

ulm university universität **UUIM**

Communication Style Modelling and Adaptation in Spoken Dialogue Systems

DISSERTATION

zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

(DR.-ING.)

der Fakultät für Ingenieurwissenschaften, Informatik und Psychologie der Universität Ulm

von

Juliana Miehle

aus Biberach an der Riß

Gutachter:

Prof. Dr. Dr.-Ing. Wolfgang Minker Prof. Kristiina Jokinen

Amtierender Dekan:

Prof. Dr.-Ing. Maurits Ortmanns

Ulm, 03.12.2021

Abstract

When communicating, people use strategies while choosing the words and the non-verbal signs in order to achieve their purpose. Hence, they do not only focus on *what* they say, but also on *how* they formulate it. The aim of this thesis is to examine the role of communication styles in human-computer interaction. This is approached from two angles: it is investigated how varying communication styles are perceived by the user and how communication styles can be integrated into spoken dialogue systems. In order to answer the first question, system requirements are analysed in a series of complex prototypes and various user evaluations are conducted to examine different user groups in diverse scenarios. The second aspect is addressed through the implementation of two new components for spoken dialogue systems.

There are numerous parameters that influence an interaction between two participants and the appropriate or preferred communication style, like the speakers' roles, their cultures, their personalities or the aim of the interaction. In order to enable adaptation through communication styles, these different aspects need to be set into relation. Therefore, the Communication Style Perception Model is presented within the scope of this thesis. It covers elements that are relevant for the selection of the system communication style as well as aspects that are influenced by the system communication style. It is based on numerous user evaluations, examining various user groups in different scenarios. Three experiments investigate the influence of numerous variables on the user's preference in the system communication style. The results show that both user traits and system properties influence the user's communication style preferences in human-computer interaction. Further experiments investigate how varying system communication styles affect the users, if they are selected according to the users' personal preferences. To examine this, different communication styles are included in various systems and applications. The results show that the system's communication style influences the user's satisfaction and the user's perception of the dialogue. For specific applications like behaviour change support systems, the communication style even has an impact on the user's behaviour. Furthermore, the results show that there is no general preference in the system's communication style. The preference appears to be individual for every person and the system needs to adapt its communication style to each user individually during every dialogue.

The second question of how communication styles can be integrated into spoken dialogue systems is addressed by the extension of the standard architecture of spoken dialogue systems. Two new components are proposed, implemented and evaluated: a communication style classifier that automatically identifies the user communication style and a communication style selection module that selects an appropriate system communication style. Both tasks are formulated as classification problems. Due to the novelty of the underlying machine learning task, a multi-lingual corpus is created, containing 258 dialogues with

annotations for the *elaborateness* and *indirectness* for each of the 7,930 dialogue acts. For the user communication style recognition, three different classifiers are compared on the task: a support vector machine classifier, a multi-layer perceptron classifier, and a custom recurrent neural network classifier. Furthermore, different feature sets are tested as input for the classifiers. All features that are used for the communication style classification can be automatically recognised in spoken dialogue systems during an ongoing interaction, without any prior annotation. The results show that for the *elaborateness*, analysing the utterance length dependent on the dialogue act contains enough information to achieve good classification performance. The *indirectness* seems to be a more difficult classification task and additional linguistic features in form of word embeddings give improvement in the classification results. Furthermore, temporal information is beneficial in this case. For the system communication style selection, a multi-layer perceptron classifier is trained and evaluated, using features that encode what the system wants to say in the current turn, what the user wants from the system and how the user talks to the system. As for the first task, the features can be automatically recognised in spoken dialogue systems. The results outperform both a majority-class classifier and a baseline which is mimicking the last user communication style for each of the four languages. When combining both components, the spoken dialogue system is enabled to recognise the user's communication style and select an appropriate communication style for the system.

Acknowledgements

This dissertation would not have been possible without the help of many people. First and foremost, I would like to express my highest gratitude to my supervisor Prof. Wolfgang Minker. His constant support and guidance have been crucial for the successful realisation of my dissertation. He has always been encouraging and offered me many opportunities.

I would like to extend my gratitude to my co-supervisor Dr.-Ing. Stefan Ultes. He always provided me with scientific advice and direction. Without his support and encouragement, I would not have started nor completed this doctorate.

Moreover, I am particularly grateful to my reviewer Prof. Kristiina Jokinen (National Institute of Advanced Industrial Science and Technology, Tokyo) for her interest in my research work. She provided valuable advice during my visits to Japan.

I would like to thank Prof. Satoshi Nakamura and Prof. Keiichi Yasumoto for the opportunity of research stays at the Nara Institute of Science and Technology (NAIST) in Japan. Many thanks also go to Prof. Koichiro Yoshino, Prof. Yuki Matsuda, Shogo Kawanaka and Zhihua Zhang for their collaboration.

Furthermore, I want to express my gratitude to my colleagues and former colleagues of the Dialogue Systems Group at the Institute of Communications Engineering at Ulm University, namely Matthias Kraus, Isabel Feustel, Annalena Aicher, Sabine Wieluch, Florian Nothdurft, Dmitrii Fedotov, Denis Dresvyanskiy, Danila Mamontov, Oleg Akhtiamov, Maxim Sidorov, Roman Sergienko and Anastasiia Spirina. A special thanks goes to Louisa Pragst for always helping me with programming issues, and to Nicolas Wagner for his support regarding SQL. Thank you for the conversations and activities that went, from time to time, beyond work, the good working environment, and the supportive atmosphere. This has been very important for me for successfully completing my doctorate. I would also like to thank our secretaries and technical staff Michaela Baumann, Fe Hägele, Heike Schewe, Werner Birkle and Werner Hack for always offering support, no matter which problem I had to solve. My thanks are extended to my former Bachelor and Master students Ilker Bagci, Nadine Gerstenlauer, Stefanie Walz, Florian Voß, Emre Inanc, Jonathan Eisenbeiss, Julia Hornauer, Patrick Lutz, Philipp Dörzenbach and Pascal Weber.

On a personal level, a very special thanks goes to my long-term office colleague Niklas Rach for the daily coffee breaks and the myriad of discussions about work, life, and everything else. He has always helped me to regain focus and keep going.

Finally, but most importantly, I would like to thank my family and my friends. My parents, Dietlinde Dom-Miehle and Peter Miehle, have always believed in me and provided constant support to pursue higher education. Similarly, I have always been encouraged by my sister Christina Miehle and my former flatmate Mona Rodewald. Many thanks for the countless and inspiring discussions about life and future plans.

Contents

List of Figures xi			xi
Lis	t of [.]	Tables	xv
1	Intro 1.1 1.2 1.3	Oduction Motivation	1 2 3 4
2	Rele 2.1 2.2 2.3 2.4 2.5 2.6	vant BackgroundSpoken Dialogue Systems2.1.1General Architecture of Spoken Dialogue Systems2.1.2Adaptation of Spoken Dialogue SystemsCommunication StylesCommunication Styles1Interactive Adaptation in Human Communication2.3.1Interactive Adaptation in Human-Human Interaction2.3.2Interactive Adaptation in Human-Computer InteractionStatistical MetricsSummary	7 7 8 10 13 14 15 17 20 22
3	Rela 3.1 3.2 3.3	ted WorkAdaptation of Communication Styles in Human-Computer Interaction3.1.1Development of Computer Personalities3.1.2Style Variation3.1.3Style AdaptationRecognition of Elaborateness and IndirectnessSummary	 23 23 23 25 26 27 28
4	Real 4.1	isation of Use Case Scenarios and Prototypes A Social Companion and Conversational Partner for the Elderly 4.1.1 Requirement Analysis 4.1.2 Implementation and Evaluation A Culturally Adaptive Information Agent for Immigrants	31 32 32 34
	4.3	 4.2.1 Challenges for Dialogue Management	35 35 37 37

Contents

		4.3.2 Implementation and Evaluation	8
	4.4	An Intelligent Digital Assistant for Clinical Operating Rooms	0
		4.4.1 Functionalities and Challenges	0
		4.4.2 Implementation and Expert Evaluation	1
	4.5	Summary	3
5	The	Communication Style Perception Model 4	5
	5.1	Cultural Communication Idiosyncrasies in Human-Computer Interaction . 4	5
		5.1.1 Integrating Cultural Communication Idiosyncrasies	6
		5.1.2 Experimental Setting	8
		5.1.3 Evaluation Results 5	0
		5.1.4 Conclusion	2
	5.2	Influences on the User's Communication Style Preferences	4
		5.2.1 Experimental Setting	4
		5.2.2 Evaluation Results 5	7
		5.2.3 Conclusion	5
	5.3	The Impact of the System's Role on Different Presentation Styles for In-	
		conclusive Situations	6
		5.3.1 Experimental Setting	7
		5.3.2 Evaluation Results	0
		5.3.3 Conclusion	3
	5.4	The Impact of Varying Communication Styles on User Satisfaction 7	4
		5.4.1 Experimental Setting	4
		5.4.2 Evaluation Results	9
		5.4.3 Conclusion	5
	5.5	The Impact of Varying Communication Styles in Gamified Participatory	
		Sensing	6
		5.5.1 Design of the Task Allocation Interfaces	6
		5.5.2 Experimental Setting	8
		5.5.3 Evaluation Results $\dots \dots \dots$	1
		5.5.4 Conclusion $\dots \dots \dots$	5
	5.6	The Impact of Varying Communication Styles in a Behaviour Change Sup-	
		port System	5
		5.6.1 Walkeeper Prototype Design	5
		5.6.2 Experimental Setting	8
		5.6.3 Evaluation Results	9
		5.6.4 Conclusion $\ldots \ldots \ldots$	2
	5.7	Summary	3
6	Stat	istical Classification Approaches to Communication Style Adaptation 10	5
	6.1	Corpus Creation	5
	6.2	Investigating the Correlation between User and System Communication Style11	2
	6.3	User Communication Style Recognition	5
		6.3.1 The Dialogue Act Features	7

Contents

	6.3.2 The Contribution of Grammatical and Linguistic Features	119
6.4	Automatic System Communication Style Selection	126
6.5	Summary	130
~		
Con	clusions and Future Directions	133
7.1	Contributions	135
	7.1.1 Theoretical \ldots	136
	7.1.2 Practical	136
	7.1.3 Experimental	137
7.2	Future Directions	137
	6.46.5Con7.17.2	6.3.2 The Contribution of Grammatical and Linguistic Features 6.4 Automatic System Communication Style Selection 6.5 Summary 6.6 Summary 7.1 Conclusions and Future Directions 7.1 Contributions 7.1.1 Theoretical 7.1.2 Practical 7.1.3 Experimental 7.2 Future Directions

List of Figures

2.1	The architecture of a spoken dialogue system consisting of speech recogni- tion, linguistic analysis, dialogue management, text generation and speech synthesis. Based on (McTear, 2004; Jokinen and McTear, 2009)	8
2.2	An example for classification in two classes using a support vector machine, showing the maximum-margin hyperplane and the support vectors (red). Based on (Cortes and Vapnik, 1995).	18
2.3	An example for a multi-layer perceptron with an input layer consisting of three nodes, one hidden layer consisting of four nodes and an output layer consisting of two nodes.	19
2.4	An example for a recurrent neural network with an input layer consisting of three nodes, one hidden layer consisting of four nodes and an output layer consisting of two nodes.	20
3.1	The standard architecture of spoken dialogue systems is extended by two components: 1) a communication style classifier that automatically iden- tifies the user communication style and 2) a communication style se- lection module that selects an appropriate system communication style. Taken from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)	29
4.1	The questionnaire contained 11 statements which had to be rated on a five-point Likert scale (1 = fully agree, 5 = fully disagree). Overall, the ratings show that the NAO robot and the dialogue were perceived very positively. Taken from (Miehle et al., 2019a, Figure 1). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Advanced Social Interaction with Agents, Lecture Notes in Electrical Engineering, Vol. 510 by Maxine Eskenazi, Laurence Devillers, Joseph Mariani (eds) ©Springer International Publishing AG, part of Springer Nature (2019).	33
4.2	Probability distribution of each parameter before training (blue) and example probability distributions of two parameters after training, representing a frequently occurring system action (green) and a rarely occurring system	20
4.3	Excerpt of a script, showing the first and the second path element. Taken from (Miehle et al., 2021a, Figure 2)	30 39

List of Figures

5.1	The Communication Style Perception Model showing 1) elements that are relevant for the selection of the system's communication style and 2) aspects that are influenced by the system communication style	46
5.2	In each dialogue turn, the participants had to watch different videos and decide which one they prefer. Taken from (Miehle et al., 2016, Figure 1), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/ 4.0).	49
5.3	Results of the user study, showing differences between German and Japanese participants. Taken from (Miehle et al., 2016, Figure 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	51
5.4	Results of the user study, showing differences between male and female participants. Taken from (Miehle et al., 2016, Figure 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)	53
5.5	The overall evaluation results, averaging across all twelve utterances and all ten user groups. Adapted from (Miehle et al., 2018c, Figure 1, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/	57
5.6	Influence of the system's role. Adapted from (Miehle et al., 2018c, Figure 2, Figure 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	59
5.7	Influence of the user's culture. Adapted from (Miehle et al., 2018c, Figure 4, Figure 5), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	61
5.8	Influence of the user's gender. Adapted from (Miehle et al., 2018c, Figure 6, Figure 7), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)	62
5.9	Influence of the user's culture and gender. Adapted from (Miehle et al., 2018c, Figure 8, Figure 9), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	63
5.10	Influence of the user's frequency of use of speech based assistants. Adapted from (Miehle et al., 2018c, Figure 10, Figure 11), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)	65
5.11	Example script for BAKING CAKE, containing alternative events (blue) and an optional event (red). Adapted from (Miehle et al., 2021d, Figure 1), © 2021 Copyright held by the owner/author(s)	67
5.12	During the study, the participants had to listen to the conversational assistant and sort the images in the correct order. Taken from (Miehle et al., 2021d, Figure 2).	68
5.13	The outcome has been presented in form of a small image. Taken from (Miehle et al., 2021d, Figure 3)	69

5.14	Likert score distribution for the user satisfaction and perceived usefulness ratings, divided by presentation style (X: exiting the dialog, Y: presenting	
	both options, YO: presenting both options and the predicted outcome, G:	
	(Miehle et al., 2021d, Figure 4).	71
5.15	Time users needed to complete a task, divided by presentation style (X: exiting the dialog, Y: presenting both options, YO: presenting both options	, -
	and the predicted outcome, G: guessing the next step) and outcome (0: bad outcome, 1: good outcome). The orange line marks the mean. Taken from (Miehle et al., 2021d, Figure 5).	73
5.16	The setting of the user study, showing a study participant talking to Alexa and afterwards filling out the on-line questionnaires on the laptop. Taken	
	from (Miehle et al., 2018b, Figure 1), with kind permission from Association for Computing Machinery	75
5.17	The results of the participants' ratings grouped by the different categories of the questionnaire, whereby the rating scale has been inverted for the negatively formulated statements so that the optimal rating is five for every	10
	statement. $M(p)$ is the mean of the ratings of all dialogues which were declared as preferred, $M(np)$ is the mean of the ratings of all dialogues which were declared as not preferred and $D = M(n) = M(np)$ is the	
	difference between $M(n)$ and $M(nn)$ Taken from (Miehle et al. 2018b	
	Figure 3), with kind permission from Association for Computing Machinery.	83
5.18	The map-based task allocation interface. Adapted from (Kawanaka et al.,	
	2021, Figure 1), licensed under CC BY 4.0 (https://creativecommons. org/licenses/by/4.0).	87
5.19	The chat-based task allocation interface. Taken from (Kawanaka et al., 2021, Figure 2), licensed under CC BY 4.0 (https://creativecommons.	
	org/licenses/by/4.0).	88
5.20	The distribution of study participants for each experimental group. Taken from (Kawanaka et al., 2021, Figure 4), licensed under CC BY 4.0 (https://www.action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/action.org/ac	
	<pre>//creativecommons.org/licenses/by/4.0)</pre>	90
5.21	Comparison of the information demand level of the completed missions with the map-based and the chat-based interface (copper = low information	
	demand, silver = medium information demand, gold = high information demand). Adapted from (Kawanaka et al., 2021, Figure 5), licensed under	
	$\label{eq:CCBY 4.0} CC BY \ 4.0 \ (\texttt{https://creativecommons.org/licenses/by/4.0}). \ . \ . \ . \ .$	91
5.22	The result of Q4, showing how many male/female participants noticed a difference in the agent's interaction. Taken from (Kawanaka et al., 2021,	
	Figure 10), licensed under CC BY 4.0 (https://creativecommons.org/	
	licenses/by/4.0)	93
5.23	The result of Q5, showing the communication style preferences of all participants $(N = 55)$. Taken from (Kawanaka et al., 2021, Figure 11), licensed	
	under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	94

5.24	The result of Q5, showing the communication style preferences of all par- ticipants who noticed a difference in the agent's interaction $(N = 26)$. Taken from (Kawanaka et al., 2021, Figure 12), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)	94
5.25	The web-based application called Walkeeper consists of three parts: the activity tracker, the contents manager and the progress reminder. Taken from (Zhang et al., 2021, Figure 1), licensed under CC BY 4.0 (https://	54
	<pre>//creativecommons.org/licenses/by/4.0)</pre>	96
5.26	The interface of Walkeeper. Adapted from (Zhang et al., 2021, Figure 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/	
	4.0)	97
61	The Communication Style Percention Model developed in Chapter 5 is ex-	
0.1	tended by the correlation between the communication style of the system	
	and the preceding communication style of the user	114
6.2	The estimated user communication style, which is classified based on fea-	
	in the dialogue management to adapt the system behaviour to the user.	
	Adapted from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0	
	(https://creativecommons.org/licenses/by/3.0)	115
6.3	The custom recurrent neural network (RNN) classifier consisting of two	
	long short-term memory (LSTM) layers followed by three subsequent fully	
	activation functions	116
6.4	The communication style selection is part of the dialogue management so	110
	that it not only decides <i>what</i> is said next, but also <i>how</i> it is said. Adapted	
	from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0 (https:	105
	//creativecommons.org/licenses/by/3.0)	126

List of Tables

2.1	The confusion matrix compares the predicted classification against the ac- tual classification in the form of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) classification. Based on (Awad and Khanna, 2015).	21
5.1	There are four different options for each proposal how it is presented to the user: (1) direct, background information, (2) direct, no background information, (3) indirect, background information, (4) indirect, no background information. Taken from (Miehle et al., 2016, Table 1), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	48
5.2	The participants' gender distribution. Adapted from (Miehle et al., 2016, Table 2), licensed under CC BY 4.0 (https://creativecommons.org/	50
5.3	The gender and age distribution of the 339 participants. Taken from (Miehle et al., 2018c, Table 1), licensed under CC BY 4.0 (https://	50
5.4	The frequency of use of speech based assistants of the 339 participants. Taken from (Miehle et al., 2018c, Table 2), licensed under CC BY 4.0	56
5.5	(https://creativecommons.org/licenses/by/4.0) The evaluation results comparing the different system roles. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https:	57
5.6	<pre>//creativecommons.org/licenses/by/4.0)</pre>	58
5.7	creativecommons.org/licenses/by/4.0) Adapted from The evaluation results comparing the different genders. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https://	60
5.8	creativecommons.org/licenses/by/4.0)	62
5.9	org/licenses/by/4.0)	64
	held by the owner/author(s). \ldots \ldots \ldots \ldots \ldots \ldots \ldots	72

5.10	The distribution of participants among the combinations of dialogue types. Adapted from (Miehle et al., 2018b, Table 1), with kind permission from Association for Computing Machinery, and a second s	78
5.11	The distribution of study participants regarding their gender, their fre- quency of use of speech based assistants as well as their age. Adapted from (Miehle et al., 2018b, Table 2), with kind permission from Association for Computing Machinery	79
5.12	The results of the participants' ratings for the categories System Response Accuracy (SRA), Likeability (L) and Cognitive Demand (CD), whereby M is the mean and σ is the standard deviation of all ratings, $M(p)$ is the mean of the ratings of all dialogues which were declared as preferred and M(np) is the mean of the ratings of all dialogues which were declared as not preferred, $D = M(p) - M(np) $ is the difference between $M(p)$ and $M(np)$, and α is the statistical significance level using the Mann-Whitney-U-Test. The differences that are statistically significant ($\alpha < 0.05$) are marked (*). Adapted from (Miehle et al., 2018b, Table 3), with kind permission from Association for Computing Machinery.	81
5.13	The results of the participants' ratings for the categories Annoyance (A), Habitability (H), System Behaviour (SB) and System Communication Style (SCS), whereby M is the mean and σ is the standard deviation of all ratings, M(p) is the mean of the ratings of all dialogues which were declared as preferred and $M(np)$ is the mean of the ratings of all dialogues which were declared as not preferred, $D = M(p) - M(np) $ is the difference between $M(p)$ and $M(np)$, and α is the statistical significance level using the Mann- Whitney-U-Test. The differences that are statistically significant ($\alpha < 0.05$) are marked (*). Adapted from (Miehle et al., 2018b, Table 3), with kind permission from Association for Computing Machinery	82
5.14	The distribution of the communication styles of the preferred dialogues showing how many study participants preferred each option. Adapted from (Miehle et al., 2018b, Table 4), with kind permission from Association for Computing Machinery.	84
5.15	The distribution of preferences of all dialogue combinations showing how many study participants preferred each option. Adapted from (Miehle et al., 2018b, Table 5), with kind permission from Association for Computing Machinery.	85
5.16	The distribution of study participants regarding their gender, their tourism experience in the experimental area and the Gamification User Types Hexad scale.	89
5.17	Comparison of the posts with the map-based and the chat-based interface.	90
5.18	The participants have been assigned randomly into one team. Adapted from (Zhang et al., 2021, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	99
	· · · · · · · · · · · · · · · · · · ·	00

xvi

5.19	The results of the weekly questionnaires containing statements that had to be rated on a five-point Likert scale $(1 = \text{fully disagree}, 5 = \text{fully agree})$. Adapted from (Zhang et al., 2021, Table 7), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)	100
5.20	The results of the one-way ANOVA showing that there are significant dif- ferences between the communication styles. Adapted from (Zhang et al., 2021, Table 4), licensed under CC BY 4.0 (https://creativecommons. org/licenses/by/4.0).	101
5.21	The results of the Tukey post-hoc analysis showing that there are significant differences between DC and IE and between DE and IE. Adapted from (Zhang et al., 2021, Table 5), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).	102
6.1	List of dialogue acts. Taken from (Miehle et al., 2022, Table 3), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).	107
6.2	Example dialogue with annotated dialogue acts. Adapted from (Miehle et al., 2022, Table 4), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).	108
6.3	Language distribution of the dialogues in the annotated corpus, whereby D is the number of dialogues and DA is the number of dialogue acts. Taken from (Miehle et al., 2022, Table 5), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).	109
6.4	Class distribution of the annotated <i>elaborateness</i> and <i>indirectness</i> scores (median of the three ratings). Adapted from (Miehle et al., 2022, Table 6), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).	109
6.5	Agreement (κ), correlation (ρ) and reliability (<i>ICC</i>) in <i>elaborateness</i> and <i>indirectness</i> of the three ratings (R1, R2, R3). All results are significant at the 0.001 level. Taken from (Miehle et al., 2022, Table 7), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).	110
6.6	Agreement (κ) and correlation (ρ) between the mean and the median of the three ratings for the <i>elaborateness</i> and the <i>indirectness</i> . All results are significant at the 0.001 level.	111
6.7	Agreement (κ) and correlation (ρ) between the mean (M) and the median (Md) of the three ratings for the <i>elaborateness</i> and the <i>indirectness</i> and the <i>indirectness</i> and the individual ratings (R1, R2, R3). All results are significant at the 0.001 level	l. 112
6.8	The correlation between the last user action U_5 and the first system action S_1 of each turn as well as the median of all user and system actions of the respective turn U_{Md} and S_{Md} in terms of Spearman's rank correlation coefficient Rho ρ . All results marked with (*) are significant at the 0.01 level. Taken from (Miehle et al., 2022, Table 8), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)	113

List of Tables

6.9	The correlation between the last user action U_5 and the median of all system actions of the respective turn S_{Md} in terms of Spearman's rank correlation coefficient Rho ρ for the different languages. All results marked with (*) are significant at the 0.01 level, all results marked with (**) are significant at the 0.05 level. Taken from (Miehle et al., 2022, Table 9), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)	113
6.10	The grid of parameter values for the user communication style recognition.	117
6.11	The classification results using the dialogue act features (DA) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ	118
6.12	The classification results using the dialogue act features as well as the grammatical features (DA+G) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .	119
6.13	The classification results using the SVM classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .	120
6.14	The classification results using the MLP classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ . Taken from (Miehle et al., 2022, Table 14), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)	121
6.15	The classification results using the RNN classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .	122
6.16	The classification results using the SVM classifier and the linguistic features encoded as word embeddings (WE) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ	123

6.17	The classification results using the MLP classifier and the linguistic fea- tures encoded as word embeddings (WE) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1- Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ . Adapted from (Miehle et al., 2022, Table 15), licensed under CC BY 3.0	
6 18	(https://creativecommons.org/licenses/by/3.0)	124
0.10	encoded as word embeddings (WE) (separately and in combination with	
	the dialogue act features and the grammatical features) in terms of the	
	Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score,	
6.19	Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ Class distribution of the annotated <i>elaborateness</i> and <i>indirectness</i> scores	125
	for the 2,880 dialogue turns. Taken from (Miehle et al., 2022, Table 16),	
	licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/	
	3.0)	127
6.20	The grid of parameter values for the system communication style selection.	
	Adapted from (Miehle et al., 2022, Table 17), licensed under CC BY 3.0	107
6 91	(https://creativecommons.org/licenses/by/3.0)	127
0.21	different feature sets in terms of the Unweighted Average Bocall (UAB)	
	the Accuracy (ACC) and the F1-Score Taken from (Miehle et al. 2022	
	Table 19) licensed under CC BY 3.0 (https://creativecommons.org/	
	licenses/by/3.0).	128
6.22	The classification results for the system communication style selection in	
	terms of the Unweighted Average Recall (UAR), the Accuracy (ACC) and	
	the F1-Score of the overall test set and the individual languages. Taken	
	from (Miehle et al., 2022, Table 20), licensed under CC BY 3.0 (https://www.action.com/actional-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-action-acti	
	<pre>//creativecommons.org/licenses/by/3.0)</pre>	129
6.23	The classification results for the system communication style selection base-	
	line which is mimicking the last user communication style U_5 or the me-	
	dian of all previous user communication styles U_{Md} in terms of the Un-	
	weighted Average Recall (UAR), the Accuracy (ACC) and the F1-Score.	
	Taken from (Miehle et al., 2022, Table 21), licensed under CC BY 3.0	100
	(https://creativecommons.org/licenses/by/3.0)	129

1 Introduction

"To effectively communicate, we must realise that we are all different in the way we perceive the world and use this understanding as a guide to our communication with others."

Anthony Robbins

In recent years, speech-based assistants have become more and more part of everyone's daily life. Since the launch of Siri in October 2011, the technology has become widespread and increasingly popular. Nowadays, there are numerous assistants from different companies available, e.g. Amazon Alexa, Microsoft Cortana, Google Assistant or the already mentioned Apple Siri, to name only some of the most prominent ones. People use these speech-based assistants on their smartphones, tablets and computers, smart speakers, or in the car. Popular applications include dictating messages or starting a phone call with a contact in the personal address book, displaying appointments or birthdays, setting the alarm clock, starting a navigation or searching for information on the internet. Speech-based assistants support hands-free operation and allow for an intuitive interaction via spoken language. Nevertheless, natural communication on a human level is not yet possible with virtual assistants. While these systems mainly focus on content (*what* is said), humans also attach significant importance to formulation (*how* is it said).

This thesis examines the role of communication styles in human-computer interaction. To this end, the following research questions are investigated:

- 1. How are varying communication styles perceived by the user?
- 2. How can communication styles be integrated into spoken dialogue systems?

In order to answer the first question, different use case scenarios are considered and various user studies are conducted. It is analysed what causes the differences in the user's communication style preferences in human-computer interaction and whether there are general preferences regarding the communication style of choice. Furthermore, the question of whether varying communication styles affect the user's satisfaction and behaviour is addressed. To respond to the second question, the user's communication style with regards to the system is examined and the correlation between the user and the system communication style is investigated. Moreover, the automatic selection of the system is implemented.

1 Introduction

1.1 Motivation

Human communication is more than words. Dimbleby and Burton (1998) describe communication as an exchange of messages. These messages contain facts, opinions or beliefs and people use them to share information, feelings and ideas with others. Each message that we give or receive has a meaning. However, there might be more than one meaning and a message might mean different things to different people. When exchanging messages and meanings with others, people use strategies in their choice of words and non-verbal signs to achieve some purpose. They do not only focus on *what* they say, but also on *how* they formulate it.

For both human-human and human-computer interaction, it has been shown that people adapt across many levels of utterance production: they use the same words (Schober, 1993; Brennan and Clark, 1996; Niederhoffer and Pennebaker, 2002; Nenkova et al., 2008; Brennan, 1996; Koulouri et al., 2016; Pearson et al., 2006; Bergmann et al., 2015), coordinate their phonetic repertoire (Jungers et al., 2002), their amplitude (Coulston et al., 2002; Suzuki and Katagiri, 2007), their sentence and pause duration (Jungers et al., 2002; Darves and Oviatt, 2002; Oviatt et al., 2004; Bell et al., 2003), the prepositional form (Levelt and Kelter, 1982) and syntactic structures (Branigan et al., 2000; Reitter et al., 2006; Branigan et al., 2003) of their utterances, and the style of their messages (Garrod and Anderson, 1987; Brennan, 1991; Brennan and Ohaeri, 1994). It can be seen that the textual elements (i.e. how to formulate the utterance) play an important role. These aspects are covered by the concept of communication styles.

Numerous studies have shown that humans use different communication styles which has an effect on their interlocutor and the conversation. For instance, a diversity of communication styles in teams improves the creative environment and facilitates product innovativeness (Pesch et al., 2015). The communication style of an advisor in an online commercial group chat influences the importance of technology attributes to customers and causes different group dynamics to develop which influence customer satisfaction (Van Dolen et al., 2007). The communication style of physicians impacts the clients' knowledge of the medication regimen, initial beliefs about the medication, satisfaction with medication and medication use, thus positively influencing the treatment outcomes (Bultman and Svarstad, 2000).

At the same time, recent development in spoken dialogue system research has shown a lot of improvements in corresponding domains with a considerable increase in complexity. Based on the extensive work on communication styles and their significant importance in human-human interaction, communication styles are a valuable possibility for adaptation in human-computer interaction which can be applied across different domains and use case scenarios. To better illustrate the high potential, the idea is illustrated in the following example: the user asks the system for today's headlines in the newspaper. User A is in the mid-40s, employed in a large company and leaves the house every day at exactly 8:00 am. He is already fully dressed for today and has prepared everything to leave the house on time. USER A: What time is it?

SYSTEM: It's 5 minutes before 8:00.

- USER A: All right, then I have five minutes left. What are today's headlines in the newspaper?
- SYSTEM: The local newspaper reports the following: a factory building burnt down tonight, there was a lot of damage to property, but no one was injured. A demonstration by the group "Climate Warriors" will take place in the city centre this afternoon. At the last municipal council meeting, it was decided that the city's footpaths will be made wheelchair-accessible. The football club won the last match and is now on a good way to win the championship title.

USER A: Thanks. Bye.

SYSTEM: Have a nice day.

Since there is no time to read the entire newspaper or have it read aloud, user A is satisfied if he gets the headlines of the day. He only wants the most important information and this should be formulated in a short and concise way. User B, on the other hand, is already retired. Since his vision is impaired, he likes to have the newspaper read aloud every morning while having breakfast. When he asks for the headlines of the day, the system should answer in more detail and provide background information.

USER B: What are today's headlines in the newspaper?

SYSTEM: In the local newspaper, the biggest headline of the day is that a factory building burnt down tonight. A fire broke out in the paper factory around midnight. However, the fire alarm system and several sprinklers prevented a possible catastrophe.

USER B: This sounds terrible!

SYSTEM: Four employees of the company had been in the building when the fire broke out, but had already escaped to safety before the fire brigade arrived. They were examined by the emergency services, but no injuries were found. However, there was a lot of damage to property.

USER B: At least no one was injured.

SYSTEM: It is further reported that ...

This example shows that varying the level of elaborateness can be used to adapt the dialogue to the user and the situation. This also applies to other communication styles. Hence, the aim of this thesis is to enable spoken dialogue systems to adapt their communication style to the user's communication idiosyncrasies. A detailed summary of the contributions made in this regard is provided in the following section.

1.2 Contributions

The task of providing spoken dialogue systems with the ability to adapt their communication style to the user's communication idiosyncrasies is divided into three parts: a

1 Introduction

thorough requirement analysis in the context of different use case scenarios and prototypes, the development of a formal model which covers elements that are relevant for the selection of the system's communication style as well as aspects that are influenced by the system communication style, and the implementation of the targeted communication style adaptation by use of statistical classification approaches.

For the requirement analysis, four different prototypes are introduced and examined: a social companion and conversational partner for the elderly, a culturally adaptive information agent for immigrants, a spoken dialogue system for indoor navigation, and an intelligent digital assistant for clinical operating rooms. These application scenarios represent a broad cross section of the current research on spoken dialogue systems, including different user groups and domains. Hence, they are a reasonable selection for the study of requirements with respect to the integration of communication styles into dialogue systems. The results demonstrate that there exist applications in which the communication style should not be varied during an interaction, but it is essential that the appropriate communication style is selected. This seems to apply for systems like the intelligent digital assistant for clinical operating rooms with strict requirements and a demand for high robustness. For companion systems and information agents, varying the system communication style is a valuable possibility for adaptation.

There are numerous parameters that influence an interaction between two participants as well as the appropriate or preferred communication style, like the speakers' roles, their cultures, their personalities or the aim of the interaction. In order to enable adaptation through communication styles, these different aspects need to be set into relation. Therefore, a formal model which covers elements that are relevant for the selection of the system communication style and aspects that are influenced by the system communication style is developed within the scope of this work. It is based on six user evaluations that investigated various user groups in different scenarios. The examined aspects are then combined in the overall *Communication Style Perception Model* which includes dependencies that are experimentally proven with human users.

Based on the results of the requirement analysis and the dependencies that are covered in the *Communication Style Perception Model*, statistical classification approaches are pursued to enable communication style adaptation in spoken dialogue systems. In order to realise these classification approaches, a corpus with annotated communication styles is created. Afterwards, the implementation is split into two steps: First, the communication style of the user is estimated. Afterwards, the communication style of the system can be selected accordingly. Classifiers to address both tasks are implemented and evaluated.

1.3 Outline

The structure of the remaining document is as follows: In Chapter 2, all relevant basics for understanding the presented work on the role of communication styles in spoken dialogue systems are explained, including spoken dialogue systems, communication styles and interactive adaptation in human-human and human-computer interaction, as well as different machine learning approaches and statistical metrics that are used throughout this thesis. Thereafter, Chapter 3 discusses relevant related work on the adaptation and recognition of communication styles in human-computer interaction. Chapter 4 examines how communication style adaptation can be incorporated into complex real-world systems. For this purpose, different prototypes are implemented and evaluated. Chapter 5 presents the Communication Style Perception Model which covers elements that are relevant for the selection of the system's communication style and aspects that are influenced by the system communication style. It is based on six user evaluations which have been conducted in order to investigate how varying communication styles are perceived by the user. Chapter 6 addresses the implementation and evaluation of the targeted communication style adaptation in a dialogue system. First, the user's communication style with regards to the system and the correlation between the user and the system communication style are examined. Then, a classification approach to automatically estimate the user communication style is presented. The estimated communication style may be used in the dialogue management to adapt the system behaviour to the user. Finally, the task of automatically selecting the system communication style during an ongoing interaction with a spoken dialogue system is addressed, which is part of the dialogue management. The thesis concludes with a summary and an outlook on future directions in Chapter 7.

