembering a Conversation – A Conversational Memory Architecture for Embodied Conversational Agents". In: *Journal of Intelligent Systems* 26.1, pp. 1–21. DOI: 10.1515/jisys-2015-0094.

Erp, J. B. F. van and A. Toet (2015). "Social Touch in Human–Computer Interaction". In: *Frontiers in Digital Humanities* 2. DOI: 10.3389/fdigh.2015.00002.

Eyben, F., F. Weninger, S. Squartini, and B. Schuller (2013). "Real-Life Voice Activity Detection with LSTM Recurrent Neural Networks and an Application to Hollywood Movies". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 483–487. DOI: 10.1109/ICASSP.2013.6637694.

Eyben, F., M. Wöllmer, and B. Schuller (2010). "Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor". In: *Proceedings of the 18th ACM International Conference on Multimedia*. Association for Computing Machinery, pp. 1459–1462. DOI: 10.1145/1873951.1874246.

Fan, H. and M. S. Poole (2006). "What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems". In: *Journal of Organizational Computing and Electronic Commerce* 16.3-4, pp. 179–202. DOI: 10.1080/10919392.2006.9681199.

Fasya, E. L. (2017). "Automatic Question Generation for Virtual Humans". MA thesis. University of Twente. 105 pp.

Ferdinando, H., T. Seppänen, and E. Alasaarela (2016). "Comparing Features from ECG Pattern and HRV Analysis for Emotion Recognition System". In: *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1–6. DOI: 10.1109/CIBCB.2016.7758108.

Floridi, L. and M. Chiriatti (2020). "GPT-3: Its Nature, Scope, Limits, and Consequences". In: *Minds and Machines* 30.4, pp. 681–694. DOI: 10.1007/s11023-020-09548-1.

Foster, M. E., B. Craenen, A. Deshmukh, O. Lemon, E. Bastianelli, C. Dondrup, I. Papaioannou, A. Vanzo, J.-M. Odobez, O. Canévet, Y. Cao, W. He, A. Martínez-González, P. Motlicek, R. Siegfried, R. Alami, K. Belhassein, G. Buisan, A. Clodic, A. Mayima, Y. Sallami, G. Sarthou, P.-T. Singamaneni, J. Waldhart, A. Mazel, M. Caniot, M. Niemelä, P. Heikkilä, H. Lammi, and A.

Tammela (2019). “MuMMER: Socially Intelligent Human-Robot Interaction in Public Spaces”. In: *Proceedings of AI-HRI 2019*. AAAI, p. 8.

Galvão Gomes da Silva, J., D. J. Kavanagh, T. Belpaeme, L. Taylor, K. Beeson, and J. Andrade (2018). “Experiences of a Motivational Interview Delivered by a Robot: Qualitative Study”. In: *Journal of Medical Internet Research* 20.5, e116. DOI: 10.2196/jmir.7737. pmid: 29724701.

Gandhe, S. and D. Traum (2007). “Creating Spoken Dialogue Characters from Corpora Without Annotations”. In: *INTERSPEECH-2007*. ISCA, pp. 2201–2204.

Ganster, T., S. Eimler, A. M. Rosenthal-von der Pütten, L. Hoffmann, and N. Krämer (2010). “Methodological Considerations for Long-Term Experience with Robots and Agents.” In: *Proceedings of EMCSR 2010*. OFAI.

Gatt, A. and E. Reiter (2009). “SimpleNLG: A Realisation Engine for Practical Applications”. In: *Proceedings of the 12th European Workshop on Natural Language Generation*. Association for Computational Linguistics, pp. 90–93. DOI: 10.5555/1610195.1610208.

Gebhard, P., G. Mehlmann, and M. Kipp (2012). “Visual SceneMaker-a Tool for Authoring Interactive Virtual Characters”. In: *Journal on Multimodal User Interfaces* 6.1-2, pp. 3–11. DOI: 10.1007/s12193-011-0077-1.

Gilmartin, E., C. Saam, C. Vogel, N. Campbell, and V. Wade (2018). “Just Talking - Modelling Casual Conversation”. In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 51–59. DOI: 10.18653/v1/w18-5006.

Glas, D. F., K. Wada, M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita (2017). “Personal Greetings: Personalizing Robot Utterances Based on Novelty of Observed Behavior”. In: *International Journal of Social Robotics* 9.2, pp. 181–198. DOI: 10.1007/s12369-016-0385-4.

Glas, N. and C. Pelachaud (2015a). “Definitions of Engagement in Human-Agent Interaction”. In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, pp. 944–949. DOI: 10.1109/ACII.2015.7344688.

Glas, N. and C. Pelachaud (2015b). “Topic Transition Strategies for an Information-Giving Agent”. In: *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*. Association for Computational Linguistics, pp. 146–155. DOI: 10.18653/v1/W15-4725.

Glas, N. and C. Pelachaud (2018). “Topic Management for an Engaging Conversational Agent”. In: *International Journal of Human-Computer Studies* 120, pp. 107–124. DOI: 10.1016/j.ijhcs.2018.07.007.

Glas, N., K. Prepin, and C. Pelachaud (2015). “Engagement Driven Topic Selection for an Information-Giving Agent”. In: *Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2015 - goDial)*.

Gockley, R., A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. Schultz, and Jue Wang (2005). "Designing Robots for Long-Term Social Interaction". In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1338–1343. DOI: 10.1109/IROS.2005.1545303.

Gouaillier, D., V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier (2009). "Mechatronic Design of NAO Humanoid". In: *2009 IEEE International Conference on Robotics and Automation*, pp. 769–774. DOI: 10.1109/ROBOT.2009.5152516.

Graaf, M. M. A. de, S. B. Allouch, and J. A. G. M. van Dijk (2019). "Why Would I Use This in My Home? A Model of Domestic Social Robot Acceptance". In: *Human–Computer Interaction* 34.2, pp. 115–173. DOI: 10.1080/07370024.2017.1312406.

Graaf, M. M. A. de, S. B. Allouch, and J. A. G. M. van Dijk (2016). "Long-Term Acceptance of Social Robots in Domestic Environments: Insights from a User's Perspective". In: *The 2016 AAAI Spring Symposium Series*. AAAI, pp. 96–103.

Grandgeorge, M. (2020). "Evaluating Human-Robot Interaction with Ethology". In: *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer International Publishing, pp. 257–268. DOI: 10.1007/978-3-030-42307-0_10.

Gratch, J., N. Wang, J. Gerten, E. Fast, and R. Duffy (2007). "Creating Rapport with Virtual Agents". In: *Intelligent Virtual Agents*. Springer Berlin Heidelberg, pp. 125–138. DOI: http://dx.doi.org/10.1007/978-3-540-74997-4_12.

Gray, J. and A. Reuter (1992). *Transaction Processing: Concepts and Techniques*. 1st. Morgan Kaufmann Publishers Inc. 1070 pp.

Grosz, B. J. (1974). "The Structure of Task Oriented Dialogs". In: *Proceedings IEEE Speech Symposium*. SRI International Menlo Park United States, p. 13.

Grosz, B. J. and C. L. Sidner (1986). "Attention, Intentions, and the Structure of Discourse". In: *Computational Linguistics* 12.3, pp. 175–204. DOI: 10.14348/molcells.2014.0104.

Gruenstein, A., I. McGraw, and I. Badr (2008). "The WAMI Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces". In: *Proceedings of the 10th International Conference on Multimodal Interfaces*. Association for Computing Machinery, pp. 141–148. DOI: 10.1145/1452392.1452420.

Guger, C., B. Z. Allison, and N. Mrachacz-Kersting (2019). "Brain-Computer Interface Research: A State-of-the-Art Summary 7". In: *Brain-Computer Interface Research: A State-of-the-Art Summary 7*. Springer International Publishing, pp. 1–9. DOI: 10.1007/978-3-030-05668-1_1.

Gundel, K. (1985). “Shared Knowledge and Topicality”. In: *Journal of Pragmatics* 9 (August 1982), pp. 83–107.

Gunson, N., W. Sieińska, C. Walsh, C. Dondrup, and O. Lemon (2020). “It’s Good to Chat? Evaluation and Design Guidelines for Combining Open-Domain Social Conversation with Task-Based Dialogue in Intelligent Buildings”. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, pp. 1–8. DOI: 10.1145/3383652.3423889.

Habib, R., S. Mariooryad, M. Shannon, E. Battenberg, R. J. Skerry-Ryan, D. Stanton, D. Kao, and T. Bagby (2019). “Semi-Supervised Generative Modeling for Controllable Speech Synthesis”. In: *International Conference on Learning Representations*, p. 18.

Han, X., C. Gao, and Y. Yu (2017). “DeepSketch2Face: A Deep Learning Based Sketching System for 3D Face and Caricature Modeling”. In: *ACM Transactions on Graphics* 36.4, 126:1–126:12. DOI: 10.1145/3072959.3073629.

Hanson, D., S. Baurmann, T. Riccio, R. Margolin, T. Dockins, M. Tavares, and K. Carpenter (2008). “Zeno: A Cognitive Character”. In: *Technical Report WS-08-08 AAAI Workshop on Mobile Robots*. AAAI Press, pp. 9–11.

Hartholt, A., D. Traum, S. C. Marsella, A. Shapiro, G. Stratou, A. Leuski, L.-P. Morency, and J. Gratch (2013). “All Together Now”. In: *Intelligent Virtual Agents*. Springer, pp. 368–381. DOI: 10.1007/978-3-642-40415-3_33.

Hayano, K. (2013). “Question Design in Conversation”. In: *The Handbook of Conversation Analysis*. John Wiley & Sons, Ltd, pp. 395–414. DOI: 10.1002/9781118325001.ch19.

Hebesberger, D., T. Koertner, C. Gisinger, J. Pripfl, and C. Dondrup (2016). “Lessons Learned from the Deployment of a Long-Term Autonomous Robot as Companion in Physical Therapy for Older Adults with Dementia a Mixed Methods Study”. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 27–34. DOI: 10.1109/HRI.2016.7451730.

Hecht, M. L. (1978). “The Conceptualization and Measurement of Interpersonal Communication Satisfaction”. In: *Human Communication Research* 4.3, pp. 253–264. DOI: 10.1111/j.1468-2958.1978.tb00614.x.

Heerink, M., B. Kröse, V. Evers, and B. Wielinga (2010). “Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model”. In: *International Journal of Social Robotics* 2.4, pp. 361–375. DOI: 10.1007/s12369-010-0068-5.

Heilman, M. and N. A. Smith (2010). “Good Question! Statistical Ranking for Question Generation”. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 609–617.

Heimerl, A., T. Baur, F. Lingenfelser, J. Wagner, and E. André (2019). "NOVA - A Tool for eXplainable Cooperative Machine Learning". In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 109–115. DOI: 10.1109/ACII.2019.8925519.

Heylen, D. K. J., B. van Dijk, and A. Nijholt (2012). "Robotic Rabbit Companions: Amusing or a Nuisance?" In: *Journal on Multimodal User Interfaces* 5.1, pp. 53–59. DOI: 10.1007/s12193-011-0083-3.

Heylen, D. K. J., S. Kopp, S. C. Marsella, C. Pelachaud, and H. Vilhjálmsson (2008). "The Next Step towards a Function Markup Language". In: *Intelligent Virtual Agents*. Springer, pp. 270–280. DOI: 10.1007/978-3-540-85483-8_28.

Higashinaka, R., T. Meguro, K. Imamura, H. Sugiyama, T. Makino, and Y. Matsuo (2014). "Evaluating Coherence in Open Domain Conversational Systems". In: *INTERSPEECH-2014*. ISCA, pp. 130–134.

Hobbs, J. R. (1990). "Topic Drift". In: *Topic Drift*. Vol. 38. Ablex Norwood, NJ, USA, pp. 3–22.

Hoegen, R., D. Aneja, D. McDuff, and M. Czerwinski (2019). "An End-to-End Conversational Style Matching Agent". In: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, pp. 111–118. DOI: 10.1145/3308532.3329473.