For understanding the presented work on the role of communication styles in spoken dialogue systems, all relevant basics are explained in this chapter. First of all, spoken dialogue systems are described in detail, including their components and common approaches for adaptation. Furthermore, communication styles and interactive adaptation in human-human and human-computer interaction are introduced. Showing that these aspects play an important role in human communication, the background for the work on communication style adaptation in spoken dialogue systems is provided. Finally, different machine learning approaches and statistical metrics that are used throughout this work are described.

2.1 Spoken Dialogue Systems

A spoken dialogue system is an interface that allows the user to access applications on a computer or other technical devices via speech. In the following, the general architecture of spoken dialogue systems and adaptation factors are discussed, as adaptive dialogue systems have been a focus of research in recent years.

2.1.1 General Architecture of Spoken Dialogue Systems

In general, a spoken dialogue system consists of five components, as can be seen in Figure 2.1. The user's spoken utterance is input to the speech recognition, where the most likely spoken word sequences are extracted from the speech signal and transformed into text. A word hypothesis graph is generated and passed to the linguistic analysis, where the recognised hypotheses are mapped to a semantic representation. Hence, the meaning of the user's input is captured in a formal structure. The semantic representation is forwarded to the dialogue management which is the core of the spoken dialogue system. It communicates with the application, updates the current state of the dialogue, tracks the dialogue history and determines how the spoken dialogue system should react, thus controlling the dialogue flow and managing the interaction between the human user and the computer. The next system action, which has been selected by the dialogue manager, is passed to the text generation. The speech synthesis then converts this text into a speech signal which is output to the user. A more detailed description can be found in (McTear, 2004) or (Jokinen and McTear, 2009).



Figure 2.1: The architecture of a spoken dialogue system consisting of speech recognition, linguistic analysis, dialogue management, text generation and speech synthesis. Based on (McTear, 2004; Jokinen and McTear, 2009).

2.1.2 Adaptation of Spoken Dialogue Systems

In recent years, adaptive dialogue systems have been a focus of research. This thesis makes contributions with regard to adaptation through communication styles. In order to provide the necessary context and to describe the theoretical foundation, relevant adaptation approaches will be presented in the following.

Adaptation can take place in any part of the spoken dialogue system. On the input side, work is being carried out to make speech recognition and linguistic analysis speaker-dependent. Kaur et al. (2017) review different techniques of speaker-dependent speech recognition. A survey of the different feature extraction techniques Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding Coefficients (LPCC), Perceptual Linear Prediction (PLP) and Relative Spectra Perceptual Linear Predictive (RASTA-PLP) is conducted and the evaluation is presented. When comparing speaker-dependent and speaker-independent speech recognition, it is shown that the recognition accuracy increases in case of the speaker-dependent mode. Kolář et al. (2010) explore speakerspecific modelling for the task of automatic segmentation of speech into dialog acts. A linear combination of speaker-dependent and speaker-independent language and prosodic models is used and trained with a speech corpus of multiparty meetings. The results show that speaker adaptation leads to a significant overall improvement and to improvements for many individual speakers.

In the field of generation, research focuses on stylistic variation of the system utterance to achieve adaptation. Mairesse and Walker (2011) describe a parameterizable generator whose parameters are based on psychological findings about the linguistic reflexes of personality and a method based on parameter estimation models that estimate generation parameters based on target stylistic scores which are used by the generator to produce the output utterance. In a human evaluation it is shown that the parameter estimation models produce stylistic variation along multiple dimensions that is recognised by humans. Janarthanam and Lemon (2014) investigate how a system can learn to choose referring expressions (i.e. linguistic expressions that are used to refer to domain objects of interest) for users with different levels of expertise when the domain knowledge is initially unknown to the system. User modelling and adaptation strategies are learned by using a small corpus of non-adaptive dialogues and user knowledge profiles and the evaluation results show that the learned strategies perform better than hand-coded baseline policies with both simulated and real users. Jokinen and Wilcock (2001) present a spoken dialogue system that enables adaptation of the system response to the confidence in speech recognition results. When the confidence in speech recognition accuracy is high, the system uses a simple answer, while it chooses a more explicit response with repetition of old information when the speech recognition confidence is low.

For speech synthesis, studies investigate phonetic variation in dialogue. Raveh et al. (2018) present a web-based spoken dialogue system with a focus on phonetic convergence and its analysis over the course of the interaction. The state of phonetic features in the user's speech may be detected and tracked. Furthermore, the system may be adapted accordingly. Using this system, mutual speech variation influences are studied. The results show three main user behaviours with respect to their tendency to change their pronunciation based on the system's stimulus input. Gessinger et al. (2019) investigate phonetic accommodation in a question-and-answer exchange with a simulated spoken dialogue system in a Wizard-of-Oz experiment. It is shown that almost every user converges to the system to a substantial degree for a subset of the examined features, but on the level of individual users, there are cases of convergence and maintenance for all examined features, as well as occasional divergence.

The most common approach for adaptation in dialogue systems is to adapt the dialogue strategy, which is done in the dialogue management. In the course of this, various aspects are taken into account, e.g. the user's knowledge or level of expertise, the emotional state of the user, the user satisfaction or the quality of the interaction, social information about the user like the personality, and context information or environmental factors. Here are some examples to illustrate this diversity: Adaptation to the user's expertise has been performed by Jokinen and Kanto (2004). A user expertise model is used in a speechbased e-mail system, where the system's assumptions about the user expertise are encoded using three levels (novice, competent, expert). The system then responds depending on the assumed competence levels of the user and varies how much extra information is given to the user in one go. Nothdurft et al. (2012) present a companion system which adapts the structure and the content of the interaction to each user's knowledge by including explanations. In doing so, the user is prepared for upcoming tasks that have to be solved together with the system. The knowledge levels are thereby represented with a probability value ranging from 0 to 1. If the probability of success is too low, the explanation manager is notified via an explanation request to generate an explanation

which enables the user to execute the dialogue step successfully. The explanation manager then selects which explanation type is appropriate for the current lack of knowledge and should therefore be given to the user in the next system output. A framework to model the user's emotional state during a dialogue and adapt the dialogue model dynamically has been presented by Griol and Molina (2015). An emotion recogniser is incorporated as an additional module, predicting the user's emotional state. The information provided by this recogniser is then considered as additional input for the dialog manager. Litman and Pan (2002) identify problematic situations during the interaction with a spoken dialogue system for retrieving online train schedules by analysing the speech recognition performance. Afterwards, they use this information to automatically adapt the dialogue strategy. A user satisfaction based dialogue strategy is learned by Ultes et al. (2017). In simulated experiments it is shown that a live user satisfaction estimation model may be applied as reward signal in a statistical spoken dialogue system, resulting in higher estimated satisfaction whilst also achieving high task success rates.

2.2 Communication Styles

This section covers the relevant background on communication styles. The goal of this work is to model the usage of communication styles, which is a human capability and therefore builds upon the extensive preliminary work that exists on this topic on humanhuman interaction. The relevant aspects are summarised to provide the formal linguistic background for the thesis. First, the general concepts are discussed, followed by a definition of the communication styles *indirectness* and *elaborateness*. Finally, it is outlined how communication depends on the context in which it occurs.

Grice (1975) describes conversation as a cooperative activity where the talk exchanges consist of a succession of connected remarks. Following his *cooperative principle* ("Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."), each speaker makes a statement in order to promote the purpose and objective of the conversation. This superordinate principle is divided into four categories, under each of which fall different maxims (Grice, 1975):

- Quantity:
 - 1. Make your contribution as informative as is required (for the current purposes of the exchange).
 - 2. Do not make your contribution more informative than is required.
- Quality: Try to make your contribution one that is true.
 - 1. Do not say what you believe to be false.
 - 2. Do not say that for which you lack adequate evidence.
- Relation: Be relevant.

• Manner: Be perspicuous.

1. Avoid obscurity of expression.

- 2. Avoid ambiguity.
- 3. Be brief (avoid unnecessary prolixity).
- 4. Be orderly.

The listener naturally assumes that an utterance follows the cooperative principle, i.e. he presumes the speaker's cooperation in the process of understanding the utterance. However, according to Kroeger (2019), the cooperative principle is no code of conduct which has to be obeyed. A speaker may also break the maxims, as long as the hearer is able to recognise it. Hence, a deliberate deviation of the principle can be used to communicate extra elements of meaning. Meaning that is derived not from the words themselves, but from the way those words are used in a particular context, is thereby called *conversational implicature* (Grice, 1975). These implications constitute an important part of our communication.

One specific type of conversational implicature is *indirectness* (Kroeger, 2019). Searle (1975) defines indirect speech acts as "cases in which one illocutionary act is performed indirectly by way of performing another". A speech act is thereby an action that is performed through speaking, e.g. greeting, making a request, giving some information or giving an order. This means that a speaker utters a sentence and means not only what he says, but also something more. In contrast, in case of a direct speech act, a speaker utters a sentence and means exactly and literally what he says. Searle provides the following example:

SPEAKER A: Let's go to the movies tonight. SPEAKER B: I have to study for an exam.

The utterance of speaker A is a direct proposal in virtue of its meaning. In contrast, the answer of speaker B is an indirect rejection of the proposal. Literally, speaker B is making a statement, but within the given context, speaker A can infer that speaker B is rejecting the proposal as he is assuming that speaker B is cooperating in the conversation according to Grice's cooperative principle. Therefore, speaker A assumes that the response of speaker B is relevant for the current conversation. As the literal statement is not an acceptance or rejection of the proposal, speaker B probably means more than he says. As speaker A knows that both studying for an exam and going to a movie takes a large amount of time relative to a single evening, he can infer that speaker B cannot do both in one evening. As he is not able to perform the proposed act, he is probably rejecting the proposal.

Similarly, Kroeger (2019) describes a direct speech act as "one that is accomplished by the literal meaning of the words that are spoken", whereas an indirect speech act is "one that is accomplished by implicature". Neuliep (2018) describes the indirect style as a "manner of speaking in which the intentions of the speaker are hidden or only hinted at

during interaction" and the direct style as a "manner of speaking in which one employs overt expressions of intention".

Another special type of conversational implication is the flouting of the first maxim of quantity (Grice, 1975), i.e. being more *elaborate* or *concise*. This is for example the case if speaker A asks for some information and speaker B responds by not only giving the requested information, but also some additional information like how certain the respective information or its evidence is. Neuliep (2018) defines three levels for the quantity of talk: the elaborate style as the "mode of speaking that emphasises rich, expressive language", the exacting style as "manner of speaking in which persons say no more or less than is needed to communicate a point" and the succinct style as "manner of concise speaking often accompanied by silence".

Neuliep (2018) defines communication as the "simultaneous encoding, decoding and interpretation of verbal and nonverbal messages between people" that is dependent on the context in which it occurs, i.e. the cultural, physical, relational, and perceptual environment. Thus, people communicate differently depending on their cultural background. This is consistent with various cultural models (Hofstede, 2009; Elliott et al., 2016; Kaplan, 1966; R. D. Lewis, 2010) and cross-cultural research has shown that people have to be perceptive and adaptable when communicating with foreign cultures (Earley and Mosakowski, 2004). According to Neuliep (2018), the direct style is often used in individualistic, low-context cultures like, for example, the United States, England, Australia and Germany. In contrast, the indirect style is often seen in collectivistic, high-context cultures like the Asian cultures. An elaborate style of communication is usually used in Arab, Middle Eastern and Afro-American cultures, whereas European Americans generally prefer an exacting style, and a succinct style can be found in Japan. China, and some Native American/American Indian cultures. However, the context of the speaker comprises more than just the culture. The message sent by a speaker is altered by where and with whom he or she interacts, what is the goal of the interaction and which effect he or she wants to achieve.

Numerous studies have shown that humans use different communication styles which have different effects on their interlocutor and the conversation. Pesch et al. (2015) presented a study on how new product development is affected by communication style diversity in teams. The results show that a diversity of communication styles in teams improves the creative environment within these teams and thus facilitates product innovativeness and speed to market of new product development. On the other hand, it also increases relationship conflicts that hamper a creative team environment. However, the beneficial effects seem to outweigh the dysfunctional effects on the team innovation performance. The study of Van Dolen et al. (2007) examined online commercial group chat and, in particular, how the communication style of the advisor influences the effects of perceived technology attributes (perceived control, reliability, speed, and ease of use) and chat group characteristics (group involvement, similarity, and receptivity) on chat session satisfaction. The advisor used a task-oriented communication style (highly goal oriented and purposeful, giving direction and information, repeating, clarifying and evaluating information) and a socially oriented communication style (more personal and social, even to the ex-

2.3 Interactive Adaptation in Human Communication

tent of sometimes ignoring the task at hand, making jokes, showing understanding, using emoticons and rewarding the input of the customers). The results show that the online chat advisor's communication style influences the importance of technology attributes to customers and causes different group dynamics to develop which influence customer satisfaction. Bultman and Svarstad (2000) examined how the communication style of physicians impacts the clients' knowledge, initial beliefs, satisfaction, and adherence behaviour of individuals who have been prescribed a new medication for depression. The results of the study show that a collaborative communication style enhances the clients' knowledge of the medication regimen, initial beliefs about the medication, satisfaction with medication and medication use, thus positively influencing the treatment outcomes. It is not required that the given information is exhaustive, but it does require that the physician clearly communicates what to take, how much and when to take the antidepressant, when one can expect to begin feeling better, potential side effects and ways to alleviate these side effects, expected length of treatment, and a general idea of how the medication works. Another interesting finding is that the physician communication style varies between the initial visit and follow-up visits, even with the same patient. The perceptions of direct and indirect speech was investigated by Holtgraves (1986). The results indicate that the perceived appropriateness of an interactant's choice regarding how to phrase a remark in a conversation may be affected by the social process of face management. Indirect replies were perceived as more likely in face-threatening than non-face-threatening situations. When the situation was face-threatening, indirect replies that were evasive were perceived as more likely and polite than direct replies, and indirect replies were more likely to be accepted rather than challenged. Madaio et al. (2017) explored the impact of peer tutors' use of indirectness with feedback and instructions as well as the impact of the interpersonal closeness between tutor and tutee on the use of indirectness. The results show that, in comparison with friend tutors, stranger tutors provide more positive feedback and use more indirect instructions. Moreover, tutees attempt and solve more problems if the stranger tutor uses indirect instructions. No such effect was found for friend tutors, indicating that relationship impacts students' collaborative learning behaviours and that interpersonal closeness reduces the face-threat of direct instructions.

2.3 Interactive Adaptation in Human Communication

In the previous section, it has been discussed which communication styles exist and how they have been characterised by existing research. The second important step is to know how humans operate with communication styles, i.e. how the communication styles evolve in conversation and whether they are changing or staying consistent. Hence, this section reviews the relevant background regarding the human perspective on interactive adaptation in human-human and human-computer interaction.

It has been shown that people adapt their interaction styles to one another across many levels of utterance production when they communicate, e.g. by matching each other's behaviour or synchronising the timing of behaviour. Burgoon et al. (1995) review a broad range of interaction adaptation theories and models and present their own interaction

adaptation theory. According to their theory, adaptation in interaction is responsive to the needs, the expectations, and the desires of the communicators. A mechanistic theory of language processing, the interactive alignment model, is outlined by Pickering and Garrod (2004). It assumes that, in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, including the phonetic representation, the phonological representation, the lexical representation, the syntactic representation, the semantic representation and the situation model. This process of alignment is a largely automatic process which simplifies production and comprehension in dialogue. In the following, some studies that have investigated the phenomenon of interactive adaptation in human-human and human-computer interaction will be presented.

2.3.1 Interactive Adaptation in Human-Human Interaction

Levelt and Kelter (1982) investigated how speakers repeat materials from previous talk in question-answering situations. The results of two experiments show that a question's surface form can affect the format of the answer given in the way that answers tend to match to the prepositional form of the question, e.g. "(At) what time do you close?" – "(At) five o'clock." The coordination of spatial descriptions has been explored by Garrod and Anderson (1987). The results of this study show that speakers adopt similar forms of descriptions, suggesting that interlocutors adapt their description styles to one another. Thus, the results indicate that language processing in dialogue is governed by local principles of interaction. Schober (1993) investigated how speakers describe the locations of objects (from their own perspective, their addressee's perspective, or some perspective that avoids choosing one or the other person) when performing a referential communication task. The results revealed that two speakers often used exactly the same or nearly identical words to describe the same display when communicating, showing that both partners actively collaborated with each other to ensure understanding. Brennan and Clark (1996) examined lexical entrainment, which describes the phenomenon that people in conversation use the same terms when referring repeatedly to the same object. After carrying out three experiments, the authors suggest that people are proposing a conceptualisation of an object when referring to it. Their addressees may or may not agree to that proposal, but once a shared conceptualisation is established, both interlocutors appeal to it in later references. Over time, speakers may simplify their conceptual pacts or abandon them for new ones. Niederhoffer and Pennebaker (2002) explored to which degree two people in conversation coordinate by matching their word use and how this coordination is related to the success or failure of the conversation. The results of their studies offer convincing evidence that individuals coordinate their word use on both the conversational level as well as on a turn-by-turn level. An unexpected finding is the lack of a relationship between the perceived interaction quality and the degree of linguistic style matching. A corpus study examining entrainment in the use of high frequency words (i.e. the most common words in the corpus) is presented in (Nenkova et al., 2008). The results show that the degree of high-frequency word entrainment is positively correlated with task success, and that entrainment in high-frequency word usage is a good indicator of the perceived naturalness of a conversation.

2.3 Interactive Adaptation in Human Communication

Syntactic adaptation has been investigated by Branigan et al. (2000). It is examined whether speakers in a dialogue tend to coordinate the syntactic structures of their contributions, irrespective of lexical and semantic content. The results reveal that, when comparing prepositional object structures and double object structures, speakers tend to produce a syntactic form that they have just heard the other dialogue participant use. Reitter et al. (2006) examined two corpora of spoken, two-person dialogues for syntactic repetitions. Positive priming effects are found in both corpora, both for within-speaker and between-speaker priming. However, the comparison of both corpora indicates that spontaneous conversation shows significantly less priming than task-oriented dialogue. Jungers et al. (2002) examined whether speakers imitate the rate of a previously heard sentence when producing a sentence of analogous structure. In their experiment, the speakers' sentence duration was significantly longer following a slow prime than a fast prime, and significantly shorter following a fast prime than a slow prime, but the speakers were also influenced by their own preferred production rate. Therefore, the authors concluded that both the preferred rate and the prime rate influence the produced rate. Phonetic convergence during conversational interaction has been investigated by Pardo (2006). By asking separate listeners to detect pronunciation similarity in a conversational speech corpus it was determined whether pairs of talkers converged in phonetic repertoire over the course of a single interaction. The results show the existence of a relatively rapid process of phonetic convergence between interacting talkers that is influenced by a talker's role and sex, and that is persisting beyond the conversation that induces it.

2.3.2 Interactive Adaptation in Human-Computer Interaction

Even if it has been shown that there exist clear differences in human-human interaction and human-computer interaction (Doran et al., 2003), numerous studies prove that interactive adaptation also occurs in the context of human-computer interaction. Brennan (1991) compared keyboard conversations involving a simulated computer partner with those involving a human partner. In a Wizard-of-Oz experiment, both the human and the simulated computer partner varied between three styles of responses: a short response containing only one or several words but no complete sentence, a sentence response, and a lexical change response without heed to the particular lexical items used in the adjacent query. The results show both differences and similarities between a simulated computer partner and a human partner. There were significantly more acknowledgements, firstperson and second-person pronouns and ellipses with the human partner. However, there was no difference in the number of third-person pronouns, showing that people expected connectedness across conversational turns between sentences and turns, independent of whether they believed they were talking to another person or a computer. Moreover, there were differences in the style of the participants' queries. The first query was always a complete sentence with human partners, whereas with simulated computer partners, half the time the first query was a phrase or key words. As the dialogue proceeded, people adapted to their partners by designing queries that were more similar to their partners' responses. In the last half of each dialogue, the mean percentage of complete sentences was not different across both kinds of partners, and was affected only by whether the

response style was short or sentential. These results indicate that the design of the user's utterances is shaped both by the initial model of the partner and also by the partner's responses. In another Wizard-of-Oz experiment, Brennan and Ohaeri (1994) compared the effect of a telegraphic, a fluent and an anthropomorphic message style. The results show no difference in the success of the participants and in their ratings about the perceived intelligence of the system. However, the language they used was shaped by the system's message style. Lexical convergence with computers has been investigated in (Brennan, 1996). It is shown that people adopt the terms of their computer partners during textbased and speech-based interaction. Lexical alignment has also been studied by Koulouri et al. (2016). In a Wizard-of-Oz experiment, it was analysed whether speakers used the same words as their partner. The results show that the vocabulary stabilised early in the dialogue, suggesting the operation of lexical alignment between speakers.

Darves and Oviatt (2002) examined whether the duration of children's interspeaker response latencies is influenced by a computer partner's speech output. Four different voices were used in a study: male extrovert, male introvert, female extrovert and female introvert. The extrovert voices had a higher utterance rate (measured in syllables per second) and a shorter dialogue response latency. The results reveal that the children's response latencies differed depending on whether they conversed with an animated character that spoke with the extrovert or introvert voice: their response latencies increased when first exposed to the extrovert voice and then to the introvert, and decreased when first exposed to the introvert voice and then to the extrovert. Coulston et al. (2002) investigated the amplitude convergence in the children's conversational speech with animated personas. It is shown that children actively adapted to the amplitude of their partner and even readapted when a new voice was was introduced. They increased their amplitude when interacting with a louder extroverted character, and dropped it with the quiet introverted one. In (Oviatt et al., 2004), it is shown that, additionally to the adaptation of the amplitude and the interspeaker response latencies, the children also accommodated their utterance duration, their utterance rate and their utterance pause structure. The average utterance duration as well as the utterance rate increased when first interacting with the extrovert voice and then with the introvert one, and decreased when first interacting with the introvert voice and then with the extrovert one. The children's average number of pauses and the total pause duration increased when the animated character's voice switched from extrovert to introvert, and decreased when it switched from introvert to extrovert. The authors conclude that the observed changes in the children's speech represented a substantial convergence towards their computer partner's voice. However, as there is no perfect match, the children are not doing mimicry. Bell et al. (2003) investigated whether people adapt their speaking rate while interacting with an animated character. The results confirm that the users adapted to the speaking rate of the system, even if the subjects afterwards stated that they have not been aware of it. Moreover, the speakers varied their speaking rate substantially in the course of the dialogue. Slower speech was used during problematic sequences where subjects had to repeat or rephrase their utterance several times. Prosodic adaptation has also been studied by Suzuki and Katagiri (2007). They found that the participants of their study aligned at least unidirectionally: the partici-
pants produced a louder voice when the system's speech amplitude was increased, and a shorter pause duration when the system's pause duration was decreased. However, no bidirectional adaptation was found.

Branigan et al. (2003) investigated syntactic alignment in typed communication via a computer. An experiment was conducted where the participants played a dialogue game in which they believed that they were interacting with either a person or a computer. The results demonstrate syntactic alignment for both conditions and suggest that it is largely an automatic process that is unmediated by consideration of the mental states of the interlocutor. In another experiment, Pearson et al. (2006) show that the users' lexical alignment is influenced by their expectations about a system. When users believe the system to be unsophisticated and restricted in capability, they adapt their language to match the system's language more than when they believe the system to be sophisticated and capable. This tendency is unaffected by the actual behaviour that the system exhibits. In (Branigan and Pearson, 2006), the findings of the studies are summarised and it is concluded that speakers tend to align both syntactically and lexically to both computer and human addressees. Moreover, alignment in human-computer interaction seems to be even more important than in human-human interaction as it involves a stronger strategic component that is designed to increase the likelihood of successful communication. Possible mechanisms that might lead to linguistic alignment in human-computer interaction are discussed in (Branigan et al., 2010). Bergmann et al. (2015) explored lexical and gestural alignment with real and virtual humans. It is shown that adaptation takes place regarding communicative features (lexical alignment) as well as features without obvious communicative function (handedness alignment).

2.4 Machine Learning Approaches

In this work, machine learning is used to recognise the user communication style and select the corresponding system communication style from annotated data. Machine learning is a branch of artificial intelligence that has the goal of synthesising the underlying relationships among data and information (Awad and Khanna, 2015). This knowledge can then be used to predict future events or scenarios that are unknown to the computer. According to Samuel (1959), machine learning algorithms enable computers to learn from data and to improve themselves without being explicitly programmed. Machine learning can be classified into three major categories:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Kubat (2017) describes supervised learning as induction from pre-classified examples. This means that the input consists of example data and associated target responses, and the goal is to learn an approximation of the mapping function. After training is finished,

2 Relevant Background



Figure 2.2: An example for classification in two classes using a support vector machine, showing the maximum-margin hyperplane and the support vectors (red). Based on (Cortes and Vapnik, 1995).

correct responses can be predicted for new (unseen) data points. Supervised learning can be used for classification (i.e. the output is a discrete-valued category) or regression (i.e. the output is a number from a continuous range). In contrast, unsupervised learning uses examples whose classes are not known in order to induce useful properties of the available data (Kubat, 2017). It can be used for clustering (i.e. learning the inherent groupings in the data) or association rule learning (i.e. learning rules that describe portions of the data). In reinforcement learning, an agent learns from interacting with its environment. The environment responds with rewards (i.e. positive or negative feedback) and the agent optimises its behaviour with the goal of maximising the cumulative reward (Sutton and Barto, 1998).

In this work, only supervised learning is used. Classification is performed using support vector machines (SVMs) and artificial neural networks (ANNs). The concept of support vector machines has been proposed by Cortes and Vapnik (1995). Given some data points that belong to two ore more classes, the objective is to find a hyperplane (or set of hyperplanes in the case of more than two classes) that separates the data points into the different classes. A good separation is achieved if the hyperplane has the maximum margin, i.e. the maximum distance to the nearest data points of all classes. These data points that are close to the hyperplane and thus influence its position and orientation are the support vectors. This principle is illustrated in Figure 2.2, showing the maximum margin hyperplane and the support vectors.

Another widely-used approach for classification is the usage of multi-layer perceptrons (MLPs) that have been presented by Rosenblatt (1958). Being a subset of artificial neural networks, it consists of three or more layers (an input layer, an output layer, and one or more hidden layers) of nodes. Each node in one layer connects with a certain weight to every node in the following layer. Figure 2.3 shows an exemplary MLP with an input

2.4 Machine Learning Approaches



Figure 2.3: An example for a multi-layer perceptron with an input layer consisting of three nodes, one hidden layer consisting of four nodes and an output layer consisting of two nodes.

layer consisting of three nodes, a hidden layer consisting of four nodes and an output layer consisting of two nodes. During training, an MLP learns to model the correlation between the input and output by adjusting the weights of the connections between the nodes. During the forward propagation, the input is moved from the input layer through the hidden layer(s) to the output layer, and the decision of the output layer is compared to the ground truth labels. The error is calculated and back-propagated, and the weights are adjusted accordingly. Once training is done, the weights are used to classify new data points.

Recurrent neural networks (RNNs) are another class of artificial neural networks that use recurrent links in order to represent time and create a dynamic memory (Elman, 1990). As illustrated in Figure 2.4, nodes are fed information not only from the previous layer, but also from themselves from the previous pass. This allows the RNN to exhibit temporal dynamic behaviour and to process variable length sequences of inputs. Classic RNNs can keep track of arbitrary long dependencies. However, Bengio et al. (1994) have shown that common approaches are inefficient in learning long-term dependencies due to a vanishing gradient when back-propagating the error. This is why Hochreiter and Schmidhuber (1997) presented the long short-term memory (LSTM), a special form of recurrent neural network architecture using gates and an explicitly defined memory cell. In the initial version, each memory cell had an input gate protecting the memory contents from perturbation by irrelevant inputs, and an output gate protecting other units from perturbation by currently irrelevant memory contents stored in the memory cell. Gers et al. (1999) introduced an additional forget gate that enables the cell to reset itself at appropriate times. These memory cells enforce a constant error flow during backpropagation, making LSTMs insensitive to gap length. Hence, LSTMs are able to bridge long time lags and preserve temporal information even if important events are separated by a longer time duration.

2 Relevant Background



Figure 2.4: An example for a recurrent neural network with an input layer consisting of three nodes, one hidden layer consisting of four nodes and an output layer consisting of two nodes.

2.5 Statistical Metrics

In order to evaluate the performance of a classification approach, different statistical metrics are used. In the following, the metrics used throughout this work are presented.

The confusion matrix visualises and compares the predicted classification against the actual classification (Awad and Khanna, 2015). Table 2.1 shows a confusion matrix for a two-class classifier. True positive (TP) means that the outcome is correctly classified as positive, whereas false positive (FP) means that the outcome is negative but incorrectly classified as positive. Similarly, true negative (TN) means that the outcome is correctly classified as negative, whereas false negative. (FN) means that the outcome is positive but incorrectly classified as negative, whereas false negative. Based on the confusion matrix, the terms Accuracy (ACC), Recall (R), Precision (P) and F-Measure (F) are defined:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$
(2.1)

$$R = \frac{TP}{TP + FN} \tag{2.2}$$

$$P = \frac{TP}{TP + FP} \tag{2.3}$$

$$F = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 \cdot P + R} \tag{2.4}$$

 β is used to control the weight assigned to P and R and has a value from 0 to infinity.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Table 2.1: The confusion matrix compares the predicted classification against the actual classification in the form of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) classification. Based on (Awad and Khanna, 2015).

Within this work, the F1-Measure is used:

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{2.5}$$

For the comparison of two sets of ratings, the following metrics are utilised. Cohen's Kappa κ measures the relative agreement between two sets of ratings and is defined as

$$\kappa = \frac{p_0 - p_e}{1 - p_e},\tag{2.6}$$

where p_0 is the observed agreement, and p_e is the chance agreement (Cohen, 1960). Hence, $\kappa = 1$ for perfect agreement and $\kappa = -1$ for perfect disagreement.

Spearman's rank correlation coefficient Rho ρ is a non-parametric measure for the rank correlation between two variables and describes how well one variable can be expressed by the other (Spearman, 1904). It is defined as

$$\rho = \frac{\sum_{i} (x_{i} - \bar{x})(y_{i} - \bar{y})}{\sqrt{\sum_{i} (x_{i} - \bar{x})^{2}(y_{i} - \bar{y})^{2}}},$$
(2.7)

where x_i and y_i are corresponding ranked ratings, while \bar{x} and \bar{y} are the mean ranks. Thus, $\rho = 1$ if observations have identical ranks and $\rho = -1$ if observations have fully opposed ranks.

The Intraclass Correlation Coefficient ICC measures the reliability between the ratings. In this work, the One-Way Random Average Measures ICC(1, k) is used as defined in (Shrout and Fleiss, 1979) with k = 3 raters who are randomly selected from a lager population of raters. It is defined as

$$ICC(1,k) = \frac{BMS - WMS}{BMS + (k-1)WMS},$$
(2.8)

where BMS is the between-targets mean sugare and WMS is the within-targets mean square. ICC = 1 indicates maximum reliability, ICC = -1 maximum unreliability.

2 Relevant Background

2.6 Summary

In this chapter, all relevant basics for understanding the presented work on the role of communication styles in spoken dialogue systems have been explained. A spoken dialogue systems is an interface that allows the user to access applications on a technical device via speech. It consists of five components, including the speech recognition, the linguistic analysis, the dialogue management, the text generation, and the speech synthesis. Adaptation can take place in each of the components. However, the most common approach for adaptation in dialogue systems is to adapt the dialogue strategy, which is done in the dialogue management.

Communication styles play an important role in human communication. The theoretical background and the definitions of communication styles in general and for the *elaborateness* and *indirectness* in particular have been introduced. These definitions are used throughout this work for annotations and classifications. Furthermore, a broad review of studies investigating the phenomenon of interactive adaptation in human-human and human-computer interaction has been provided. It has been shown that people adapt their interaction styles to one another across many levels of utterance production when they communicate: they use the same words, coordinate their phonetic repertoire, their amplitude, their sentence and pause duration, the prepositional form and syntactic structures of their utterances, and the style of their messages—both when communicating with a human and a computer interaction partner. As the textual elements (i.e. how to formulate the utterance) are covered by the concept of communication styles, this thesis concentrates on this aspect.

Machine learning algorithms enable computers to learn from data and to improve themselves without being explicitly programmed. In this work, supervised learning is applied and classification is performed using support vector machines and artificial neural networks. In order to evaluate the classification performance, different statistical metrics are used, including the confusion matrix, Cohen's Kappa κ , Spearman's rank correlation coefficient Rho ρ and the Intraclass Correlation Coefficient *ICC*.

3 Related Work

This chapter summarises related work from the different fields considered within this thesis. In Section 3.1, approaches to the adaptation of communication styles in humanmachine interaction are presented. The focus is no longer on how the human adapts to the system, but on approaches to adapt the communication style of the system to the human user. Section 3.2 reviews approaches for the recognition of *elaborateness* and *indirectness* in the context of related applications.

3.1 Adaptation of Communication Styles in Human-Computer Interaction

Various studies suggest to adapt spoken dialogue systems to the users in a similar way that people adapt to their interlocutors. For example, Stenchikova and Stent (2007) proposed two new approaches for measuring adaptation between dialogues and used these measures to study adaptation in a corpus of spoken dialogues. As these measures can identify features that exhibit variation and can be used to evaluate adaptation, it is proposed to incorporate models of adaptation to syntactic and lexical choice into spoken dialogue systems to enable the adaptation of these systems. By adapting the system's behaviour to the user, the conversation agent may appear more familiar and trustworthy and the dialogue may be more effective. So far, communication styles have been used to create computer personalities, and approaches for stylistic variation as well as for stylistic adaptation have been investigated, which is elaborated in the following sections.

3.1.1 Development of Computer Personalities

Communication styles are a widely used medium to create computer personalities. Nass et al. (1995) endowed their system with properties associated with a dominant or submissive personality. While the dominant version displayed high confidence and used strong language, assertions and commands, the submissive version displayed a low confidence level and used weaker language, questions and suggestions. The fundamental information conveyed by the system was thereby not changed. The results of a user study showed that the users recognised the computer's personality. Moreover, they preferred the system that displayed the personality that is similar to their own personality and were more satisfied with the interaction with this system in comparison to the system that used the dissimilar personality. In (Moon and Nass, 1996), it was additionally investigated how changes in the system's dominance/submissiveness were perceived by the users. The results showed that changes in the direction towards a similar personality generated greater attraction

3 Related Work

than consistent similarity. Isbister and Nass (2000) created an extrovert and an introvert version of a computer character by use of verbal and non-verbal cues. The extroverted character used strong and friendly language in form of confident assertions that were relatively lengthy, poses with the limbs spread wide from its body, and postures that made the character seem to have moved closer to the participant. In contrast, the introverted character used weaker language in form of questions and suggestions that were relatively short, poses with the limbs closer to its body, and did not appear to approach the participant. Again, the fundamental information conveyed by the system was not changed, only the style of communicating the information. After conducting a user study, the results showed that the participants were able to identify both the verbal and the non-verbal personality cues. However, contrary to the previous studies, the participants preferred a character that had a personality that is complementary to their own personality, instead of a similar one. Tapus and Mataric (2008) also focused on the level of extroversion/introversion. The introverted version of a socially assistive therapist robot used vocal content that was nurturing and contained gentle and supportive language, as well as low pitch and volume. For the extroverted personality, a challenging language and high pitch and volume were used. The experimental results showed preference for a robot personality that matched the personality of the respective user. André et al. (2000) introduced animated presentation teams with different character settings for the personality dimensions agreeableness, extroversion and openness. Personality was conveyed by the choice of dialogue acts, the linguistic style (verbosity, specificity, force, formality, floridity, and bias), the choice of semantic content, syntactic form, and acoustical realisation. Feedback from users showed that they were able to identify the different personalities. Smestad and Volden (2019) designed a chatbot with an agreeable personality and one with a conscientious personality. Both chatbots interacted through written input and output and were equal in all regards expect their personalities. The differences in personality were displayed through the choice of language and tone of voice. The experimental results showed that the personality affected the user experience of the chatbots. Irfan et al. (2020) modelled the emotional state of users and an agent to dynamically adapt the dialogue utterance selection of a system in multiparty interactions. A proof of concept user study demonstrated that the system can deliver and maintain distinct agent personalities.

Mairesse and Walker (2010) presented a parameterizable language generator that provides a large number of parameters to support different linguistic styles in order to produce utterances matching particular personality profiles. These personality profiles were assigned fixed parameter values. An evaluation with human judges showed that the generated personality cues were reliably interpreted by humans. In (Mairesse and Walker, 2011), the same language generator has been used with parameter estimation models trained using personality-annotated data. Thus, generation parameters were estimated given target stylistic scores, which were then used by the generator to produce the output utterance. The results of a human evaluation showed that the trained models produced recognisable system personalities. Oraby et al. (2018) used the generator to synthesise a new corpus of over 88,000 restaurant domain utterances whose linguistic style varies according to the personality models. This corpus has then been used to train three neural models. An

3.1 Adaptation of Communication Styles in Human-Computer Interaction

evaluation of these trained models showed that they both preserve semantic fidelity and exhibit distinguishable personality styles. Aly and Tapus (2016) used the generator in a humanoid robot and additionally explored the usage of gestures. The introverted robot used gestures that were narrow, slow and executed at a low rate, while the extroverted gestures were broad, quick and executed at a high rate. Moreover, the generated speech content was adapted so that the robot gave more details in the extroverted condition than in the introverted condition. Experimental results showed that the participants found the robot that adapted both the speech and the gestures more engaging than the robot that adapted only the speech. Moreover, the majority of extroverted users preferred the extroverted robot, while the majority of introverted users preferred the introverted version. However, there were also some contrary preferences, even if they were not dominant. This variance in the perception of the robot behaviour reveals the difficulty in setting up clear borders and rules for the decision when which personality is preferred.

3.1.2 Style Variation

Obviously, there exist other applications than computer personalities. In the following, more general approaches to style variation are described. Whittaker et al. (2003) investigated how conciseness can be realised in spoken dialogue systems. Conciseness was thereby implemented by the number of attributes included in one option: concise descriptions mentioned only the highest weighted attribute, sufficient descriptions mentioned the top three weighted attributes, and verbose descriptions mentioned five attributes. Kruijff-Korbayová et al. (2008) described a multimodal in-car dialogue system with a template-based generator that generates and controls personal and impersonal style variation in the output. The dichotomy of the personal/impersonal style was defined in such a way that it primarily reflected a distinction in terms of agent activity: the personal style involved the explicit realisation of an agent (e.g. "I've found three songs."), while the impersonal style avoided it (e.g. "Three songs have been found.").

Porayska-Pomsta and Mellish (2004) defined a natural language model for a tutoring system with strategies for a positive or negative face. A positive face was thereby defined as a person's need to be approved of by others, while a negative face was defined as a person's need for autonomy from others. The strategies differed in the amount of content specificity (i.e. how specific and how structured the feedback is) and illocutionary specificity (i.e. how explicitly accepting or rejecting the tutor's feedback is). They were characterised in terms of the degree to which each of them accommodates for the user's need for autonomy and approval and selected based on these dimensions. Another tutoring system that models politeness was presented by Johnson et al. (2004). Natural language templates were defined and assigned positive and negative politeness values. During an interaction, the template matching the target politeness values most closely was selected. A Wizardof-Oz experiment to evaluate the interaction tactics where the participants were randomly assigned to either a polite or a direct treatment was conducted in (Wang et al., 2005). The results showed that the polite agent had a positive impact on the students' learning gains. Wilkie et al. (2005) integrated politeness strategies for system-initiated digressions in a mass-market telephone banking dialogue. Templates for a positive face redress were

3 Related Work

optimistic, informal, intensifying interest with the addressee, exaggerating approval with the addressee, presupposing common ground, showing concern for the addressee's wants, offering and promising, giving or asking for reasons. Templates for a negative face redress were pessimistic, indirect, apologising, stating the face-threatening act as a general rule, impersonalising the speaker and the addressee, giving deference, going on record as not indebting the addressee. In contrast to these templates used to mitigate positive and negative face threats, the bald templates were direct and concise. Experimental results showed no general preference for one of the strategies. Gupta et al. (2007) presented a system combining a spoken language generator with an artificial intelligence planner to model politeness in collaborative task-oriented dialogue. A direct strategy (e.g. "Do X."), an approval strategy (e.g. "Could you please do X mate?"), an autonomy strategy (e.g. "Could you possibly do X for me?") and an indirect strategy (e.g. "X is not done yet.") were used to model different levels of politeness, and different linguistic forms were defined to model each strategy. These politeness strategies have also been used in the conversational agent described in (De Jong et al., 2008) and (Hofs et al., 2010) that can help users to find their way in a virtual environment, while adapting its politeness to that of the user. In each turn, a pre-generated sentence template with politeness tags was selected depending on the politeness value of the system that is calculated based on the system's previous politeness level and the user's politeness level.

3.1.3 Style Adaptation

Besides the realisation of style variation, approaches to adaptation were examined. Walker et al. (2007) presented a two-stage sentence planner for providing restaurant information in different styles. It randomly generates multiple alternative realisations of an information presentation which differ in how the content is allocated into sentences, how the sentences are ordered, and which discourse cues are used to express the relationships between content elements. These alternative realisations are ranked using a statistical model trained on human feedback. Brockmann et al. (2005) used an approach for ranking alternative utterance candidates to simulate the effect of syntactic alignment in natural language generation. Ball and Breese (2000) presented an architecture that uses models of emotions and personality encoded as Bayesian networks. One is used to diagnose the emotions and personality of the user, and a second one to generate an appropriate behaviour for the agent by selecting scripted paraphrases that are related to its emotional state and personality. However, the agent's mood and personality might only match that of the user or be the exact opposite of the user. Buschmeier et al. (2009) presented an alignment-capable microplanner that models the interactive alignment behaviour of human speakers for different microplanning tasks (lexical choice, syntactic choice, referring expression generation and aggregation). The alignment behaviour is calculated based on the recency of use by the system itself, the recency of use by the interlocutor, the frequency of use by the system itself and the frequency of use by the interlocutor. Hoegen et al. (2019) developed an end-to-end voice-based conversational agent that is able to align with the interlocutor's conversational style. The conversational style is categorised on an axis ranging from high consideration to high involvement. The agent uses content variables (pronoun use, repetition, and utterance length) and acoustic variables (speech rate, pitch, and loudness) to calculate the user's conversational style and to match the participant on these conversational style variables. Hu et al. (2018) proposed an adaptation measure which can model adaptation on any subset of linguistic features and can be applied on a turn by turn basis during the dialogue to control adaptation in natural language generation. The method was applied to multiple corpora to investigate how the dialog situation and speaker roles influenced the level and type of adaptation to the interlocutor. It was shown that the adaptation varied depending on the feature sets, the conversational situations, the dialogue initiative and the course of the dialogue. However, the application of the measure to natural language generation was left to future work.

3.2 Recognition of Elaborateness and Indirectness

Previous work has already explored approaches for the recognition of *elaborateness* and *indirectness* in the context of related applications. Di Buccio et al. (2014) proposed a methodology to automatically detect and process verbose queries submitted to search engines. It was shown that the information retrieval effectiveness can be significantly improved by considering the query verbosity. Moreover, Gharouit and Nfaoui (2017) suggested to use BabelNet as knowledge base in the detection of verbose queries and then presented a comparative study between different algorithms to classify queries into two classes, verbose or succinct. However, both papers deal with the classification of queries submitted to search engines. To the best of our knowledge, there exists no previous work in the field of elaborateness classification for spoken language.

Goel et al. (2018) explored different supervised machine learning approaches to automatically detect indirectness in tutoring conversations. The authors collected a corpus of tutoring dialogues from 12 American-English speaking pairs of teenagers whereby the conversations included social interaction as well as tutoring periods. They annotated four types of indirectness for the tutoring periods, namely apologising (e.g. "Sorry, its negative 2."), hedging language (e.g. "You just add 5 to both sides."), the use of vague category extenders (e.g. "You have to multiply and stuff.") and subjectivising (e.g. "I think you divide by 3 here."). Each utterance was then classified as direct or indirect based on its inclusion in any of these categories. Afterwards, they used different classification approaches to detect indirectness based on textual and visual features, reaching an F1 sore of 62%. However, the literature presented in Section 2.2 suggests that there are more aspects than the four types of indirectness annotated in this corpus and that indirectness cannot be broken down to rather simple key word spotting (e.g. "sorry", "just", "and stuff", "I think"). In this work, the definition of Neuliep (2018) is used which describes the indirect style as a "manner of speaking in which the intentions of the speaker are hidden or only hinted at during interaction" (see Section 2.2) and the directness/indirectness is annotated and classified in a global way and not based on fixed structures or key words.

Other work in this field only focused on specific phenomena of indirect speech, like hedge detection (Prokofieva and Hirschberg, 2014; Ulinski et al., 2018), politeness detection (Danescu-Niculescu-Mizil et al., 2013; Aubakirova and Bansal, 2016) and uncertainty

3 Related Work

detection (Liscombe et al., 2005; Dral et al., 2008; Forbes-Riley and Litman, 2011; Adel and Schütze, 2017).

3.3 Summary

In this chapter, related work on communication styles in human-computer interaction has been presented. This includes both work on communication style adaptation as well as *elaborateness* and *indirectness* recognition.

Regarding the adaptation of communication styles in human-computer interaction, so far, work has focused on alignment and on the realisation of communication style variation in the natural language generation, both for general variation and for the development of computer personalities. However, it has also been shown that alignment is not always the appropriate system reaction. Depending on numerous parameters that influence an interaction between two participants, like the speakers' roles, their cultures, their personalities or the aim of the interaction, the appropriate or preferred speaking style or system personality differ. In this work, it is further investigated what influences the communication style preferences and how varying communication styles of a spoken dialogue system are perceived by the user. The state of the art is advanced by new insights and a formal model capturing elements that are relevant for the selection of the system's communication style as well as aspects that are influenced by the system communication style. This model demonstrates that communication styles play an important role in human-computer interaction. Based on these insights it is argued that the decision which communication style is to be used by a spoken dialogue system at which time needs to be covered by the dialogue management to ensure that the relevant parameters can be included in the decision process. In order to do so, a novel module for the communication style selection in spoken dialogue systems is implemented and evaluated.

Regarding the recognition of *elaborateness* and *indirectness* in spoken dialogue systems, only little previous work has been done. For the *elaborateness*, only queries submitted to search engines have been examined, and for the *indirectness*, merely different categories have been explored. In contrast, in this work, the *indirectness* is classified in a more global way and not based on fixed structures or key words, and the *elaborateness* is, for the first time, classified in spoken language. In this way, this thesis contributes to the state of the art of communication style classification. As a result, a novel module for the automatic user communication style recognition in spoken dialogue systems is presented. Both the user communication style classification module and the system communication style selection module extend the standard architecture of spoken dialogue systems, as can be seen in Figure 3.1. Both modules are discussed in depth in Chapter 6.

3.3 Summary



Figure 3.1: The standard architecture of spoken dialogue systems is extended by two components: 1) a communication style classifier that automatically identifies the user communication style and 2) a communication style selection module that selects an appropriate system communication style. Taken from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

4 Realisation of Use Case Scenarios and Prototypes

A first step in the development of dialogue systems that adapt their communication style is a thorough requirement analysis in the context of different use case scenarios and prototypes. The discussion of related work in Chapter 3 has shown that variation and adaptation of communication styles are used to model various system properties. For example, a spoken dialogue system can be endowed with a personality or mood, it can be shown that the system is involved in the conversation, or the system can be provided with different politeness strategies. This chapter examines how communication style adaptation can be incorporated into complex real-world systems. For this purpose, different prototypes are implemented and examined: a social companion and conversational partner for the elderly, a culturally adaptive information agent for immigrants, a spoken dialogue system for indoor navigation, and an intelligent digital assistant for clinical operating rooms. These application scenarios represent a broad cross section of the current research on spoken dialogue systems, including different user groups and domains. Hence, they are a reasonable selection for the study of requirements with respect to the integration of communication styles into dialogue systems. The results demonstrate that there exist applications in which the communication style should not be varied during an interaction, but it is essential that the appropriate communication style is selected. This seems to apply for systems like the intelligent digital assistant for clinical operating rooms, which have to comply with strict requirements and presuppose high robustness. For companion systems and information agents, varying the system communication style is a valuable possibility for adaptation.

4.1 A Social Companion and Conversational Partner for the Elderly

An important use case scenario for adaptive communication styles is a social companion and conversational partner for the specific user group of elderly persons who tells the elderly about news and current topics and allows them to talk about their past. Depending on the individual preferences, the companion could report on news and provide entertainment in an elaborate way, or keep in the background and be more concise, allowing the user to talk and take over the conversation. In (Miehle et al., 2019a), this scenario has been set up in a prototypical implementation and the evaluation shows that such individualisation is relevant. In the following, the development and evaluation are presented.

4.1.1 Requirement Analysis

With the intention of getting an insight in the specific needs and requests of ageing adults, discussions with the residents of a nursing home in the south of Germany as well as with the social managers and caregivers have been carried out. The major topic was the loneliness of the elderly living in a nursing home. The fact that those people feel lonesome can be easily explained. Nearly all of the nursing home residents are widowed and therefore do not have a partner any more. Moreover, their children reached adulthood. They have a job where they pursue a career, their friends, their hobbies and usually they have raised their own family. Even if they do not aim to leave their parents alone, they often do not have the time for many visits. Commonly, children come to visit their parents in the nursing home only on weekends. Friends of the nursing home residents are usually of the same age. Some of them have already died, others suffer from physical disabilities and therefore meetings with friends are also nearly impossible. The feeling of loneliness often leads to a perception of neglect. Therefore, most of the nursing home residents wished to have a contact person who talks to them, and even more importantly, who listens to them. The interviews with the elderly revealed that the ideal companion would not talk about topics like the person's fear of isolation, psychological or physical complaints, diseases or experiences of loss. In contrast, the companion should talk about news and current topics while at the same time allowing the elderly to tell about their past.

However, when designing a prototype of a social companion for elderly persons, not only the needs and wishes of the user group have to be taken into account, but also specific difficulties which might occur due to the person's physical and mental condition. The interviews with the social managers and caregivers in the nursing home indicated that dementia, Alzheimer's disease, depressions and apoplectic strokes lead to a reduction of the person's cognitive abilities to produce speech. Furthermore, senior adults may have difficulty in breathing due to various diseases which leads to pronunciation problems. On the other hand, elderly persons tend to be hard of hearing.

In summary, the requirement analysis revealed that the companion should talk to the elderly about news and current topics in an appropriate volume and pace so that the elderly can easily listen and keep up with the conversation. Moreover, it should be a good listener allowing the elderly to tell about their past.

4.1.2 Implementation and Evaluation

Based on the results and impressions obtained during the interviews at the nursing home, a prototype of a social companion for the elderly has been designed and a user study with a small group of retired seniors living at home or in a nursing home has been conducted. Due to the fact that the elderly expressed the wish to have a companion that talks and listens to them, the priorities have been set on the verbal interaction between the elderly and the social agent and it has been decided to use the well-known NAO robot as an off-the-shelf solution for the platform. During the design of the dialogue, the aim was to respond to the desires defined during the interviews. As, in general, elderly persons are not used to talking to any technical device and as all of the interviewed persons stated that



4.1 A Social Companion and Conversational Partner for the Elderly

Figure 4.1: The questionnaire contained 11 statements which had to be rated on a fivepoint Likert scale (1 = fully agree, 5 = fully disagree). Overall, the ratings show that the NAO robot and the dialogue were perceived very positively. Taken from (Miehle et al., 2019a, Figure 1). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Advanced Social Interaction with Agents, Lecture Notes in Electrical Engineering, Vol. 510 by Maxine Eskenazi, Laurence Devillers, Joseph Mariani (eds) ©Springer International Publishing AG, part of Springer Nature (2019).

they have never seen a robot before, the NAO robot started with singing a well-known German folk song to break the ice. While singing, the robot started to make eye contact and waved his hand. After greeting the user and asking for his or her well-being, the robot asked whether he should read out aloud some news. The user was able to chose between the fields of sport, politics and economy. After each newspaper article, the robot asked some personal questions where the user could tell about his or her past. The NAO robot thereby sat down and listened to the elderly as long as they were talking. As a good listener, he just nodded from time to time and kept eye contact. In the end, the robot said goodbye and after some good wishes he ended by singing another part of the folk song. Due to the fact that elderly persons tend to be hard of hearing, the speaking rate was slowed down, the volume was increased and the utterances were repeated if needed.

After implementing the dialogue, a user study has been conducted with a small group of retired seniors living at home or in a nursing home. In total, 16 persons participated in

4 Realisation of Use Case Scenarios and Prototypes

the survey, six of them lived in a nursing home. The participants living at home together with their spouse were aged between 50 and 75, whereas the participants living in the nursing home were widowed and aged between 75 and 98. It has been quite hard to find elderly persons who wanted to talk to a robot. Moreover, three participants terminated the study right after the beginning, one of them due to hearing problems, the other two changed their mind when they saw the robot and did not want to talk to him. The course of the survey was as follows: at first, the participants had a conversation with the NAO robot. Afterwards, they filled in the questionnaire containing statements which had to be rated on a five-point Likert scale (1 = fully agree, 5 = fully disagree) as well as open questions on what might be improved and which kind of robot the elderly would like to use. The evaluation results are depicted in Figure 4.1. Overall, the ratings show that the NAO robot and the dialogue were perceived very positively. The users stated that they liked the overall concept (M = 1.38) and that they found the NAO robot pleasant (M = 1.31). Especially the gestures, the eye contact and the broad knowledge of the robot were emphasised. Moreover, the elderly liked the dialogue (M = 1.69) and its topics (M = 1.69). The participants felt that the robot understood what they said (M = 1.46)and that they understood what the robot said (M = 1.38). Most of the elderly perceived the NAO robot as a pleasant dialogue partner (M = 1.69) and would like to talk to him again (M = 1.38).

This leads to the conclusion that a social companion as conversational partner and good listener for the elderly is indeed in demand by lonely seniors of advanced age. However, an extension of the dialogue towards individualisation would be expedient and desirable, as mentioned in response to the open questions on what might be improved and which kind of robot the elderly would like to use. One option for individualising the dialogue is the usage of different communication styles. Depending on whether the elderly user wishes their companion to provide them with news and entertain them or to be a good listener, the communication style of the conversational partner could be adapted to a more elaborate or concise one. In this way, the system can be tailored to the user's needs and his or her desired goal of interaction. To determine which communication style is appropriate, the system needs to identify relevant aspects like the system's current role or the user's background and take them into account in the decision-making process. Chapter 5 investigates what influences the decision about which communication style to choose in the current turn.

4.2 A Culturally Adaptive Information Agent for Immigrants

Another meaningful use case scenario for adaptive communication styles is a socially competent and culturally adaptive information agent that can assist immigrants in getting information about health care related questions. As described in Section 2.2, one approach to achieve cultural adaptivity is the use of varying communication styles.

4.2.1 Challenges for Dialogue Management

In (Pragst et al., 2017a), the challenges for adaptive dialogue management have been described. At the core of the agent, the dialogue manager is responsible for choosing the most suitable system action, taking into account the dialogue history as well as the user's culture and emotional state. The chosen system actions impact the user's perception of the agent to a significant degree and need to convey the impression of a trustworthy, socially competent and culturally adapted dialogue partner. There are several challenges that the dialogue manager needs to address to achieve this goal: various use cases require a large dialogue domain to be handled and considerably different target cultures result in culturally adaptive dialogue strategies being utilised. In order to address these challenges, a knowledge integration component is utilised as integral part of the dialogue manager. This allows handling a large dialogue domain while creating culturally adapted system actions from existing system actions and knowledge base content automatically without predefining culturally adapted variants of system actions. The knowledge integration component is relied upon to provide information regarding a multitude of user questions, gathering them from a knowledge base containing relevant personal information as well as trusted online sources. The dialogue strategy takes into account general features, such as very broad dialogue acts (e.g. statement or request) and its topics (e.g. weather, today), instead of predefined actions. Such features can be extracted for every system action regardless of whether it has been anticipated by the developers or not. Unforeseen topics can be correlated with known inputs to determine the most suitable system action. This approach offers further benefits: the knowledge integration can include new information dynamically, always providing the latest insights to the user, without the need to update the dialogue manager itself. Furthermore, if the user asks unforeseen questions, the knowledge integration is able to search trusted websites for relevant information and provide a suitable answer.

The knowledge integration as main source of system actions is mainly concerned with providing factually correct answers to user requests. Moreover, it allows the dialogue manager to select culturally appropriate variants of those answers. On the one hand, if the knowledge integration provides several possible system dialogue acts, the dialogue manager can select the action that is most suitable for the prevailing culture. On the other hand, the communication style of the information agent's actions can be varied by creating culturally adapted system actions from existing system actions and knowledge base content.

4.2.2 Implementation of a Culture-Aware Dialogue Manager

The challenge of culturally adaptive dialogue strategies has been investigated in (Miehle et al., 2021c). With the aim of designing a culture-aware dialogue manager which allows communication in accordance with the user's cultural idiosyncrasies, it has been examined whether culture-specific parameters may be trained by use of a supervised learning approach. For the implementation, the open-source software toolkit OpenDial (Lison and Kennington, 2016) has been used. It combines the benefits of logical and statistical meth-

4 Realisation of Use Case Scenarios and Prototypes



Figure 4.2: Probability distribution of each parameter before training (blue) and example probability distributions of two parameters after training, representing a frequently occurring system action (green) and a rarely occurring system action (red). Taken from (Miehle et al., 2021c, Figure 4).

ods to dialogue modelling by adopting a hybrid approach. Probabilistic rules represent the domain model in a structured format and allow system designers to integrate their domain knowledge. These rules contain unknown parameters that can be estimated from dialogue data using supervised learning. Thus, this hybrid concept allows the system designers to integrate domain-dependent constraints into a probabilistic context. The probabilistic rules formalism is described in (Lison, 2015). Practically, they are defined as *if...then...else* constructs that map logical conditions to a distribution over possible effects. For the action selection, OpenDial provides utility rules that associate utility values to system decisions. They can be used to find the action with the highest expected utility in the current state. The utility rules have been derived from the database described in Section 6.1, containing spoken dialogues from four European cultures (German, Polish, Spanish and Turkish). All possible system actions in response to a user action have been extracted, regardless of culture, and a rule has been implemented for each of the seven user actions. Afterwards, the supervised learning approach based on the so-called Wizard-of-Oz learning provided within the OpenDial toolkit has been used in order to estimate the parameters. This learning approach allows not only to learn from Wizard-of-Oz experiments, but also from dialogue transcripts. As the corpus contains dialogue interactions between two participants where one is taking the role of the system while the other one is taking the role of the user of that system, thus resembling the situation of Wizard-of-Oz experiments, transcripts of these dialogues have been created as input for the parameter estimation. Using these transcript files, four different culture-specific domains have been trained. Proceeding from the initial probability distribution (Gaussian distribution, $\mu = 5$, $\sigma^2 = 1$),

each parameter has been trained based on the appearance of the corresponding system action in the data set. Since the parameters are updated after each *user action-system action* tuple, a more frequent occurrence of a system action in the database causes the shifting of the mean value to a higher value. In contrast, a rare occurrence correlates with a lower mean value, reducing the probability that such a system action is selected. This effect is illustrated in Figure 4.2. Afterwards, it has been evaluated whether the trained parameters vary among the different cultures and therefore represent cultural patterns. The evaluation results have shown that the different characteristics of the cultures result in different parameters with highest mean values. Hence, the system response to a user action varies depending on the culture.

This shows that cultural adaptivity in the dialogue management can be achieved by selecting the appropriate system action. However, an approach to achieve even more cultural adaptivity is the use of varying communication styles. The discussion of communication styles in Section 2.2 has shown that people use different communication styles depending on their cultural background. Hence, varying the communication style of the information agent's actions seems to be generally applicable and culturally relevant. Chapter 5 investigates whether cultural communication idiosyncrasies found in human-human interaction may also be observed during human-computer interaction in a spoken dialogue system context.

4.3 A Spoken Dialogue System for Indoor Navigation

In (Miehle et al., 2021a), an indoor navigation system that is based on natural spoken interaction has been presented. This is another interesting use case for adaptive communication styles. Depending on whether the user knows the building in which he is navigating or not, the communication style of the navigation system could be a more concise or elaborate one. Moreover, the directness could be increased in case of communication problems, leading to more precise and accurate route descriptions.

4.3.1 Functionalities

The system navigates the user through the Ulm University based on scripts, supporting three different routes. Moreover, it can cope with incomplete scripts and inconclusive situations by passing the dialogue initiative to the user. The user can perform the following actions:

- Request Navigation: The user can ask the system for navigation (e.g. "Where do I have to go?", "What's next?").
- Offer Navigation: The user can give route descriptions in case the system has an incomplete script (e.g. "I turn left at the stairs.").
- End Navigation: During the dialogue, the user can end the navigation at any point if he is not satisfied with the interaction (e.g. "I want to end the navigation.").

4 Realisation of Use Case Scenarios and Prototypes

Besides these navigation actions, the user can accept or acknowledge the system output or ask the system to repeat the previous description, e.g. in case he did not understand the last output. The system supports four different communication styles for the route descriptions:

- Elaborate, direct (e.g. "Go straight and turn left near the stand containing magazines in order to reach the stairs.")
- Concise, direct (e.g. "Go straight to the stairs.")
- Elaborate, indirect (e.g. "Find the stairs to the left of the stand containing magazines.")
- Concise, indirect (e.g. "Find the stairs.")

Depending on whether the user knows the building in which he is navigating or not, the communication style of the navigation system could be a more concise or elaborate one. Moreover, the directness could be increased in case of communication problems, leading to more precise and accurate route descriptions.

4.3.2 Implementation and Evaluation

In order to create the scripts for the indoor navigation dialogue system, route descriptions have been collected from native speakers based on videos. The routes contain different points of interest and path elements like the cafeteria, an elevator or stairs. Overall, 74 complete scripts have been collected for each route. These descriptions were then used to create the XML scripts for the dialogue system. An excerpt of a script is shown in Figure 4.3, including examples for the different communication styles.

The system consists of two main components: the *dialogue handler* and the *dialogue manager*. The dialogue handler initiates the dialogue with the request for the first system action to the dialogue manager. The dialogue manager has a stack of system actions which was set within the initialisation. The stack is filled with the actions extracted from the XML script. The dialogue manager returns the next system action to the dialogue handler and the dialogue handler activates the text-to-speech synthesis. In parallel, a background job is running that listens to the next user input and enables barge-ins. As soon as a speech signal is detected, the voice is recorded and the audio file is sent to the Google Cloud Speech-to-Text API¹. It returns the text which is passed to the machine-learning based RASA Natural Language Understanding component². The user's intention is classified and assigned to a user dialogue manager to return a dialogue act asking the user to repeat the previous input. Otherwise, the user action is passed to the dialogue manager with the request for the next system action. The dialogue manager then decides whether the next description is taken from the system actions stack, a request is created

¹https://cloud.google.com/speech-to-text

²https://rasa.com/docs/nlu

```
<path>
  <pathelement id='1' nextElement='2'>
     <concise_direct>
        Go straight to the stairs.
     </concise_direct>
     <elaborate_direct>
        Go straight and turn left near the stand containing magazines
        in order to reach the stairs.
     </elaborate_direct>
     <concise_indirect>
        Find the stairs.
     </concise_indirect>
     <elaborate_indirect>
        Find the stairs to the left of the stand containing magazines.
     </elaborate_indirect>
  </pathelement>
  <pathelement id='2' nextElement='3'>
     <concise_direct>
        Go down the stairs.
     </concise_direct>
     <elaborate_direct>
        Go down the stairs and pass the sign saying "Hochschulsport".
     </elaborate_direct>
     <concise_indirect>
        The path continues down the stairs.
     </concise_indirect>
     <elaborate_indirect>
        The path continues down the stairs, passing a sign saying
        "Hochschulsport".
     </elaborate_indirect>
  </pathelement>
</path>
```

Figure 4.3: Excerpt of a script, showing the first and the second path element. Taken from

(Miehle et al., 2021a, Figure 2).

4 Realisation of Use Case Scenarios and Prototypes

(e.g. ask the user to go on with the navigation as the script is incomplete) or the dialogue is terminated.

In order to evaluate the indoor navigation dialogue system, a user study has been conducted where 30 participants used the system to navigate around the Ulm University. Using the SASSI questionnaire (Hone and Graham, 2000), which has been developed in order to measure the subjective assessment of speech system interfaces, an overall user satisfaction of 3.74 from a maximum of 5 has been obtained. This user satisfaction might be increased by adapting the system's communication style to the user. Depending on whether the user knows the building in which he is navigating (e.g. students or faculty members of the university) or not (e.g. guests or visitors), the communication style of the navigation system could be adapted to a more concise or elaborate one. Moreover, the directness could be increased in case of communication problems, leading to more precise and accurate route descriptions.

4.4 An Intelligent Digital Assistant for Clinical Operating Rooms

There are also applications where the spoken dialogue system has to comply with strict requirements and presupposes a high level of robustness. In such systems, the communication style should not be varied, but it seems to be essential that the appropriate communication style is selected. To demonstrate this, an intelligent digital assistant for clinical operating rooms has been implemented and evaluated (Miehle et al., 2017a; Miehle et al., 2018a). The concept of the assistant has been described in (Miehle et al., 2017b).

4.4.1 Functionalities and Challenges

The system provides the surgeon assistance in many different situations before and during an ongoing surgery. It allows speech-based interaction as speech is the modality used by the surgeon to communicate with the staff and therefore does not pose an additional mental burden if it is used to control surgical devices. In order to increase productivity and reduce the workload for the operating staff, the system acts active-cooperatively and supports the surgeon autonomously during the surgery. It escorts the surgery team throughout the entire procedure and provides assistance where necessary. The main functionalities of the presented speech-based assistant for a clinical operation room include:

- Providing data about surgery type, operating team, general patient data, laboratory data, pre-diseases and medical treatment
- Saving preferred device settings for each surgeon, reading and changing the presettings as well as transmitting the parameters to the operation room devices (e.g. table, room light, insufflator, suction and irrigation unit)
- Automatically controlling surgical devices (e.g. starting the insufflator, increasing the gas insufflation, turning off and on the light, tilting the table)

- Tracking the usage of surgical material (e.g. trocars, different types of clips, suturing material) and warning if the usage differs from the predicted surgical workflow
- Emergency mode for unforeseen incidents during a surgery, which includes a *silent* option to prevent further distractions by the system

Enabling an intelligent operating assistance system to follow a surgery and control surgical devices automatically bears several challenges. For keeping track of the procedure and automatically controlling surgical devices, the system needs to know when to perform which action on which device and when to stay in the background. Therefore, it has to be aware of the whole context of the surgery, i.e. the current point of the procedure and all past and future actions. This means that a reliable method for tracking the course of the surgery needs to be developed, thus allowing to detect unscheduled events. Moreover, it has to be clearly defined how the system is supposed to react in tenuous situations. For this purpose, standardised surgeries need to be described in detail, allowing the system to compare the actual course of the procedure to the schedule (Feußner and Wilhelm, 2016). Using this medical domain knowledge, exact models of the complex surgery structure need to be created which are then applied to the voice interaction system. Additionally, an interface needs to be designed and implemented which allows intercommunication between the voice interaction system and the surgical devices as well as the clinical information system. With respect to patient safety, appropriate strategies need to be defined in order to maintain full control of the medical devices even if the system is allowed to perform some predefined actions during the surgery and control devices automatically.

4.4.2 Implementation and Expert Evaluation

For the implementation of the intelligent digital assistant, the ontology-based dialogue management system OwlSpeak developed by Heinroth et al. (2010) and further extended by Ultes and Minker (2014) has been used. It has been connected to the hospital information database which acts as the interface between the dialogue manager and the intelligent operating room, thus allowing OwlSpeak to access necessary data and to control surgical devices. Keeping track of the surgery is done by tracking the tool usage. Therefore, variables for all kinds of instruments and assistance actions have been introduced. The system listens to each of the surgeon's instructions and increments the variables after each user utterance corresponding to its specific purpose. The workflow and hence the current part of the operation are then derived from the history of used tools at any point of the surgical intervention. The observed course of the procedure is compared to the surgery schedule which has been modelled in OwlSpeak. In case of a deviation from the schedule, the system reacts proactively and utters a warning. The surgeon can then correct the amount of used material or tell the system that the expected usage has to be adapted for the rest of the procedure. For the emergency mode, an Agenda³ without any system

 $^{^{3}\}mathrm{Concept}$ used by OwlSpeak to bundle several actions that belong to a specific dialogue turn (Ultes and Minker, 2014).

4 Realisation of Use Case Scenarios and Prototypes

action and only one possible user action, which is the user giving the command to deactivate this mode, has been introduced. For the entire dialogue, the system utterances are concise and direct, meaning that the requested information is output very concretely and without any additional information that might be inappropriate.

The system has been implemented in an experimental set-up in order to get an expert evaluation from medical specialists, as described in (Miehle et al., 2018a). As a first prototype, a laparoscopic cholecystectomy has been modelled. According to Cuschieri (1999), this is the gold standard for the treatment of gallstones and a highly standardised surgical procedure which can be segmented into ten procedural tasks. Each procedural part comprises several steps which are directly linked with the usage of certain material and instruments. As these instruments and materials, which are necessary to perform each procedural task, are clearly defined, it is possible to predict the surgeon's utterances during each step. The knowledge about these utterances has been used in order to define a dialogue corresponding to the surgical workflow. For example, the first procedural task of the laparoscopic cholecystectomy, the insertion of a Veress needle, comprises four steps:

- 1. Incision with a scalpel
- 2. Application of the first Backhaus clip
- 3. Application of the second Backhaus clip
- 4. Insertion of the Veress needle

Hence, the system needs to track "scalpel", "Backhaus clip", "Backhaus clip" and "Veress needle" in order to complete the first part of the surgery and to move on to the second procedural task which is the creation of the pneumoperitoneum. In order to begin with this part, the gas insufflation needs to be started. Therefore, the system asks the surgeon whether this action should be performed. The surgeon can then confirm the execution or tell the system to wait until he explicitly utters to do so. The resulting dialogue excerpt looks as follows:

SURGEON: Scalpel, please.SURGEON: Backhaus clip.SURGEON: Please give me another Backhaus clip.SURGEON: Now the Veress needle.SYSTEM: Shall I start the gas insufflation?SURGEON: Yes, please.SYSTEM: The gas insufflation has been started.