Hone, K. S. and R. Graham (2001). "Subjective Assessment of Speech-System Interface Usability". In: *7th European Conference on Speech Communication and Technology (EUROSPEECH 2001)*. ISCA, pp. 2083–2086.

Horvath, A. O. and L. S. Greenberg (1989). "Development and Validation of the Working Alliance Inventory". In: *Journal of Counseling Psychology* 36.2, pp. 223–233. DOI: 10.1037/0022-0167.36.2.223.

Hu, W., B. Liu, J. Ma, D. Zhao, and R. Yan (2018). "Aspect-Based Question Generation". In: *Proceedings of the Workshop Track at the Sixth International Conference on Learning Representations*. OpenReview.net, p. 10.

Huang, K., M. Yeomans, A. W. Brooks, J. Minson, F. Gino, K. Huang, M. Yeomans, A. W. Brooks, J. Minson, and F. Gino (2017). "It Doesn' t Hurt to Ask : Question-Asking Increases Liking". In: *Journal of Personality and Social Psychology* 113.3, pp. 430–452.

Huang, Y.-M., R. Shadiev, and W.-Y. Hwang (2016). "Investigating the Effectiveness of Speech-to-Text Recognition Applications on Learning Performance and Cognitive Load". In: *Computers & Education* 101, pp. 15–28. DOI: 10.1016/j.compedu.2016.05.011.

Huber, M., J. André Knottnerus, L. Green, H. Van Der Horst, A. R. Jadad, D. Kromhout, B. Leonard, K. Lorig, M. I. Loureiro, J. W. Van Der Meer, P. Schnabel, R. Smith, C. Van Weel, and H. Smid (2011). "How Should We Define Health?" In: *BMJ (Online)* 343.7817, pp. 1–3. DOI: 10.1136/bmj.d4163. pmid: 21791490.

Huisman, G. (2017). "Social Touch Technology: A Survey of Haptic Technology for Social Touch". In: *IEEE Transactions on Haptics* 10.3, pp. 391–408. DOI: 10.1109/TOH.2017.2650221.

Huisman, G., M. Bruijnes, J. Kolkmeier, M. Jung, A. Darriba Frederiks, and Y. Rybarczyk (2014). "Touching Virtual Agents: Embodiment and Mind". In: *Innovative and Creative Developments in Multimodal Interaction Systems*. Springer, pp. 114–138. DOI: 10.1007/978-3-642-55143-7_5.

Huisman, G., A. D. Frederiks, B. V. Dijk, D. Hevlen, and B. Kröse (2013). "The TaSST: Tactile Sleeve for Social Touch". In: *2013 World Haptics Conference (WHC)*, pp. 211–216. DOI: 10.1109/WHC.2013.6548410.

Hunt, A. J. and A. Black (1996). "Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database". In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 1. IEEE, 373–376 vol. 1. DOI: 10.1109/ICASSP.1996.541110.

Huttenrauch, H. and K. S. Eklundh (2002). "Fetch-and-Carry with CERO: Observations from a Long-Term User Study with a Service Robot". In: *11th IEEE International Workshop on Robot and Human Interactive Communication Proceedings*. IEEE, pp. 158–163. DOI: 10.1109/ROMAN.2002.1045615.

Hutto, C. J. and E. Gilbert (2014). "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". In: *Proceedings of the Eight International Conference on Weblogs and Social Media*. AAAI Press, p. 10.

Inaba, M. and K. Takahashi (2018). "Estimating User Interest from Open-Domain Dialogue". In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 32–40. DOI: 10.18653/v1/w18-5004.

Inoue, K., K. Hara, D. Lala, K. Yamamoto, S. Nakamura, K. Takanashi, and T. Kawahara (2020). "Job Interviewer Android with Elaborate Follow-up Question Generation". In: *Proceedings of the 2020 International Conference on Multimodal Interaction*. Association for Computing Machinery, pp. 324–332. DOI: 10.1145/3382507.3418839.

Irfan, B., M. Hellou, A. Mazel, and T. Belpaeme (2020). "Challenges of a Real-World HRI Study with Non-Native English Speakers: Can Personalisation Save the Day?" In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, pp. 272–274. DOI: 10.1145/3371382.3378278.

Irfan, B., A. Ramachandran, S. Spaulding, D. F. Glas, I. Leite, and K. L. Koay (2019). "Personalization in Long-Term Human-Robot Interaction". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 685–686. DOI: 10.1109/HRI.2019.8673076.

Isonishi, S., K. Inoue, D. Lala, K. Takanashi, and T. Kawahara (2021). "Response Generation to Out-of-Database Questions for Example-Based Dialogue Systems". In: *Conversational Dialogue Systems for the Next Decade*. Vol. 704. Springer, pp. 303–317. DOI: 10.1007/978-981-15-8395-7_23.

Jaber, R. and D. McMillan (2020). "Conversational User Interfaces on Mobile Devices: Survey". In: *Proceedings of the 2nd Conference on Conversational User Interfaces*. Association for Computing Machinery, pp. 1–11. DOI: 10.1145/3405755.3406130.

Jaiswal, S. and M. Valstar (2016). "Deep Learning the Dynamic Appearance and Shape of Facial Action Units". In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 1–8. DOI: 10.1109/WACV.2016.7477625.

Jaiswal, S., M. Valstar, K. Kusumam, and C. Greenhalgh (2019). "Virtual Human Questionnaire for Analysis of Depression, Anxiety and Personality". In: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, pp. 81–87. DOI: 10.1145/3308532.3329469.

James, J., B. T. Balamurali, C. I. Watson, and B. MacDonald (2020). "Empathetic Speech Synthesis and Testing for Healthcare Robots". In: *International Journal of Social Robotics*. DOI: 10.1007/s12369-020-00691-4.

Jong, R. de and M. Theune (2018). "Going Dutch: Creating SimpleNLG-NL". In: *Proceedings of the 11th International Conference on Natural Language Generation*. Association for Computational Linguistics, pp. 73–78. DOI: 10.18653/v1/W18-6508.

Jourard, S. M. (1961). "Self-Disclosure Patterns in British and American College Females". In: *The Journal of Social Psychology* 54.2, pp. 315–320. DOI: 10.1080/00224545.1961.9922143.

Jung, M. M., X. L. Cang, M. Poel, and K. E. MacLean (2015). "Touch Challenge '15: Recognizing Social Touch Gestures". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. Association for Computing Machinery, pp. 387–390. DOI: 10.1145/2818346.2829993.

Jung, M. M., M. Poel, D. Reidsma, and D. K. J. Heylen (2017). "A First Step toward the Automatic Understanding of Social Touch for Naturalistic Human–Robot Interaction". In: *Frontiers in ICT* 4. DOI: 10.3389/fict.2017.00003.

Jung, Y. and K. M. Lee (2004). "Effects of Physical Embodiment on Social Presence of Social Robots". In: *Proceedings of the 7th Annual International Workshop on Presence*. International Society for Presence Research, pp. 80–87.

Jurafsky, D. and J. H. Martin (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. 2nd ed. Pearson Prentice Hall, DOI: 10.4324/9780203461891-3.

Kahn, J. H., R. M. Tobin, A. E. Massey, and J. A. Anderson (2007). "Measuring Emotional Expression with the Linguistic Inquiry and Word Count". In: *The American Journal of Psychology* 120.2, pp. 263–286. DOI: 10.2307/20445398.

Kanda, T., M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita (2010). "A Communication Robot in a Shopping Mall". In: *IEEE Transactions on Robotics* 26.5, pp. 897–913. DOI: 10.1109/TRO.2010.2062550.

Kardol, T. (2015). "Zora, Een Zelflerende Robot in de Praktijk". In: *Geron* 17.2, pp. 43–45. DOI: 10.1007/s40718-015-0041-3.

Karg, M., A.-A. Samadani, R. Gorbet, K. Kühnlenz, J. Hoey, and D. Kulić (2013). "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation". In: *IEEE Transactions on Affective Computing* 4.4, pp. 341–359. DOI: 10.1109/T-AFFC.2013.29.

Kellermann, K. (2007). "Persuasive Question Asking: How Question Wording Influences Answers". In: *Annual Meeting of the State Bar Association of California*. Social Science Research Network.

Kennedy, J., I. Leite, A. '. Pereira, M. Sun, B. Li, R. Jain, R. Cheng, E. Pincus, E. J. Carter, and J. F. Lehman (2017). "Learning and Reusing Dialog for Repeated Interactions with a Situated Social Agent". In: *International Conference on Intelligent Virtual Agents*. Vol. 10498 LNAI. Springer International Publishing, pp. 192–204. DOI: 10.1007/978-3-319-67401-8_22.

Kennington, C., D. Moro, L. Marchand, J. Carns, and D. McNeill (2020). "rrSDS: Towards a Robot-Ready Spoken Dialogue System". In: *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pp. 132–135.

Kim, J., K. P. Truong, G. Englebienne, and V. Evers (2017). "Learning Spectro-Temporal Features with 3D CNNs for Speech Emotion Recognition". In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 383–388. DOI: 10.1109/ACII.2017.8273628.

Kim, Y., J. Bang, J. Choi, S. Ryu, S. Koo, and G. B. Lee (2014). "Acquisition and Use of Long-Term Memory for Personalized Dialog Systems". In: *International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*. Vol. 8757. Springer International Publishing, pp. 78–87. DOI: 10.1007/978-3-319-15557-9.

Kirakowski, J. and M. Corbett (1993). "SUMI: The Software Usability Measurement Inventory". In: *British Journal of Educational Technology* 24.3, pp. 210–212. DOI: 10.1111/j.1467-8535.1993.tb00076.x.

Kleinsmith, A. and N. Bianchi-Berthouze (2013). "Affective Body Expression Perception and Recognition: A Survey". In: *IEEE Transactions on Affective Computing* 4.1, pp. 15–33. DOI: 10.1109/T-AFFC.2012.16.

Knijnenburg, B. P., A. Kobsa, and H. Jin (2013). "Dimensionality of Information Disclosure Behavior". In: *International Journal of Human-Computer Studies* 71.12, pp. 1144–1162. DOI: 10.1016/j.ijhcs.2013.06.003.

Kobsa, A. (1989). "A Taxonomy of Beliefs and Goals for User Models in Dialog Systems". In: *User Models in Dialog Systems*. Springer, pp. 52–68. DOI: 10.1007/978-3-642-83230-7_3.

Koenderink, S., K. Truong, D. Nazareth, J. van Waterschoot, and R. de Jong (2019). "Twentse toepassingen van TST in de zorg". In: *DIXIT: Taal en Spraaktechnologie en Gezondheidszorg* 16 (December 2019), pp. 12–13.

Kolkmeier, J., M. Bruijnes, and D. Reidsma (2017a). "A Demonstration of the ASAP Realizer-Unity3D Bridge for Virtual and Mixed Reality Applications". In: *Intelligent Virtual Agents*. Vol. 10498. Springer, Cham, pp. 223–226. DOI: 10.1007/978-3-319-67401-8_26.

Kolkmeier, J., M. Lee, and D. Heylen (2017b). "Moral Conflicts in VR: Addressing Grade Disputes with a Virtual Trainer". In: *Intelligent Virtual Agents*. Springer International Publishing, pp. 231–234. DOI: 10.1007/978-3-319-67401-8_28.

Kolomiyets, O. and M. F. Moens (2011). "A Survey on Question Answering Technology from an Information Retrieval Perspective". In: *Information Sciences* 181.24, pp. 5412–5434. DOI: 10.1016/j.ins.2011.07.047.

Kopp, S., L. Gesellensetter, N. C. Krämer, and I. Wachsmuth (2005). "A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application". In: *Intelligent Virtual Agents*. Springer, pp. 329–343. DOI: 10.1007/11550617_28.

Kopp, S., B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. Vilhjálmsson (2006). "Towards a Common Framework for Multimodal Generation: The Behavior Markup Language". In: *Intelligent Virtual Agents*. Springer, pp. 205–217. DOI: 10.1007/11821830_17.