The implemented prototype was evaluated by specialists in the field of minimally invasive surgery where it received good feedback. The speech interface and the dialogue were perceived very positively. The system is designed not to annoy the surgeon and the operating staff with inappropriate behaviour and unnecessary system prompts during the surgery. Therefore, the system utterances are concise and direct. This communication style has been assessed well-suited for the underlying dialogue scenario in an operating room. The experts have emphasised that it is crucial that a digital assistant for clinical operating rooms provides the requested information in a very concrete way and without additional, possibly inappropriate information.

4.5 Summary

In this chapter, different use case scenarios and prototypes were examined with regard to incorporating communication style variation and adaptation. The presented systems fulfil complex tasks within diverse domains, each of which has its own requirements. Together they represent a broad cross-section of current research on dialog systems. Hence, they provide valuable insights for the integration of communication styles into dialogue systems. The results of the extensive requirement analysis demonstrate that there exist applications in which the communication style should not be varied during an interaction, but it is essential that the appropriate communication style is selected. This applies for systems with strict requirements and a demand for high robustness. For example, the intelligent digital assistant for clinical operating rooms provides the surgeon assistance during an ongoing surgery. Therefore, the system must be designed in a way such that it does not to disturb the surgeon and the operating staff with inappropriate behaviour and unnecessary system prompts during the surgery. In contrast, for companion systems and information agents, varying the system communication style seems to be a valuable possibility for adaptation. In this way, the system can be tailored to the users' needs, their cultural background, their knowledge about the current situation, or their desired goal of interaction, to name only some of the presented dimensions. Building upon these results, in Chapter 5 it will be investigated which aspects are relevant for the selection of the system communication style.

5 The Communication Style Perception Model

This chapter addresses the research question of how varying communication styles are perceived by the user. The discussion of related work on communication style adaptation in Chapter 3 and the requirement analysis based on the implemented use case scenarios and prototypes in Chapter 4 have shown that there are numerous parameters that influence an interaction between two participants as well as the appropriate or preferred communication style. Examples in this regard are the speakers' roles, their cultures, their personalities or the aim of the interaction. In order to enable adaptation through communication styles, these different aspects need to be set into relation. Therefore, a formal model which covers elements that are relevant for the selection of the system's communication style and aspects that are influenced by the system communication style is developed within the scope of this chapter. In order to do so, six user evaluations are presented, investigating various user groups in different scenarios. The examined aspects are then combined in an overall model. The resulting *Communication Style Perception Model* can be seen in Figure 5.1. It shows dependencies that are experimentally proven with human users. Sections 5.1, 5.2 and 5.3 provide evidence that both user traits and system properties influence the user's communication style preferences in human-computer interaction. Sections 5.4, 5.5 and 5.6 address the question of how varying communication styles affect the user's satisfaction and behaviour. To examine this, different communication styles are included in various systems and applications.

5.1 Cultural Communication Idiosyncrasies in Human-Computer Interaction

One specific user trait is the cultural background. The discussion of communication styles in Section 2.2 has shown that people use different communication styles depending on their cultural background. Therefore, the first experiment (Miehle et al., 2016) has investigated whether cultural communication idiosyncrasies found in human-human interaction may be transferred to human-computer interaction. Since Japanese and Germans have particularly contrasting cultural traits, this experiment has been conducted in cooperation with the Augmented Human Communication laboratory at the Nara Institute of Science and Technology (NAIST) in Japan. To examine whether the cultural background of the user influences his or her preference in the system communication style, a user study has been designed and conducted, creating a dialogue in the healthcare domain that has the potential capacity to reveal cultural differences. The dialogue contained different options

5 The Communication Style Perception Model



Figure 5.1: The *Communication Style Perception Model* showing 1) elements that are relevant for the selection of the system's communication style and 2) aspects that are influenced by the system communication style.

for the system output according to cultural differences. In every dialogue turn, the study participants from Germany and Japan had to indicate their preference concerning the system output. With the findings of the study, it is demonstrated that there are different communication style preferences in human-computer interaction depending on the cultural background of the user.

5.1.1 Integrating Cultural Communication Idiosyncrasies

In a culturally aware intelligent conversation agent, the dialogue management, sitting at the core of a dialogue system, has to be aware of cultural interaction idiosyncrasies to generate culturally appropriate output. Hence, the dialogue management is not only responsible for what is said next, but also for how it is said. This is what makes the difference to generic dialogue management where the two main tasks are to track the dialogue state and to select the next system action, i.e. what is uttered by the system. According to various cultural models (Hofstede, 2009; Elliott et al., 2016; Kaplan, 1966; R. D. Lewis, 2010; Qingxue, 2003), different cultures prefer different communication styles and various cultural differences are expected to exist between Germany and Japan. There are four dimensions which are considered relevant for dialogue management:

5.1 Cultural Communication Idiosyncrasies in Human-Computer Interaction

Animation/Emotion

The display of emotions and the apparent involvement in a topic can be perceived very differently across cultures. While in some cultures the people are likely to express their emotions, in other cultures this is quite unusual. However, both Germans and Japanese are not expected to be emotionally expressive. According to Elliott et al. (2016), both cultures avoid intensely emotional interactions as they may lead to a loss of self-control. R. D. Lewis (2010) affirms the fact that both Germans and Japanese don't like losing their face. Hence, emotionally expressive communication is not a preferred mode and the people try to preserve a friendly appearance.

Directness/Indirectness

Information provided for the user has to be presented suitable so that the user is more likely to accept it. It has to be decided whether the intent is directly expressed (e.g. "Drink more water.") or if an indirect communication style is chosen (e.g. "Drinking more water may help with headaches.") whereby the listener has to deduce the intent from the context. Elliott et al. (2016) and R. D. Lewis (2010) suppose differences between Germany and Japan in their cultural model. While Germans tend to speak very direct about certain things, Japanese prefer an implicit and indirect communication.

Identity Orientation

Internalised self-perception and certain values influence the decisions of humans which depend on their culture. Hence, arguments addressing these values may be constructed based on the user's culture. In some cultures, the people are individualistically oriented which means that the peoples' personal goals take priority over their allegiance to groups or group goals and decisions are made individualistically. In other cultures, the people are collectivistically oriented which means that there is a greater emphasis on the views, needs, and goals for the group rather than oneself and decisions are often made in relation to obligations to the group (e.g. family). According to (Hofstede, 2009; Elliott et al., 2016; R. D. Lewis, 2010; Qingxue, 2003), the identity orientation is expected to be different for Germans and Japanese. Germans are supposed to be rather individualistically oriented and the personal goals take priority over the allegiance to groups or group goals. In contrast, Japanese are more collectivistically oriented and often make their decisions in relation to obligations to their family or other groups. They tend to be people-oriented and the self is often subordinated in the interests of harmony.

Thought Patterns and Rhetorical Style

Different cultures use different argumentation styles (e.g. linear, parallel, circular or digressive). In a discussion, the way arguments are presented helps to provide necessary information to the user in an appropriate way. Additionally, some cultures have lowcontext communication whereas other cultures have high-context communication. In lowcontext communication, there is a low use of non-verbal communication. Therefore, the

5 The Communication Style Perception Model

Option	Formulation
1	Offer him tea instead of water. It tastes good
	and is not as bad as soft drinks.
2	Offer him tea instead of water.
3	Offering tea instead of water can help. It tastes
	good and is not as bad as soft drinks.
4	Offering tea instead of water can help.

Table 5.1: There are four different options for each proposal how it is presented to the user: (1) direct, background information, (2) direct, no background information, (3) indirect, background information, (4) indirect, no background information. Taken from (Miehle et al., 2016, Table 1), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

people need background information and expect messages to be detailed. In contrast, in high-context communication, there is a high use of non-verbal communication and the people do not require, nor do they expect much in-depth background information. Taking these facts into account means that the dialogue management has to make a very detailed decision about how to present the information to the user. The cultural models suppose various differences between Germans and Japanese. First of all, Qingxue (2003) states that Germans have a low-context communication while Japanese have a high-context communication. Therefore, Germans need background information and expect messages to be detailed. In contrast, Japanese provide a lot of information through gestures, the use of space, and even silence. Most of the information is not explicitly transmitted in the verbal part of the message. Furthermore, according to Elliott et al. (2016), the two cultures are expected to use different argumentation styles. For Germans, directness in stating the point, purpose, or conclusion of a communication is the preferred style while for Japanese this is not considered appropriate.

5.1.2 Experimental Setting

Based on the cultural differences in the dimensions *Directness/Indirectness, Identity Ori*entation and *Thought Patterns and Rhetorical Style*, a study has been designed in order to investigate if these differences may be transferred to human-computer interaction. Four hypotheses have been formulated:

- 1. Germans choose options with direct communication more often than Japanese do.
- 2. Japanese choose options with motivation using group oriented arguments more often than Germans do.
- 3. Germans choose options with background information more often than Japanese do.
- 4. There are differences in the selection of argumentation styles.

5.1 Cultural Communication Idiosyncrasies in Human-Computer Interaction



わかりました。それでは、水分摂取量を増やすための方法が4つあります。まず、あなたのお父さんは水がおいしくないので好きではないのでしょう。水のかわりに紅茶を飲むことを提案してみてはどうでしょうか。 Okay. Ich habe vier Ideen, wie man die Flüssigkeitsaufnahme steigern kann. Vielleicht mag dein Vater den Geschmack von Wasser nicht. Es könnte helfen, ihm Tee statt Wasser anzubieten.

Which one do you prefer?

どの選択肢がもっとも良いと感じましたか? Welche Option bevorzugen Sie?

- Figure 5.2: In each dialogue turn, the participants had to watch different videos and decide which one they prefer. Taken from (Miehle et al., 2016, Figure 1), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

For the study, a scripted dialogue in the healthcare domain has been created. This domain has the potential capacity to reveal such differences as very sensitive topics are covered. For every system output, different variations have been formulated. Each of them has been adapted according to the supposed cultural differences. The participants assumed the role of a caregiver who is caring for their father.

In the beginning of the dialogue, the system greets the user. The user also greets it and tells that their father doesn't drink enough. The system asks how much he usually drinks and the answer is that he drinks only one cup of tea after breakfast. Afterwards, different possibilities for the system's output are presented. The first one doesn't contain any background information: "You're right, that's not enough. Do you know why your father doesn't drink enough?" In contrast, the other four options include some background information why it is important for an adult to drink at least 1.5 litres of water per day. However, they differ in the argumentation style (parallel, linear, circular, digressive). The user answers that he/she doesn't know why their father doesn't drink enough. Then, the system has different proposals how the water-intake may be increased and there are four different options for each proposal how it is presented to the user. The first option contains background information and expresses the content directly. The second option is also direct but doesn't give any background information. For the third and the fourth op-

5 The Communication Style Perception Model

	German	Japanese
male	23	38
female	42	8

Table 5.2: The participants' gender distribution. Adapted from (Miehle et al., 2016, Table 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/ 4.0).

tions an indirect communication style is chosen, whereby one option contains background information and the other doesn't. An example for the different options can be found in Table 5.1.

In the end of the dialogue, the system tries to motivate the user. Two different kinds of motivation are formulated and presented by the system. The first one uses individualistically oriented arguments ("You're really doing a great job! It's impressive that you are able to handle all of this.") whereas the second one uses group oriented arguments ("You're really a big help for your family!"). Afterwards, the system and the user say goodbye and the dialogue ends.

The survey has been conducted on-line. A video for each possible system output has been created using a spoken dialogue system with an animated agent. For all recordings, the same system and the same agent have been used. In each dialogue turn, the participants had to watch videos representing the different variants of the system output and decide which one they prefer. An example of this web page is shown in Figure 5.2. During the survey, all descriptions have been provided in English, German and Japanese. The videos have been recorded in English and subtitled in German and Japanese. The translations have been made by German and Japanese native speakers who were instructed to be aware of the linguistic features and details of the cultural differences to assure equivalence in the translations.

Altogether, 65 Germans and 46 Japanese participated in the study. They have been recruited using mailing lists and social networks. The participants were aged between 15 and 62 years. The average age of the Germans was 25.7 years while the average age of the Japanese participants was 27.9 years. The gender distribution of the participants is shown in Table 5.2. It can be seen that 65% of the German and only 17% of the Japanese participants were female.

5.1.3 Evaluation Results

The evaluation of the survey confirms the main hypothesis that Germans and Japanese have different preferences in communication style in human-computer interaction. The first hypotheses says that Germans choose options with direct communication more often than Japanese do. The study contains four questions where the participants have to choose between direct and indirect options. Figure 5.3a shows the mean of how often Germans (dark grey) and Japanese (light grey) selected the direct option. The German mean is





(a) On average, Germans (dark) choose options with direct communication significantly more often than Japanese (light) do $(M_{Ger} = 1.89, M_{Jap} = 1.17, p < 0.001).$





- (c) On average, both Germans (dark) and Japanese (light) prefer options with background information ($M_{Ger} = 3.77, M_{Jap} = 3.67$). There is no significant difference.
- Figure 5.3: Results of the user study, showing differences between German and Japanese participants. Taken from (Miehle et al., 2016, Figure 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

5 The Communication Style Perception Model

with 1.89 significantly higher than the Japanese mean (p < 0.001 using the T-Test), thus confirming our hypothesis.

The second hypotheses says that Japanese choose options with motivation using group oriented arguments more often than Germans do. The survey includes one system action where the agent motivates the user. Figure 5.3b shows the mean of how often Germans (dark grey) and Japanese (light grey) selected the motivation with group oriented arguments. It can be seen that the opposite of the hypothesised effect occurred. On average, the Germans chose the option with group oriented arguments more often than the Japanese (p < 0.05 using the T-Test). An explanation for this result might be that motivation may be dependent on the topic of the dialogue. In our case, the dialogue is in the healthcare domain and caring for a family member is inherently group oriented. Therefore, it is most likely that motivating using group oriented arguments is more preferred for individualistically oriented people. However, if for someone it is natural to care for a family member because he is group oriented, then motivation using group oriented arguments is not needed and individualistically oriented arguments seem to be favoured.

The third hypotheses says that Germans choose options with background information more often than Japanese do. The survey comprises five questions where the participants could select between system outputs with and without background information. Figure 5.3c shows the mean of how often Germans (dark grey) and Japanese (light grey) selected the option with background information. On average, both Germans and Japanese preferred the options with background information. This suggests that there is no nonverbal communication in this kind of human-computer interaction which is only based on speech and does not include other modalities (the agent in the videos does not produce any output but the speech). In this case, Japanese tend to miss the non-verbal communication which they use to have in human-human interaction and therefore need verbal background information.

Our last hypotheses says that there are differences in the selection of argumentation styles. The survey contains one system output where the participants have to choose between different argumentation styles. However, no significant difference could be found.

Due to the difference in the gender distribution, it is important to investigate whether this has an effect on the overall results. As can be seen in Figure 5.4, only for *Thought Patterns and Rhetorical Style*, a significant difference has been found: on average, women chose options with background information more often than men. However, as the majority of both genders and both cultures chose the options with background information $(M_m >$ 2.5, $M_w > 2.5$, $M_{Ger} > 2.5$, $M_{Jap} > 2.5$), the difference between the genders is not supposed to effect the result based on the culture.

5.1.4 Conclusion

The presented user study has examined whether the cultural background of the user, which is one specific user trait, influences his or her communication style preference in human-computer interaction. In order to so so, it has investigated whether cultural communication idiosyncrasies found in human-human interaction may also be observed during human-computer interaction in a spoken dialogue system context. In an on-line survey,






(a) On average, both men (dark) and women (light) prefer options with indirect communication ($M_m = 1.52$, $M_w = 1.68$). There is no significant difference.

(b) On average, both men (dark) and women (light) prefer options with motivation using group oriented arguments ($M_m = 0.52$, $M_w = 0.62$). There is no significant difference.



- (c) On average, women (light) choose options with background information significantly more often than men (dark) do $(M_m = 3.52, M_w = 3.98, p < 0.05)$.
- Figure 5.4: Results of the user study, showing differences between male and female participants. Taken from (Miehle et al., 2016, Figure 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

German and Japanese participants have indicated their preference concerning different options for the system output. This multicultural study has shown that not all results are consistent with the existing cultural models for human-human interaction and not all cultural idiosyncrasies that occur in human-human interaction may be applied for human-computer interaction. This suggests that the communication patterns are not only influenced by the culture, but also by system properties and other user traits. However, there are differences between Germans and Japanese concerning their preferences in the communication patterns of the system output. Hence, it is concluded that the cultural background of the user has a direct influence on his or her preference in the system communication style.

5.2 Influences on the User's Communication Style Preferences

The last section has already shown that the cultural background has an influence on the user's communication style preference in human-computer interaction. The results further indicated that the communication patterns are also influenced by system properties and other user traits. Thus, further influencing variables were investigated in the second experiment (Miehle et al., 2018c). It has explored not only the influence of the user's culture, but also the influence of the gender, the frequency of use of speech based assistants as well as the system's role. Moreover, the cultural aspect was investigated in more detail. Hence, five European cultures were examined whose communication styles are much more alike than the German and Japanese communication idiosyncrasies. In order to investigate what causes the differences in communication style preferences, a web-based user study with 339 participants from Germany, Russia, Poland, Spain and the United Kingdom has been designed and conducted where the participants had to imagine that they are talking to a digital agent. For three different dialogues, the study participants had to indicate their preference concerning the system output in every dialogue turn. For the system output, the *elaborateness* and *indirectness* have been varied as Pragst et al. (2017b) have shown that these dimensions influence the user's perception of a dialogue and are therefore valuable possibilities for adaptive dialogue management. With the findings of this study, it is demonstrated which of the examined dimensions influence the differences in the user's communication style preferences in human-computer interaction in terms of the *elaborateness* and the *indirectness* of the system output.

5.2.1 Experimental Setting

A multicultural user study has been designed to investigate what causes the differences in the user's communication style preferences in terms of the system's *elaborateness* and *indirectness*. To do this, three dialogues have been created where the digital agent assumed three different roles. For every dialogue step, four options of how the agent talks to the user have been formulated:

- Direct, elaborate (D, E)
- Direct, concise (D, C)
- Indirect, elaborate (I, E)
- Indirect, concise (I, C)

As described in Section 2.2, the *elaborateness* refers to the amount of additional information provided to the user and the *indirectness* describes how concretely the information that is to be conveyed is addressed by the speaker. If the user, for example, asks the digital agent whether it can tell him about today's weather, the four variations of the system output look as follows:

- It will be cloudy mostly and it might rain during the afternoon. (D, E)
- It will rain. (D, C)
- I would propose to take an umbrella. A scarf and gloves would be good as well. (I, E)
- You should take an umbrella. (I, C)

This example shows that in the concise version of a system utterance only the requested information is given to the user, while the elaborate version of the same utterance results in giving a more detailed weather forecast. Moreover, the direct option gives an accurate description of the weather, whereas the indirect approach to answering that question is the advise to take an umbrella. In this case, the weather is not stated directly but can be inferred from the given information.

The study has been conducted on-line. The participants had to imagine that they are talking to a digital assistant called Kristina. They were shown the three dialogues, including the user input and different options for the system output. An exemplary dialogue turn (as it has been presented to the study participant) can be seen in the following:

YOU: Kristina, can you tell me about the weather today?

- KRISTINA: I would propose to take an umbrella. A scarf and gloves would be good as well.
- KRISTINA: It will be cloudy mostly and it might rain during the afternoon.
- KRISTINA: You should take an umbrella.
- KRISTINA: It will rain.

For each dialogue turn, the participants had to read the four different variants of the system output carefully and decide afterwards which one they prefer. All descriptions and all dialogues have been provided in the participants' mother tongues (German, English, Polish, Russian and Spanish). The translations have been made by native speakers who

Culture	Gender	#Participants	Average age
	male	32	30.75
German	female	34	30.62
De eliste	male	35	29.26
English	female	35	30.09
Daliah	male	34	27.09
Polisii	female	34	29.24
Dugaian	male	33	28.42
Russian	female	35	28.83
Craniah	male	32	25.44
Spanish	female	35	28.94

5 The Communication Style Perception Model

Table 5.3: The gender and age distribution of the 339 participants. Taken from (Miehle et al., 2018c, Table 1), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

were instructed to be aware of the linguistic features and details of the differences to assure equivalence in the translations. Moreover, the quality of the translations has been assured by the use of backward translations.

The application of our digital agent Kristina is to help people in European countries get health-related information. For the user study, three dialogues have been created where Kristina assumed three different roles. In the first dialogue, the agent took the role of a social companion. The dialogue is chat-oriented and Kristina and the user make small talk about the weather and the user's mood. The study participants had to imagine that they are talking to Kristina about the weather as they want to go swimming later on. For the second dialogue, the participants had to put themselves in the shoes of a parent who asks Kristina for help to bathe their baby. Kristina assumed the role of an expert providing the user with facts and descriptions regarding baby care. In contrast to the first dialogue, this one is task-oriented with the purpose of information retrieval. During the third dialogue, Kristina acted as a personal assistant. The users had to imagine that they are looking after Eugene who is in need of care and ask Kristina about his sleeping habits. Kristina retrieved useful information about the sleeping routine of the care recipient Eugene from personal profile data and provided it to the user.

Altogether, 339 persons from Germany, Russia, Poland, Spain and the United Kingdom participated in the user study. They have been recruited and paid using the Clickworker Survey Service¹ where the target group can be defined according to demographic data. The participants were aged between 18 and 55 years, 166 of them have been male (48.97%), 173 have been female (51.03%). The participants' detailed gender and age distribution is shown in Table 5.3. It can be seen that the number of participants is evenly distributed among the five different cultures. Moreover, the average age of the ten different groups ranges between 25.44 and 30.75 years. The participants were asked how often they use a

¹www.clickworker.com

Usage	#Participants	%
every day	40	11.80
several times a week	84	24.78
a couple of times a month	121	35.69
never	94	27.73

5.2 Influences on the User's Communication Style Preferences





Communication Style	#Participants	%
direct, elaborate	2060	50.64
direct, concise	661	16.25
indirect, elaborate	921	22.64
indirect, concise	426	10.47

speech based assistant like Apple Siri, Google Assistant or Microsoft Cortana. The results are shown in Table 5.4. It can be seen that 35.69% stated that they use speech based assistants a couple of times a month, 27.73% indicated that they never use such systems, 24.78% use them several times a week and only 11.80% use them every day.

5.2.2 Evaluation Results

In this section, the results of the user study are described. For the evaluation of the results, the first and the last dialogue turn of every dialogue are not taken into account as these system outputs have been used for greeting and leave-taking in order to have complete dialogues for the survey. This results in three dialogues, the first one containing four system utterances, the second one containing three utterances and the third one containing five utterances.

The overall evaluation results, i.e. the average across all twelve utterances and all ten user groups shown, can be seen in Figure 5.5. It is shown that 50.64% selected the direct and elaborate (D, E) version of the system utterances, while 16.25% selected the direct

Figure 5.5: The overall evaluation results, averaging across all twelve utterances and all ten user groups. Adapted from (Miehle et al., 2018c, Figure 1, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

	Communication Style	#Participants	%
	direct, elaborate	797	58.78
Dialogue 1	direct, concise	212	15.63
Dialogue 1	indirect, elaborate	191	14.09
	indirect, concise	156	11.50
	direct, elaborate	438	43.07
Dialogue 9	direct, concise	84	8.26
Dialogue 2	indirect, elaborate	438	43.07
	indirect, concise	57	5.60
	direct, elaborate	825	48.67
Dialoguo 2	direct, concise	365	21.53
Dialogue 5	indirect, elaborate	292	17.23
	indirect, concise	213	12.57

Table 5.5: The evaluation results comparing the different system roles. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

and concise (D, C) version, 22.64% selected the indirect and elaborate (I, E) version and only 10.47% selected the indirect and concise (I, C) version. Isolating the two dimensions *indirectness* and *elaborateness*, it can be seen that 66.89% selected the direct (D) version of the system utterances, while only 33.11% chose the indirect (I) version. Moreover, 73.28% selected the elaborate (E) and 26.72% the concise (C) version of the system utterances. This indicates that the participants tend to prefer the direct and elaborate variations. Furthermore, the *elaborateness* of the system has a greater influence on the user's preference than the *indirectness*.

Following the same approach, the influence of the system's role as well as the user's culture, gender and frequency of use of speech based assistants have been investigated. In the following, the detailed discussion of the results is presented.

Influence of the system's role

First of all, the influence of the system's role is examined. In the first dialogue, the agent took the role of a social companion. For the second dialogue, it assumed the role of an expert providing the user with facts and descriptions regarding baby care. During the third dialogue, the system acted as a personal assistant. The results concerning the *elaborateness* are shown in Figure 5.6a, those concerning the *indirectness* are depicted in Figure 5.6b. The exact values can be seen in Table 5.5. In Dialogue 1, where the agent's role was the social companion, 72.86% chose the elaborate and 74.41% chose the direct versions of the system utterance. This leads to significantly more direct choices than the average over all three dialogues (66.89%). In Dialogue 2, where the agent acted as an expert, 86.14% chose the elaborate and 51.33% chose the direct versions of the system utterance.



(a) In Dialogue 2, the elaborate options (dark) have been chosen significantly (p < 0.001) more often than the average over all three dialogues (black line). In Dialogue 3, the concise options (light) have been chosen significantly (p < 0.001) more often than the average over all three dialogues. In Dialogue 1, there is no significant difference to the average over all three dialogues.



- (b) In Dialogue 1, the direct options (dark) have been chosen significantly (p < 0.001) more often than the average over all three dialogues (black line). In Dialogue 2, the indirect options (light) have been chosen significantly (p < 0.001) more often than the average over all three dialogues. In Dialogue 3, the direct options (dark) have been chosen significantly (p < 0.05) more often than the average over all three dialogues.
- Figure 5.6: Influence of the system's role. Adapted from (Miehle et al., 2018c, Figure 2, Figure 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

This means that the elaborate options have been chosen significantly more often than the average over all three dialogues (73.28%). Moreover, the direct options have been chosen significantly less often than the average over all three dialogues (66.89%). In Dialogue 3, where the agent assumed the role of a personal assistant, 65.90% chose the elaborate and 70.21% chose the direct versions of the system utterance. This means that the elaborate options have been preferred significantly less often than the average over all three dialogues (73.28%). Moreover, the direct options have been chosen significantly more often than the elaborate options have been preferred significantly less often than the average over all three dialogues (73.28%). Moreover, the direct options have been chosen significantly more often than the average over all three dialogues (66.89%). This leads to the conclusion that the system's role significantly influences the user's preference in the system's communication style. The largest differences to the average over all three dialogues occur when the system acts as an expert (Dialogue 2). In this case, the elaborate and indirect options have been selected most often.

	Communication Style	#Participants	%
	direct, elaborate	370	46.72
Cormon	direct, concise	148	18.69
German	indirect, elaborate	165	20.83
	indirect, concise	109	13.76
	direct, elaborate	421	50.12
Fnglich	direct, concise	143	17.02
English	indirect, elaborate	193	22.98
	indirect, concise	83	9.88
	direct, elaborate	399	48.90
Dolich	$\begin{array}{c} \mbox{German} & \mbox{direct, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{direct, concise} & 1 \\ & \mbox{direct, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 \\ & \mbox{indirect, elaborate} & 1 \\ & \mbox{indirect, concise} & 1 $	158	19.36
1 OHSH		179	21.94
	indirect, concise	80	9.80
	direct, elaborate	420	51.47
Duccion	direct, concise	124	15.20
Russian	lirect, elaborate370lirect, concise148ndirect, concise165ndirect, concise109lirect, elaborate421lirect, elaborate143ndirect, elaborate193ndirect, elaborate193ndirect, concise83lirect, elaborate158ndirect, elaborate179ndirect, concise80lirect, elaborate179ndirect, elaborate172ndirect, elaborate172ndirect, elaborate172ndirect, elaborate172ndirect, concise100lirect, elaborate450lirect, elaborate450lirect, elaborate212ndirect, elaborate54	21.08	
	indirect, concise	100	12.25
	direct, elaborate	450	55.97
Spanich	direct, concise	88	10.95
opanish	indirect, elaborate	212	26.37
	indirect, concise	54	6.72

Table 5.6: The evaluation results comparing the different cultures. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

Influence of the user's culture

In this section, the influence of the user's culture is discussed based on the evaluation results presented in Table 5.6. Figure 5.7a shows that the participants of all cultures selected the elaborate versions (dark) of the system utterances significantly more often than the concise versions (light). Hence, all five cultures seem to prefer an elaborate communication style. However, there are small differences among the investigated cultures. While the German participants selected the elaborate options significantly less often than the average (73.28%), the Spanish participants selected the elaborate options significantly more often than the average. The results concerning the *indirectness*, which are depicted in Figure 5.7b, show that the participants of all cultures selected the direct versions (dark) of the system utterances significantly more often than the indirect versions (light). This indicates that all five cultures prefer a direct communication style. Moreover, there are no significant differences to the average (66.89%). This leads to the conclusion that there is no difference between the investigated European cultures concerning the *indirectness*.





(a) Among all cultures, the elaborate versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the concise versions (light). Moreover, there are significant differences to the average (black line) for German (p < 0.05) and Spanish (p < 0.001). There are no significant differences to the average for English, Polish and Russian.



- (b) Among all cultures, the direct versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the indirect versions (light). There are no significant differences to the average (black line).
- Figure 5.7: Influence of the user's culture. Adapted from (Miehle et al., 2018c, Figure 4, Figure 5), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

of the system's output. In contrast, there are indeed significant differences on the user's preference of the system's *elaborateness*.

Influence of the user's gender

In the following, the influence of the user's gender is investigated. The results are presented in Table 5.7 and Figure 5.8. It can be seen that both the male and the female participants selected the elaborate and direct versions (dark) of the system utterances significantly more often than the concise and indirect versions (light) and that there are no significant differences to the averages. Moreover, no significant difference between the two groups could be found concerning the system's *indirectness* whereas the women selected the elaborate options significantly more often than the men. This leads to the conclusion that the gender does not influence the user's preference concerning the *indirectness* of a system utterance. In contrast, the gender seems to influence the preference concerning the *elaborateness*. Even if both genders prefer the elaborate options over the concise options, the female participants selected the elaborate options significantly more often than the male participants did.

	Communication Style	#Participants	%
male	direct, elaborate	978	49.10
	direct, concise	342	17.17
	indirect, elaborate	450	22.59
	indirect, concise	222	11.14
female	direct, elaborate	1082	52.12
	direct, concise	319	15.37
	indirect, elaborate	471	22.69
	indirect, concise	204	9.83

Table 5.7: The evaluation results comparing the different genders. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).



(a) Both the male and the female participants selected the elaborate versions (dark) of the system utterances significantly (p < 0.001) more often than the concise versions (light). Moreover, there is a significant (p < 0.05) difference between the two groups. There are no significant differences to the average (black line).



- (b) Both the male and the female participants selected the direct versions (dark) of the system utterances significantly (p < 0.001) more often than the indirect versions (light). There is no significant difference between the two groups and there are no significant differences to the average (black line).
- Figure 5.8: Influence of the user's gender. Adapted from (Miehle et al., 2018c, Figure 6, Figure 7), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).



(a) Among all cultures and genders, the elaborate versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the concise versions (light). Moreover, there are significant differences between men and women for German and Polish (p < 0.005). There are no significant differences between men and women for English, Russian and Spanish.



- (b) Among all cultures and genders, the direct versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the indirect versions (light). Moreover, there is a significant difference between men and women for Spanish (p < 0.05). There are no significant differences between men and women for German, English, Polish and Russian.
- Figure 5.9: Influence of the user's culture and gender. Adapted from (Miehle et al., 2018c, Figure 8, Figure 9), licensed under CC BY 4.0 (https://creativecommons. org/licenses/by/4.0).

Influence of the user's culture and gender

In the previous sections, the participants' culture and gender have been considered separately. In the following, it is examined whether there are gender differences within the cultures. The results concerning the *elaborateness* are shown in Figure 5.9a. The German female participants selected the elaborate options significantly more often than the German male participants did and the Polish female participants selected the elaborate options significantly more often than the Polish male participants did. These results support the conclusion drawn from the results depicted in Figure 5.8a that the gender may influence the user's preference concerning the *elaborateness* of the system utterances. In contrast, there are no significant differences between men and women for English, Russian and Spanish, what leads to the conclusion that it depends on the culture whether there are gender differences concerning the *elaborateness*. The results concerning the *indirect*ness, which are depicted in Figure 5.9b, show that there is a significant difference between men and women for Spanish: the Spanish female participants selected the direct options significantly more often than the Spanish male participants did. There are no significant differences between men and women for German, English, Polish and Russian. This shows that in some cultures the gender may indeed influence the user's preference concerning

	Communication Style	#Participants	%
	direct, elaborate	237	49.38
ovoru dou	direct, concise	83	17.29
every day	indirect, elaborate	111	23.13
	indirect, concise	49	10.21
	direct, elaborate	500	49.60
several times	direct, concise	168	16.67
a week	indirect, elaborate	231	22.92
	indirect, enaborate 251 indirect, concise 109	109	10.81
	direct, elaborate	715	49.24
a couple of times	direct, concise	232	15.98
a month	indirect, elaborate	346	23.83
	indirect, concise	159	10.95
	direct, elaborate	608	53.90
novor	direct, concise	178	15.78
nevel	indirect, elaborate	233	20.66
	indirect, concise	109	9.66

Table 5.8: The evaluation results comparing the different user groups based on the frequency of use of speech based assistants. Adapted from (Miehle et al., 2018c, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/ licenses/by/4.0).

the *indirectness* of the system utterances and that the conclusion drawn from the results depicted in Figure 5.8b are not valid for all cultures.

Influence of the user's frequency of use of speech based assistants

In this section, the influence of the user's frequency of use of speech based assistants like Apple Siri, Google Assistant or Microsoft Cortana is discussed. The results concerning the *elaborateness* are shown in Figure 5.10a, those concerning the *indirectness* are depicted in Figure 5.10b. The exact values can be seen in Table 5.8. It can be seen that among all four groups, the elaborate and the direct versions (dark) of the system utterances have been chosen significantly more often than the concise and the indirect versions (light). Moreover, for both the *elaborateness* and the *indirectness*, there are no significant differences to the averages. Therefore, it can be concluded that the user's frequency of use of speech based assistants does not influence their preference in the system's communication style.

Summary of the findings

Summing up the results of the user study which have been described in the previous sections, it is concluded:

5.2 Influences on the User's Communication Style Preferences



(a) Among all four groups, the elaborate versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the concise versions (light). There are no significant differences to the average (black line).