Korte, M. de, J. Kim, and E. Klabbers (2020). "Efficient Neural Speech Synthesis for Low-Resource Languages Through Multilingual Modeling". In: *Interspeech 2020*. ISCA, pp. 2967–2971. DOI: 10.21437/Interspeech.2020-2664.

Krägeloh, C. U., J. Bharatharaj, S. K. Sasthan Kutty, P. R. Nirmala, and L. Huang (2019). "Questionnaires to Measure Acceptability of Social Robots: A Critical Review". In: *Robotics* 8.4 (4), p. 88. DOI: 10.3390/robotics8040088.

Kroczek, L. O. H., M. Pfaller, B. Lange, M. Müller, and A. Mühlberger (2020). "Interpersonal Distance During Real-Time Social Interaction: Insights From Subjective Experience, Behavior, and Physiology". In: *Frontiers in Psychiatry* 11. DOI: 10.3389/fpsyt.2020.00561.

Kucherenko, T., P. Jonell, S. van Waveren, G. E. Henter, S. Alexandersson, I. Leite, and H. Kjellström (2020). "Gesticulator: A Framework for Semantically-Aware Speech-Driven Gesture Generation". In: *Proceedings of the 2020 Inter-*

national Conference on Multimodal Interaction. Association for Computing Machinery, pp. 242–250. DOI: 10.1145/3382507.3418815.

Landauer, T. K., D. S. McNamara, S. Dennis, and W. Kintsch (2013). *Handbook of Latent Semantic Analysis*. Psychology Press. 545 pp.

Langlet, C. and C. Clavel (2016). "Grounding the Detection of the User's Likes and Dislikes on the Topic Structure of Human-Agent Interactions". In: *Knowledge-Based Systems* 106, pp. 116–124. DOI: 10.1016/j.knosys.2016.05.038.

Larsson, S. and D. R. Traum (2000). "Information State and Dialogue Management in the TRINDI Dialogue Move Engine Toolkit". In: *Natural Language Engineering* 6.3-4, pp. 323–340. DOI: 10.1017/S1351324900002539.

Le, Q. A. and C. Pelachaud (2012). "Generating Co-Speech Gestures for the Humanoid Robot NAO through BML". In: *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. Springer, pp. 228–237. DOI: 10.1007/978-3-642-34182-3_21.

Leite, I., G. Castellano, A. '. Pereira, C. Martinho, and A. Paiva (2014). "Empathic Robots for Long-Term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children". In: *International Journal of Social Robotics* 6.3, pp. 329–341. DOI: 10.1007/s12369-014-0227-1.

Leite, I., C. Martinho, and A. Paiva (2013). "Social Robots for Long-Term Interaction: A Survey". In: *International Journal of Social Robotics* 5.2, pp. 291–308. DOI: 10.1007/s12369-013-0178-y.

Lemon, O., L. Cavedon, and B. Kelly (2003). "Managing Dialogue Interaction: A Multi-Layered Approach". In: *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*. Association for Computational Linguistics, pp. 168–177.

Leuski, A. and D. Traum (2010). "NPCEditor: A Tool for Building Question-Answering Characters". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), pp. 2463–2470.

Li, C.-Y., D. Ortega, D. Väth, F. Lux, L. Vanderlyn, M. Schmidt, M. Neumann, M. Völkel, P. Denisov, S. Jenne, Z. Kacarevic, and N. T. Vu (2020). "ADVISER: A Toolkit for Developing Multi-Modal, Multi-Domain and Socially-Engaged Conversational Agents". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 279–286. DOI: 10.18653/v1/2020.acl-demos.31.

Li, S. (2018). *Topic Modeling and Latent Dirichlet Allocation (LDA) in Python*. Medium. URL: https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24 (visited on 11/10/2021).

Lietz, R., M. Harraghy, D. Calderon, J. Brady, E. Becker, and F. Makedon (2019). "Survey of Mood Detection through Various Input Modes". In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. Association for Computing Machinery, pp. 28–31. DOI: 10.1145/3316782.3321543.

Likert, R. (1932). "A Technique for the Measurement of Attitudes". In: *Archives of Psychology* 22 140, pp. 55–55.

Lim, M. Y. (2012). "Memory Models for Intelligent Social Companions". In: *Human-Computer Interaction: The Agency Perspective*. Springer, pp. 241–262. DOI: 10.1007/978-3-642-25691-2_10.

Lim, M. Y., R. Aylett, W. C. Ho, S. Enz, and P. Vargas (2009). "A Socially-Aware Memory for Companion Agents". In: *Intelligent Virtual Agents*. Springer, pp. 20–26. DOI: 10.1007/978-3-642-04380-2_5.

Lison, P. and C. Kennington (2016). "OpenDial: A Toolkit for Developing Spoken Dialogue Systems with Probabilistic Rules". In: *Proceedings of ACL-2016 System Demonstrations*. Association for Computational Linguistics, pp. 67–72. DOI: 10.18653/v1/P16-4012.

Liu, Q., Y. Chen, B. Chen, J.-G. Lou, Z. Chen, B. Zhou, and D. Zhang (2020). "You Impress Me: Dialogue Generation via Mutual Persona Perception". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 1417–1427. DOI: 10.18653/v1/2020.acl-main.131.

Logacheva, V., M. Burtsev, V. Malykh, V. Polulyakh, and A. Seliverstov (2018). "ConvAI Dataset of Topic-Oriented Human-to-Chatbot Dialogues". In: *The NIPS '17 Competition: Building Intelligent Systems*. Springer International Publishing, pp. 47–57. DOI: 10.1007/978-3-319-94042-7_3.

Lubold, N., E. Walker, and H. Pon-Barry (2020). "Effects of Adapting to User Pitch on Rapport Perception, Behavior, and State with a Social Robotic Learning Companion". In: *User Modeling and User-Adapted Interaction*. DOI: 10.1007/s11257-020-09267-3.

Lundell Vinkler, M. and P. Yu (2020). "Conversational Chatbots with Memory-Based Question and Answer Generation". MA thesis. Linköping University. 120 pp.

Maat, M. ter and D. Heylen (2011). "Flipper: An Information State Component for Spoken Dialogue Systems". In: *Intelligent Virtual Agents*. Vol. 1. Springer Berlin Heidelberg, pp. 470–472. DOI: 10.1007/978-3-642-23974-8_67.

Macias-Galindo, D., W. Wong, L. Cavedon, and J. Thangarajah (2012). "Coherent Topic Transition in a Conversational Agent". In: *13th Annual Conference of the International Speech Communication Association*. ISCA, pp. 743–746.

Mancini, M. and C. Pelachaud (2008). "The FML - APML Language". In: *AAMAS '08*. ACM, pp. 1–5.

Mandasari, Y. (2019). "Follow-up Question Generation". MA thesis. Universtiy of Twente. 1-79.

Manning, C., M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky (2014). "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 55–60. DOI: 10.3115/v1/P14-5010.

Marmion, V., D. E. Millard, E. H. Gerding, and S. V. Stevenage (2019). "The Willingness of Crowds: Cohort Disclosure Preferences for Personally Identifying Information". In: *Proceedings of the International AAAI Conference on Web and Social Media* 13, pp. 358–368.

Martinez, B., M. F. Valstar, B. Jiang, and M. Pantic (2019). "Automatic Analysis of Facial Actions: A Survey". In: *IEEE Transactions on Affective Computing* 10.3, pp. 325–347. DOI: 10.1109/TAFFC.2017.2731763.

Martinez-Martin, E. and A. P. del Pobil (2018). "Personal Robot Assistants for Elderly Care: An Overview". In: *Personal Assistants: Emerging Computational Technologies*. First. Springer International Publishing, pp. 77–91. DOI: 10.1007/978-3-319-62530-0_5.

Matsuyama, Y., A. Bhardwaj, R. Zhao, O. J. Romero, S. A. Akoju, and J. Cassell (2016). "Socially-Aware Animated Intelligent Personal Assistant Agent". In: *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. ACL, pp. 224–227.

Mattar, N. and I. Wachsmuth (2012). "Who Are You? On the Acquisition of Information about People for an Agent That Remembers". In: *Proceedings of the 4th International Conference on Agents and Artificial Intelligence (ICAART-2012)*. Vol. 2. SciTePress, pp. 98–105. DOI: 10.5220/0003710900980105.

Mattar, N. and I. Wachsmuth (2013). "Strangers and Friends". In: *Human-Computer Interaction. Towards Intelligent and Implicit Interaction*. Springer, pp. 102–111. DOI: 10.1007/978-3-642-39342-6_12.

Mattar, N. and I. Wachsmuth (2014). "Let's Get Personal". In: *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*. Springer International Publishing, pp. 450–461. DOI: 10.1007/978-3-319-07230-2_43.

Maynard, D. W. (1980). "Placement of Topic Changes in Conversation". In: *Semiotica* 30.3-4, pp. 263–290. DOI: 10.1515/semi.1980.30.3-4.263.

McFee, B., C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto (2015). "Librosa: Audio and Music Signal Analysis in Python". In: *Proceedings of the 14th Python in Science Conference*, pp. 18–24. DOI: 10.25080/Majora-7b98e3ed-003.

McKeown, G., M. Valstar, R. Cowie, M. Pantic, and M. Schroder (2012). "The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored

Conversations between a Person and a Limited Agent". In: *IEEE Transactions on Affective Computing* 3.1, pp. 5–17. DOI: 10.1109/T-AFFC.2011.20.

Mead, R., D. H. Grollman, A. Lim, C. Yeung, A. Stout, and W. B. Knox (2018). "HRI 2018 Workshop: Social Robots in the Wild". In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, pp. 399–400. DOI: 10.1145/3173386.3173569.

Mendelson, M. J. and F. E. Aboud (1999). "Measuring Friendship Quality in Late Adolescents and Young Adults: McGill Friendship Questionnaires." In: *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportoment* 31.2, pp. 130–132. DOI: 10.1037/h0087080.

Meschtscherjakov, A., M. Gärtner, A. Mirnig, C. Rödel, and M. Tscheligi (2016). "The Persuasive Potential Questionnaire (PPQ): Challenges, Drawbacks, and Lessons Learned". In: *Persuasive Technology*. Springer International Publishing, pp. 162–175. DOI: 10.1007/978-3-319-31510-2_14.

Michael, T. (2020). "Retico: An Incremental Framework for Spoken Dialogue Systems". In: *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pp. 49–52.

Ministerie van Volksgezondheid, W. e. S. (2016). *Het Nederlandse zorgstelsel*. Brochure. Rijksoverheid, pp. 1–28.

Montenegro, J. L. Z., C. A. da Costa, and R. da Rosa Righi (2019). "Survey of Conversational Agents in Health". In: *Expert Systems with Applications* 129, pp. 56–67. DOI: 10.1016/j.eswa.2019.03.054.

Morbini, F., D. DeVault, K. Sagae, J. Gerten, A. Nazarian, and D. Traum (2014). "FLoReS: A Forward Looking, Reward Seeking, Dialogue Manager". In: *Natural Interaction with Robots, Knowbots and Smartphones*, pp. 313–325. DOI: 10.1007/978-1-4614-8280-2_28.

Morbini, F., E. Forbell, D. DeVault, K. Sagae, D. R. Traum, and A. A. Rizzo (2012). "A Mixed-Initiative Conversational Dialogue System for Healthcare". In: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pp. 137–139.

Mousa, A. E.-D. and B. Schüller (2016). "Deep Bidirectional Long Short-Term Memory Recurrent Neural Networks for Grapheme-to-Phoneme Conversion Utilizing Complex Many-to-Many Alignments". In: *Interspeech 2016*. ISCA, pp. 2836–2840. DOI: 10.21437/Interspeech.2016-1229.

Nakano, Y. I. and R. Ishii (2010). "Estimating User's Engagement from Eye-Gaze Behaviors in Human-Agent Conversations". In: *Proceedings of the 15th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, pp. 139–148. DOI: 10.1145/1719970.1719990.