- (b) Among all four groups, the direct versions (dark) of the system utterances have been chosen significantly (p < 0.001) more often than the indirect versions (light). There are no significant differences to the average (black line).
- Figure 5.10: Influence of the user's frequency of use of speech based assistants. Adapted from (Miehle et al., 2018c, Figure 10, Figure 11), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).
 - The system's role significantly influences the user's preference in the system's communication style.
 - There is no difference between the investigated European cultures concerning the *indirectness* of the system's output.
 - In contrast, there are cultural differences on the user's preference of the system's *elaborateness*.
 - It depends on the culture whether there are gender differences concerning the *elab*orateness and *indirectness* of the system utterances.
 - The user's frequency of use of speech based assistants does not influence their preference in the system's communication style.

5.2.3 Conclusion

The presented multicultural user study has investigated whether different user traits and system properties influence the user's preference in the system communication style. In order to do so, the impact of the user's culture and gender, the frequency of use of speech based assistants as well as the system's role on the communication styles *indirectness* and *elaborateness* has been explored. The results have shown that the user's culture

significantly influences the user's preference in the system's communication style even though five European cultures were examined whose communication styles are very alike. Depending on the culture, there are also gender differences. The user's frequency of use of speech based assistants seems to have no influence, but the system's role significantly influences the user's preference in the *indirectness* and *elaborateness* of the system. Thus, it is concluded that both user traits and system properties influence the user's preference in the system communication style. Consequently, these dependencies are included in the *Communication Style Perception Model*.

5.3 The Impact of the System's Role on Different Presentation Styles for Inconclusive Situations

The user evaluations described in Sections 5.1 and 5.2 have shown that user traits and system properties influence the user's preference in the system communication style. The third experiment (Miehle et al., 2021d) has investigated whether the system's role has an impact on user preferences that goes beyond the communication styles *indirectness* and *elaborateness*. It addresses the question of how a script knowledge based conversational assistant should act in situations of inconclusive information.

Typical everyday activities are standardised event sequences that may be represented in scripts. For example, the script for BAKING CAKE contains the events "get a bowl", "add flour, butter, sugar, eggs", "mix", etc. as can be seen in Figure 5.11. These events occur in a certain order, some might be optional (e.g. you can add chocolate chunks or leave them out), others might have an alternative (e.g. you can add cocoa or vanilla). Knowledge about these sequences of events or actions, describing a particular task or situation, is called *script knowledge* (Schank and Abelson, 1977). A script knowledge based conversational assistant is able to help an inexperienced user to fulfil the task at hand (e.g. baking a cake) by use of natural spoken language. The assistant guides the user through the script explaining step by step what to do. However, if a script is incomplete or contains multiple correct paths through the script (e.g. if there are alternative or optional events), the system may find itself in a situation with inconclusive information. This is why the question of how the conversational assistant should act in such a situation is addressed. In order to do so, a user study has been conducted, comparing and evaluating four different presentation styles to handle conflicting script information.

In previous work, scripts have already been used for various applications, e.g. text understanding (Cullingford, 1978), information extraction (Rau et al., 1989) or story telling (Swanson and Gordon, 2008). Script knowledge is often represented in graphs since this allows for a flexible structure. Common question answering systems such as the dialog system presented by Hixon et al. (2015) likewise use knowledge graphs. This is why it is suggested that script knowledge may also be used as basis for conversational assistants.

5.3 The Impact of the System's Role on Different Presentation Styles for Inconclusive Situations



Figure 5.11: Example script for BAKING CAKE, containing alternative events (blue) and an optional event (red). Adapted from (Miehle et al., 2021d, Figure 1), © 2021 Copyright held by the owner/author(s).

5.3.1 Experimental Setting

A user study has been designed in order to evaluate different situations and variables when presenting conflicting information to the user. Four possible presentation styles for such a situation have been developed:

- Exiting the dialog (X): When arriving at the inconclusive point in the script, the assistant states that he is not able to further help the user and exits the dialog (e.g. "I am very sorry but I cannot help you further.").
- Presenting both options (Y): The assistant presents both options to the user, thus delegating the decision (e.g. "Some users take the right path, while others take the left.").
- Presenting both options and the predicted outcome (YO): In certain situations, the assistant might be able to predict different consequences and to present them along with the event options (e.g. "Some users take the right path, while others take the left. The right path takes about 30 minutes longer to walk.").
- Guessing the next step (G): The assistant does not mention the options, decides by guessing which is the best next step and thus provides a direct instruction (e.g. "At the fork, take the left path."). In this study, the guessing has been implemented by choosing randomly between the options.

Moreover, the influence of the following variables has been investigated:

Task: 1/4 Please listen to the audio file to hear you agents suggestions. If you completed the task, the "Next"- button will be activated. Your available options are on the left side. Drag them to the right side in the correct order. Not all options are needed and some might not be available.	
▶ 0:00 / 0:16 ● ● ●	

- Figure 5.12: During the study, the participants had to listen to the conversational assistant and sort the images in the correct order. Taken from (Miehle et al., 2021d, Figure 2).
 - Alternative/optional event: Inconclusive situations can refer to events which are either optional (e.g. putting salt onto your hard boiled egg or not) or alternative (e.g. walking down the right or left path).
 - *Small/large impact:* Choosing an event can have a small (e.g. adding salt makes the egg more tasty but it still tastes good without salt) or large (e.g. taking the right path adds 30 minutes to the walking route) impact on the outcome.
 - *Good/bad outcome:* A situation can have a good or a bad outcome (e.g. baking a cake can either be successful or not).

The study has been conducted on-line. As shown in Figure 5.12, the study participants had to imagine that they are listening to a conversational assistant. They have been presented different images and their task has been to choose the correct images and sort them in the correct order, so that the images represented the correct path through the script. The participants had to listen to the conversational assistant and consider his instructions while solving a task. They have been able to repeat the instructions as often as they wanted and could only proceed if the image order was correct. Each participant had to solve four tasks:

• MAKING COFFEE (optional event: add a pinch of salt or not; small impact: the coffee does not taste that great without salt but you can still drink it)

5.3 The Impact of the System's Role on Different Presentation Styles for Inconclusive Situations



- Figure 5.13: The outcome has been presented in form of a small image. Taken from (Miehle et al., 2021d, Figure 3).
 - BAKING CAKE (optional event: add baking powder or not; large impact: the cake does not rise without baking powder and you have to throw it away)
 - FINDING THE ROUTE TO MAIN STATION (alternative event: go right or left at the crossroads; small impact: if you choose the wrong way, you will arrive three minutes later)
 - FINDING THE ROUTE TO CAMPING GROUND (alternative event: go right or left at the crossroads; large impact: if you choose the wrong way, it will take you so long that you get in the rain and are completely wet)

These tasks have been constructed so that each contained one inconclusive situation, where one of the presentation styles (X, Y, YO, G) was used to support the participant, and a good or a bad outcome. Presentation style and outcome were counterbalanced using Latin Square (Grant, 1948).

If the participants have been tested on the bad outcome, they certainly would not choose the event leading to the bad outcome voluntarily, especially if both event options and the possible outcome were presented (YO). Therefore, the study website has only accepted the script path with the bad outcome as correct solution in these cases. Even though the participants have been forced to choose the "bad" event, these situations can occur in real life situations, e.g. if the fast street is blocked because of construction work.

After each task, the outcome has been presented in form of a small image as can be seen in Figure 5.13 and a short explanation. Afterwards, the participants have been asked to answer the *After Scenario Questionnaire* (ASQ) (J. R. Lewis, 1995) measuring user satisfaction and the six questions of perceived usefulness of the questionnaire *Perceived Usefulness and Ease of Use* (PUEU) (Davis, 1989):

- **ASQ1** Overall, I am satisfied with the ease of completing the tasks in this scenario.
- **ASQ2** Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.
- **ASQ3** Overall, I am satisfied with the support information (messages, documentation) when completing the tasks.
- **PUEU1** Using the system in my daily life would enable me to accomplish tasks more quickly.
- **PUEU2** Using the system would improve my daily life performance.

PUEU3 Using the system in my daily life would increase my productivity.

PUEU4 Using the system would enhance my effectiveness in my daily life.

PUEU5 Using the system would make my daily life easier.

PUEU6 I would find the system useful in my daily life.

In the PUEU questionnaire, the wording "in my job" has been changed to "in my daily life" to better fit the scenarios. Both questionnaires have been rated on a seven-point Likert scale (1 =strongly disagree, 7 =strongly agree). Additionally, the time needed to complete each task has been recorded. At the end, the participants have been asked for their age, gender and frequency of use of speech based assistants.

Altogether, 182 persons from Canada, the USA and the UK participated in the user study. They have been recruited and paid using the Clickworker Survey Service². The participants' age ranged from 18 to 72 with a mean of 36 years. The gender distribution was very equal with 90 participants each identified as "female" or "male" and one each stated "non-binary" or "other". 119 participants stated to never use speech based assistants, 25 participants use them at least once a month, 22 at least once a week and 16 at least once a day.

5.3.2 Evaluation Results

The Likert score distribution for the user satisfaction and perceived usefulness ratings can be seen in Figure 5.14, the results in terms of user satisfaction, perceived usefulness and time needed to complete a task are presented in Table 5.9. For all following significance tests, the Mann-Whitney-U-Test with a significance level of 0.05 has been used. It can be seen that guessing the next step (G) results in the highest user satisfaction (US) with a mean Likert score of M(G) = 5.89. All other presentation styles perform significantly worse with M(Y) = 5.25 and M(YO) = 5.23. Far behind is the exit presentation style with M(X) = 3.82. There is no significant difference between Y and YO. The results for the perceived usefulness (PU) show a similar ranking. Again, guessing the next step (G) performs best with M(G) = 4.85. The three remaining presentation styles perform significantly worse with M(Y) = 4.42, M(YO) = 4.3 and M(X) = 3.21. Again, there is no significant difference between Y and YO. The results in terms of the mean time needed to complete a task show that the tasks solved with G were completed the fastest with M(G) = 47.71s. The other presentation styles perform significantly slower with M(Y) = 60.3s, M(YO) = 67.81s and M(X) = 82.39s. Again, YO shows no significant difference to Y.

If the assistant is guessing and thus providing a direct instruction, the user is not aware of the inconclusive situation in the assistant's script. With a good outcome, the user thus experiences a perfect conversation, and with a bad outcome, he is not aware that a better outcome exists. Thus, he is highly satisfied in both cases. Moreover, the assistant

 $^{^2}$ www.clickworker.com



Figure 5.14: Likert score distribution for the user satisfaction and perceived usefulness ratings, divided by presentation style (X: exiting the dialog, Y: presenting both options, YO: presenting both options and the predicted outcome, G: guessing the next step). The orange line marks the median. Taken from (Miehle et al., 2021d, Figure 4).

is perceived useful as all of the assistant's suggestions may directly be used to solve the task. In case of Y and YO, both possible events are presented what delegates the decision to the user. This lowers the satisfaction and the perceived usefulness. The reason might be that the user seeks help from the conversational assistant and as he does not get as much help as anticipated, the satisfaction decreases. Furthermore, the lower score for the perceived usefulness might result from the fact that the system does not completely fulfil its main task, which is helping the user through the script. The reason for the low scores of X might be similar. The assistant stated that he is unsure about the next step and quits the dialog. This leaves the user confused and he now needs to choose not only from two suggested events but from all other available events. This additional workload results in a lower satisfaction score. Moreover, the user receives no further help as the dialogue is terminated. This is why it appears least useful.

The increased time derives from the time the user needs to decide which event to choose. In case of Y and YO, the user needs to decide between two events. With X, users take the longest time to complete a task, as the assistant quits the dialog during the ongoing

5 The Communication Style Perception Model

Style	US	PU	Time
Guessing the next step (G)	5.89	4.85	47.71s
Presenting both options (Y)	5.25	4.42	60.3s
Presenting both options and the predicted outcome (YO)	5.23	4.43	67.81s
Exiting the dialog (X)	3.82	3.21	82.39s

Table 5.9: Mean Likert score for user satisfaction (US), perceived usefulness (PU) and mean time. Adapted from (Miehle et al., 2021d, Table 1), © 2021 Copyright held by the owner/author(s).

task. From there on, the user needs to try out by himself which events to choose and in what order. This consumes a lot of time.

Figure 5.15 shows the results for the presentation styles split by outcome (0: bad outcome, 1: good outcome). In case of a bad outcome, YO has a huge variance compared to the good outcome. This can be explained by the fact that the assistant mentions the possible outcomes to the two events to choose from. Obviously, the user wants to choose the event with a better outcome but he can not. Hence, he tries desperately to prevent the bad ending until he gives up and chooses the event leading to the bad outcome. This effect can not be observed for Y as in this case two events are presented without possible outcomes. Thus, both options appear neutral to the user.

Afterwards, the questionnaire answers from tasks with good and bad outcome have been compared. Both user satisfaction and perceived usefulness have a significantly worse result if the outcome is bad. On average, users experiencing a bad outcome assigned 0.45 Likert points less for user satisfaction and 0.4 points less for perceived usefulness. Both differences are statistically significant. This result confirms the assumption that users are less satisfied if their actions lead to a worse result. In addition, it has been investigated whether users feel differently when the negative outcome is particularly bad (i.e. if it has a large impact). The evaluation shows that, in case of a bad outcome, users rate significantly lower if the outcome has a large impact. On average, user satisfaction is rated 0.4 Likert points lower and perceived usefulness is rated 0.3 Likert points lower than a negative outcome with small impact. Both differences are statistically significant. Therefore, a user is less satisfied if his effort turns into a worse result. In addition, users find the assistant less useful in such a situation.

The evaluation regarding the different event types (alternative/optional), age and gender shows no significant difference. They do not seem to influence the user perception of the assistant in terms of user satisfaction and perceived usefulness. However, a significant difference between participants who regularly use speech based assistants ("at least once a day", "at least once a week" or "at least once a month") and participants who stated that they never use such systems has been found. Non-users give lower scores for both the user satisfaction (0.2 Likert points) and the perceived usefulness (0.55 Likert points). One reason for this might be that regular users can compare the assistant against other systems and therefore realise that the assistant is behaving differently from state-of-the-



Figure 5.15: Time users needed to complete a task, divided by presentation style (X: exiting the dialog, Y: presenting both options, YO: presenting both options and the predicted outcome, G: guessing the next step) and outcome (0: bad outcome, 1: good outcome). The orange line marks the mean. Taken from (Miehle et al., 2021d, Figure 5).

art conversational assistants, whereas non-users cannot value the different approaches compared to a state-of-the-art conversational assistant. Another explanation might be that non-users do not enjoy interacting with a conversational assistant (after all, they are non-users) and therefore give worse ratings.

Overall, the results show that the users prefer if the assistant guesses the next step and provides a direct instruction (even if this might not be the best option), instead of delegating the decision to the user. Guessing the next step, and thus maintaining competence, leads to better results in terms of user satisfaction, perceived usefulness and time needed to complete the task.

5.3.3 Conclusion

The presented study has investigated how a script knowledge based conversational assistant should act in situations of inconclusive information, i.e. where two events can be chosen as the next step. In order to do so, different presentation styles have been evaluated. The results have shown that there is a clear user preference for this particular

application: the assistant should guess the next step and provide a direct instruction instead of delegating the decision to the user. Since the system's role is to provide assistance with the task at hand, it has to act accordingly. This shows that the system's role has an impact on user preferences that goes beyond the communication styles *indirectness* and *elaborateness*.

5.4 The Impact of Varying Communication Styles on User Satisfaction

In Sections 5.1, 5.2 and 5.3, it has been investigated what influences the user's communication style preference in human-computer interaction and shown that both user traits and system properties play an important role. The second step is to know how varying system communication styles affect the user's satisfaction and behaviour if they are selected according to the user's personal preference. In order to investigate the impact of varying communication styles on user satisfaction, the fourth experiment (Miehle et al., 2018b) has addressed the question of how varying communication styles of a spoken user interface are perceived by users. In addition, it has been examined whether there exist global preferences in the communication styles *elaborateness* and *indirectness*. A total of 60 participants had two conversations each with Amazon's Alexa where Alexa used varying wordings for its output. In a post-survey, the participants had to rate statements to subjectively assess each dialogue. Moreover, they had to indicate which dialogue they preferred. The results show that the system's communication style has a direct influence on the user's satisfaction level as well as the user's perception of the dialogue and imply that the preference in the system's communication style is individual for every person. This emphasises the need for adaptive user interfaces.

5.4.1 Experimental Setting

In order to investigate the impact of the communication styles *elaborateness* and *indirectness*, a user study has been designed where the participants had two conversations each with Amazon's Alexa which has been used as front-end for the self-implemented dialogue. The setting can be seen in Figure 5.16. The dialogue has been created using the Alexa Skills Kit³, a set of APIs and tools that enables the development of new Alexa Skills. Hence, specific dialogues have been created that contained pre-scripted answers to the questions the study participants posed during their interaction with Alexa. During both dialogues, each participant had to complete the same task. In order to ensure that every participant asked the same questions, the following task description was provided:

You want to plan your day and ask Alexa for help. First of all, you want to know what the weather is like today. You are hoping for sunshine, as you are already looking forward to go to an outdoor pool. If the weather forecast is not promising, you will be disappointed because you wanted to go swimming.

³https://developer.amazon.com/alexa-skills-kit

5.4 The Impact of Varying Communication Styles on User Satisfaction



Figure 5.16: The setting of the user study, showing a study participant talking to Alexa and afterwards filling out the on-line questionnaires on the laptop. Taken from (Miehle et al., 2018b, Figure 1), with kind permission from Association for Computing Machinery.

In this case, Alexa will of course try to suggest alternatives. Indoor swimming pool is not your thing, so you wouldn't want to go there if Alexa encourages it. However, if she suggests a visit to a museum, you will be thrilled because you like going to museums. In this case, you would like to know what exhibition there is in Ulm today. You also want to go to a restaurant nearby. Finally, you thank for the information and say goodbye to Alexa.

However, within this scenario, the users could formulate their utterances freely. Both the *elaborateness* and the *indirectness* have been used as binary dimensions. For every dialogue, the system selected one out of four different communication styles which has been used during the entire dialogue. According to the communication styles, four different wordings have been formulated for every system output:

- Direct, elaborate
- Direct, concise
- Indirect, elaborate
- Indirect, concise

The resulting dialogue types are as follows, where the user input is exemplary (as the users could formulate their utterances freely), but the system utterances are pre-scripted:

Direct, elaborate:

SYSTEM: Hello, what can I do for you?

USER: Hello, can you tell me what the weather's gonna be like today?

SYSTEM: Most of the time it is cloudy and in the afternoon it will rain.

USER: Oh, that's too bad. I wanted to go swimming this afternoon.

SYSTEM: Why don't you go to the indoor pool? The "Donaubad" in Neu-Ulm is the largest adventure pool in the region and offers a lot of fun and action, but

also peace and relaxation with its sauna and thermal bath.

USER: I don't like indoor swimming pools.

SYSTEM: Then how about a visit to a museum? In Ulm, there are numerous offers of different museums and galleries.

USER: I like museums. What's in Ulm today?

SYSTEM: Today, an exhibition on the subject of "Curiosity Cabinet and Chamber of Wonders" is on display in the Ulm Museum. Here, works by contemporary artists resurrect the atmosphere of the early chambers of curiosities and reveal the secrets of our world to visitors.

USER: That sounds interesting. Is there a good restaurant nearby?

SYSTEM: In the museum there is a small café and opposite is the restaurant Enchilada. The café in the museum offers a selection of small dishes and is open all day long, while the restaurant Enchilada serves Mexican dishes from 6 pm.

USER: Great, thanks for the info! Bye.

SYSTEM: You're welcome. Just let me know if there's anything else I can do for you. See you soon!

Indirect, elaborate:

SYSTEM: Hello, what can I do for you?

USER: Hello, can you tell me what the weather's gonna be like today?

SYSTEM: Today is a good day for cosy activities at home. In the afternoon you could get wet outside.

USER: Oh, that's too bad. I wanted to go swimming this afternoon.

SYSTEM: The "Donaubad" in Neu-Ulm is the largest indoor swimming pool in the region and offers a lot of fun and action, but also peace and relaxation with its sauna and thermal bath.

USER: I don't like indoor swimming pools.

SYSTEM: Visits to museums are very popular when the weather is bad. In Ulm, there are numerous offers of different museums and galleries.

USER: I like museums. What's in Ulm today?

SYSTEM: The exhibition on the subject of "Curiosity Cabinet and Chamber of Wonders" in the Ulm Museum shows works by contemporary artists. These resurrect the atmosphere of the early chambers of curiosities and reveal the secrets of our world to visitors.

USER: That sounds interesting. Is there a good restaurant nearby?

SYSTEM: The café in the museum offers a selection of small dishes and is open all day long, while the restaurant Enchilada serves Mexican dishes from 6 pm.

USER: Great, thanks for the info! Bye.

SYSTEM: You're welcome. I'm always happy if I can help you. See you soon!

Direct, concise:

SYSTEM: Hello.

USER: Hello, can you tell me what the weather's gonna be like today? SYSTEM: It will rain.

USER: Oh, that's too bad. I wanted to go swimming this afternoon.

SYSTEM: Why don't you go to the indoor pool?

USER: I don't like indoor swimming pools.

SYSTEM: Then how about a visit to a museum?

USER: I like museums. What's in Ulm today?

SYSTEM: Today, an exhibition on the subject of "Curiosity Cabinet and Chamber of Wonders" is on display in the Ulm Museum.

USER: That sounds interesting. Is there a good restaurant nearby?

SYSTEM: In the museum there is a small café and opposite is the restaurant Enchilada.

USER: Great, thanks for the info! Bye.

SYSTEM: You're welcome. See you soon!

Indirect, concise:

SYSTEM: Hello.

USER: Hello, can you tell me what the weather's gonna be like today?

SYSTEM: Today is a good day for cosy activities at home.

USER: Oh, that's too bad. I wanted to go swimming this afternoon.

SYSTEM: The "Donaubad" in Neu-Ulm is the largest indoor swimming pool in the region.

USER: I don't like indoor swimming pools.

SYSTEM: Visits to museums are very popular when the weather is bad.

USER: I like museums. What's in Ulm today?

SYSTEM: The exhibition on the subject of "Curiosity Cabinet and Chamber of Wonders" in the Ulm Museum shows works by contemporary artists.

USER: That sounds interesting. Is there a good restaurant nearby?

SYSTEM: The café in the museum offers a selection of small dishes, while the restaurant Enchilada serves Mexican dishes.

USER: Great, thanks for the info! Bye.

SYSTEM: You're welcome. See you soon!

Dialogue 1	Dialogue 2	#Participants
direct, elaborate	direct, concise	5
direct, elaborate	indirect, elaborate	5
direct, elaborate	indirect, concise	5
direct, concise	direct, elaborate	5
direct, concise	indirect, elaborate	5
direct, concise	indirect, concise	5
indirect, elaborate	direct, elaborate	5
indirect, elaborate	direct, concise	5
indirect, elaborate	indirect, concise	5
indirect, concise	direct, elaborate	5
indirect, concise	direct, concise	5
indirect, concise	indirect, elaborate	5

5 The Communication Style Perception Model

Table 5.10: The distribution of participants among the combinations of dialogue types. Adapted from (Miehle et al., 2018b, Table 1), with kind permission from Association for Computing Machinery.

To avoid differences related to the ordering and the combination of the dialogue types, the participants where evenly distributed among the different wording options, as can be seen in Table 5.10. After each dialogue, the participants had to rate 38 statements about the two dialogue types on a five-point Likert scale in an on-line questionnaire on the laptop. Most of the statements were taken from the SASSI questionnaire (Hone and Graham, 2000) which has been developed in order to measure the subjective assessment of speech system interfaces. More precisely, the categories System Response Accuracy, Likeability, Cognitive Demand, Annoyance and Habitability have been used. Moreover, some statements about the special perception of the system behaviour and the system communication style have been added, as can be seen in Tables 5.12 and 5.13. In the end, the participants were asked to complete a third questionnaire about their age, their gender as well as how often they use a speech based assistant like Apple Siri, Amazon Alexa, Google Assistant or Microsoft Cortana. Furthermore, they had to state which dialogue they would prefer and answer the following open questions:

- Have you noticed any differences between the dialogues? If so, which ones?
- Do you have any additional general remarks about the dialogues?

The participants were recruited via flyers and mailing lists at the university and received an expense allowance of 5 EUR. This amount is based on an hourly wage of 10 EUR and an expected duration of 30 minutes. Overall, 60 people participated in the study. All of them were German native speakers and the dialogues as well as the questionnaires were kept in German to avoid any kind of misunderstandings based on the participants' language skills. As can be seen in Table 5.11, 34 of them were male and 26 were female, most of the participants have never used speech-based assistants and the average age was

	#Participants
overall	60
male	34
female	26
every day	5
several times a week	6
several times a month	14
never	35
< 20 years	8
20-29 years	45
30-39 years	4
≥ 40 years	3

5.4 The Impact of Varying Communication Styles on User Satisfaction

Table 5.11: The distribution of study participants regarding their gender, their frequency of use of speech based assistants as well as their age. Adapted from (Miehle et al., 2018b, Table 2), with kind permission from Association for Computing Machinery.

25.6 years. Before the study, every participant signed a consent form declaring voluntary participation and consent to the use of their data.

5.4.2 Evaluation Results

In the following, the results of the user study are described. First, the impact of the varying communication styles *elaborateness* and *indirectness* on the satisfaction of the users will be examined and the question "How are varying communication styles of a spoken dialogue system perceived by the user?" will be answered. Afterwards, the preferences regarding the two communication styles among the group of study participants will be investigated and the question "Are there general preferences regarding the communication style of choice?" will be answered.

Impact of Elaborateness and Indirectness on User Satisfaction

The study participants had to rate 38 statements about the dialogues on a five-point Likert scale (1 = fully disagree, 5 = fully agree) for each dialogue. The statements were divided into 7 categories:

- System Response Accuracy (SRA)
- Likeability (L)
- Cognitive Demand (CD)

- Annoyance (A)
- Habitability (H)
- System Behaviour (SB)
- System Communication Style (SCS)

Moreover, the participants were asked which of the two dialogues they would prefer. The results of the ratings can be seen in Tables 5.12 and 5.13, whereby M is the mean and σ is the standard deviation of all ratings, M(p) is the mean of the ratings of all dialogues which were declared as preferred and M(np) is the mean of the ratings of all dialogues which were declared as not preferred, D = |M(p) - M(np)| is the difference between M(p) and M(np), and α is the statistical significance level using the Mann-Whitney-U-Test. The differences that are statistically significant ($\alpha < 0.05$) are marked (*). It should be noted that some of the statements were formulated positively while others were formulated negatively. Thus, the optimal rating is five for positive formulations (e.g. The system is accurate.) and one for negative formulations (e.g. The system is unreliable.). It can be seen that for all 38 statements, the ratings are better for the preferred dialogues. For 28 statements, the differences are statistically significant. In the following, the results of the differences will be discussed.

SRA During the preferred dialogue, the system was perceived as more accurate, reliable and dependable. Moreover, the interaction was rated as more consistent and more efficient. During the dialogue that was not preferred, the participants had the feeling that the system didn't always do what they expected, even if the system was doing exactly the same as during the dialogue that was preferred, but used different wordings for its output.

L In this category, all differences are statistically significant, showing that the likeability was significantly higher during the preferred dialogue. The system was perceived as more useful, pleasant and friendly. Moreover, the participants had more enjoyment using the system and stated that it is clearer how to speak to the system and easier to learn to use the system. They had the feeling of having more control of the interaction and stated that it is more likely that they would use this system.

CD The study participants stated that the system was easier to use and that they themselves were calmer during the preferred dialogue. However, there is no significant difference in the level of concentration required when using the system, the participants' confidence and tensity when using the system.

A During the dialogue that was not preferred, the interaction was rated as more irritating and frustrating and the system was perceived as less flexible. However, there is no significant difference in the repetitiveness and the boringness of the interaction.

Cat.	Statement	M	σ	M(p)	M(np)	D	α
	The system is accurate.	3.85	1.07	4.10	3.60	0.50	0.016*
	The system is unreliable.	1.98	1.11	1.70	2.27	0.57	0.008*
	The interaction with the system is unpredictable.	2.18	1.00	2.05	2.32	0.27	0.273
	The system didn't always do what I wanted.	2.63	1.36	2.40	2.87	0.47	0.056
SRA	The system didn't always do what I expected.	2.90	1.39	2.60	3.20	0.60	0.018*
	The system is dependable.	3.66	1.08	3.98	3.33	0.65	0.002^{*}
	The system makes few errors.	3.53	1.27	3.75	3.32	0.43	0.061
Cat.	The interaction with the system is consistent.	3.72	0.95	3.87	3.57	0.30	0.034*
	The interaction with the system is efficient.	3.51	1.03	3.87	3.15	0.72	0.000*
	The system is useful.	4.12	0.89	4.32	3.92	0.40	0.025*
	The system is pleasant.	3.98	0.93	4.37	3.60	0.77	0.000*
	The system is friendly.	4.27	1.01	4.43	4.10	0.33	0.040^{*}
	I enjoyed using the system.	4.22	0.90	4.47	3.97	0.50	0.001^{*}
L	It is clear how to speak to the system.	4.22	0.87	4.42	4.02	0.40	0.031*
	It is easy to learn to use the system.	4.58	0.73	4.73	4.42	0.32	0.023*
	I would use this system.	3.64	1.19	3.98	3.30	0.68	0.001^{*}
SRA L CD	I felt in control of the interaction with the system.	3.52	1.07	3.83	3.20	0.63	0.002*
	I felt confident using the system.	3.93	0.87	4.05	3.80	0.25	0.129
	I felt tense using the system.	2.03	1.00	1.87	2.18	0.32	0.097
CD	I felt calm using the system.	4.02	0.90	4.22	3.82	0.40	0.015^{*}
UD	A high level of concentration is required when using the system.	2.31	1.04	2.22	2.40	0.18	0.366
	The system is easy to use.	4.43	0.75	4.57	4.30	0.27	0.027^{*}

5.4 The Impact of Varying Communication Styles on User Satisfaction

Table 5.12: The results of the participants' ratings for the categories System Response Accuracy (SRA), Likeability (L) and Cognitive Demand (CD), whereby M is the mean and σ is the standard deviation of all ratings, M(p) is the mean of the ratings of all dialogues which were declared as preferred and M(np) is the mean of the ratings of all dialogues which were declared as not preferred, D = |M(p) - M(np)| is the difference between M(p) and M(np), and α is the statistical significance level using the Mann-Whitney-U-Test. The differences that are statistically significant ($\alpha < 0.05$) are marked (*). Adapted from (Miehle et al., 2018b, Table 3), with kind permission from Association for Computing Machinery.

Cat.	Statement	M	σ	M(p)	M(np)	D	α
A	The interaction with the system is repetitive.	2.35	1.30	2.15	2.55	0.40	0.131
	The interaction with the system is boring.	1.95	1.03	1.87	2.03	0.17	0.762
	The interaction with the system is irritating.	2.13	1.16	1.68	2.58	0.90	0.000*
	The interaction with the system is frustrating.	1.82	0.94	1.43	2.20	0.77	0.000*
	The system is too inflexible.	2.54	1.22	2.12	2.97	0.85	0.000*
Н	I sometimes wondered if I was using the right word.	2.85	1.36	2.65	3.05	0.40	0.114
	I always knew what to say to the system.	3.70	1.14	3.95	3.45	0.50	0.021*
	I was not always sure what the system was doing.	2.47	1.24	2.17	2.77	0.60	0.010*
	It is easy to lose track of where you are in an interaction with the system.	2.17	1.09	1.97	2.37	0.40	0.054
SB	The system seemed polite.	4.26	0.99	4.45	4.07	0.38	0.035^{*}
	The system seemed courteous.	3.75	1.11	3.98	3.52	0.47	0.019*
	The system seemed friendly.	4.23	1.07	4.47	3.98	0.48	0.011*
	in its speaking style.	4.17	0.92	4.37	3.97	0.40	0.031*
SCS	The system provided the right amount of information.	3.34	1.16	3.83	2.85	0.98	0.000*
	I got the information I wanted.	4.08	1.04	4.40	3.77	0.63	0.000*
	I always knew what the system wanted to tell me.	4.08	1.08	4.33	3.82	0.52	0.019*

Table 5.13: The results of the participants' ratings for the categories Annoyance (A), Habitability (H), System Behaviour (SB) and System Communication Style (SCS), whereby M is the mean and σ is the standard deviation of all ratings, M(p)is the mean of the ratings of all dialogues which were declared as preferred and M(np) is the mean of the ratings of all dialogues which were declared as not preferred, D = |M(p) - M(np)| is the difference between M(p) and M(np), and α is the statistical significance level using the Mann-Whitney-U-Test. The differences that are statistically significant ($\alpha < 0.05$) are marked (*). Adapted from (Miehle et al., 2018b, Table 3), with kind permission from Association for Computing Machinery.



5.4 The Impact of Varying Communication Styles on User Satisfaction

Figure 5.17: The results of the participants' ratings grouped by the different categories of the questionnaire, whereby the rating scale has been inverted for the negatively formulated statements so that the optimal rating is five for every statement. M(p) is the mean of the ratings of all dialogues which were declared as preferred, M(np) is the mean of the ratings of all dialogues which were declared as not preferred and D = |M(p) - M(np)| is the difference between M(p) and M(np). Taken from (Miehle et al., 2018b, Figure 3), with kind permission from Association for Computing Machinery.

H For the preferred dialogue, the ratings are significantly higher that the participants knew what to say to the system and that they knew what the system was doing.

SB Regarding the system behaviour, the participants rated that the system seemed more polite, courteous, friendly and professional in its speaking style during the preferred dialogue.

SCS The ratings in the last category show that during the preferred dialogue, the system provided the right amount of information, the participants got the information they wanted and they knew what the system wanted to tell them significantly more often than during the dialogue that was not preferred.

If the rating scale for the negatively formulated statements is inverted so that the optimal rating is five for every statement, an overall score for every category of the questionnaire can be calculated by taking the average of all ratings of the corresponding category. The

	elaborate	concise	total
direct indirect	15 14	$\begin{array}{c} 17\\14 \end{array}$	$\begin{vmatrix} 32 \\ 28 \end{vmatrix}$
total	29	31	60

Table 5.14: The distribution of the communication styles of the preferred dialogues showing how many study participants preferred each option. Adapted from (Miehle et al., 2018b, Table 4), with kind permission from Association for Computing Machinery.

resulting scores are shown in Figure 5.17. Moreover, a score for the overall user satisfaction US is obtained by calculating the average of all ratings of the SASSI questionnaire which includes the categories System Response Accuracy (SRA), Likeability (L), Cognitive Demand (CD), Annoyance (A) and Habitability (H). It can be seen that the user satisfaction for the dialogues which were declared as preferred is US(p) = 4.07 whereas the user satisfaction for the dialogues which were declared as not preferred is US(np) = 3.58.

Moreover, the results of the ratings are in line with the participant's statements to the open questions they had to answer. Here, the participants commented that during the preferred dialogue, they had a more pleasant conversation and that Alexa appeared more friendly and polite. The interaction appeared more natural and more human-like. In contrast, during the disliked dialogue, Alexa was described as rude, presumptuous, irritating and rigid. Furthermore, the study participants stated that they noticed differences between the dialogues in how Alexa talked to them. The elaborate dialogues were described as more detailed, verbose and extensive, whereas the concise dialogues were characterised as succinct. The direct dialogues were described as more precise, accurate and exact, whereas indirectness was perceived as more improper.

These results show that the system's communication style indeed influences the user's satisfaction and the user's perception of the dialogue. As nothing but the system's communication style changed between the preferred and the disliked dialogue, it can be concluded that the change of the user satisfaction depends on the system's communication style. Regarding the initial question of how varying communication styles of a spoken dialogue system are perceived by the user, it can be seen that the users are more satisfied with the system if it is using the preferred communication style.

General Preferences Regarding Elaborateness and Indirectness

In order to determine any general preferences regarding the *elaborateness* and *indirectness* among the study participants, the communication styles of the dialogues that were marked as preferred have been analysed. The overall distribution of these 60 dialogues is shown in Table 5.14. It can be seen that 32 participants (53.33%) selected the direct versions, whereas 28 participants (46.66%) preferred the indirect ones. Moreover, 29 participants (48.33%) opted for the elaborate and 31 participants (51.66%) for the concise dialogues.

Dialogue A	Dialogue B	#A	#B
direct, elaborate	direct, concise	4	6
direct, elaborate	indirect, concise	5	5
direct, elaborate	indirect, elaborate	6	4
direct, concise	indirect, concise	6	4
direct, concise	indirect, elaborate	5	5
indirect, concise	indirect, elaborate	5	5

5.4 The Impact of Varying Communication Styles on User Satisfaction

Table 5.15: The distribution of preferences of all dialogue combinations showing how many study participants preferred each option. Adapted from (Miehle et al., 2018b, Table 5), with kind permission from Association for Computing Machinery.

Table 5.15 shows the exact dialogue combinations. It can be seen that there are only slight tendencies regarding the direct and concise options. Overall, there are no significant differences and no general preference is visible among the study participants. As all study participants were Germans, it can be inferred that there is no general preference related to German culture in this particular set-up. In order to determine any preference related to the user's gender, their frequency of use of speech based assistants or their age, the different sub-groups shown in Table 5.11 have been analysed. The same result is obtained: there is no general preference for any of the groups investigated within this set-up. This leads to the conclusion that it is not possible to detect any general preferences which can be integrated into a spoken dialogue system by means of predefined rules on fixed user characteristics such as the user's age, culture, gender or frequency of use of speech based assistants. In contrast, the preference appears to be individual for every person. It seems that a spoken dialogue system has to take into account more subtle characteristics such as the user's emotional state or the current situation which are not fixed in advance but have to be determined during the ongoing dialogue.

5.4.3 Conclusion

The presented study has investigated how varying communication styles of a spoken user interface are perceived by users and whether there exist global preferences in the communication styles *elaborateness* and *indirectness*. In order to do so, the participants had two conversations each with Amazon's Alexa where Alexa used varying communication styles for its output. In a post-survey, the participants had to rate statements to subjectively assess each dialogue. Moreover, they had to indicate which dialogue they preferred. The results have shown that there is no general preference in the system's communication style and therefore the preference appears to be individual for every person. Furthermore, it has been shown that the system's communication style has a direct influence on the user's satisfaction. Hence, the conclusion is drawn that a higher level of user satisfaction can be achieved if the system selects its communication style according to the user's preference. This dependency is included in the *Communication Style Perception Model*.

5.5 The Impact of Varying Communication Styles in Gamified Participatory Sensing

Since the last section has already shown that the system communication style has an influence on the user's satisfaction, the impact on the user's behaviour is investigated in the following. The fifth experiment (Kawanaka et al., 2021), which has been done in cooperation with the Ubiquitous Computing Systems laboratory at the Nara Institute of Science and Technology (NAIST) in Japan, has explored the impact of different communication styles in gamified participatory sensing for tourism.

With the spread of information technology, people can easily post their tourism experiences. This kind of information generated by consumers is accepted as an effective and reliable source of information (Sigala et al., 2012). One approach to collecting such information is participatory sensing, which uses mobile devices such as smartphones owned by the public as sensing devices. While it has the property of collecting data with high spatio-temporal resolution at low cost, the amount and quality of the data collected is dependent on the contribution of the participating users. In a pre-study (Kawanaka et al., 2020), which has also been conducted in cooperation with the Ubiquitous Computing Systems laboratory at the Nara Institute of Science and Technology (NAIST) in Japan, the effects of different task allocation interfaces on the efficiency of tourism information collection, tourism behaviour and user satisfaction have been investigated. Two types of task allocation interfaces (map-based and chat-based) were designed and implemented, and a sightseeing experiment with ten participants was conducted in Nara, Japan. The results have shown that there is no difference in the users' sightseeing satisfaction between the interfaces. However, there were differences in the data that has been collected. While there was more data collected with the map-based interface, the data collection with the chat-based interface was more efficient in collecting data for spots with higher information demand. In the subsequent experiment, different levels of *elaborateness* and *indirectness* have been introduced into the chat-based interface.

5.5.1 Design of the Task Allocation Interfaces

Both task allocation interfaces have been implemented into the participatory sensing platform application called Parmosense (Matsuda et al., 2016). The map-based interface is shown in Figure 5.18. All sightseeing spots are displayed on the map by pins which are coloured according to the information demand level (gold = high information demand, silver = medium information demand, copper = low information demand). The user can tap on the pins in order to get detailed information about the sightseeing spot and about the points that can be obtained when posting information about this spot. When arriving at the sightseeing spot, the user is able to check-in, take a photo and post it together with a comment. After completing the post, the user gets the points.

The chat-based interface is shown in Figure 5.19. The main screen (middle) shows a chat interaction with the agent who asks the user to do a mission. For the agent's recommendation, the ten closest sightseeing spots from the user's current location are



5.5 The Impact of Varying Communication Styles in Gamified Participatory Sensing

Figure 5.18: The map-based task allocation interface. Adapted from (Kawanaka et al., 2021, Figure 1), licensed under CC BY 4.0 (https://creativecommons.org/ licenses/by/4.0).

selected and sorted by the points the user can get, whereby the spot with the highest information demand gives the most points and thus is in the first place. The user then has three choices (blue buttons at the bottom of the screen): accepting to do the recommended mission, asking for alternatives and asking for details about the sightseeing spot. In addition, the user can see the location of the spot on the map by clicking the "Check Map" button. After selecting a mission, the user walks to the sightseeing spot, takes a photo and posts it together with a comment in order to get the points. In order to investigate the influence of different communication styles, different templates for the agent's recommendations have been implemented:

- Direct, elaborate (DE): Go for mission <spot name>. You can get <100> points and it is just about <5> minutes from here. It is the <closest> spot from your current location and the one where you can get the <most> points.
- Direct, concise (DC): Go for mission <spot name>. You can get <100> points and it is just about <5> minutes from here.
- Indirect, elaborate (IE): You can get <100> points for mission <spot name> which is just about <5> minutes from here. It is the <closest> spot from your current location and the one where you can get the <most> points.
- Indirect, concise (IC): You can get <100> points for mission <spot name> which is just about <5> minutes from here.



Figure 5.19: The chat-based task allocation interface. Taken from (Kawanaka et al., 2021, Figure 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

The colour of the agent's clothing changes with the communication style (DE: green, DC: yellow, IE: red, IC: blue).

In addition, both interfaces provide the possibility to make free posts about any information the participants find useful or interesting during their sightseeing tour. In order to do so, the participants have to tap the camera button at the bottom of the screen.

5.5.2 Experimental Setting

To evaluate the influence of the different levels of *elaborateness* and *indirectness*, a user study has been conducted in Nara, Japan. Overall, 108 people participated in the study. They have been hired by a recruitment company. All participants were Japanese and living outside Nara Prefecture. Before the experiment started, they have been asked for their age, gender, previous tourism experience in the experimental area, and user types based on the Gamification User Types Hexad scale (Tondello et al., 2019). The participants were aged between 19 and 71 with an average age of 41.0 years. The distribution of participants is shown in Table 5.16. 50 participants were male, 58 were female. With regard to their tourism experience in Nara, six people stated that they have never visited the area before, 28 people have visited it once, 26 people have visited it twice and 49 people have visited it three or more times. Using the the Gamification User Types Hexad scale, 49 participants
	#Participants
overall	108
male	50
female	58
never visited Nara	6
visited it once	28
visited it twice	26
visited it three or more times	49
free spirit	49
philanthropist	46
player	21
socialiser	17
achiever	14
disrupter	0

5.5 The Impact of Varying Communication Styles in Gamified Participatory Sensing

Table 5.16: The distribution of study participants regarding their gender, their tourism experience in the experimental area and the Gamification User Types Hexad scale.

are classified as free spirit, 46 people as philanthropist, 21 as player, 17 as socialiser and 14 as achiever (as some participants have multiple user types, the total amount of user types exceeds 108).

The participants have been divided into two experimental groups: the 53 participants in group A used the map-based interface, while the 55 participants in group B used the chat-based interface. The study participants' distribution of both groups can be seen in Figure 5.20.

The participants were paid 8,000 Yen (around \$80) as reward for joining the study. This amount included the transportation fee to the venue and entrance fees during their sightseeing tour. Before the participants started their sightseeing tour, they have been introduced to the experiment. The experimental application was installed on the participants' smartphone and explained with the help of documents and videos. The duration of the experiment was four hours. During that time, the participants did sightseeing around the experimental area in Nara. They were on foot and alone while accomplishing missions and earning points. After finishing their sightseeing tour, the participants were asked to complete a questionnaire about their satisfaction, their communication style preference (only group B), the usability of the application and their overall impressions of the experiment.







(b) Group B: chat-based interface

Figure 5.20: The distribution of study participants for each experimental group. Taken from (Kawanaka et al., 2021, Figure 4), licensed under CC BY 4.0 (https: //creativecommons.org/licenses/by/4.0).

	#Posts with map-based interface	#Posts with chat-based interface
overall	1,802	1,346
missions free posts	1,512 290	956 390
high demand missions medium demand missions low demand missions	$480 \\ 470 \\ 562$	532 258 166

Table 5.17: Comparison of the posts with the map-based and the chat-based interface.

5.5 The Impact of Varying Communication Styles in Gamified Participatory Sensing



Figure 5.21: Comparison of the information demand level of the completed missions with the map-based and the chat-based interface (copper = low information demand, silver = medium information demand, gold = high information demand). Adapted from (Kawanaka et al., 2021, Figure 5), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

5.5.3 Evaluation Results

Overall, 3,148 posts containing photos and comments have been collected during the experiment. In a first step, it has been evaluated how many post have been created with each interface. The results are shown in Table 5.17. It can be seen that, overall, 1,802 posts were obtained with the map-based interface, and 1,346 posts were obtained with the chat-based interface. While there were more missions completed with the map-based interface, the participants made more free posts with the chat-based interface. When comparing how many missions have been completed by each participant using the Mann-Whitney-U-Test, there is a statistically significant difference. The comparison of the information demand level of the sightseeing spots that have been visited, which is shown in Figure 5.21, shows that most of the missions that have been completed with the map-based interface (37%) have a low information-demand level, while the majority of the missions that have been completed with the chat-based interface (37%) have a low information-demand level, while the majority of the missions that have been completed with the chat-based interface (37%) have a low information-demand level, while the majority of the missions that have been completed with the chat-based interface (36%) have a high information demand level. These findings are in line with the result of the pre-study (Kawanaka et al., 2020). When comparing the different communication styles of the chat-based interface, there is no significant difference between the missions that have been completed.

In the next step, the results of the questionnaire have been evaluated. The general part that had to be answered by all participants contained the following items (translated from Japanese as the original questionnaire has been provided in the native language of the participants):

- 5 The Communication Style Perception Model
- **Q1** Did you prioritise sightseeing or completing the mission? (1 = sightseeing, 5 = mission)
- **Q2** Did the application make you have fun doing the sightseeing tour? (1 = not at all, 5 = a lot of fun)
- **Q3** Did you like the interface? (1 = don't like it at all, 5 = like it very much)

Furthermore, the participants of group B were asked about their communication style preferences:

- Q4 Did you notice any differences in the agent's interaction? (yes/no)
- Q5 Which communication style did you like most? (selection based on screenshots)

The evaluation of the questionnaires yielded the following results:

Q1 The mean M and the median Md for the map-based interface are M(m) = 3.44 and Md(m) = 4.00, the results for the chat-based interface are M(c) = 4.04 and Md(c) = 4.00. These results show that both groups prioritised the mission over sightseeing. However, the difference between M(m) and M(c) is still statistically significant (p = 0.04 using the Mann-Whitney-U-Test), showing that the participants tended to prioritise the mission significantly more often with the chat-based interface. As main reason for prioritising the mission, the participants stated that they wanted to earn as much points as possible and saw their sightseeing tour as a game. Moreover, for the chat-based interface, the participants stated that they followed the recommendations of the agent which were based on the missions.

Q2 The results for the map-based interface are M(m) = 3.87 and Md(m) = 4.00, the results for the chat-based interface are M(c) = 3.58 and Md(c) = 4.00, showing that the participants liked both interfaces. There is no significant difference between the map-based and the chat-based interface (p = 0.10 using the Mann-Whitney-U-Test). The main reason given by the participants was that they visited places they did not know or they have not been aware of before using the application.

Q3 The results for the map-based interface are M(m) = 3.69 and Md(m) = 4.00, the results for the chat-based interface are M(c) = 3.29 and Md(c) = 3.00. There is a statistically significant difference between M(m) and M(c) (p = 0.03 using the Mann-Whitney-U-Test), showing that the participants prefer the map-based interface over the chat-based interface.



Figure 5.22: The result of Q4, showing how many male/female participants noticed a difference in the agent's interaction. Taken from (Kawanaka et al., 2021, Figure 10), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

Q4 26 participants (47%) noticed a difference in the agent's interaction, while 29 participants (53%) did not notice any difference. The gender distribution is shown in Figure 5.22. This result indicates that the female participants were more sensitive to the changes in the agent's communication style. However, there is no significant difference.

Q5 Figure 5.23 shows the communication style preferences of all participants (N = 55) and Figure 5.24 shows the communication style preferences of all participants who noticed a difference in the agent's interaction (N = 26). It can be seen that there is a clear preference in the *elaborateness* dimension, both when looking at all participants and when looking at only those participants who noticed a difference in the agent's interaction. A Chi-Squared Test shows that the difference between elaborate and concise is statistically significant for both groups (p < 0, 05). Hence, the participants significantly preferred the elaborate communication style over the concise one. This shows that, in general, more detailed information is preferred for the task at hand. However, it has to be noted that this does not apply to all participants. Some of them clearly stated that they find a short and simple text easier to read on the smartphone while walking. Moreover, there is no preference for the *indirectness* dimension. This is in line with the results presented in Section 5.4, that there is no general preference in the system's communication style and therefore the preference appears to be individual for every person.



Figure 5.23: The result of Q5, showing the communication style preferences of all participants (N = 55). Taken from (Kawanaka et al., 2021, Figure 11), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).



Figure 5.24: The result of Q5, showing the communication style preferences of all participants who noticed a difference in the agent's interaction (N = 26). Taken from (Kawanaka et al., 2021, Figure 12), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

5.6 The Impact of Varying Communication Styles in a Behaviour Change Support System

5.5.4 Conclusion

The presented user study has explored the impact of different communication styles in gamified participatory sensing for tourism. In order to do so, different levels of *elaborateness* and *indirectness* have been implemented in a chat-based interface. 47% of the participants noticed a difference in the agent's communication style, showing that a large number of participants were aware of these subtle changes. However, even if the results have shown that there are clear preferences in the communication styles, both when looking at all participants and when looking at only those participants who noticed a difference in the agent's interaction, no impact on the amount or type (in terms of the information demand level) of collected data could be detected.

5.6 The Impact of Varying Communication Styles in a Behaviour Change Support System

The sixth experiment (Zhang et al., 2021), which has also been done in cooperation with the Ubiquitous Computing Systems laboratory at the Nara Institute of Science and Technology (NAIST) in Japan, has explored the impact of different communication styles for interpretations of quantitative data in a behaviour change support system. Numerous technologies exist for promoting a healthier lifestyle. However, while the majority of the existing apps use a quantitative data representation, it has been shown that this approach might harm the users' motivation and lead to a failure of promoting behaviour change since it is hard to understand the meaning behind the data (Daskalova et al., 2017; Lupton, 2013; Crum and Langer, 2007). Therefore, it is necessary to provide the interpretation of the quantitative data as supplement. However, different descriptions of the same data may lead to different outcomes. This is why the web-based application called Walkeeper, which provides interpretation of the users' daily amount of steps using different levels of *elaborateness* and *indirectness* to promote the users to walk, has been developed and evaluated.

5.6.1 Walkeeper Prototype Design

In cooperation with the Ubiquitous Computing Systems laboratory at the Nara Institute of Science and Technology (NAIST), the web-based application called Walkeeper has been developed. It provides interpretations of the users' daily amount of steps based on different levels of *elaborateness* and *indirectness* to examine how the communication style of the interpretation affects the walking performance of the users. As can be seen in Figure 5.25, Walkeeper consists of three parts. The activity tracker tracks the users' activity and sets the daily step goal for each user. Based on the data collected by the activity tracker, the contents manager decides about the contents to be provided to the users including the content of the interpretation, the face expression of the emoji, the graph of history records, and the group ranking table. The progress reminder sends messages to the users before lunch time and off hours every week day, trying to promote the users to walk more



Figure 5.25: The web-based application called Walkeeper consists of three parts: the activity tracker, the contents manager and the progress reminder. Taken from (Zhang et al., 2021, Figure 1), licensed under CC BY 4.0 (https: //creativecommons.org/licenses/by/4.0).

by reminding them of their walking progress. In the following, the main features are described.

Activity Tracking and Goal Setting

To track the number of steps of each user, the mobile application called Walkus⁴ has been used, which is able to track the users' activity and upload the data to its cloud server. The activity tracker accesses the API of Walkus every five minutes to get the data at that time via the users' Walkus IDs, and stores the data in the Walkeeper database. Afterwards, the activity tracker calculates the average amount of steps of the whole group and stores it into the database as the group performance at that time. At the end of each day, the activity tracker checks whether the amount of steps of each user has reached the daily goal. If the user has reached the daily goal, the activity tracker increases the goal for the next day by 500. Otherwise, the activity tracker keeps the current goal for the next day.

Progress Reporting

Walkeeper provides a unique URL for each user so that the users can check their walking progress whenever they want by accessing the URL. After accessing the URL, each user can see the interface of Walkeeper, containing an interpretation of their current amount of steps, an emoji face expression, a graph of history records, and a ranking table, as can

 $^{{}^{4}\}mathrm{App}\ \mathrm{Store:}\ \mathtt{https://apps.apple.com/us/app/walkus/id1273735006}$





Figure 5.26: The interface of Walkeeper. Adapted from (Zhang et al., 2021, Figure 2), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/ 4.0).

be seen in Figure 5.26. The contents manager decides which contents are provided to the user when opening Walkeeper. Every time when a user accesses the URL, the contents manager performs the following steps: (1) identifying the user by the URL, (2) calculating the walking progress of the user, (3) checking the communication style setting for the user, (4) formulating the interpretation based on the communication style and the walking progress, (5) selecting the face expression of the emoji, the graph of history records, and the ranking table based on the walking progress of the user and the group performance. Walkeeper provides four different communications styles for both the situation that the user has reached the goal and the situation that the user has not reached the goal:

- The user has reached the goal:
 - Direct, elaborate (DE): You have reached your daily goal. You already did
 <X> steps today. Well done! You improve your health with physical activity.
 - Direct, concise (DC): You have reached your daily goal. Well done!
 - Indirect, elaborate (IE): You did it. <X> steps are fantastic! Your health is benefiting from physical activity.
 - Indirect, concise (IC): You did it. Fantastic!
- The user has not reached the goal:
 - Direct, elaborate (DE): You haven't reached your daily goal yet. You only did
 <X> steps. Come on! You need <Y> steps to reach your goal.

- Direct, concise (DC): You haven't reached your daily goal yet. Come on!
- Indirect, elaborate (IE): You still need to do some steps today. <X> steps are a good start. Get going! <Y> are all that is needed.
- Indirect, concise (IC): You still need to do some steps today. Get going!

The interpretation is displayed as a speech bubble of an emoji, as shown in Figure 5.26a. Eleven different emoji face expressions are used in order to represent the user's walking progress, as can be seen in Figure 5.26b. To maintain the users' motivation, Walkeeper also provides a graph showing the daily step records, the daily step goal, the average amount of daily steps of the whole group of the past seven days (see Figure 5.26c), and an overall ranking table containing the user's rank, nickname, and current amount of steps (see Figure 5.26d). For Japanese speaking participants, all contents are provided in Japanese.

Progress Reminding

The progress reminder sends messages to users two times per day to remind the users of their walking progress. In order to do so, a chat bot has been developed in Workplace Chat⁵. The message sent by the progress reminder contains a text message ("Please access the URL and check your performance.") and the unique URL for each user. For Japanese speaking participants, the reminder is provided in Japanese. The first message is sent at 11:30 am, the second one is sent at 16:30 pm. Hence, the users get their walking progress before lunch time and before leaving the office. In this way, they can decide to have an active lunch break (e.g. use the stairs instead of the elevator when going to the canteen or take a walk after having lunch) or plan to walk more in the evening (e.g. walk to the train station instead of taking the bus or do some sport after arriving at home).

5.6.2 Experimental Setting

Walkeeper has been evaluated during a six-week user study in order to explore the following research questions: (1) How do people react to Walkeeper? (2) How do different communication styles affect the people's attitude and behaviour?

24 members of the Ubiquitous Computing Systems laboratory at the Nara Institute of Science and Technology (NAIST) participated in the evaluation (21 male, 3 female). They were aged between 22 and 52 with an average age of 29 years. 18 participants were Japanese speakers, the remaining 6 participants were English speakers. After the participants provided consent, they were introduced to the experiment without mentioning the experimental details (communication styles, contents). Moreover, they got help with the installation and set up of Walkus, Workplace and Walkeeper.

During the first two weeks, it was ensured that the participants' steps were tracked continuously. Moreover, the overall activity level of the participants was investigated in order to set an appropriate daily goal for each participant and to obtain a comparative

⁵https://www.workplace.com

Team	#Participants	Week 1	Week 2	Week 3	Week 4
1	7	DE	IC	IE	DC
2	6	DC	IE	IC	DE
3	4	IE	DC	DE	IC
4	7	IC	DE	DC	IE

5.6 The Impact of Varying Communication Styles in a Behaviour Change Support System

Table 5.18: The participants have been assigned randomly into one team. Adapted from (Zhang et al., 2021, Table 3), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

value for the evaluation. During this experimental stage, Walkeeper did not send any messages or provide URLs to the participants to check their progress, but only collected the step data. Using this data, the average amount of steps was calculated and set as the initial daily step goal for each participant.

Afterwards, a four-week study has been conducted. The participants have been assigned randomly into one of four teams, whereby each team head a different communication style in each week, as can be seen in Table 5.18. To ensure the same conditions for the data collection during the course of the study, Walkeeper interacted with the participants only on the working days (Monday to Friday). Moreover, the daily step goal of each participant was reset to the initial daily step goal at the beginning of each week so that the starting point was identical for each week. At the end of each week, the participants completed a short survey about their opinions on this week's interface and interpretations. Additionally, analytic data about the timing of accessing Walkeeper and the duration of browsing the web page of Walkeeper has been captured. At the end of the study, the participants were asked to complete a final questionnaire about their overall experience with the system. Afterwards, the participants received a 2,000 Yen (around \$19) gift card as reward for joining the study.

5.6.3 Evaluation Results

The participants' overall impressions about the system were obtained by use of the weekly questionnaires which contained statements that had to be rated on a five-point Likert scale (1 = fully disagree, 5 = fully agree). Table 5.19 shows the results grouped by communication styles. It can be seen that the different communication styles did not have a significant influence on the ratings. For each communication style, the participants rated that the system was polite and friendly. The system provided the right amount of information and the information was precise. The participants had the feeling that they got the information they wanted and that they always knew what the system wanted to tell them. Moreover, the participants were asked whether they have noticed any changes in the interpretations. Out of the 24 participants, 18 stated that they have noticed differences. They reported that they noticed a change in whether the exact numbers have been mentioned or not and a change in the tone of the message.

Statement	Metric	DE	DC	IE	IC
The system seemed polite.	Median Average	4 3.81	4 3.73	4 3.65	$4 \\ 3.69 \\ 0.74$
The system seemed unfriendly.	Median Average SD	$ \begin{array}{r} 2 \\ 2.23 \\ 0.99 \end{array} $	2 2.15 1.05	2 2.04 1.08	2 2.00 0.75
The system seemed professional.	Median Average SD	3 3.00 0.80	3 3.15 0.83	3 3.00 1.10	3 3.00 0.94
The system provided more information than I needed.	Median Average SD	3 2.85 0.83	$3 \\ 2.62 \\ 1.02$	2 2.42 0.99	3 2.69 1.09
The system provided the right amount of information.	Median Average SD	4 3.54 0.99	4 3.50 0.96	4 3.50 1.10	4 3.46 1.03
I would have preferred to get more details.	Median Average SD	$3.5 \\ 3.58 \\ 0.95$	3 3.27 1.12	3.5 3.46 1.24	3.5 3.38 1.17
I got the information I wanted.	Median Average SD	3 3.38 0.94	4 3.54 0.99	4 3.38 0.85	3 3.19 1.06
I always knew what the system wanted to tell me.	Median Average SD	4 4.19 0.75	4 4.23 0.82	4 4.12 0.99	4 4.04 0.82
The system provided imprecise information.	Median Average SD	2 2.12 0.86	1 1.96 1.09	2 2.42 1.30	$2 \\ 2.00 \\ 1.02$

Table 5.19: The results of the weekly questionnaires containing statements that had to be rated on a five-point Likert scale (1 = fully disagree, 5 = fully agree). Adapted from (Zhang et al., 2021, Table 7), licensed under CC BY 4.0 (https: //creativecommons.org/licenses/by/4.0).

	Sample Size	Average SIR	<i>p</i> -value (ANOVA)
DE	96	4.82%	
DC	99	5.07%	0.017*
IE	103	32.79%	0.017
IC	102	13.38%	

5.6 The Impact of Varying Communication Styles in a Behaviour Change Support System

In order to find out whether there exists a correlation between the amount of steps and the communication style, a statistical analysis has been conducted. To ensure the reliability of the data, incorrect data due to technical issues has been filtered out according to the following conditions:

- 1. The daily amount of steps has to be at least 500.
- 2. For each participant, at least three data records have to be available for the first experimental phase (i.e. the average amount of steps for this phase is reliable).

Any data that did not meet these conditions was excluded in the subsequent analysis. After data filtering, a total of 89 data records have been obtained for the first experimental phase and 400 for the second experimental phase (DE: 96, DC: 99, IE: 103, IC: 102). In order to compare the participants' amount of steps, each participant's average amount of steps during the first experimental phase M and each participant's step increase rate (SIR) for each day of the second experimental phase has been calculated:

$$SIR = \frac{\#DailySteps - M}{M} \tag{5.1}$$

It represents the change of the participants' amount of daily steps compared to their average amount of steps before getting reminders by the Walkeeper system. On days with a positive step increase rate, the user walked more than the average amount of steps during the first experimental phase; when the increase rate is negative, the participant walked less than the average amount of steps during the first experimental phase.

The average step increase rate for each communication style can be seen in Table 5.20. A positive step increase rate has been obtained for each communication style, showing that the participants overall reacted positively to the Walkeeper system and walked more during the second experimental phase than during the first experimental phase when only the steps have been tracked but Walkeeper did not send any messages or provide URLs to the participants to check their progress. The highest average step increase rate of 32.79% has been obtained for IE, whereas the lowest average step increase rates of 4.82% and 5.07% have been obtained for DE and DC.

Table 5.20: The results of the one-way ANOVA showing that there are significant differences between the communication styles. Adapted from (Zhang et al., 2021, Table 4), licensed under CC BY 4.0 (https://creativecommons.org/ licenses/by/4.0).

DE	DC	IE	IC
-	1.000	0.030*	0.833
1.000	-	0.030^{*}	0.842
0.030^{*}	0.030^{*}	-	0.209
0.833	0.842	0.209	-
	DE - 1.000 0.030* 0.833	DE DC - 1.000 1.000 - 0.030* 0.030* 0.833 0.842	DE DC IE - 1.000 0.030* 1.000 - 0.030* 0.030* 0.030* - 0.833 0.842 0.209

Table 5.21: The results of the Tukey post-hoc analysis showing that there are significant differences between DC and IE and between DE and IE. Adapted from (Zhang et al., 2021, Table 5), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0).

A one-way ANOVA has been conducted to assess the effects of the different communication styles. The results of the statistical analysis show that there are statistically significant differences between the communication styles (p = 0.017). The Tukey post-hoc analysis revealed a significant difference between DC and IE (p = 0.030) and between DE and IE (p = 0.030), as can be seen in Table 5.21.

To further explore the impact of the different communication styles on the participants' behaviour, in the following the two dimensions *elaborateness* and *indirectness* are considered independently. Using a T-Test, it is shown that the participants had a significantly higher step increase rate with an indirect communication style than with a direct communication style (p = 0.011), but there was no significant difference between the elaborate and concise communication style (p = 0.164).

These results lead to the conclusion that the communication style of the interpretations of the Walkus system influences the behaviour of the user. Among the four communication styles, the indirect and elaborate one had the highest step increase rate. This reflects two issues that have also been emphasised by the users in the questionnaire: (1) In comparison to the concise style, the elaborate one includes some more details (e.g. the exact amount of steps that has already been achieved at the current time and some information about the benefits of being more active and doing more sports). This helps the users to better understand their progress and to have a clearer motivation to achieve their goal. (2) The indirect style is usually described as being more polite than the direct style as it uses a softer tone and does not concretely address (negative) information. If the Walkus system has to address the issue that the user does not move enough, it is easier for the user to accept it if the system does not mention it directly. If these aspects are considered when designing interpretations of quantitative data for a behaviour change support system, the user's behaviour change can be influenced and it can be achieve that the user moves more.

5.6.4 Conclusion

The presented behaviour change support system called Walkeeper provides interpretations of the users' daily amount of steps using different levels of *elaborateness* and *indirectness*. It has been evaluated in a six-week user study with 24 participants. The results have shown

that Walkeeper has a positive effect on increasing the users' amount of steps. Moreover, the communication style of the interpretations of the quantitative data has an impact on the users' performance. The indirect and elaborate communication style has led to the highest step increase rate (32.79%), while the direct and elaborate communication style and the direct and concise communication style were related to the lowest step increase rates. Therefore, it is concluded that the communication style of the system has a direct influence on the user's behaviour. This dependency is included in the *Communication Style Perception Model*.

5.7 Summary

In this chapter, the *Communication Style Perception Model* (see Figure 5.1) has been presented. It covers elements that influence the user's preference in the system communication style, thus being relevant for the selection of the system communication style, as well as aspects that are influenced by the system communication style. In order to substantiate the relations in the model, different user evaluations were conducted, examining various user groups in different scenarios.

The first experiment has examined whether the cultural background of the user, which is one specific user trait, influences his or her communication style preference in humancomputer interaction. In order to do so, it has investigated whether cultural communication idiosyncrasies found in human-human interaction may also be observed during human-computer interaction in a spoken dialogue system context. The results of the study have shown that there are differences between Germans and Japanese concerning their preferences in the communication patterns of the system output. Hence, it is concluded that the cultural background of the user has a direct influence on his or her preference in the system communication style. Further influencing variables were investigated in the second experiment. In a multicultural study, the influence of the user's culture and gender, the frequency of use of speech based assistants as well as the system's role on differences in the user's communication style preferences in human-computer interaction have been explored. The results have shown that the user's culture significantly influences the user's preference in the system's communication style even though five European cultures were examined whose communication styles are very alike. Depending on the culture, there are also gender differences. The user's frequency of use of speech based assistants seems to have no influence, but the system's role significantly influences the user's preference in the *indirectness* and *elaborateness* of the system. The third experiment has shown that the system's role has an impact on user preferences that goes beyond the communication styles *indirectness* and *elaborateness*. Summarising the results of these experiments, it is concluded that both user traits and system properties influence the user's communication style preferences in human-computer interaction. This forms the first part of the Communication Style Perception Model.

The remaining experiments investigated how varying system communication styles affect the users, if they are selected according to the users' personal preferences. To examine this, different communication styles have been included in various systems and applica-

tions. The results have shown that the system communication style influences the user's satisfaction and the user's perception of the dialogue. For specific applications like behaviour change support systems, the communication style even has an impact on the user's behaviour. Thus, it is concluded that the system communication style has a direct influence on the user's satisfaction and behaviour, which forms the second part of the *Communication Style Perception Model*.

Although the model makes no claim to be complete in its current state, it already represents a variety of important relationships investigated in this thesis and can be supplemented based on new results and dependencies. The presented *Communication Style Perception Model* serves as the basis for the technical realisation and evaluation of the communication style adaptation in the following chapter, which also demonstrates the practical applicability of the model.

6 Statistical Classification Approaches to Communication Style Adaptation

The previous chapter has demonstrated that communication styles play an important role in human-computer interaction and influence the user's satisfaction as well as the user's perception of the dialogue. As the results have shown that there is no general preference in the system communication style, but the preference appears to be individual for every person, a solution has to be found to adapt the system communication style to each user individually during every dialogue. This chapter explores the research question of how communication styles can be integrated into spoken dialogue systems, addressing the implementation and evaluation of the targeted communication style adaptation. Based on the Communication Style Perception Model, which has been developed in Chapter 5, statistical classification approaches are pursued to enable communication style adaptation in spoken dialogue systems. In order to realise these classification approaches, a corpus with annotated communication styles is required. Since there are no publicly available corpora with annotated communication styles, a suitable corpus is created in Section 6.1. With the aid of this corpus, further investigations into the interrelationships between the communication style of the user and the communication style of the system are carried out in Section 6.2. This results in an extension of the Communication Style Perception Model by the correlation between the user and the system communication style. Based on this, two statistical classification approaches are presented. The first one is described in Section 6.3 and targets the estimation of the user communication style, which can then be used in the dialogue management to adapt the system behaviour to the user. The second classification approach is presented in Section 6.4 and addresses the task of automatically selecting the system communication style during an ongoing interaction with a spoken dialogue system, which is part of the dialogue management so that it not only decides what is said next, but also how.

6.1 Corpus Creation

In order to realise the classification-based communication style adaptation, a corpus with annotated communication styles is required. Since there are no publicly available corpora with annotated communication styles, a suitable corpus¹ has been created with annotations for the *elaborateness* and *indirectness* for each user and system dialogue act (Miehle et al., 2020; Miehle et al., 2021c; Miehle et al., 2022). For the communication style

¹The dialogues have been recorded within the KRISTINA project (http://kristina-project.eu). Unfortunately, it is not possible to publish the corpus due to privacy reasons.

6 Statistical Classification Approaches to Communication Style Adaptation

annotations, the definitions presented in Section 2.2 have been used.