Nazareth, D. S., C. Burghardt, A. Capra, P. Cristoforetti, W. Lam, J. B. van Waterschoot, G. J. Westerhof, and K. P. Truong (2019a). “Babbelbord: A Personalized Conversational Game for People with Dementia”. In: *Dementia Lab 2019. Making Design Work: Engaging with Dementia in Context*. Springer International Publishing, pp. 169–173. DOI: 10.1007/978-3-030-33540-3_18.

Nazareth, D. S., M.-P. Jansen, K. P. Truong, G. J. Westerhof, and D. Heylen (2019b). “MEMOA: Introducing the Multi-Modal Emotional Memories of Older Adults Database”. In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 697–703. DOI: 10.1109/ACII.2019.8925462.

Nguyen, T., M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng (2016). “MS MARCO: A Human Generated MAchine Reading COmprehension Dataset”. In: *Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches at NIPS 2016*. NIPS, pp. 96–105. arXiv: 1611.09268.

Nguyen, V.-A., J. Boyd-Graber, P. Resnik, D. A. Cai, J. E. Midberry, and Y. Wang (2014). “Modeling Topic Control to Detect Influence in Conversations Using Nonparametric Topic Models”. In: *Machine Learning* 95.3, pp. 381–421. DOI: 10.1007/s10994-013-5417-9.

Nomura, T., T. Suzuki, T. Kanda, and K. Kato (2006). “Measurement of Negative Attitudes toward Robots”. In: *Interaction Studies* 7.3, pp. 437–454. DOI: 10.1075/is.7.3.14nom.

Noroozi, F., D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera, and G. Anbarjafari (2018). “Survey on Emotional Body Gesture Recognition”. In: *IEEE Transactions on Affective Computing*, pp. 1–1. DOI: 10.1109/TAFFC.2018.2874986.

Norouzi, N., K. Kim, J. Hochreiter, M. Lee, S. Daher, G. Bruder, and G. Welch (2018). “A Systematic Survey of 15 Years of User Studies Published in the Intelligent Virtual Agents Conference”. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 17–22. DOI: 10.1145/3267851.3267901.

Oertel, C., F. Cummins, J. Edlund, P. Wagner, and N. Campbell (2013). “D64: A Corpus of Richly Recorded Conversational Interaction”. In: *Journal on Multimodal User Interfaces* 7.1, pp. 19–28. DOI: 10.1007/s12193-012-0108-6.

Olafsson, S., T. Bickmore, S. Olafsson, T. Bickmore, H. Koiso, Y. Tanaka, R. Watanabe, Y. Den, S. Luz, N. Campbell, F. Haider, K. Hiovain, K. Jokinen, K. E. Haddad, H. Cakmak, T. Dutoit, E. Gilmartin, K. Su, Y. Huang, K. E. Haddad, C. Elias, B. R. Cowan, and N. Campbell (2016). “That Reminds Me...: Towards a Computational Model of Topic Development Within and

Across Conversations”. In: *Proceedings of the Just Talking - Casual Talk among Humans and Machines at LREC 2016*. LREC, pp. 5–8.

Östlund, B., E. Olander, O. Jonsson, and S. Frennert (2015). “STS-Inspired Design to Meet the Challenges of Modern Aging. Welfare Technology as a Tool to Promote User Driven Innovations or Another Way to Keep Older Users Hostage?” In: *Technological Forecasting and Social Change* 93, pp. 82–90. DOI: 10.1016/j.techfore.2014.04.012.

Pak, A. and P. Paroubek (2010). “Twitter as a Corpus for Sentiment Analysis and Opinion Mining”. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*. European Language Resources Association (ELRA), pp. 1320–1326.

Pan, M. K. X. J., E. A. Croft, and G. Niemeyer (2017). “Validation of the Robot Social Attributes Scale (RoSAS) for Human-Robot Interaction through a Human-to-Robot Handover Use Case”. In: *IROS 2017 Workshop Human-Robot Interaction in Collaborative Manufacturing Environments (HRI-CME)*. IEEE, p. 2.

Pandey, A. K. and R. Gelin (2018). “A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind”. In: *IEEE Robotics Automation Magazine* 25.3, pp. 40–48. DOI: 10.1109/MRA.2018.2833157.

Paranjape, A., A. See, K. Kenealy, H. Li, A. Hardy, P. Qi, K. R. Sadagopan, N. M. Phu, D. Soylu, and C. D. Manning (2020). “Neural Generation Meets Real People: Towards Emotionally Engaging Mixed-Initiative Conversations”. In: *Third Proceedings of the Alexa Prize*. Amazon. arXiv: 2008.12348.

Pas, M. E. te, W. G. M. M. Rutten, R. A. Bouwman, and M. P. Buise (2020). “User Experience of a Chatbot Questionnaire Versus a Regular Computer Questionnaire: Prospective Comparative Study”. In: *JMIR Medical Informatics* 8.12, e21982. DOI: 10.2196/21982.

Paul, M. J. (2016). “Interpretable Machine Learning: Lessons from Topic Modeling”. In: *Proceedings Of*. ACM, p. 6.

Picard, R. W. (1998). “Toward Agents That Recognize Emotion”. In: *Actes Proceedings IMAGINA*, pp. 153–165.

Pickard, M. D., C. A. Roster, and Y. Chen (2016). “Revealing Sensitive Information in Personal Interviews: Is Self-Disclosure Easier with Humans or Avatars and under What Conditions?” In: *Computers in Human Behavior* 65, pp. 23–30. DOI: 10.1016/j.chb.2016.08.004.

Pires Kusumota, V. L., R. Vidal Aroca, and F. N. Martins (2018). “An Open Source Framework for Educational Applications Using Cozmo Mobile Robot”. In: *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*, pp. 569–576. DOI: 10.1109/LARS/SBR/WRE.2018.00104.

Poggi, I. (2007). *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. 1st ed. Weidler Buchverlag Berlin. 414 pp.

Poggi, I., C. Pelachaud, F. de Rosis, V. Carofiglio, and B. De Carolis (2005). “Greta. A Believable Embodied Conversational Agent”. In: *Multimodal Intelligent Information Presentation*. Springer Netherlands, pp. 3–25. DOI: 10.1007/1-4020-3051-7_1.

Povey, D., A. Ghoshal, G. Boulianne, L. Burget, O. e. Glembek, N. Goel, M. Hannemann, P. Motlıcˇ ek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely (2011). “The Kaldi Speech Recognition Toolkit”. In: *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society.

Prabhakaran, V., A. Arora, and O. Rambow (2014). “Staying on Topic: An Indicator of Power in Political Debates”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1481–1486.

Prendinger, H., C. Becker, and M. Ishizuka (2006). “A Study in Users’ Physiological Response to an Empathic Interface Agent”. In: *International Journal of Humanoid Robotics* 03.03, pp. 371–391. DOI: 10.1142/S0219843606000801.

Price, P. J. (1990). “Evaluation of Spoken Language Systems: The ATIS Domain”. In: *Proceedings of the Workshop on Speech and Natural Language*. Association for Computational Linguistics, pp. 91–95. DOI: 10.3115/116580.116612.

Pripfl, J., T. Körtner, D. Batko-Klein, D. Hebesberger, M. Weninger, C. Gisinger, S. Frennert, H. Eftring, M. Antona, I. Adami, A. Weiss, M. Bajones, and M. Vincze (2016). “Results of a Real World Trial with a Mobile Social Service Robot for Older Adults”. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 497–498. DOI: 10.1109/HRI.2016.7451824.

Qu, C., L. Yang, M. Qiu, W. B. Croft, Y. Zhang, and M. Iyyer (2019). “BERT with History Answer Embedding for Conversational Question Answering”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, pp. 1133–1136. DOI: 10.1145/3331184.3331341.

Raad, S.-E. (2020). *Zorg voor de toekomst: Over de toekomstbestendigheid van de zorg*. Exploration 02. Sociaal-Economische Raad, p. 240.

Radlinski, F., K. Balog, B. Byrne, and K. Krishnamoorthi (2019). “Coached Conversational Preference Elicitation: A Case Study in Understanding Movie Preferences”. In: *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 353–360. DOI: 10.18653/v1/w19-5941.

Radziwill, N. and M. Benton (2017). "Quality of Chatbots and Intelligent Conversational Agents". In: *Software Quality Professional* 19.3, pp. 25–36.

Rahman, W., M. K. Hasan, S. Lee, A. Bagher Zadeh, C. Mao, L.-P. Morency, and E. Hoque (2020). "Integrating Multimodal Information in Large Pretrained Transformers". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 2359–2369. DOI: `10.18653/v1/2020.acl-main.214`.

Rajpurkar, P., J. Zhang, K. Lopyrev, and P. Liang (2016). "SQuAD: 100,000+ Questions for Machine Comprehension of Text". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 2383–2392. DOI: `10.18653/v1/D16-1264`.

Ramanarayanan, V., D. Suendermann-Oeft, A. V. Ivanov, and K. Evanini (2015). "A Distributed Cloud-Based Dialog System for Conversational Application Development". In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pp. 432–434. DOI: `10.18653/v1/W15-4658`.

Rapp, A. and F. Cena (2016). "Personal Informatics for Everyday Life: How Users without Prior Self-Tracking Experience Engage with Personal Data". In: *International Journal of Human-Computer Studies* 94, pp. 1–17. DOI: `10.1016/j.ijhcs.2016.05.006`.

Rats, M. M. (1994). "Topic-Comment Structures in Information Dialogues". In: *Focus and Natural Language Processing. Proceedings of a Conference in Celebration of the 10th Anniversary of the Journal of Semantics*. Heidelberg, Germany, pp. 591–599.

Rats, M. M. (1995). "Referring to Topics: A Corpus-Based Study". In: *Proceedings of the Ninth Twente Workshop on Language Technology*. Vol. 9. University of Twente, pp. 51–66.

Rats, M. M. (1996). "Topic Management in Information Dialogues". PhD thesis. Tilburg University. 148 pp.

Ravenet, B., C. Pelachaud, C. Clavel, and S. Marsella (2018). "Automating the Production of Communicative Gestures in Embodied Characters". In: *Frontiers in Psychology* 9. DOI: `10.3389/fpsyg.2018.01144`.

Ravichander, A. and A. W. Black (2019). "An Empirical Study of Self-Disclosure in Spoken Dialogue Systems". In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, pp. 253–263. DOI: `10.18653/v1/w18-5030`.

Razavi, S. Z., L. K. Schubert, M. R. Ali, and M. E. Hoque (2017). "Managing Casual Spoken Dialogue Using Flexible Schemas, Pattern Transduction Trees, and Gist Clauses". In: *Proceedings of the Fifth Annual Conference on Advances in Cognitive Systems*, p. 15.

Razavi, S. Z., L. K. Schubert, B. Kane, M. R. Ali, K. A. V. Orden, and T. Ma (2019). "Dialogue Design and Management for Multi-Session Casual Conversation with Older Adults". In: *Joint Proceedings of the ACM IUI 2019 Workshops*. ACM, p. 9.

Řehůřek, R. and P. Sojka (2010). "Software Framework for Topic Modelling with Large Corpora". In: *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. LREC, pp. 46–50.

Reidsma, D., V. Charisi, D. Davison, F. Wijnen, J. van der Meij, V. Evers, D. Cameron, S. Fernando, R. Moore, T. Prescott, D. Mazzei, M. Pieroni, L. Cominelli, R. Garofalo, D. De Rossi, V. Vouloutsi, R. Zucca, K. Grechuta, M. Blancas, and P. Verschure (2016). "The EASEL Project: Towards Educational Human-Robot Symbiotic Interaction". In: *Biomimetic and Biohybrid Systems*. Springer International Publishing, pp. 297–306. DOI: 10.1007/978-3-319-42417-0_27.

Rich, C. and C. L. Sidner (2012). "Using Collaborative Discourse Theory to Partially Automate Dialogue Tree Authoring". In: *Intelligent Virtual Agents*. Springer Berlin Heidelberg, pp. 327–340. DOI: 10.1007/978-3-642-33197-8_34.

Richards, D. and K. Bransky (2014). "ForgetMeNot: What and How Users Expect Intelligent Virtual Agents to Recall and Forget Personal Conversational Content". In: *International Journal of Human-Computer Studies* 72.5, pp. 460–476. DOI: 10.1016/j.ijhcs.2014.01.005.

Ritschel, H., I. Aslan, S. Mertes, A. Seiderer, and E. André (2019). "Personalized Synthesis of Intentional and Emotional Non-Verbal Sounds for Social Robots". In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7. DOI: 10.1109/ACII.2019.8925487.

Ritter, A., C. Cherry, and W. B. Dolan (2011). "Data-Driven Response Generation in Social Media". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 583–593.

Rosenthal-von der Pütten, A. M., N. C. Krämer, L. Hoffmann, S. Sobieraj, and S. C. Eimler (2013). "An Experimental Study on Emotional Reactions Towards a Robot". In: *International Journal of Social Robotics* 5.1, pp. 17–34. DOI: 10.1007/s12369-012-0173-8.

Rosenthal-von der Pütten, A. M., A. Weiss, and S. Šabanović (2016). "The Challenge (Not) to Go Wild! Challenges and Best Practices to Study HRI in Natural Interaction Settings". In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, pp. 583–584.

Rothe, A., B. M. Lake, and T. M. Gureckis (2018). "Do People Ask Good Questions?" In: *Computational Brain & Behavior* 1.1, pp. 69–89. DOI: 10.1007/s42113-018-0005-5.

Rueben, M., S. A. Elprama, D. Chrysostomou, and A. Jacobs (2020). "Introduction to (Re)Using Questionnaires in Human-Robot Interaction Research". In: *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer International Publishing, pp. 125–144. DOI: 10.1007/978-3-030-42307-0_5.

Ruhland, K., C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell (2015). "A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception". In: *Computer Graphics Forum* 34.6, pp. 299–326. DOI: 10.1111/cgf.12603.

Rus, V., P. Piwek, S. Stoyanchev, B. Wyse, M. Lintean, and C. Moldovan (2011). "Question Generation Shared Task and Evaluation Challenge: Status Report". In: *Proceedings of the 13th European Workshop on Natural Language Generation*. Association for Computational Linguistics, pp. 318–320.

Sabanovic, S., M. Michalowski, and R. Simmons (2006). "Robots in the Wild: Observing Human-Robot Social Interaction Outside the Lab". In: *9th IEEE International Workshop on Advanced Motion Control (AMC 2006)*. IEEE, pp. 596–601. DOI: 10.1109/AMC.2006.1631758.

Sánchez-Lozano, E., B. Martinez, and M. F. Valstar (2016). "Cascaded Regression with Sparsified Feature Covariance Matrix for Facial Landmark Detection". In: *Pattern Recognition Letters* 73, pp. 19–25. DOI: 10.1016/j.patrec.2015.11.014.

Santhanam, S., A. Karduni, and S. Shaikh (2020). "Studying the Effects of Cognitive Biases in Evaluation of Conversational Agents". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, pp. 1–13. DOI: 10.1145/3313831.3376318.

Schlangen, D. and G. Skantze (2009). "A General, Abstract Model of Incremental Dialogue Processing". In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 710–718. DOI: 10.5087/dad.2011.105.

Schneider, K. P. (1988). *Small Talk: Analyzing Phatic Discourse*. Hitzeroth. 368 pp.

Schröder, M., E. Bevacqua, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, S. Pammi1, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. Valstar, and M. Wöllmer (2009). "A Demonstration of Audiovisual Sensitive Artificial Listeners". In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–2. DOI: 10.1109/ACII.2009.5349530.

Schröder, M., M. Charfuelan, S. Pammi, and I. Steiner (2011). "Open Source Voice Creation Toolkit for the MARY TTS Platform". In: *INTERSPEECH-2011*. ISCA, pp. 3253–3256.

Schuller, B. W., M. Valstar, E. André, A. E.-D. Mousa, T. Baur, J. Wagner, B. Martinez, S. Jaiswal, T. Almaev, and Y. Zhang (2015). *Deliverable 2.1 (D2.1): Implementation of Cross-Domain, Context-Sensitive Speech Analysis*. Deliverable 2.1. European Commission, p. 18.

Schulman, D. (2013). "Embodied Agents for Long-Term Interaction". PhD thesis. Northeastern University. 346 pp.

Segura, E. M., M. Kriegel, R. Aylett, A. Deshmukh, and H. Cramer (2012). "How Do You Like Me in This: User Embodiment Preferences for Companion Agents". In: *Intelligent Virtual Agents*. Springer, pp. 112–125. DOI: 10.1007/978-3-642-33197-8_12.

Seligman, M. E. P. (2002). "Positive Psychology, Positive Prevention, and Positive Therapy". In: *Handbook of Positive Psychology*. Oxford University Press, pp. 3–9. DOI: 10.1017/CBO9781107415324.004. pmid: 25246403.

Seligman, M. E. P. (2012). *Flourish: A Visionary New Understanding of Happiness and Well-Being*. Simon and Schuster. 370 pp.

Sen, S., M. Lesicko, A. Weiland, R. Gold, Y. Li, B. Hillmann, T. J.-J. Li, and B. Hecht (2014). "WikiBrain: Democratizing Computation on Wikipedia". In: *Proceedings of the 10th International Symposium on Open Collaboration*, F4. DOI: 10.1145/2641580.2641615.

Severinson-Eklundh, K., A. Green, and H. Hüttenrauch (2003). "Social and Collaborative Aspects of Interaction with a Service Robot". In: *Robotics and Autonomous Systems* 42.3, pp. 223–234. DOI: 10.1016/S0921-8890(02)00377-9.

Shechtman, S. and A. Sorin (2019). "Sequence to Sequence Neural Speech Synthesis with Prosody Modification Capabilities". In: *Proc. 10th ISCA Speech Synthesis Workshop*. ISCA, pp. 275–280. DOI: 10.21437/SSW.2019-49.

Short, J., E. Williams, and B. Christie (1976). *The Social Psychology of Telecommunications*. Wiley. 216 pp.

Shuang, Z. W. and D. Burnett (2010). *Speech Synthesis Markup Language (SSML) Version 1.1*. W3C Recommendation. W3C.

Skantze, G. (2007). "Error Handling in Spoken Dialogue Systems: Managing Uncertainty, Grounding and Miscommunication". PhD thesis. Kungliga Tekniska Högskolan. 197 pp. DOI: 10.1016/j.specom.2004.11.003.

Skantze, G. (2021). "Turn-Taking in Conversational Systems and Human-Robot Interaction: A Review". In: *Computer Speech & Language* 67, p. 101178. DOI: 10.1016/j.csl.2020.101178.

Skantze, G. and S. Al Moubayed (2012). "IrisTK: A Statechart-Based Toolkit for Multi-Party Face-to-Face Interaction". In: *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, pp. 69–76. DOI: 10.1145/2388676.2388698.

Skjuve, M. and P. B. Brandzaeg (2019). "Measuring User Experience in Chatbots: An Approach to Interpersonal Communication Competence". In: *Internet Science*. Springer International Publishing, pp. 113–120. DOI: 10.1007/978-3-030-17705-8_10.

Speer, R., J. Chin, and C. Havasi (2017). "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge". In: *AAAI Conference on Artificial Intelligence (AAAI-17)*. AAAI Press, pp. 4444–4451. arXiv: 1612.03975.

Stede, M. and D. Schlangen (2004). "Information-Seeking Chat: Dialogue Management by Topic-Structure". In: *Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*, pp. 117–124.

Strapparava, C. and A. Valitutti (2004). "WordNet Affect: An Affective Extension of WordNet". In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. European Language Resources Association (ELRA).

Strik, H., A. Russel, H. van den Heuvel, C. Cucchiarini, and L. Boves (1997). "A Spoken Dialog System for the Dutch Public Transport Information Service". In: *International Journal of Speech Technology* 2.2, pp. 121–131. DOI: 10.1007/BF02208824.

Su, M.-H., C.-H. Wu, and Y. Chang (2019). "Follow-Up Question Generation Using Neural Tensor Network-Based Domain Ontology Population in an Interview Coaching System". In: *Interspeech 2019*. ISCA, pp. 4185–4189. DOI: 10.21437/Interspeech.2019-1300.

Su, M.-H., C.-H. Wu, K.-Y. Huang, Q.-B. Hong, and H.-H. Huang (2018). "Follow-up Question Generation Using Pattern-Based Seq2seq with a Small Corpus for Interview Coaching". In: *Interspeech 2018*. ISCA, pp. 1006–1010. DOI: 10.21437/Interspeech.2018-1007.

Sugiyama, H., T. Meguro, R. Higashinaka, and Y. Minami (2013). "Open-Domain Utterance Generation for Conversational Dialogue Systems Using Web-Scale Dependency Structures". In: *Proceedings of the SIGDIAL 2013 Conference*. Association for Computational Linguistics, pp. 334–338.

Sugiyama, H., T. Meguro, R. Higashinaka, and Y. Minami (2014a). "Large-Scale Collection and Analysis of Personal Question-Answer Pairs for Conversational Agents". In: *Intelligent Virtual Agents*. Springer International Publishing, pp. 420–433. DOI: 10.1007/978-3-319-09767-1_53.

Sugiyama, H., T. Meguro, R. Higashinaka, and Y. Minami (2014b). "Open-Domain Utterance Generation Using Phrase Pairs Based on Dependency Relations". In: *2014 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, pp. 60–65. DOI: 10.1109/SLT.2014.7078550.

Sun, X., J. Liu, Y. Lyu, W. He, Y. Ma, and S. Wang (2018). "Answer-Focused and Position-Aware Neural Question Generation". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 3930–3939. DOI: 10.18653/v1/D18-1427.

Sung, J., R. E. Grinter, and H. I. Christensen (2010). "Domestic Robot Ecology". In: *International Journal of Social Robotics* 2.4, pp. 417–429. DOI: 10.1007/s12369-010-0065-8.

Svennevig, J. (2000). *Getting Acquainted in Conversation: A Study of Initial Interactions*. John Benjamins Publishing. 398 pp.

Taylor, P. and A. Isard (1997). "SSML: A Speech Synthesis Markup Language". In: *Speech Communication* 21.1, pp. 123–133. DOI: 10.1016/S0167-6393(96)00068-4.

Theune, M., E. Klabbers, J. R. De Pijper, E. Krahmer, and J. Odijk (2001). "From Data to Speech: A General Approach". In: *Natural Language Engineering* 7.1, pp. 47–86.

Theune, M., E. A. M. Klabbers, J. E. J. M. Odijk, and d. J. R. Pijper (1997). "Computing Prosodic Properties in a Data-to-Speech System". In: *Proceedings of the Workshop on Concept to Speech Generation Systems*, pp. 39–45.

Theune, M., D. Wiltenburg, M. Bode, and J. Linssen (2017). "R3D3 in the Wild: Using A Robot for Turn Management in Multi-Party Interaction with a Virtual Human". In: *Proceedings of 1st Interaction with Agents and Robots 2017: Different Embodiments, Common Challenges: Workshop in Conjunction with the 17th International Conference on Intelligent Virtual Agents*.

Thiebaux, M., S. Marsella, A. N. Marshall, and M. Kallmann (2008). "SmartBody: Behavior Realization for Embodied Conversational Agents". In: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 151–158.

Tickle-degnen, L. and R. Rosenthal (1990). "The Nature of Rapport and Its Nonverbal Correlates". In: *Psychological Inquiry* 1.4, pp. 285–293. DOI: 10.1207/s15327965pli0104_1.

Tidwell, L. C. and J. B. Walther (2002). "Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time". In: *Human Communication Research* 28.3, pp. 317–348. DOI: 10.1111/j.1468-2958.2002.tb00811.x.