The data set is based on recordings on health care topics containing spontaneous interactions in dialogue format between two participants: one is taking the role of the system while the other one is taking the role of the user of that system. The audio and video files of each participant have been recorded using the open-source Social Signal Interpretation (SSI) framework (Wagner et al., 2013). The raw log files containing the text of the audio files' utterances have been transferred to structured spoken dialogues with sequential dialogue acts. Each dialogue has been allocated with a unique dialogue ID and each dialogue act has been assigned the following attributes:

- A dialogue act number
- A participant
- A speaker
- A predefined dialogue act
- The utterance in the original language

The dialogue act number counts from 1 to n starting with the first dialogue act, n = the number of dialogue acts of the respective dialogue. The participant specifies the two roles system and user and the speaker indicates which of the predefined speakers was talking. Each speaker is identified by an anonymous speaker ID and a separate table contains profile information about each speaker, including the gender, the culture, the age, the country of birth and the current country of residence. Each dialogue turn contains one or more user dialogue acts followed by one or more system dialogue acts. These dialogue acts are chosen out of a set of 47 distinct dialogue acts which have been predefined. A list of all dialogue acts can be found in Table 6.1 and an example dialogue with the annotated dialogue acts is shown in Table 6.2. Along with the dialogue acts, the respective utterances (in the original language) and the topics being talked about (in English) are also added to the data set. Moreover, for each dialogue, the system's role is specified. The available system roles are defined as *social companion*, *nursing assistant* and *health expert*. Overall, the corpus covers 258 dialogues containing 2,880 turns and 7,930 annotated dialogue acts. The dialogues are in four different languages: German, Polish, Spanish and Turkish. The language distribution is shown in Table 6.3. It can be seen that the distribution of dialogue acts per dialogue (DA/D) varies among the languages: while German and Turkish are very similar, there is a difference compared to Polish and Spanish. This is the case even though the task and the familiarity between the speakers have been identical for the different languages. The pairs of speakers did not know each other and were swapped (i.e. speaker A did not always talk to speaker B, but also to other speakers). Hence, we conclude that the differences in the distribution are due to differences in the languages/cultures.

Each dialogue act has been annotated with the two communication styles *indirectness* and *elaborateness*. Both are assigned scores between 1 and 5 which have been defined as follows: 1 means that the utterance is extremely direct/concise, i.e. the speaker used

Dialogue acts				
Accept	PersonalApologise			
Acknowledge	PersonalGreet			
Advise	PersonalSayGoodbye			
AfternoonGreet	PersonalThank			
AfternoonSayGoodbye	ReadNewspaper			
AnswerThank	Reject			
AskMood	RepeatPreviousUtterance			
AskPlans	RephrasePreviousUtterance			
AskTask	Request			
AskWellBeing	RequestAdditionalInformation			
CheerUp	${ m Request Missing Information}$			
Console	RequestNewspaper			
Declare	Request Reason For Emotion			
EveningGreet	RequestRepeat			
EveningSayGoodbye	Request Rephrase			
ExplicitlyConfirmRecognisedInput	RequestWeather			
Implicitly Confirm Recognised Input	ShareJoy			
$\label{eq:constraint} Individual is tically Oriented Motivate$	ShowWeather			
MeetAgainSayGoodbye	SimpleApologise			
MorningGreet	SimpleGreet			
MorningSayGoodbye	SimpleMotivate			
Obligate	SimpleSayGoodbye			
Order	SimpleThank			
PersonalAnswerThank				

Table 6.1: List of dialogue acts. Taken from (Miehle et al., 2022, Table 3), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

Participant	Utterance	Dialogue act
User	Hello Kristina.	PersonalGreet
System	Hi, nice to meet you.	SimpleGreet
User	I have got a problem.	Declare
System	You can ask me everything.	Declare
User	Okay.	Acknowledge
	I have heard that old people should drink a lot	Declare
	and I have the impression that my father does	
	not drink enough.	
System	How much does he drink on average every day?	RequestMissingInformation
User	In the morning he likes a cup of coffee, at	Declare
	lunchtime I always give him a glass of water or	
	sometimes juice, but when I come back after he	
	has eaten, there's usually still half of it in the	
	glass. Tes, and in the evening, sometimes a glass	
System	Do you have the impression that he has yory dry	Boquest MissingInformation
System	skin?	requestionsingrinormation
User	Not really.	Reject
	Well. I always put cream on his hands and face	Declare
	and also on his head, so I couldn't say that now,	
	but I would have to see how his arms and legs	
	are.	
	Would that be a sign that he is not drinking	Request
	enough?	
System	It is a sign that he is not drinking enough.	Accept
User	Okay.	Acknowledge
System	Is he very dry in the mouth?	RequestMissingInformation
User	I've never paid attention to that before.	Declare
System	It is said that old people should drink about 1.3	Declare
Usor	Ittes a day. I den't think he drinks that much	Doclaro
System	Try to give him something he enjoys drinking	Advise
User	Okay I will try that	Acknowledge
System	It might also be nice if you drink together. For	Advise
System	example, you could have a cup of tea together in	
	the afternoon.	
User	That's a great idea!	Acknowledge
	Thank you very much!	SimpleThank
System	You're welcome.	AnswerThank
User	See you next time!	MeetAgainSayGoodbye
System	Bye.	SimpleSayGoodbye

Table 6.2: Example dialogue with annotated dialogue acts. Adapted from (Miehle et al., 2022, Table 4), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

6.1 Corpus Creation

	D	DA	DA/D
German	135	4,887	36.20
Spanish	52	1,002	19.27
Polish	42	1,017	24.21
Turkish	29	$1,\!024$	35.31
Overall	258	7,930	30.74

Table 6.3: Language distribution of the dialogues in the annotated corpus, whereby D is the number of dialogues and DA is the number of dialogue acts. Taken from (Miehle et al., 2022, Table 5), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

		Elab	pratenes	s			Indire	ctness		
	1	2	3	4	5	1	2	3	4	5
German	1,782	1,850	795	312	148	3,825	840	142	78	2
Spanish	295	242	139	118	208	681	296	8	17	0
Polish	273	383	198	95	68	744	249	4	20	0
Turkish	323	391	179	76	55	777	216	12	19	0
Overall	$2,\!673$	2,866	1,311	601	479	6,027	$1,\!601$	166	134	2

Table 6.4: Class distribution of the annotated *elaborateness* and *indirectness* scores (median of the three ratings). Adapted from (Miehle et al., 2022, Table 6), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

the most direct/concise option to give the requested information. For the *indirectness* dimension, this means that the information is conveyed very concretely and the listener can understand it literally and does not have to imply anything. For the *elaborateness* dimension, this means that only the most important information is given by use of as few words as possible. For example, a response to the question about tomorrow's weather forecast rated with 1 for the *indirectness* and *elaborateness* would be: "It will rain." The higher the rating for the *indirectness*, the more hidden are the intentions of the speaker (2 = slightly indirect, 5 = extremely indirect). The higher the rating for the *elaborateness*, the more additional information is given (2 = slightly elaborate, 5 = extremely elaborate). For instance, an indirect response to the question about tomorrow's weather forecast would be an advice to take an umbrella, and an elaborate response would result in providing the weather forecast for the next few days.

Each dialogue act has been annotated by three different raters. They have been instructed with annotated sample dialogues. Moreover, uncertainties have been discussed in a weekly meeting. The class distribution of the annotated *elaborateness* and *indirectness* scores (median of the three ratings) is shown in Table 6.4. It can be seen that the

Elaborateness (5 classes)						
	R1/R2	R1/R3	R2/R3	Av.		
κ	0.560	0.515	0.516	0.530		
ho	0.848	0.813	0.799	0.820		
ICC				0.934		
	Elabora	ateness (3	classes)			
	R1/R2	R1/R3	R2/R3	Av.		
κ	0.670	0.612	0.608	0.630		
ho	0.826	0.794	0.767	0.796		
ICC				0.916		
	Indire	ctness (5	classes)			
	R1/R2	R1/R3	R2/R3	Av.		
κ	0.315	0.423	0.368	0.369		
ho	0.387	0.504	0.442	0.444		
ICC				0.686		
	Indire	ctness (3	classes)			
	R1/R2	R1/R3	R2/R3	Av.		
κ	0.335	0.439	0.382	0.385		
ho	0.387	0.504	0.441	0.444		
ICC				0.695		
Indirectness (2 classes)						
	R1/R2	R1/R3	R2/R3	Av.		
κ	0.376	0.499	0.440	0.438		
ho	0.377	0.500	0.440	0.439		
ICC				0.701		

Table 6.5: Agreement (κ), correlation (ρ) and reliability (*ICC*) in *elaborateness* and *indirectness* of the three ratings (R1, R2, R3). All results are significant at the 0.001 level. Taken from (Miehle et al., 2022, Table 7), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

	Elaborateness	Indirectness
	Mean/Median	Mean/Median
κ	0.960	0.849
ρ	0.993	0.912

Table 6.6: Agreement (κ) and correlation (ρ) between the mean and the median of the three ratings for the *elaborateness* and the *indirectness*. All results are significant at the 0.001 level.

classes 1 and 2 are the most common for both the *elaborateness* and the *indirectness*. The classes 3, 4, and 5 contain utterances which are elaborate/indirect to a greater or lesser extent and the weekly meetings with the annotators revealed that it is quite hard to distinguish between different levels of *elaborateness* and *indirectness*. Hence, we have combined the classes 3, 4, and 5 to one new class, reducing the corpus to three classes. For the *indirectness*, the annotation has shown that it even makes sense to see it as a binary decision between direct/indirect utterances. As the classes 2-5 contain different degrees of *indirectness* (from slightly indirect to extremely indirect), we additionally combined these classes to one indirect class for binary classification.

In order to analyse the quality of the annotated scores, the following measures have been used: Cohen's Kappa κ , Spearman's rank correlation coefficient Rho ρ and the Intraclass Correlation Coefficient *ICC*. The results can be seen in Table 6.5. The original ratings (five classes) achieve an overall inter-rater agreement of $\kappa = 0.53$ for *elaborateness* and $\kappa = 0.37$ for *indirectness*, a correlation of $\rho = 0.82$ for *elaborateness* and $\rho = 0.44$ for *indirectness* and a inter-rater reliability of *ICC* = 0.93 for *elaborateness* and *ICC* = 0.69 for *indirectness*. If the classes are reduced to three or two (in case of the *indirectness*), a higher agreement is obtained while the correlation and the inter-rater reliability do not change significantly. Overall, it is a good inter-rater reliability for both communication styles given the difficulty of the annotation task.

To use the communication style annotations as target for the classification tasks, a final score has to be calculated from the three ratings. Typical candidates are the mean and the median. To decide which one to use, the mean and the median of the three ratings have been calculated and it has been analysed which of the two metrics better reflects the individual ratings. The comparison of the mean and the median in terms of Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ is shown in Table 6.6. A strong correlation ρ can be observed. The values of κ indicate a high level of agreement, but still there seem to be some differences, especially for the *indirectness*. Therefore, the agreement and correlation between the mean and the median of the three ratings have been compared with the individual ratings (see Table 6.7). The values of κ indicate that the median better reflects the individual ratings, which might be explained by the fact that an ordinal scale has been applied. Therefore, the median of the ratings is used in the following sections.

Elaborateness					
	M/R1	M/R2	M/R3	Av.	
κ	0.756	0.725	0.691	0.724	
ρ	0.926	0.925	0.896	0.916	
	Md/R1	Md/R2	Md/R3	Av.	
κ	0.787	0.751	0.713	0.750	
ρ	0.930	0.922	0.889	0.914	
		Indirectr	ness		
	M/R1	Indirectr M/R2	${ m mess}$ M/R3	Av.	
 κ	M/R1 0.570	Indirectr M/R2 0.542	ness M/R3 0.626	Av. 0.579	
 ρ	M/R1 0.570 0.701	Indirectr M/R2 0.542 0.706	ness M/R3 0.626 0.749	Av. 0.579 0.719	
κ ρ	M/R1 0.570 0.701 Md/R1	Indirectr M/R2 0.542 0.706 Md/R2	ness M/R3 0.626 0.749 Md/R3	Av. 0.579 0.719 Av.	
κ ρ κ	M/R1 0.570 0.701 Md/R1 0.661	Indirectr M/R2 0.542 0.706 Md/R2 0.596	ness M/R3 0.626 0.749 Md/R3 0.714	Av. 0.579 0.719 Av. 0.657	

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.7: Agreement (κ) and correlation (ρ) between the mean (M) and the median (Md) of the three ratings for the *elaborateness* and the *indirectness* and the individual ratings (R1, R2, R3). All results are significant at the 0.001 level.

6.2 Investigating the Correlation between User and System Communication Style

With the aid of the corpus described in Section 6.1, further investigations into the interrelationships between the communication style of the user and the communication style of the system are carried out in this section (Miehle et al., 2022). In order to do so, it is analysed whether the communication style of the speaker who assumed the role of the system (hereafter referred to as system) is correlated with the communication style of speaker who assumed the role of the user (hereafter referred to as user). The purpose of this is to find out whether the system should take into account the user's communication style when selecting its communication style. Section 2.3 provided evidence that humans adapt their communication styles during an interaction. However, it is to be shown that this also applies to the current setting. In order to do so, the 2,880 user-system exchanges (i.e. the single turns where the system responds to a user inquiry) and the respective *elaborateness* and *indirectness* annotations are extracted from the corpus. One turn contains up to five consecutive user actions U and up to four consecutive system actions S. Therefore, the correlation between the last user action U_5 and the first system action S_1 of each turn as well as the median (Md) of all user and system actions of the respective turn U_{Md} and S_{Md} are analysed. The results in terms of Spearman's rank correlation coefficient Rho

	Elaborateness (5 classes)	Elaborateness (3 classes)	Indirectness (5 classes)	Indirectness (3 classes)	Indirectness (2 classes)
U_5/S_1	0.202*	0.184*	0.107^{*}	0.107^{*}	0.096^{*}
U_5/S_{Md}	0.243^{*}	0.219^{*}	0.144*	0.143*	0.138^{*}
U_{Md}/S_1	0.175^{*}	0.154^{*}	0.089^{*}	0.087^{*}	0.080^{*}
U_{Md}/S_{Md}	0.219*	0.189^{*}	0.132^{*}	0.131*	0.128^{*}

6.2 Investigating the Correlation between User and System Communication Style

Table 6.8: The correlation between the last user action U_5 and the first system action S_1 of each turn as well as the median of all user and system actions of the respective turn U_{Md} and S_{Md} in terms of Spearman's rank correlation coefficient Rho ρ . All results marked with (*) are significant at the 0.01 level. Taken from (Miehle et al., 2022, Table 8), licensed under CC BY 3.0 (https://creativecommons. org/licenses/by/3.0).

	Elaborateness (5 classes)	Elaborateness (3 classes)	Indirectness (5 classes)	Indirectness (3 classes)	Indirectness (2 classes)
German	0.138*	0.137*	0.128*	0.127^{*}	0.124*
Spanish	0.378^{*}	0.368*	0.140^{*}	0.138^{*}	0.115^{**}
Polish	0.240^{*}	0.235^{*}	0.235^{*}	0.233^{*}	0.223^{*}
Turkish	0.354^{*}	0.320*	0.104^{**}	0.103**	0.104^{**}

Table 6.9: The correlation between the last user action U_5 and the median of all system actions of the respective turn S_{Md} in terms of Spearman's rank correlation coefficient Rho ρ for the different languages. All results marked with (*) are significant at the 0.01 level, all results marked with (**) are significant at the 0.05 level. Taken from (Miehle et al., 2022, Table 9), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

 ρ for both the *elaborateness* and the *indirectness* in five, three and two classes can be seen in Table 6.8. All results are significant at the 0.01 level which shows that there is a significant correlation between the communication style of the system and the preceding communication style of the user. Moreover, the results show that the highest correlation is between the last user action U_5 and the median of the subsequent system actions S_{Md} . The correlation between the last user action U_5 and the median of all system actions of the respective turn S_{Md} for the different languages is shown in Table 6.9. It can be seen that there is a significant correlation for both the *elaborateness* and the *indirectness* for all four languages. However, the effect size for the *elaborateness* varies between the languages. While there is a small correlation for German and Polish, there is a medium correlation for Spanish and Turkish (according to Cohen (1977)). As the task and the familiarity between the speakers have been identical for the different languages, it is concluded that the discrepancy is due to differences in the languages/cultures.

In summary, it can be concluded that the communication style of the speaker who

6 Statistical Classification Approaches to Communication Style Adaptation



Figure 6.1: The *Communication Style Perception Model* developed in Chapter 5 is extended by the correlation between the communication style of the system and the preceding communication style of the user.

6.3 User Communication Style Recognition



Figure 6.2: The estimated user communication style, which is classified based on features from the speech recognition and the linguistic analysis, can be used in the dialogue management to adapt the system behaviour to the user. Adapted from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0 (https:// creativecommons.org/licenses/by/3.0).

assumed the role of the user has a direct influence on the subsequent communication style of the speaker who assumed the role of the system. The communication style is thereby a dynamic variable that changes during the interaction. Hence, a spoken dialogue system has to dynamically adapt its communication style to the preceding communication style of the user if it is supposed to imitate this human capacity. Therefore, the *Communication Style Perception Model* (Figure 5.1) which has been developed in Chapter 5 is extended by this dependency. The resulting model can be seen in Figure 6.1. It shows the interconnection between the communication style of the user and the communication style of the system (in red) in addition to the static dependencies that have been experimentally proven with human users (see Chapter 5). Based on this model, the adaptation has to be split into two steps. First, the communication style of the user needs to be estimated. Afterwards, the communication style of the system can be selected accordingly. Statistical classification approaches for both tasks are presented in the following sections.

6.3 User Communication Style Recognition

The last section has shown that the communication style of the user influences the selection of the communication style of the system. It is therefore a useful feature for turn-wise adaptation. However, the communication style of the user is dynamically changing during the interaction and needs to be automatically recognised. In order to do so, a separate

6 Statistical Classification Approaches to Communication Style Adaptation



Figure 6.3: The custom recurrent neural network (RNN) classifier consisting of two long short-term memory (LSTM) layers followed by three subsequent fully connected perceptron layers, as well as sigmoid (sig) and softmax (soft) activation functions.

module for spoken dialogue systems that copes with this task is implemented. The output of this module is an estimation of the user's communication style which can be used in the dialogue management to adapt the system behaviour to the user, as depicted in Figure 6.2. For the implementation of this communication style recognition module, a statistical classification approach has been presented in (Miehle et al., 2020). In the scope of this work, multiple classifiers were compared on the task including:

- A support vector machine (SVM) classifier
- A multi-layer perceptron (MLP) classifier with one hidden layer
- A custom recurrent neural network (RNN) classifier

The RNN architecture was proposed to cover possible temporal dependencies in the task. Since temporal features tend to be important for other adaptation approaches, they are likely to have an influence on the task at hand. In order to extract those temporal features automatically, an architecture consisting of long short-term memory (LSTM) layers is proposed as they have proven to be a powerful method for similar tasks (Rach et al., 2017). In the following, the overall architecture which can be seen in Figure 6.3 is discussed in detail. It contains two stacked LSTM layers to learn the temporal dependencies from the data, and three subsequent fully connected perceptron layers which are responsible for the classification. The output of the second LSTM layer is normalised to a probability distribution by use of a softmax activation function. The following two hidden perceptron layers each use a sigmoid activation function, while the activation function of the output layer is selected during grid search. Furthermore, both LSTM layers are regularised by randomly dropping out nodes during training to reduce overfitting. The dropout fraction is also selected during grid search. Likewise, the amount of nodes of all layers, the amount of training epochs and the training configuration (i.e. the optimiser and the loss function) are altered in the grid search.

	Parameter	Grid
SVM	$\begin{array}{c} \mathrm{C} \\ \gamma \end{array}$	1, 10, 100, 1000, 10000, 100000 0.01, 0.1, 1, 10, 100
MLP	#nodes #epochs optimiser output function loss function	3, 6, 12, 24, 48, 96, 144, 192 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000 adadelta, adam, nadam, adagrad, sgd, rmsprop sigmoid, softmax categorical crossentropy, mean squared error
RNN	#nodes #epochs dropout optimiser output function loss function	5, 10, 25, 35, 50, 70, 100 350, 500, 1000 0.1, 0.3, 0.5 adadelta, adam, nadam, adagrad, sgd, rmsprop sigmoid, softmax categorical crossentropy, mean squared error

Table 6.10: The grid of parameter values for the user communication style recognition.

The classifiers are trained and evaluated with a 10-fold cross-validation setting on the German part of the corpus described in Section 6.1. Grid search is used to find the best set of hyper parameters (SVM: C and γ ; MLP: the amount of nodes, the amount of epochs, the optimiser, the output function and the loss function; RNN: the amount of nodes of the LSTM layers, the dropout, the amount of nodes of the hidden perceptron layers, the amount of epochs, the optimiser, the output function and the loss function). The grid of parameter values can be found in Table 6.10. To take account for the imbalanced data during the grid search optimisation, the unweighted average recall (UAR) is used, which is the arithmetic average of all class-wise recalls.

6.3.1 The Dialogue Act Features

As a first approach, the classifiers are trained using only *dialogue act features* (DA) that can directly be derived from the data. These features contain the dialogue act and the amount of words in the corresponding utterance. Note that the dialogue act is the output of the linguistic analysis while the text representation of the utterance is the output of the speech recogniser (see Figure 6.2). Hence, both features in this feature set can be automatically derived during an ongoing interaction in every spoken dialogue system and no annotation is necessary. During the experiments, additional annotated features (the amount of topics being talked about in the current utterance, the speaker's culture, gender, age, year of birth, country of birth, country of residence and whether he/she played the role of the user or the system, as well as the system role and the number of the dialogue act in the current dialogue) have also been tested, but this led to worse results. The results are shown in Table 6.11.

For the classification of the 3-class elaborateness and the 3-class indirectness, the SVM

		Elaborateness	Indirectness	Indirectness
		(5 classes)	(5 classes)	(2 classes)
	UAR	0.842	0.553	0.735
	ACC	0.839	0.829	0.849
SVM	F1	0.839	0.588	0.753
	κ	0.752	0.438	0.510
	ρ	0.863	0.487	0.525
	UAR	0.840	0.555	0.753
	ACC	0.838	0.832	0.848
MLP	F1	0.838	0.582	0.761
	κ	0.749	0.467	0.527
	ρ	0.862	0.523	0.541
	UAR	0.805	0.507	0.765
	ACC	0.794	0.778	0.821
RNN	F1	0.797	0.466	0.747
	κ	0.686	0.409	0.496
	ρ	0.822	0.465	0.502

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.11: The classification results using the dialogue act features (DA) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .

and the MLP classifiers yield the best results. However, for the estimation of the binary *indirectness*, the RNN classifier outperforms the other classifiers. This shows that temporal information, which is extracted and stored by the LSTM layers, is beneficial for the estimation of the *indirectness*, but not for the estimation of the *elaborateness*.

Classification of the 3-class *elaborateness* reaches an UAR of 84% only using dialogue act features, which is quite promising. Classification of the 3-class *indirectness* results in an UAR of 56%, and the binary *indirectness* reaches an UAR of 77%. The results for the *indirectness* clearly show the difficulty of the task, which has already been revealed by the corpus creation. There, it has been quite hard for the annotators to distinguish between different levels of *indirectness* so that the class distribution of the *indirectness* is sub-optimal for the classification task. However, comparing the results to a majority-class classifier clearly shows that there is still a lot of information encoded in the DA feature set achieving higher UAR. The majority-class classifier always predicts the most frequent class in the training set and achieves an UAR of 33% for three classes and an UAR of 50% for two classes. Furthermore, comparing the achieved κ and ρ with the results obtained for the human annotators (Table 6.5) shows that all, the results of the classifier for the *elaborateness* (three classes) and the results of the classifiers for the *indirectness* (three human annotators.

		$\begin{array}{c} Elaborateness\\ (3 \text{ classes}) \end{array}$	Indirectness (3 classes)	Indirectness (2 classes)
	UAR	0.843	0.534	0.749
	ACC	0.841	0.831	0.818
SVM	F1	0.840	0.569	0.739
	κ	0.754	0.422	0.480
	ho	0.864	0.493	0.485
	UAR	0.841	0.558	0.753
	ACC	0.840	0.834	0.848
MLP	F1	0.839	0.588	0.761
	κ	0.753	0.470	0.526
	ho	0.864	0.521	0.540
	UAR	0.811	0.511	0.771
	ACC	0.800	0.789	0.817
RNN	F1	0.799	0.472	0.746
	κ	0.696	0.432	0.495
	ρ	0.837	0.500	0.503

Table 6.12: The classification results using the dialogue act features as well as the grammatical features (DA+G) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .

6.3.2 The Contribution of Grammatical and Linguistic Features

To address the question of whether grammatical features improve the estimation of the communication style, a second feature set is used containing the dialogue act features as well as grammatical features. For the grammatical features (G), Part-of-speech (POS) tags are assigned to the utterances using the RDRPOSTagger (Nguyen et al., 2014) and the number of each POS tag per utterance is counted. As the utterance is the output of the speech recognition and this tagger can be used online during an ongoing interaction, there is also no annotation necessary for this feature set. The results are shown in Table 6.12. It can be seen that there is no improvement in comparison to using only the dialogue act features.

In addition to grammatical features, *linguistic features* may majorly contribute to the overall classification performance. In order to encode the linguistic features, a Bag-of-Words (BoW) approach is used in combination with unigrams (U), unigrams and bigrams (UB) and word embeddings (WE). Using BoW and the corpus presented in Section 6.1, two distinct vocabularies are created:

• The BoW-U vocabulary contains every word occurring in the database of spoken dialogues.

		Elaborateness	Indirectness	Indirectness
		(3 classes)	(3 classes)	(2 classes)
	UAR	0.757	0.518	0.717
	ACC	0.763	0.778	0.826
U	F1	0.754	0.519	0.726
	κ	0.635	0.364	0.454
	ho	0.791	0.406	0.460
	UAR	0.816	0.529	0.742
	ACC	0.815	0.801	0.819
U+DA	F1	0.812	0.534	0.735
	κ	0.716	0.417	0.471
	ho	0.835	0.470	0.476
	UAR	0.818	0.530	0.744
	ACC	0.817	0.804	0.821
U+DA+G	F1	0.813	0.536	0.736
	κ	0.719	0.421	0.474
	ho	0.839	0.471	0.479
	UAR	0.762	0.503	0.734
	ACC	0.759	0.811	0.827
UB	F1	0.752	0.519	0.736
	κ	0.633	0.396	0.474
	ρ	0.806	0.440	0.479
	UAR	0.813	0.528	0.760
	ACC	0.809	0.815	0.840
UB+DA	F1	0.805	0.541	0.758
	κ	0.708	0.434	0.518
	ρ	0.843	0.483	0.523
	UAR	0.814	0.529	0.759
	ACC	0.809	0.816	0.840
UB+DA+G	F1	0.805	0.543	0.758
	κ	0.708	0.438	0.518
	ho	0.843	0.486	0.523

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.13: The classification results using the SVM classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .

		Elaborateness	Indirectness	Indirectness
		(3 classes)	(3 classes)	(2 classes)
	UAR	0.747	0.485	0.729
	ACC	0.752	0.822	0.842
U	F1	0.742	0.478	0.744
	κ	0.618	0.430	0.492
	ho	0.779	0.490	0.503
	UAR	0.809	0.484	0.743
	ACC	0.811	0.823	0.846
U+DA	F1	0.807	0.477	0.755
	κ	0.708	0.433	0.512
	ho	0.831	0.507	0.522
	UAR	0.817	0.484	0.746
	ACC	0.818	0.822	0.846
U+DA+G	F1	0.814	0.476	0.757
	κ	0.719	0.431	0.516
	ho	0.841	0.505	0.524
	UAR	0.745	0.520	0.748
	ACC	0.742	0.751	0.822
UB	F1	0.734	0.497	0.740
	κ	0.607	0.354	0.481
	ho	0.776	0.411	0.485
	UAR	0.786	0.533	0.748
	ACC	0.785	0.761	0.826
UB+DA	F1	0.781	0.511	0.742
	κ	0.669	0.387	0.485
	ho	0.811	0.452	0.490
	UAR	0.799	0.542	0.756
	ACC	0.796	0.757	0.827
UB+DA+G	F1	0.793	0.513	0.747
	κ	0.687	0.391	0.495
	ho	0.827	0.458	0.500

Table 6.14: The classification results using the MLP classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ . Taken from (Miehle et al., 2022, Table 14), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

		Elaborateness	Indirectness	Indirectness (2 classes)
	TTAD			(2 classes)
	UAR	0.657	0.478	0.719
	ACC	0.643	0.784	0.813
U	F1	0.631	0.456	0.717
	κ	0.463	0.371	0.436
	ho	0.671	0.409	0.440
	UAR	0.688	0.486	0.721
	ACC	0.670	0.796	0.816
U+DA	F1	0.663	0.464	0.719
	κ	0.504	0.340	0.441
	ρ	0.697	0.454	0.448
	UAR	0.699	0.489	0.722
	ACC	0.687	0.791	0.818
U+DA+G	F1	0.677	0.464	0.723
	κ	0.527	0.401	0.447
	ρ	0.716	0.456	0.449
	UAR	0.691	0.489	0.726
	ACC	0.685	0.774	0.806
UB	F1	0.679	0.456	0.717
	κ	0.521	0.381	0.436
	ρ	0.723	0.437	0.438
	UAR	0.696	0.497	0.728
	ACC	0.694	0.780	0.812
UB+DA	F1	0.686	0.462	0.719
	κ	0.534	0.401	0.441
	ρ	0.721	0.462	0.449
	UAR	0.723	0.498	0.724
	ACC	0.719	0.775	0.816
UB+DA+G	F1	0.714	0.460	0.722
	κ	0.573	0.395	0.445
	ρ	0.746	0.454	0.450

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.15: The classification results using the RNN classifier and the linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .

		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
	UAR	0.748	0.512	0.743
	ACC	0.748	0.801	0.818
WE	F1	0.741	0.511	0.735
	κ	0.614	0.415	0.471
	ρ	0.777	0.458	0.475
	UAR	0.820	0.544	0.762
	ACC	0.815	0.807	0.835
WE+DA	F1	0.813	0.554	0.757
	κ	0.717	0.417	0.515
	ρ	0.847	0.473	0.520
	UAR	0.815	0.551	0.762
	ACC	0.810	0.811	0.835
WE+DA+G	F1	0.808	0.562	0.757
	κ	0.709	0.426	0.515
	ho	0.842	0.481	0.520

- Table 6.16: The classification results using the SVM classifier and the linguistic features encoded as word embeddings (WE) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .
 - The BoW-UB vocabulary contains the BoW-U vocabulary (single words) as well as every two-word-sequence in the database.

These vocabularies and the combination with word embeddings lead to three different linguistic feature sets:

- U: This feature set contains a BoW-U vector for each utterance, thus encoding the number of times each word (of the overall vocabulary) appears in the corresponding utterance.
- UB: This feature set contains a BoW-UB vector for each utterance, thus encoding the number of times each word and each two-word-sequence (of the overall vocabulary) appear in the corresponding utterance.
- WE: For this feature set, the BoW-U vocabulary has been combined with the German pre-trained fastText word vectors by Grave et al. (2018). During the experiments, self-trained word vectors have also been tested, but this led to worse results. Matrix X of dimension $u \times w$ contains the BoW-U vectors (dimension $1 \times w$ with w the amount of words in vocabulary BoW-U) for each utterance, where u is the total

		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
	UAR	0.757	0.493	0.727
	ACC	0.755	0.783	0.828
WE	F1	0.749	0.495	0.729
	κ	0.626	0.364	0.464
	ho	0.786	0.414	0.479
	UAR	0.825	0.589	0.762
	ACC	0.821	0.803	0.842
WE+DA	F1	0.819	0.589	0.759
	κ	0.726	0.443	0.522
	ho	0.855	0.498	0.535
	UAR	0.827	0.594	0.765
	ACC	0.823	0.794	0.843
WE+DA+G	F1	0.821	0.588	0.762
	κ	0.729	0.432	0.528
	ho	0.857	0.480	0.544

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.17: The classification results using the MLP classifier and the linguistic features encoded as word embeddings (WE) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ . Adapted from (Miehle et al., 2022, Table 15), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

number of utterances. Matrix W of dimension $w \times p$ contains the fastText word vectors (dimension $1 \times p$ with p the length of each word vector) for each word. By multiplying these matrices a new matrix $Z = X \cdot W$ of dimension $u \times p$ is obtained, containing a vector representation for each utterance. These utterance vectors of dimension $1 \times p$ can then be used as feature vectors for the classification task.

In addition to using these linguistic feature sets individually, they are used in combination with the dialogue act features and the grammatical features. When using the U and UB feature sets, the results show no improvement (see Tables 6.13, 6.14 and 6.15). The classification results using the SVM classifier and the linguistic features encoded as word embeddings (WE) are shown in Table 6.16. It can be seen that there is no improvement for either the *elaborateness* and the *indirectness*. The classification results using the MLP classifier can be found in Table 6.17. There is again no improvement for the *elaborateness* classification. However, the performance of the 3-class *indirectness* could be improved by using linguistic information encoded as word embeddings. This in combination with grammatical and dialogue act features (WE+DA+G) leads to an UAR of 59%. The classi-
		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
	UAR	0.731	0.502	0.736
	ACC	0.743	0.794	0.806
WE	F1	0.734	0.473	0.723
	κ	0.604	0.416	0.447
	ho	0.781	0.454	0.451
	UAR	0.808	0.513	0.770
WE+DA	ACC	0.807	0.808	0.838
	F1	0.806	0.483	0.762
	κ	0.704	0.456	0.527
	ρ	0.837	0.523	0.534
WE+DA+G	UAR	0.814	0.521	0.782
	ACC	0.814	0.806	0.843
	F1	0.814	0.485	0.772
	κ	0.714	0.462	0.545
	ho	0.842	0.521	0.550

Table 6.18: The classification results using the RNN classifier and the linguistic features encoded as word embeddings (WE) (separately and in combination with the dialogue act features and the grammatical features) in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC), the F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ .

fication results using the RNN classifier are shown in Table 6.18. There is no improvement for the *elaborateness* and the 3-class *indirectness*, but for the binary *indirectness* classification. The combination of linguistic features encoded as word embeddings, dialogue act features and grammatical features (WE+DA+G) leads to an UAR of 78%.

To sum up, for the *elaborateness*, the best results are achieved by only using the dialogue act feature set (i.e. the dialogue act and the amount of words in the utterance). Grammatical and linguistic features do not seem to have any effect on the classification performance. This leads to the conclusion that for the *elaborateness*, analysing the utterance length dependent on the dialogue act seems to contain enough information to achieve good classification performance. For the *indirectness*, the overall performance could be improved by using linguistic information encoded as word embeddings. This in combination with grammatical and dialogue act features (WE+DA+G) leads to UARs of 59% and 78% for the estimation of *indirectness* using three classes and two classes, respectively. Furthermore, for the estimation of the binary *indirectness*, the RNN classifier outperforms the SVM and the MLP classifiers, showing that temporal information is beneficial in this case. Overall, the task of recognising the user's communication style is performed by use of the presented classifiers. Hence, in the next section, the task of automatically selecting the subsequent system communication style is addressed.

6 Statistical Classification Approaches to Communication Style Adaptation



Figure 6.4: The communication style selection is part of the dialogue management so that it not only decides what is said next, but also how it is said. Adapted from (Miehle et al., 2022, Figure 1), licensed under CC BY 3.0 (https: //creativecommons.org/licenses/by/3.0).

6.4 Automatic System Communication Style Selection

Section 6.2 has shown that the communication style adaptation has to be split into two steps. The task of recognising the user's communication style has been presented in Section 6.3. Building upon these results, this section addresses the task of automatically selecting the system communication style during an ongoing interaction with a spoken dialogue system, which has been presented in (Miehle et al., 2022). As depicted in Figure 6.4, this is part of the dialogue management so that it not only decides *what* is said next, but also *how*. The *Communication Style Perception Model* suggests that the system communication style depends on two components: 1) the content of the system dialogue act (what the system wants to say in the current turn) and 2) the reaction to the user (what the user wants from and how the user talks to the system).