Touretzky, D. S. and C. Gardner-McCune (2018). "Calypso for Cozmo: Robotic AI for Everyone (Abstract Only)". In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, p. 1110. DOI: 10.1145/3159450.3162200.

Toussaint, W., A. G. Ortega, J. Vroon, J. Harty, G. Solmaz, O. Kudina, E. Peltonen, J. Bourgeois, and A. Y. Ding (2021). *Design Considerations for Data Daemons: Co-Creating Design Futures to Explore Ethical Personal Data Management*. arXiv: 2106.14975.

Traum, D. R. and S. Larsson (2003). "The Information State Approach to Dialogue Management". In: *Current and New Directions in Discourse and Dialogue*. Springer Dordrecht, pp. 325–353. DOI: 10.1007/978-94-010-0019-2_15.

Trinh, H., A. Shamekhi, E. Kimani, and T. W. Bickmore (2018). "Predicting User Engagement in Longitudinal Interventions with Virtual Agents". In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 9–16. DOI: 10.1145/3267851.3267909.

Tsiourti, C., M. Ben Moussa, J. Quintas, B. Loke, I. Jochem, J. A. Lopes, and D. Konstantas (2018). "A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application". In: *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*. Springer International Publishing, pp. 1014–1033. DOI: 10.1007/978-3-319-56994-9_69.

Tulving, E. (1972). "Episodic and Semantic Memory". In: *Organization of Memory*. Academic Press, pp. xiii, 423–xiii, 423.

UHL, H. W. (2016). *Working for Health and Growth: Investing in the Health Workforce*. World Health Organization, p. 74.

Ultes, S., L. M. Rojas-Barahona, P.-H. Su, D. Vandyke, D. Kim, I. Casanueva, P. l. Budzianowski, N. Mrkšić, T.-H. Wen, M. Gašić, and S. Young (2017). "PyDial: A Multi-Domain Statistical Dialogue System Toolkit". In: *Proceedings of ACL 2017, System Demonstrations*. Association for Computational Linguistics, pp. 73–78.

Valstar, M., S. Dermouche, C. Pelachaud, E. Coutinho, B. Schuller, Y. Zhang, D. Heylen, M. Theune, J. van Waterschoot, T. Baur, A. Cafaro, A. Ghitulescu, B. Potard, J. Wagner, E. André, L. Durieu, and M. Aylett (2016). "Ask Alice: An Artificial Retrieval of Information Agent". In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*. ACM Press, pp. 419–420. DOI: 10.1145/2993148.2998535.

Valstar, M., C. Pelachaud, E. André, B. Schuller, D. Heylen, M. Theune, E. Coutinho, S. Dermouche, Y. Zhang, J. van Waterschoot, T. Baur, B. Potard, L. Durieu, and M. Aylett (2018). *Deliverable (D6.4): Evaluation Report*. Deliverable 6.4. European Commission, p. 31.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). "Attention Is All You Need". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., pp. 6000–6010.

Venkatesh, A., C. Khatri, A. Ram, F. Guo, R. Gabriel, A. Nagar, R. Prasad, M. Cheng, B. Hedayatnia, A. Metallinou, R. Goel, S. Yang, and A. Raju (2018). "On Evaluating and Comparing Conversational Agents". In: *Conversational AI Workshop at the 31st Conference on Neural Information Processing Systems (NIPS 2017)*. NIPS, pp. 1–10. arXiv: 1801.03625.

Ventola, E. (1979). "The Structure of Casual Conversation in English". In: *Journal of Pragmatics* 3.3, pp. 267–298. DOI: 10.1016/0378-2166(79)90034-1.

Vilhjálmsson, H., N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf (2007). "The Behavior Markup Language: Recent Developments and Challenges". In: *Intelligent Virtual Agents*. Springer, pp. 99–111. DOI: 10.1007/978-3-540-74997-4_10.

Völske, M., M. Potthast, S. Syed, and B. Stein (2017). "TL;DR: Mining Reddit to Learn Automatic Summarization". In: *Proceedings of the Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, pp. 59–63. DOI: 10.18653/v1/W17-4508.

Vrandečić, D. and M. Krötzsch (2014). "Wikidata". In: *Communications of the ACM* 57.10, pp. 78–85. DOI: 10.1145/2629489.

Vroon, J., C. Zaga, D. Davison, J. Kolkmeier, and J. Linssen (2017). "Snoozle – A Robotic Pillow That Helps You Go to Sleep: HRI 2017 Student Design Competition". In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, pp. 399–400. DOI: 10.1145/3029798.3034949.

Wagner, J., F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André (2013). "The Social Signal Interpretation (SSI) Framework: Multimodal Signal Processing and Recognition in Real-Time". In: *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, pp. 831–834. DOI: 10.1145/2502081.2502223.

Wagner, J., F. Lingenfelser, N. Bee, and E. Andr (2011). "Social Signal Interpretation (SSI): A Framework for Real-Time Sensing of Affective and Social Signals". In: *KI - Künstliche Intelligenz* 25.3, pp. 251–256. DOI: 10.1007/s13218-011-0115-x.

Wahlster, W. and A. Kobsa (1989). "User Models in Dialog Systems". In: *User Models in Dialog Systems*. Springer, pp. 4–34. DOI: 10.1007/978-3-642-83230-7_1.

Walker, M. A., D. J. Litman, C. A. Kamm, and A. Abella (1997). "PARADISE: A Framework for Evaluating Spoken Dialogue Agents". In: *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 271–280. DOI: 10.3115/976909.979652.

Wallace, R. S. (2009). "The Anatomy of A.L.I.C.E." In: *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer Netherlands, pp. 181–210. DOI: 10.1007/978-1-4020-6710-5_13.

Wallström, J. and J. Lindblom (2020). "Design and Development of the USUS Goals Evaluation Framework". In: *Human-Robot Interaction: Evaluation*

Methods and Their Standardization. Springer International Publishing, pp. 177–201. DOI: 10.1007/978-3-030-42307-0_7.

Waltinger, U. and A. Mehler (2009). “Social Semantics and Its Evaluation by Means of Semantic Relatedness and Open Topic Models”. In: *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. Vol. 1, pp. 42–49. DOI: 10.1109/WI-IAT.2009.14.

Wang, Y., C. Liu, M. Huang, and L. Nie (2018). “Learning to Ask Questions in Open-Domain Conversational Systems with Typed Decoders”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pp. 2193–2203. DOI: 10.18653/v1/P18-1204.

Wang, Y., R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. Le, Y. Agiomyrgiannakis, R. Clark, and R. A. Saurous (2017). *Tacotron: Towards End-to-End Speech Synthesis*. arXiv: 1703.10135.

Waterschoot, J. van, M. Bruijnes, J. Flokstra, D. Reidsma, D. Davison, M. Theune, and D. Heylen (2018a). “Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents”. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 43–50. DOI: 10.1145/3267851.3267882.

Waterschoot, J. van, G. Dubuisson Duplessis, L. Gatti, M. Bruijnes, and D. Heylen (2018b). “An Information-Providing Closed-Domain Human-Agent Interaction Corpus”. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), pp. 2746–2752.

Waterschoot, J. van, I. Hendrickx, A. Khan, E. Klabbers, M. de Korte, H. Strik, C. Cucchiarini, and M. Theune (2020a). “BLISS: An Agent for Collecting Spoken Dialogue Data about Health and Well-Being”. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, pp. 449–458.

Watson, D., L. Anna, and A. Tellegen (1988). “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales”. In: *Journal of Personality and Social Psychology* 54.6, pp. 1063–1070.

Weiss, A., R. Bernhaupt, and M. Tscheligi (2011). “The USUS Evaluation Framework for User-Centered HRI”. In: *New Frontiers in Human–Robot Interaction*, pp. 89–110.

Weiss, A. and G. Hannibal (2018). “What Makes People Accept or Reject Companion Robots? A Research Agenda”. In: *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. Association for Computing Machinery, pp. 397–404. DOI: 10.1145/3197768.3203177.

Weiss, R. J., R. J. Skerry-Ryan, E. Battenberg, S. Mariooryad, and D. P. Kingma (2021). *Wave-Tacotron: Spectrogram-Free End-to-End Text-to-Speech Synthesis*. arXiv: 2011.03568.

Weizenbaum, J. (1966). "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine". In: *Communications of the ACM* 9.1, pp. 36–45. DOI: 10.1145/365153.365168.

Welbergen, H. van, R. Yaghoubzadeh, and S. Kopp (2014). "AsapRealizer 2.0: The Next Steps in Fluent Behavior Realization for ECAs". In: *Intelligent Virtual Agents*. Springer, Cham, pp. 449–462. DOI: 10.1007/978-3-319-09767-1_56.

Werner, F. (2020). "A Survey on Current Practices in User Evaluation of Companion Robots". In: *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer International Publishing, pp. 65–88. DOI: 10.1007/978-3-030-42307-0_3.

Wittenburg, P., H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes (2006). "ELAN: A Professional Framework for Multimodality Research". In: *Proceedings of the 5th International Conference on Language Resources and Evaluation*. LREC, pp. 1556–1559.

Wu, J., S. Ghosh, M. Chollet, S. Ly, S. Mozgai, and S. Scherer (2018). "NADiA - Towards Neural Network Driven Virtual Human Conversation Agents". In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 2262–2264.

Wu, P., Z. Ling, L. Liu, Y. Jiang, H. Wu, and L. Dai (2019). "End-to-End Emotional Speech Synthesis Using Style Tokens and Semi-Supervised Training". In: *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 623–627. DOI: 10.1109/APSIPAASC47483.2019.9023186.

Xiao, Z., M. X. Zhou, V. Q. Liao, G. J. Mark, C. Chi, W. Chen, and H. Yang (2020). "Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-Ended Questions". In: *ACM Transactions on Computer-Human Interaction* 27.3, 15:1–15:37. DOI: 10.1145/3381804.

Xing, C., W. Wu, Y. Wu, J. Liu, Y. Huang, M. Zhou, and W.-Y. Ma (2017). "Topic Aware Neural Response Generation". In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI Press, pp. 3351–3357.

Yao, X., E. Tosch, G. Chen, E. Nouri, R. Artstein, A. Leuski, K. Sagae, and D. Traum (2012). "Creating Conversational Characters Using Question Generation Tools". In: *Dialogue & Discourse* 3.2, pp. 125–146.

Yee, J. L. and D. Niemeier (1996). *Advantages and Disadvantages: Longitudinal vs. Repeated Cross-Section Surveys*. Discussion Paper. Federal Highway Administration, pp. 16–22.

Yeh, J. F., Y. S. Tan, and C. H. Lee (2016). "Topic Detection and Tracking for Conversational Content by Using Conceptual Dynamic Latent Dirichlet Allocation". In: *Neurocomputing* 216, pp. 310–318. DOI: 10.1016/j.neucom.2016.08.017.

Zboja, J. J., R. A. Clark, and D. L. Haytko (2016). "An Offer You Can't Refuse: Consumer Perceptions of Sales Pressure". In: *Journal of the Academy of Marketing Science* 44.6, pp. 822–823. DOI: 10.1007/s11747-016-0480-y.

Zhang, S., E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston (2018). "Personalizing Dialogue Agents: I Have a Dog, Do You Have Pets Too?" In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pp. 2204–2213. DOI: 10.18653/v1/P18-1205.

Zhao, R., A. Papangelis, and J. Cassell (2014). "Towards a Dyadic Computational Model of Rapport Management for Human-Virtual Agent Interaction". In: *Intelligent Virtual Agents*. Springer International Publishing, pp. 514–527. DOI: 10.1007/978-3-319-09767-1_62.

Zheng, Y., G. Chen, M. Huang, S. Liu, and X. Zhu (2020). *Personalized Dialogue Generation with Diversified Traits*. arXiv: 1901.09672.

Zhu, Q., Z. Zhang, Y. Fang, X. Li, R. Takanobu, J. Li, B. Peng, J. Gao, X. Zhu, and M. Huang (2020). "ConvLab-2: An Open-Source Toolkit for Building, Evaluating, and Diagnosing Dialogue Systems". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, pp. 142–149. DOI: 10.18653/v1/2020.acl-demos.19.

Zhu, W., A. Chowanda, and M. Valstar (2016). "Topic Switch Models for Dialogue Management in Virtual Humans". In: *International Conference on Intelligent Virtual Agents*. Vol. 10011. Springer, pp. 407–411. DOI: 10.1007/978-3-319-47665-0_43.

SIKS dissertation list

2011 01 Botond Cseke (RUN), Variational Algorithms for Bayesian Inference in Latent Gaussian Models

02 Nick Tinnemeier (UU), Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language

03 Jan Martijn van der Werf (TU/e), Compositional Design and Verification of Component-Based Information Systems

04 Hado van Hasselt (UU), Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference

05 Bas van der Raadt (VU), Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

06 Yiwen Wang (TU/e), Semantically-Enhanced Recommendations in Cultural Heritage

07 Yujia Cao (UT), Multimodal Information Presentation for High Load Human Computer Interaction

08 Nieske Vergunst (UU), BDI-based Generation of Robust Task-Oriented Dialogues

09 Tim de Jong (OU), Contextualised Mobile Media for Learning

10 Bart Bogaert (UvT), Cloud Content Contention

11 Dhaval Vyas (UT), Designing for Awareness: An Experience-focused HCI Perspective

12 Carmen Bratosin (TU/e), Grid Architecture for Distributed Process Mining

13 Xiaoyu Mao (UvT), Airport under Control. Multiagent Scheduling for Airport Ground Handling

14 Milan Lovric (EUR), Behavioral Finance and Agent-Based Artificial Markets

15 Marijn Koolen (UvA), The Meaning of Structure: the Value of Link Evidence for Information Retrieval

16 Maarten Schadd (UM), Selective Search in Games of Different Complexity

17 Jiyin He (UvA), Exploring Topic Structure: Coherence, Diversity and Relatedness

18 Mark Ponsen (UM), Strategic Decision-Making in complex games

19 Ellen Rusman (OU), The Mind's Eye on Personal Profiles

20 Qing Gu (VU), Guiding service-oriented software engineering - A view-based approach

21 Linda Terlouw (TUD), Modularization and Specification of Service-Oriented Systems

22 Junte Zhang (UvA), System Evaluation of Archival Description and Access

23 Wouter Weerkamp (UvA), Finding People and their Utterances in Social Media

24 Herwin van Welbergen (UT), Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior

25 Syed Waqar ul Qounain Jaffry (VU), Analysis and Validation of Models for Trust Dynamics

26 Matthijs Aart Pontier (VU), Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots
27 Aniel Bhulai (VU), Dynamic website optimization through autonomous management of design patterns
28 Rianne Kaptein (UvA), Effective Focused Retrieval by Exploiting Query Context and Document Structure
29 Faisal Kamiran (TU/e), Discrimination-aware Classification
30 Egon van den Broek (UT), Affective Signal Processing (ASP): Unraveling the mystery of emotions
31 Ludo Waltman (EUR), Computational and Game-Theoretic Approaches for Modeling Bounded Rationality
32 Nees-Jan van Eck (EUR), Methodological Advances in Bibliometric Mapping of Science
33 Tom van der Weide (UU), Arguing to Motivate Decisions
34 Paolo Turrini (UU), Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations
35 Maaike Harbers (UU), Explaining Agent Behavior in Virtual Training
36 Erik van der Spek (UU), Experiments in serious game design: a cognitive approach
37 Adriana Burlutiu (RUN), Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference
38 Nyree Lemmens (UM), Bee-inspired Distributed Optimization
39 Joost Westra (UU), Organizing Adaptation using Agents in Serious Games
40 Viktor Clerc (VU), Architectural Knowledge Management in Global Software Development
41 Luan Ibraimi (UT), Cryptographically Enforced Distributed Data Access Control
42 Michal Sindlar (UU), Explaining Behavior through Mental State Attribution
43 Henk van der Schuur (UU), Process Improvement through Software Operation Knowledge
44 Boris Reuderink (UT), Robust Brain-Computer Interfaces
45 Herman Stehouwer (UvT), Statistical Language Models for Alternative Sequence Selection
46 Beibei Hu (TUD), Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
47 Azizi Bin Ab Aziz (VU), Exploring Computational Models for Intelligent Support of Persons with Depression
48 Mark Ter Maat (UT), Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
49 Andreea Niculescu (UT), Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

---

2012 01 Terry Kakeeto (UvT), Relationship Marketing for SMEs in Uganda
02 Muhammad Umair (VU), Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
03 Adam Vanya (VU), Supporting Architecture Evolution by Mining Software Repositories
04 Jurriaan Souer (UU), Development of Content Management System-based Web Applications
05 Marijn Plomp (UU), Maturing Interorganisational Information Systems
06 Wolfgang Reinhardt (OU), Awareness Support for Knowledge Workers in Research Networks
07 Rianne van Lambalgen (VU), When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

08 Gerben de Vries (UvA), Kernel Methods for Vessel Trajectories
09 Ricardo Neisse (UT), Trust and Privacy Management Support for Context-Aware Service Platforms
10 David Smits (TU/e), Towards a Generic Distributed Adaptive Hypermedia Environment
11 J.C.B. Rantham Prabhakara (TU/e), Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
12 Kees van der Sluijs (TU/e), Model Driven Design and Data Integration in Semantic Web Information Systems
13 Suleman Shahid (UvT), Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
14 Evgeny Knutov (TU/e), Generic Adaptation Framework for Unifying Adaptive Web-based Systems
15 Natalie van der Wal (VU), Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.
16 Fiemke Both (VU), Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
17 Amal Elgammal (UvT), Towards a Comprehensive Framework for Business Process Compliance
18 Eltjo Poort (VU), Improving Solution Architecting Practices
19 Helen Schonenberg (TU/e), What's Next? Operational Support for Business Process Execution
20 Ali Bahramisharif (RUN), Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
21 Roberto Cornacchia (TUD), Querying Sparse Matrices for Information Retrieval
22 Thijs Vis (UvT), Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
23 Christian Muehl (UT), Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
24 Laurens van der Werff (UT), Evaluation of Noisy Transcripts for Spoken Document Retrieval
25 Silja Eckartz (UT), Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application
26 Emile de Maat (UvA), Making Sense of Legal Text
27 Hayrettin Gurkok (UT), Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
28 Nancy Pascall (UvT), Engendering Technology Empowering Women
29 Almer Tigelaar (UT), Peer-to-Peer Information Retrieval
30 Alina Pommeranz (TUD), Designing Human-Centered Systems for Reflective Decision Making
31 Emily Bagarukayo (RUN), A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure
32 Wietske Visser (TUD), Qualitative multi-criteria preference representation and reasoning
33 Rory Sie (OUN), Coalitions in Cooperation Networks (COCOON)
34 Pavol Jancura (RUN), Evolutionary analysis in PPI networks and applications
35 Evert Haasdijk (VU), Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics
36 Denis Ssebugwawo (RUN), Analysis and Evaluation of Collaborative Modeling Processes
37 Agnes Nakakawa (RUN), A Collaboration Process for Enterprise Architecture Creation
38 Selmar Smit (VU), Parameter Tuning and Scientific Testing in Evolutionary Algorithms

39 Hassan Fatemi (UT), Risk-aware design of value and coordination networks
40 Agus Gunawan (UvT), Information Access for SMEs in Indonesia
41 Sebastian Kelle (OU), Game Design Patterns for Learning
42 Dominique Verpoorten (OU), Reflection Amplifiers in self-regulated Learning
43 Withdrawn
44 Anna Tordai (VU), On Combining Alignment Techniques
45 Benedikt Kratz (UvT), A Model and Language for Business-aware Transactions
46 Simon Carter (UvA), Exploration and Exploitation of Multilingual Data for Statistical Machine Translation
47 Manos Tsagkias (UvA), Mining Social Media: Tracking Content and Predicting Behavior
48 Jorn Bakker (TU/e), Handling Abrupt Changes in Evolving Time-series Data
49 Michael Kaisers (UM), Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions
50 Steven van Kervel (TUD), Ontologogy driven Enterprise Information Systems Engineering
51 Jeroen de Jong (TUD), Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

---

2013 01 Viorel Milea (EUR), News Analytics for Financial Decision Support
02 Erietta Liarou (CWI), MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing
03 Szymon Klarman (VU), Reasoning with Contexts in Description Logics
04 Chetan Yadati (TUD), Coordinating autonomous planning and scheduling
05 Dulce Pumareja (UT), Groupware Requirements Evolutions Patterns
06 Romulo Goncalves (CWI), The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience
07 Giel van Lankveld (UvT), Quantifying Individual Player Differences
08 Robbert-Jan Merk (VU), Making enemies: cognitive modeling for opponent agents in fighter pilot simulators
09 Fabio Gori (RUN), Metagenomic Data Analysis: Computational Methods and Applications
10 Jeewanie Jayasinghe Arachchige (UvT), A Unified Modeling Framework for Service Design.
11 Evangelos Pournaras (TUD), Multi-level Reconfigurable Self-organization in Overlay Services
12 Marian Razavian (VU), Knowledge-driven Migration to Services
13 Mohammad Safiri (UT), Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
14 Jafar Tanha (UvA), Ensemble Approaches to Semi-Supervised Learning
15 Daniel Hennes (UM), Multiagent Learning - Dynamic Games and Applications
16 Eric Kok (UU), Exploring the practical benefits of argumentation in multi-agent deliberation
17 Koen Kok (VU), The PowerMatcher: Smart Coordination for the Smart Electricity Grid
18 Jeroen Janssens (UvT), Outlier Selection and One-Class Classification
19 Renze Steenhuizen (TUD), Coordinated Multi-Agent Planning and Scheduling
20 Katja Hofmann (UvA), Fast and Reliable Online Learning to Rank for Information Retrieval
21 Sander Wubben (UvT), Text-to-text generation by monolingual machine translation
22 Tom Claassen (RUN), Causal Discovery and Logic
23 Patricio de Alencar Silva (UvT), Value Activity Monitoring

24 Haitham Bou Ammar (UM), Automated Transfer in Reinforcement Learning
25 Agnieszka Anna Latoszek-Berendsen (UM), Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
26 Alireza Zarghami (UT), Architectural Support for Dynamic Homecare Service Provisioning
27 Mohammad Huq (UT), Inference-based Framework Managing Data Provenance
28 Frans van der Sluis (UT), When Complexity becomes Interesting: An Inquiry into the Information eXperience
29 Iwan de Kok (UT), Listening Heads
30 Joyce Nakatumba (TU/e), Resource-Aware Business Process Management: Analysis and Support
31 Dinh Khoa Nguyen (UvT), Blueprint Model and Language for Engineering Cloud Applications
32 Kamakshi Rajagopal (OUN), Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development
33 Qi Gao (TUD), User Modeling and Personalization in the Microblogging Sphere
34 Kien Tjin-Kam-Jet (UT), Distributed Deep Web Search
35 Abdallah El Ali (UvA), Minimal Mobile Human Computer Interaction
36 Than Lam Hoang (TU/e), Pattern Mining in Data Streams
37 Dirk Börner (OUN), Ambient Learning Displays
38 Eelco den Heijer (VU), Autonomous Evolutionary Art
39 Joop de Jong (TUD), A Method for Enterprise Ontology based Design of Enterprise Information Systems
40 Pim Nijssen (UM), Monte-Carlo Tree Search for Multi-Player Games
41 Jochem Liem (UvA), Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
42 Léon Planken (TUD), Algorithms for Simple Temporal Reasoning
43 Marc Bron (UvA), Exploration and Contextualization through Interaction and Concepts

---

2014 01 Nicola Barile (UU), Studies in Learning Monotone Models from Data
02 Fiona Tuliyano (RUN), Combining System Dynamics with a Domain Modeling Method
03 Sergio Raul Duarte Torres (UT), Information Retrieval for Children: Search Behavior and Solutions
04 Hanna Jochmann-Mannak (UT), Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation
05 Jurriaan van Reijsen (UU), Knowledge Perspectives on Advancing Dynamic Capability
06 Damian Tamburri (VU), Supporting Networked Software Development
07 Arya Adriansyah (TU/e), Aligning Observed and Modeled Behavior
08 Samur Araujo (TUD), Data Integration over Distributed and Heterogeneous Data Endpoints
09 Philip Jackson (UvT), Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language
10 Ivan Salvador Razo Zapata (VU), Service Value Networks
11 Janneke van der Zwaan (TUD), An Empathic Virtual Buddy for Social Support
12 Willem van Willigen (VU), Look Ma, No Hands: Aspects of Autonomous Vehicle Control
13 Arlette van Wissen (VU), Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
14 Yangyang Shi (TUD), Language Models With Meta-information

15 Natalya Mogles (VU), Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
16 Krystyna Milian (VU), Supporting trial recruitment and design by automatically interpreting eligibility criteria
17 Kathrin Dentler (VU), Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
18 Mattijs Ghijsen (UvA), Methods and Models for the Design and Study of Dynamic Agent Organizations
19 Vinicius Ramos (TU/e), Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
20 Mena Habib (UT), Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
21 Kassidy Clark (TUD), Negotiation and Monitoring in Open Environments
22 Marieke Peeters (UU), Personalized Educational Games - Developing agent-supported scenario-based training
23 Eleftherios Sidirourgos (UvA/CWI), Space Efficient Indexes for the Big Data Era
24 Davide Ceolin (VU), Trusting Semi-structured Web Data
25 Martijn Lappenschaar (RUN), New network models for the analysis of disease interaction
26 Tim Baarslag (TUD), What to Bid and When to Stop
27 Rui Jorge Almeida (EUR), Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
28 Anna Chmielowiec (VU), Decentralized k-Clique Matching
29 Jaap Kabbedijk (UU), Variability in Multi-Tenant Enterprise Software
30 Peter de Cock (UvT), Anticipating Criminal Behaviour
31 Leo van Moergestel (UU), Agent Technology in Agile Multiparallel Manufacturing and Product Support
32 Naser Ayat (UvA), On Entity Resolution in Probabilistic Data
33 Tesfa Tegegne (RUN), Service Discovery in eHealth
34 Christina Manteli (VU), The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.
35 Joost van Ooijen (UU), Cognitive Agents in Virtual Worlds: A Middleware Design Approach
36 Joos Buijs (TU/e), Flexible Evolutionary Algorithms for Mining Structured Process Models
37 Maral Dadvar (UT), Experts and Machines United Against Cyberbullying
38 Danny Plass-Oude Bos (UT), Making brain-computer interfaces better: improving usability through post-processing.
39 Jasmina Maric (UvT), Web Communities, Immigration, and Social Capital
40 Walter Omona (RUN), A Framework for Knowledge Management Using ICT in Higher Education
41 Frederic Hogenboom (EUR), Automated Detection of Financial Events in News Text
42 Carsten Eijckhof (CWI/TUD), Contextual Multidimensional Relevance Models
43 Kevin Vlaanderen (UU), Supporting Process Improvement using Method Increments
44 Paulien Meesters (UvT), Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.
45 Birgit Schmitz (OUN), Mobile Games for Learning: A Pattern-Based Approach
46 Ke Tao (TUD), Social Web Data Analytics: Relevance, Redundancy, Diversity
47 Shangsong Liang (UvA), Fusion and Diversification in Information Retrieval

---

2015 01 Niels Netten (UvA), Machine Learning for Relevance of Information in Crisis Response

02 Faiza Bukhsh (UvT), Smart auditing: Innovative Compliance Checking in Customs Controls
03 Twan van Laarhoven (RUN), Machine learning for network data
04 Howard Spoelstra (OUN), Collaborations in Open Learning Environments
05 Christoph Bösch (UT), Cryptographically Enforced Search Pattern Hiding
06 Farideh Heidari (TUD), Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes
07 Maria-Hendrike Peetz (UvA), Time-Aware Online Reputation Analysis
08 Jie Jiang (TUD), Organizational Compliance: An agent-based model for designing and evaluating organizational interactions
09 Randy Klaassen (UT), HCI Perspectives on Behavior Change Support Systems
10 Henry Hermans (OUN), OpenU: design of an integrated system to support lifelong learning
11 Yongming Luo (TU/e), Designing algorithms for big graph datasets: A study of computing bisimulation and joins
12 Julie M. Birkholz (VU), Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks
13 Giuseppe Procaccianti (VU), Energy-Efficient Software
14 Bart van Straalen (UT), A cognitive approach to modeling bad news conversations
15 Klaas Andries de Graaf (VU), Ontology-based Software Architecture Documentation
16 Changyun Wei (UT), Cognitive Coordination for Cooperative Multi-Robot Teamwork
17 André van Cleeff (UT), Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
18 Holger Pirk (CWI), Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
19 Bernardo Tabuenca (OUN), Ubiquitous Technology for Lifelong Learners
20 Lois Vanhée (UU), Using Culture and Values to Support Flexible Coordination
21 Sibren Fetter (OUN), Using Peer-Support to Expand and Stabilize Online Learning
22 Zhemin Zhu (UT), Co-occurrence Rate Networks
23 Luit Gazendam (VU), Cataloguer Support in Cultural Heritage
24 Richard Berendsen (UvA), Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
25 Steven Woudenberg (UU), Bayesian Tools for Early Disease Detection
26 Alexander Hogenboom (EUR), Sentiment Analysis of Text Guided by Semantics and Structure
27 Sándor Héman (CWI), Updating compressed column stores
28 Janet Bagorogoza (TiU), Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO
29 Hendrik Baier (UM), Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains
30 Kiavash Bahreini (OU), Real-time Multimodal Emotion Recognition in E-Learning
31 Yakup Koç (TUD), On the robustness of Power Grids
32 Jerome Gard (UL), Corporate Venture Management in SMEs
33 Frederik Schadd (TUD), Ontology Mapping with Auxiliary Resources
34 Victor de Graaf (UT), Gesocial Recommender Systems
35 Jungxao Xu (TUD), Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction

---

2016 01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow

03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VU), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UvA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VU), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VU), Trusting Crowdsourced Information on Cultural Artefacts
10 George Karafotias (VU), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UvA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VU), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UvA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VU), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VU), Refining Statistical Data on the Web
19 Julia Efremova (TU/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UvA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VU), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UvA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakarathne (VU), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (UvT), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UvA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TU/e), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UvA), Monitoring Social Media: Summarization, Classification and Recommendation

36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry
38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UvA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora
43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UvA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis
50 Yan Wang (UvT), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

---

2017 01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UvA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VU), Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TU/e), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsky (UvT), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction
16 Aleksandr Chuklin (UvA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UvA), Entity Associations for Search

19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
22 Sara Magliacane (VU), Logics for causal inference under uncertainty
23 David Graus (UvA), Entities of Interest — Discovery in Digital Traces
24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning
25 Veruska Zamborlini (VU), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch
27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
28 John Klein (VU), Architecture Practices for Complex Contexts
29 Adel Alhuraibi (UvT), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
30 Wilma Latuny (UvT), The Power of Facial Expressions
31 Ben Ruijl (UL), Advances in computational methods for QFT calculations
32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives
33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics
35 Martine de Vos (VU), Interpreting natural science spreadsheets
36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging
37 Alejandro Montes Garcia (TU/e), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
38 Alex Kayal (TUD), Normative Social Applications
39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
40 Altaf Hussain Abro (VU), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
41 Adnan Manzoor (VU), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets
43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval
44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement
46 Jan Schneider (OU), Sensor-based Learning Support
47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration
48 Angel Suarez (OU), Collaborative inquiry-based learning

2018 01 Han van der Aa (VU), Comparing and Aligning Process Representations
02 Felix Mannhardt (TU/e), Multi-perspective Process Mining
03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction

04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
05 Hugo Huurdeman (UvA), Supporting the Complex Dynamics of the Information Seeking Process
06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems
07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems
08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems
09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations
10 Julienka Mollee (VU), Moving forward: supporting physical activity behavior change through intelligent technology
11 Mahdi Sargolzaei (UvA), Enabling Framework for Service-oriented Collaborative Networks
12 Xixi Lu (TU/e), Using behavioral context in process mining
13 Seyed Amin Tabatabaei (VU), Computing a Sustainable Future
14 Bart Joosten (UvT), Detecting Social Signals with Spatiotemporal Gabor Filters
15 Naser Davarzani (UM), Biomarker discovery in heart failure
16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17 Jianpeng Zhang (TU/e), On Graph Sample Clustering
18 Henriette Nakad (UL), De Notaris en Private Rechtspraak
19 Minh Duc Pham (VU), Emergent relational schemas for RDF
20 Manxia Liu (RUN), Time and Bayesian Networks
21 Aad Slootmaker (OUN), EMERGO: a generic platform for authoring and playing scenario-based serious games
22 Eric Fernandes de Mello Araujo (VU), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25 Riste Gligorov (VU), Serious Games in Audio-Visual Collections
26 Roelof Anne Jelle de Vries (UT), Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
27 Maikel Leemans (TU/e), Hierarchical Process Mining for Scalable Software Analysis
28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel
29 Yu Gu (UvT), Emotion Recognition from Mandarin Speech
30 Wouter Beek, The “K” in “semantic web” stands for “knowledge”: scaling semantics to the web

---

2019 01 Rob van Eijk (UL), Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TU/e), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TU/e), Process Mining with Streaming Data
06 Chris Dijkshoorn (VU), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes

09 Fahimeh Alizadeh Moghaddam (UvA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs
12 Jacqueline Heinerman (VU), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TU/e), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN), Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TU/e), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VU), Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VU), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VU), Pragmatic factors in (automatic) image description
26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OUN), The Gamification Design Process applied to (Massive) Open Online Courses
28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VU), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VU), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TU/e), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TU/e), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OUN), Gamification with digital badges in learning programming
36 Kevin Ackermans (OUN), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OUN), Learning visually grounded and multilingual representations

---

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding

04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TU/e), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OUN), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL), Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TU/e), In-database Preprocessing for Process Mining
11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
12 Ward van Breda (VU), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OUN), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OUN), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VU), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VU), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nobrega Medeiros (VU/RUN), How are you feeling, human? Towards emotionally supportive chatbots
25 Xin Du (TU/e), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality
29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference
30 Bob Zadok Blok (UL), Creatief, Creatieve, Creatiefst
31 Gongjin Lan (VU), Learning better – From Baby to Better
32 Jason Rhuggenaath (TU/e), Revenue management in online markets: pricing and online advertising
33 Rick Gilsing (TU/e), Supporting service-dominant business model evaluation in the context of business model innovation
34 Anna Bon (MU), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

2021 01 Francisco Xavier Dos Santos Fonseca (TUD), Location-based Games for Social Interaction in Public Space
02 Rijk Mercuur (TUD), Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models
03 Seyyed Hadi Hashemi (UvA), Modeling Users Interacting with Smart Devices
04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning
05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems
06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot
07 Armel Lefebvre (UU), Research data management for open science
08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking
09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play
10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning
11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision
12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs
13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning
14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support
15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm
16 Esam A. H. Ghaleb (UM), BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES
17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks
18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks
19 Roberto Verdecchia (VU), Architectural Technical Debt: Identification and Management
20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems
21 Pedro Thiago Timbó Holanda (CWI), Progressive Indexes
22 Sihang Qiu (TUD), Conversational Crowdsourcing
23 Hugo Manuel Proença (LIACS), Robust rules for prediction and description
24 Kaijie Zhu (TU/e), On Efficient Temporal Subgraph Query Processing
25 Eoin Martino Grua (VU), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications
26 Benno Kruit (CWI & VU), Reading the Grid: Extending Knowledge Bases from Human-readable Tables

Alphabetical Index

ISBN 978-90-365-5269-1