For the classification of the user communication styles, promising results have been received by use of a supervised learning approach with a multi-layer perceptron (see Section 6.3). This approach might also be suitable for the task at hand. Hence, a multi-layer perceptron (MLP) classifier with one hidden layer is utilised. The neural net is trained and evaluated with a 10-fold cross-validation setting on the 2,880 turns of the corpus described in Section 6.2. The class distribution for both communication styles is shown in Table 6.19. Grid search is used to find the best set of hyper parameters (i.e. the amount of nodes, the amount of epochs, the optimiser, the output function and the loss function). The grid of parameter values can be found in Table 6.20. To take account for the imbalanced data

	1	2	3
Elaborateness Indirectness	$736 \\ 1,973$	$1,310 \\ 817$	834 90

Table 6.19: Class distribution of the annotated *elaborateness* and *indirectness* scores for the 2,880 dialogue turns. Taken from (Miehle et al., 2022, Table 16), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

Parameter	Grid
#Nodes	10, 25, 50
$\#\mathrm{Epochs}$	10, 50, 100, 200, 500
Optimiser	adadelta, adam, nadam, adagrad
Output function	sigmoid, softmax
Loss function	categorical crossentropy, mean squared error

Table 6.20: The grid of parameter values for the system communication style selection. Adapted from (Miehle et al., 2022, Table 17), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

during the grid search optimisation, the UAR is used.

For each of the 2,880 dialogue turns, the following features have been extracted:

- The system dialogue acts (S)
- The user dialogue acts (U)
- The amount of words in the utterance of the corresponding user dialogue acts (W)
- The user communication styles (CS)
- The language (German, Polish, Spanish or Turkish) (L)

During the experiments, part-of-speech tags and sentence embeddings (based on the respective utterances) have also been tested, though without improvement of the results. Note that all features can be automatically derived during an ongoing interaction in any spoken dialogue system and no annotation is necessary. The user dialogue acts are the output of the linguistic analysis while the text representation of the utterance is the output of the speech recogniser. The system dialogue acts are the output of the dialogue manager and the user communication styles may be classified by use of the communication style classifier described in Section 6.3 (see Figure 6.4).

The results are shown in Table 6.21. It can be seen that both the system dialogue act (S+L) and the information about the user (W+U+CS+L) contain relevant information for the selection of the system communication style. Overall, classification of the 3-class elaborateness reaches an UAR of 63%. Classification of the 3-class indirectness results in an

		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
S+L	UAR ACC F1	$0.625 \\ 0.651 \\ 0.636$	0.495 0.745 0.523	$0.673 \\ 0.760 \\ 0.686$
W+U+CS+L	UAR ACC F1	$\begin{array}{c} 0.535 \\ 0.560 \\ 0.542 \end{array}$	$0.409 \\ 0.702 \\ 0.406$	$0.617 \\ 0.708 \\ 0.622$
S+CS+L	UAR ACC F1	0.634 0.660 0.644	$0.484 \\ 0.731 \\ 0.499$	$0.675 \\ 0.756 \\ 0.686$
S+W+U+CS+L	UAR ACC F1	$0.627 \\ 0.647 \\ 0.635$	$0.471 \\ 0.724 \\ 0.486$	0.684 0.756 0.694

6 Statistical Classification Approaches to Communication Style Adaptation

Table 6.21: The classification results for the system communication style selection using different feature sets in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC) and the F1-Score. Taken from (Miehle et al., 2022, Table 19), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

UAR of 50%, and the binary *indirectness* reaches an UAR of 68%. The comparatively poor results of the 3-class *indirectness* classification can be explained by the data distribution. For the 2-class *indirectness*, the combination of the system dialogue act and all available user information provides the best result. For the 3-class *elaborateness*, the best result is obtained by use of the system dialogue act in combination with the user communication style (S+CS+L) and there is no improvement when adding the user dialogue act and the amount of words of the respective utterance. This shows that all relevant information about the user is covered by the user communication style.

When dividing the test set based on the languages, it can be seen that the classification works differently for the individual languages (see Table 6.22). For the 3-class *elaborate*ness, an UAR of 57% for German, 58% for Polish, 77% for Spanish and 51% for Turkish is achieved. For the 2-class *indirectness*, the classification results in an UAR of 65% for German, 64% for Polish, 82% for Spanish and 62% for Turkish. The differences between the languages indicate a cultural difference, as already revealed by studies like those described in Sections 5.1 and 5.2. However, it might also be due to the limited data. This needs to be investigated in future work.

Comparing the results to a majority-class classifier clearly shows that there is a lot of information encoded. Moreover, a baseline classifier which is mimicking the user communication style reaches an UAR of 42% for the 3-class *elaborateness*, 40% for the 3-class *indirectness* and 57% for the binary *indirectness* when using the communication style of

		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
Overall	UAR ACC F1	$0.634 \\ 0.660 \\ 0.644$	$0.495 \\ 0.745 \\ 0.523$	$0.684 \\ 0.756 \\ 0.694$
German	UAR ACC F1	$0.567 \\ 0.649 \\ 0.579$	$0.465 \\ 0.745 \\ 0.493$	$0.649 \\ 0.743 \\ 0.659$
Polish	UAR ACC F1	$0.584 \\ 0.615 \\ 0.591$	$0.439 \\ 0.715 \\ 0.439$	$0.643 \\ 0.725 \\ 0.650$
Spanish	UAR ACC F1	$0.766 \\ 0.805 \\ 0.768$	$0.552 \\ 0.797 \\ 0.535$	0.818 0.797 0.797
Turkish	UAR ACC F1	$0.506 \\ 0.586 \\ 0.520$	$0.539 \\ 0.742 \\ 0.563$	$0.619 \\ 0.760 \\ 0.630$

6.4 Automatic System Communication Style Selection

Table 6.22: The classification results for the system communication style selection in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC) and the F1-Score of the overall test set and the individual languages. Taken from (Miehle et al., 2022, Table 20), licensed under CC BY 3.0 (https://creativecommons.org/licenses/by/3.0).

		Elaborateness (3 classes)	Indirectness (3 classes)	Indirectness (2 classes)
U_5	UAR ACC F1	$0.416 \\ 0.412 \\ 0.409$	$0.398 \\ 0.586 \\ 0.389$	$\begin{array}{c} 0.571 \\ 0.618 \\ 0.569 \end{array}$
U_{Md}	UAR ACC F1	$0.399 \\ 0.406 \\ 0.398$	$0.396 \\ 0.582 \\ 0.391$	$0.566 \\ 0.609 \\ 0.563$

Table 6.23: The classification results for the system communication style selection baseline which is mimicking the last user communication style U_5 or the median of all previous user communication styles U_{Md} in terms of the Unweighted Average Recall (UAR), the Accuracy (ACC) and the F1-Score. Taken from (Miehle et al., 2022, Table 21), licensed under CC BY 3.0 (https://creativecommons. org/licenses/by/3.0).

6 Statistical Classification Approaches to Communication Style Adaptation

the last user action U_5 of the current turn. When using the median communication style of all user actions U_{Md} of the current turn, the results are even worse, as can be seen in Table 6.23. Hence, the trained system communication style selection module clearly outperforms a model which is just mimicking the user communication style at each turn.

6.5 Summary

This chapter has addressed the research question of how communication styles can be integrated into spoken dialogue systems. To this end, a supervised learning based communication style adaptation was implemented and evaluated. Due to the novelty of the underlying machine learning task, a multi-lingual corpus has been created, containing 258 dialogues with annotations for the *elaborateness* and *indirectness* for each user and system dialogue act. Using this data set, it was shown that there is a significant correlation between the communication style of the system and the preceding communication style of the user. This dependency has been included in the *Communication Style Perception Model*. Based on the resulting model, the adaptation has been split into two steps and the standard architecture of spoken dialogue systems has been extended by two components: 1) a communication style classifier that automatically identifies the user communication style and 2) a communication style selection module that selects an appropriate communication style for the system response.

The user communication style recognition model was trained to estimate the user's elaborateness and indirectness on the previously mentioned corpus. Three different classifiers have been investigated in this regard: a support vector machine (SVM) classifier, a multi-layer perceptron (MLP) classifier with one hidden layer, and a custom recurrent neural network (RNN) classifier consisting of two long short-term memory (LSTM) layers followed by three fully connected perceptron layers. Furthermore, different feature sets were tested as input for the classifiers. All features used for the communication style classification can be automatically recognised in spoken dialogue systems during an ongoing interaction, without any prior annotation. The results show that the *elaborateness* can be classified quite well by only using the dialogue act and the amount of words contained in the corresponding utterance in combination with the SVM and MLP classifiers, reaching an UAR of 84%. Grammatical and linguistic features do not seem to have any effect on the classification performance. This leads to the conclusion that for the *elaborateness*, analysing the utterance length dependent on the dialogue act contains enough information to achieve good classification performance. The *indirectness* appears to be a more difficult classification task and additional linguistic features in the form of word embeddings improve the classification results. This in combination with grammatical and dialogue act features led to UARs of 59% and 78% for the estimation of *indirectness* using three and two classes, respectively. Furthermore, for the estimation of the binary indirectness, the RNN classifier outperforms the SVM and the MLP classifiers, showing that temporal information is beneficial in this case.

For the system communication style selection, a similar supervised learning approach has been utilised. Using features that encode what the system wants to say in the current turn (i.e. the system dialogue acts), what the user wants from the system (i.e. the user dialogue acts) and how the user talks to the system (i.e. the amount of words in the utterance of the corresponding user dialogue acts, the user communication styles and the language), a multi-layer perceptron (MLP) classifier with one hidden layer has been trained and evaluated. As for the first task, the features can be automatically recognised during an interaction in spoken dialogue systems. The results outperform both a majority-class classifier and a baseline which is mimicking the last user communication style for each of the four languages, reaching an UAR of 63% for the classification of the 3-class elaborateness and an UAR of 68% for the 2-class indirectness.

7 Conclusions and Future Directions

This thesis has investigated the role of communication styles in spoken dialogue systems. This has been approached from two angles: it was investigated how varying communication styles are perceived by the user and how communication styles can be integrated into spoken dialogue systems.

In order to answer the first research question, the Communication Style Perception Model has been developed. Moreover, an extensive literature review as well as a requirement analysis by use of complex real-world systems have been conducted. The literature review has demonstrated that communication styles play an important role in human communication. Studies investigating the phenomenon of interactive adaptation in humanhuman and human-computer interaction have shown that people adapt their interaction styles to one another across many levels of utterance production when they communicate: they use the same words, coordinate their phonetic repertoire, their amplitude, their sentence and pause duration, the prepositional form and syntactic structures of their utterances, and the style of their messages-both when communicating with a human and a computer interaction partner. As the textual elements (i.e. how to formulate the utterance) are covered by the concept of communication styles, this work has concentrated on this aspect. For the requirement analysis, four different prototypes have been implemented and evaluated: a social companion and conversational partner for the elderly (Miehle et al., 2019a), a knowledge-based information agent with social competence and human interaction capabilities (Pragst et al., 2017a; Miehle et al., 2021c), a spoken dialogue system for indoor navigation (Miehle et al., 2021a), and an intelligent digital assistant for clinical operating rooms (Miehle et al., 2017a; Miehle et al., 2017b; Miehle et al., 2018a). These application scenarios represent a broad cross section of the current research on spoken dialogue systems, including different user groups and domains. The results have shown that there exist applications in which the communication style should not be varied during an interaction, but it is essential that the appropriate communication style is selected. This seems to apply for systems like the intelligent digital assistant for clinical operating rooms, which have to comply with strict requirements and presuppose high robustness. For companion systems and information agents, varying the system communication style is a valuable possibility for adaptation.

The literature review and the requirement analysis have also shown that the interaction between two participants as well as the appropriate or preferred communication style are influenced by numerous parameters. Examples in this regard are the speakers' roles, their cultures, their personalities or the aim of the interaction. In order to enable adaptation through communication styles, these different aspects need to be set into relation. Therefore, the *Communication Style Perception Model* has been developed within the scope of this work. It provides an answer to the first research question of how varying

7 Conclusions and Future Directions

communication styles are perceived by the user and covers elements that are relevant for the selection of the system communication style as well as aspects that are influenced by the system communication style. It is based on six user evaluations, examining various user groups in different scenarios. The first experiment (Miehle et al., 2016) has examined whether the cultural background of the user, which is one specific user trait, influences his or her communication style preference in human-computer interaction. In order to do so, it has investigated whether cultural communication idiosyncrasies found in human-human interaction may also be observed during human-computer interaction in a spoken dialogue system context. The results of the study have shown that there are differences between Germans and Japanese concerning their preferences in the communication patterns of the system output. Hence, it is concluded that the cultural background of the user has a direct influence on his or her preference in the system communication style. Further influencing variables were investigated in the second experiment (Miehle et al., 2018c). In a multicultural study, the influence of the user's culture and gender, the frequency of use of speech based assistants as well as the system's role on differences in the user's communication style preferences in human-computer interaction have been explored. The results have shown that the user's culture significantly influences the user's preference in the system's communication style even though five European cultures were examined whose communication styles are very alike. Depending on the culture, there are also gender differences. The user's frequency of use of speech based assistants seems to have no influence, but the system's role significantly influences the user's preference in the indirectness and elaborateness of the system. The third experiment (Miehle et al., 2021d) has shown that the system's role has an impact on user preferences that goes beyond the communication styles *indirectness* and *elaborateness*. Summarising the results of these experiments, it is concluded that both user traits and system properties influence the user's communication style preferences in human-computer interaction. This forms the first part of the Communication Style Perception Model. The remaining experiments (Miehle et al., 2018b; Kawanaka et al., 2020; Kawanaka et al., 2021; Zhang et al., 2021) investigated how varying system communication styles affect the users, if they are selected according to the users' personal preferences. To examine this, different communication styles have been included in various systems and applications. The results have shown that the system's communication style influences the user's satisfaction and the user's perception of the dialogue. For specific applications like behaviour change support systems, the communication style even has an impact on the user's behaviour. Thus, it is concluded that the system communication style has a direct influence on the user's satisfaction and behaviour, which forms the second part of the Communication Style Perception Model.

The second research question of how communication styles can be integrated into spoken dialogue systems has been addressed by the extension of the standard architecture of spoken dialogue systems. Two new components have been proposed, implemented an evaluated: a communication style classifier that automatically identifies the user communication style and a communication style selection module that selects an appropriate system communication style. Both tasks have been formulated as classification problems. Due to the novelty of the underlying machine learning task, a multi-lingual corpus has been created, containing 258 dialogues with annotations for the *elaborateness* and *indirect*ness for each of the 7,930 dialogue acts (Miehle et al., 2020; Miehle et al., 2021c; Miehle et al., 2022). Using this corpus, it has been shown that there is a significant correlation between the communication style of the system and the preceding communication style of the user (Miehle et al., 2022).

For the user communication style recognition (Miehle et al., 2020), three different classifiers have been compared on the task: a support vector machine (SVM) classifier, a multi-layer perceptron (MLP) classifier with one hidden layer, and a custom recurrent neural network (RNN) classifier consisting of two long short-term memory (LSTM) layers followed by three subsequent fully connected perceptron layers. Furthermore, different feature sets have been tested as input for the classifiers. All features that have been used for the communication style classification can be automatically recognised in spoken dialogue systems during an ongoing interaction, without any prior annotation. The results have shown that the *elaborateness* can be classified quite well by only using the dialogue act and the amount of words contained in the corresponding utterance in combination with the SVM and MLP classifiers, reaching an UAR of 84%. Grammatical and linguistic features do not seem to have any effect on the classification performance. This leads to the conclusion that for the *elaborateness*, analysing the utterance length dependent on the dialogue act seems to contain enough information to achieve good classification performance. The *indirectness* seems to be a more difficult classification task and additional linguistic features in form of word embeddings give improvement in the classification results. This in combination with grammatical and dialogue act features led to UARs of 59% and 78% for the estimation of *indirectness* using three classes and two classes, respectively. Furthermore, for the estimation of the binary *indirectness*, the RNN classifier has outperformed the SVM and the MLP classifiers, showing that temporal information is beneficial in this case.

For the system communication style selection (Miehle et al., 2022), a multi-layer perceptron (MLP) classifier with one hidden layer has been trained and evaluated, using features that encode what the system wants to say in the current turn (i.e. the system dialogue acts), what the user wants from the system (i.e. the user dialogue acts) and how the user talks to the system (i.e. the amount of words in the utterance of the corresponding user dialogue acts, the user communication styles and the language). As for the first task, the features can be automatically recognised in spoken dialogue systems during an interaction. The results have outperformed both a majority-class classifier and a baseline which is mimicking the last user communication style for each of the four languages, reaching an UAR of 63% for the classification of the 3-class *elaborateness* and an UAR of 68% for the 2-class *indirectness*.

7.1 Contributions

During the work on this thesis, several contributions have been achieved. They are divided into theoretical, practical, and experimental contributions and will be described in the following.

7 Conclusions and Future Directions

7.1.1 Theoretical

The first theoretical contribution is the transfer of communication aspects that exist in human-human interaction to human-computer interaction. This has been done in an extensive literature review on communication styles and interactive adaptation, and evaluated in user studies. Moreover, the user's communication style preferences in humancomputer interaction, the effects of varying system communication styles, and the correlation between the user and the system communication style have been analysed. As result, a formal model has been created that shows which elements are relevant for the selection of the system's communication style and which aspects are influenced by the system communication style. This model serves as a starting point for further investigations on the role of communication styles in human-computer interaction.

To recognise the user communication style during an ongoing interaction with a spoken dialogue system, a supervised learning task has been formulated. By use of dialogue act features as well as grammatical and linguistic features that are calculated based on the user utterance, the current communication style employed by the user can be estimated. This has been realised for the two communication styles *elaborateness* and *indirectness* (Miehle et al., 2020). Furthermore, a supervised learning task has been formulated for the automatic system communication style selection in spoken dialogue systems. It utilises features that encode what the system wants to say in the current turn (i.e. the system dialogue acts), what the user wants from the system (i.e. the user dialogue acts) and how the user talks to the system (i.e. the amount of words in the utterance of the corresponding user dialogue acts, the user communication styles and the language). This has also been realised for the communication styles *elaborateness* (Miehle et al., 2022).

7.1.2 Practical

In order to show that there are different applications where varying the system communication style is a valuable possibility for adaptation in spoken dialogue systems, different use case scenarios have been realised. To this end, a social companion and conversational partner for the elderly (Miehle et al., 2019a) and a spoken dialogue system for indoor navigation (Miehle et al., 2021a) have been implemented. Moreover, a culture-aware dialogue manager for a socially competent and culturally adaptive information agent for immigrants has been implemented and trained for four different cultures (German, Polish, Spanish and Turkish) (Pragst et al., 2017a; Miehle et al., 2021c). An intelligent digital assistant for clinical operating rooms has been set up to demonstrate that there are also application scenarios in which the communication style should not be varied, but it is essential that the right communication style is selected (Miehle et al., 2017a; Miehle et al., 2017b; Miehle et al., 2018a). For the experiments on user communication style recognition and system communication style selection, a corpus has been created with annotations for the *elaborateness* and *indirectness* for each user and system dialogue act (Miehle et al., 2020; Miehle et al., 2021c; Miehle et al., 2022). Based on the theoretical contributions, a user communication style recognition module as well as a system communication style selection module have been implemented (Miehle et al., 2020; Miehle et al., 2022).

7.1.3 Experimental

Within the scope of this thesis, six user studies have been conducted. They have investigated which communication idiosyncrasies found in human-human interaction may be transferred to human-computer interaction and what causes the differences in the user's communication style preferences in human-computer interaction. Moreover, it has been examined how varying communication styles are perceived by the users and how they affect the user's behaviour (Miehle et al., 2016; Miehle et al., 2018b; Miehle et al., 2018c; Miehle et al., 2021d; Kawanaka et al., 2020; Kawanaka et al., 2021; Zhang et al., 2021).

Furthermore, the implemented modules for the user communication style recognition and the system communication style selection have been trained and evaluated using the presented corpus with annotated communication styles. The experimental evaluation has shown that both modules yield solid results (Miehle et al., 2020; Miehle et al., 2022).

7.2 Future Directions

When combining both components presented in the scope of this thesis, the spoken dialogue system is enabled to recognise the user's communication style and select an appropriate communication style for the system. So far, both components have been evaluated separately and the results have shown that each module yields solid results. In future work, it would be interesting to integrate both components into one specific spoken dialogue system and conduct an evaluation of the overall system with real users. In this way, it can be verified whether the objective evaluation carried out as part of this work corresponds with the subjective evaluation of potential users. To further investigate how varying communication styles are perceived by the users and how they affect the user's behaviour, the implemented components might be integrated into different systems with diverse applications.

Within this thesis, both components have been trained an evaluated by use of one corpus. Applying other corpora could demonstrate the general validity of the approach. Moreover, it could be further investigated whether additional features improve the classification performance. For the system communication style selection, it would be interesting to consider a reinforcement learning approach instead of the herein presented supervised learning approach, as the system communication style selection in spoken dialogue systems might also depend on what the system and the user want to achieve in the long run.

Finally, future work could extend the presented approaches to other communication style dimensions than the *elaborateness* and *indirectness*. This work has focused on these dimensions as there is a verified influence on user satisfaction (Miehle et al., 2018b). However, depending on the domain and the application of the spoken dialogue system, other dimensions might also be relevant. Furthermore, contextual paraphrasing could be applied to alter the surface realisation of sentences as investigated by Pragst et al., 2020. In this way, templates are no longer required as the sentences can be changed automatically after the dialogue manager has made the decision which communication style is selected for the system.

- Adel, H. and Schütze, H. (2017). "Exploring Different Dimensions of Attention for Uncertainty Detection". In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp. 22–34.
- Aly, A. and Tapus, A. (2016). "Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human-robot interaction". In: Autonomous Robots 40.2, pp. 193–209.
- André, E., Rist, T., Van Mulken, S., Klesen, M., and Baldes, S. (2000). "The automated design of believable dialogues for animated presentation teams". In: *Embodied conver*sational agents, pp. 220–255.
- Aubakirova, M. and Bansal, M. (2016). "Interpreting neural networks to improve politeness comprehension". In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2035–2041.
- Awad, M. and Khanna, R. (2015). "Machine Learning". In: Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Berkeley, CA: Apress, pp. 1–18. DOI: https://doi.org/10.1007/978-1-4302-5990-9_1.
- Ball, G. and Breese, J. (2000). "Emotion and personality in a conversational agent". In: *Embodied conversational agents*, pp. 189–219.
- Bell, L., Gustafson, J., and Heldner, M. (2003). "Prosodic adaptation in human-computer interaction". In: *Proceedings of ICPHS*. Vol. 3. Citeseer, pp. 833–836.
- Bengio, Y., Simard, P., and Frasconi, P. (1994). "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks* 5.2, pp. 157–166.
- Bergmann, K., Branigan, H. P., and Kopp, S. (2015). "Exploring the alignment spacelexical and gestural alignment with real and virtual humans". In: Frontiers in ICT 2, p. 7.
- Branigan, H. P. and Pearson, J. (2006). "Alignment in human-computer interaction". In: How people talk to computers, robots, and other artificial communication partners, pp. 140–156.
- Branigan, H. P., Pickering, M. J., and Cleland, A. A. (2000). "Syntactic co-ordination in dialogue". In: Cognition 75.2, B13–B25.
- Branigan, H. P., Pickering, M. J., Pearson, J., and McLean, J. F. (2010). "Linguistic alignment between people and computers". In: *Journal of pragmatics* 42.9, pp. 2355– 2368.
- Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., and Nass, C. I. (2003). "Syntactic alignment between computers and people: The role of belief about mental states". In: *Proceedings of the 25th annual conference of the cognitive science society*. Lawrence Erlbaum Associates, pp. 186–191.

- Brennan, S. E. (1991). "Conversation with and through computers". In: User modeling and user-adapted interaction 1.1, pp. 67–86.
- Brennan, S. E. (1996). "Lexical entrainment in spontaneous dialog". In: Proceedings of ISSD 96, pp. 41–44.
- Brennan, S. E. and Clark, H. H. (1996). "Conceptual pacts and lexical choice in conversation." In: Journal of Experimental Psychology: Learning, Memory, and Cognition 22.6, p. 1482.
- Brennan, S. E. and Ohaeri, J. O. (1994). "Effects of message style on users' attributions toward agents". In: Conference companion on Human factors in computing systems, pp. 281–282.
- Brockmann, C., Isard, A., Oberlander, J., and White, M. (2005). "Modelling alignment for affective dialogue". In: Workshop on adapting the interaction style to affective factors at the 10th international conference on user modeling (UM-05).
- Bultman, D. C. and Svarstad, B. L. (2000). "Effects of physician communication style on client medication beliefs and adherence with antidepressant treatment". In: *Patient Education and Counseling* 40.2, pp. 173–185. ISSN: 0738-3991. DOI: https://doi.org/ 10.1016/S0738-3991(99)00083-X.
- Burgoon, J. K., Stern, L. A., and Dillman, L. (1995). Interpersonal Adaptation: Dyadic Interaction Patterns. Cambridge University Press. DOI: https://doi.org/10.1017/ CB09780511720314.
- Buschmeier, H., Bergmann, K., and Kopp, S. (2009). "An alignment-capable microplanner for natural language generation". In: *Proceedings of the 12th European Workshop on Natural Language Generation*.
- Cohen, J. (1960). "A coefficient of agreement for nominal scales". In: *Educational and* psychological measurement 20.1, pp. 37–46.
- Cohen, J. (1977). Statistical power analysis for the behavioral sciences. Academic Press.
- Cortes, C. and Vapnik, V. (1995). "Support-vector networks". In: Machine learning 20.3, pp. 273–297.
- Coulston, R., Oviatt, S., and Darves, C. (2002). "Amplitude convergence in children's conversational speech with animated personas". In: *Seventh International Conference on Spoken Language Processing*.
- Crum, A. J. and Langer, E. J. (2007). "Mind-set matters: Exercise and the placebo effect". In: Psychological Science 18.2, pp. 165–171.
- Cullingford, R. E. (1978). Script application: computer understanding of newspaper stories. Tech. rep. Yale University New Haven, Dept. of Computer Science.
- Cuschieri, A. (1999). "Laparoscopic cholecystectomy". In: Journal of the Royal College of Surgeons of Edinburgh 44.3, pp. 187–192.
- Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., and Potts, C.-t. (2013). "A computational approach to politeness with application to social factors". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers). Sofia, Bulgaria: Association for Computational Linguistics, pp. 250–259.

- Darves, C. and Oviatt, S. (2002). "Adaptation of users' spoken dialogue patterns in a conversational interface". In: Seventh International Conference on Spoken Language Processing.
- Daskalova, N., Desingh, K., Papoutsaki, A., Schulze, D., Sha, H., and Huang, J. (2017). "Lessons learned from two cohorts of personal informatics self-experiments". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3, pp. 1–22.
- Davis, F. D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology". In: *MIS quarterly*, pp. 319–340.
- De Jong, M., Theune, M., and Hofs, D. (2008). "Politeness and alignment in dialogues with a virtual guide". In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pp. 207–214.
- Di Buccio, E., Melucci, M., and Moro, F. (2014). "Detecting verbose queries and improving information retrieval". In: *Information Processing & Management* 50.2, pp. 342–360.
- Dimbleby, R. and Burton, G. (1998). More than words: An introduction to communication. Routledge.
- Doran, C., Aberdeen, J., Damianos, L., and Hirschman, L. (2003). "Comparing several aspects of human-computer and human-human dialogues". In: Current and new directions in discourse and dialogue. Springer, pp. 133–159.
- Dral, J., Heylen, D., and Akker, R. op den (2008). "Detecting Uncertainty in Spoken Dialogues: An explorative research to the automatic detection of a speakers' uncertainty by using prosodic markers". In: *listener (answering his question)* 8, p. 277.
- Earley, P. C. and Mosakowski, E. (2004). "Cultural intelligence". In: Harvard business review 82.10, pp. 139–146.
- Elliott, C., Adams, R. J., and Sockalingam, S. (2016). *Multicultural Toolkit: Toolkit for Cross-Cultural Collaboration*. Awesome Library. Accessed: 2016-05-01. URL: http://www.awesomelibrary.org/multiculturaltoolkit.html.
- Elman, J. L. (1990). "Finding structure in time". In: Cognitive science 14.2, pp. 179–211.
- Feußner, H. and Wilhelm, D. (2016). "Minimalinvasive Chirurgie und "robotic surgery": Chirurgie 4.0?" In: Der Chirurg 87.3, pp. 189–194.
- Forbes-Riley, K. and Litman, D. J. (2011). "Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor". In: Speech Communication 53.9, pp. 1115–1136. ISSN: 0167-6393.
- Garrod, S. and Anderson, A. (1987). "Saying what you mean in dialogue: A study in conceptual and semantic co-ordination". In: *Cognition* 27.2, pp. 181–218.
- Gers, F. A., Schmidhuber, J., and Cummins, F. (1999). "Learning to forget: continual prediction with LSTM". In: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470). Vol. 2, 850–855 vol.2. DOI: https://doi.org/10.1049/cp:19991218.
- Gessinger, I., Möbius, B., Andreeva, B., Raveh, E., and Steiner, I. (2019). "Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments". In: *INTER-SPEECH*, pp. 301–305.

- Gharouit, K. and Nfaoui, E. H. (2017). "A comparison of classification algorithms for verbose queries detection using BabelNet". In: 2017 Intelligent Systems and Computer Vision (ISCV). IEEE, pp. 1–5.
- Goel, P., Matsuyama, Y., Madaio, M., and Cassell, J. (2018). ""I think it might help if we multiply, and not add": Detecting Indirectness in Conversation". In: Proceedings of the International Workshop Series on Spoken Dialogue Systems Technology (IWSDS).
- Grant, D. A. (1948). "The latin square principle in the design and analysis of psychological experiments." In: *Psychological bulletin* 45.5, p. 427.
- Grave, E., Bojanowski, P., Bojanowski, P., Joulin, A., and Miko-lov, T. (2018). "Learning Word Vectors for 157 Languages". In: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018).
- Grice, H. P. (1975). "Logic and conversation". In: Speech acts. Brill, pp. 41–58.
- Griol, D. and Molina, J. M. (2015). "Modeling users emotional state for an enhanced human-machine interaction". In: International Conference on Hybrid Artificial Intelligence Systems. Springer, pp. 357–368.
- Gupta, S., Walker, M. A., and Romano, D. M. (2007). "How rude are you?: Evaluating politeness and affect in interaction". In: International Conference on Affective Computing and Intelligent Interaction. Springer, pp. 203–217.
- Heinroth, T., Denich, D., and Schmitt, A. (2010). "OwlSpeak Adaptive Spoken Dialogue within Intelligent Environments". In: *Pervasive Computing and Communications* Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on. IEEE, pp. 666–671.
- Hixon, B., Clark, P., and Hajishirzi, H. (2015). "Learning knowledge graphs for question answering through conversational dialog". In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 851–861.
- Hochreiter, S. and Schmidhuber, J. (1997). "Long short-term memory". In: Neural computation 9.8, pp. 1735–1780.
- Hoegen, R., Aneja, D., McDuff, D., and Czerwinski, M. (2019). "An end-to-end conversational style matching agent". In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, pp. 111–118.
- Hofs, D., Theune, M., and Akker, R. op den (2010). "Natural interaction with a virtual guide in a virtual environment". In: *Journal on Multimodal User Interfaces* 3.1-2, pp. 141–153.
- Hofstede, G. (2009). Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Sage.
- Holtgraves, T. (1986). "Language structure in social interaction: Perceptions of direct and indirect speech acts and interactants who use them". In: *Journal of personality and social psychology* 51.2, p. 305.
- Hone, K. S. and Graham, R. (2000). "Towards a tool for the subjective assessment of speech system interfaces (SASSI)". In: *Natural Language Engineering* 6.3-4, pp. 287– 303.

- Hu, Z., Tree, J. E. F., and Walker, M. A. (2018). "Modeling Linguistic and Personality Adaptation for Natural Language Generation". In: *Proceedings of the 19th Annual* SIGdial Meeting on Discourse and Dialogue, pp. 20–31.
- Irfan, B., Narayanan, A., and Kennedy, J. (2020). "Dynamic Emotional Language Adaptation in Multiparty Interactions with Agents". In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, pp. 1–8.
- Isbister, K. and Nass, C. (2000). "Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics". In: *International journal of human*computer studies 53.2, pp. 251–267.
- Janarthanam, S. and Lemon, O. (2014). "Adaptive Generation in Dialogue Systems Using Dynamic User Modeling". In: Computational Linguistics 40.4, pp. 883–920.
- Johnson, W. L., Rizzo, P., Bosma, W., Kole, S., Ghijsen, M., and Welbergen, H. van (2004). "Generating Socially Appropriate Tutorial Dialog". In: Affective Dialogue Systems. Springer Berlin Heidelberg, pp. 254–264. DOI: https://doi.org/10.1007/978-3-540-24842-2_27.
- Jokinen, K. and Kanto, K. (2004). "User expertise modeling and adaptivity in a speechbased e-mail system". In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), pp. 87–94.
- Jokinen, K. and McTear, M. (2009). "Spoken dialogue systems". In: Synthesis Lectures on Human Language Technologies 2.1, pp. 1–151.
- Jokinen, K. and Wilcock, G. (2001). "Confidence-based adaptivity in response generation for a spoken dialogue system". In: *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*.
- Jungers, M. K., Palmer, C., and Speer, S. R. (2002). "Time after time: The coordinating influence of tempo in music and speech". In: *Cognitive Processing* 1.2, pp. 21–35.
- Kaplan, R. B. (1966). "Cultural thought patterns in inter-cultural education". In: Language learning 16.1-2, pp. 1–20.
- Kaur, G., Srivastava, M., and Kumar, A. (2017). "Analysis of feature extraction methods for speaker dependent speech recognition". In: *International journal of engineering and technology innovation* 7.2, p. 78.
- Kawanaka, S., Miehle, J., Matsuda, Y., Suwa, H., Yasumoto, K., and Minker, W. (2020). "Design and Evaluation on Task Allocation Interfaces in Gamified Participatory Sensing for Tourism". In: *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. New York, NY, USA: Association for Computing Machinery, pp. 458–463. DOI: https://doi.org/10.1145/ 3448891.3448957.
- Kawanaka, S., Miehle, J., Matsuda, Y., Suwa, H., Yasumoto, K., and Minker, W. (2021).
 "Task allocation interface design and personalization in gamified participatory sensing for tourism". In: arXiv:2105.12032. URL: https://arxiv.org/abs/2105.12032.
- Kolář, J., Liu, Y., and Shriberg, E. (2010). "Speaker adaptation of language and prosodic models for automatic dialog act segmentation of speech". In: Speech Communication 52.3, pp. 236–245.

- Koulouri, T., Lauria, S., and Macredie, R. D. (2016). "Do (and say) as I say: Linguistic adaptation in human–computer dialogs". In: *Human–Computer Interaction* 31.1, pp. 59– 95.
- Kroeger, P. R. (2019). Analyzing meaning: An introduction to semantics and pragmatics. Second corrected and slightly revised edition. Language Science Press.
- Kruijff-Korbayová, I., Kukina, C., Olga, G., and Schehl, J. (2008). "Generation of output style variation in the SAMMIE dialogue system". In: Proceedings of the Fifth International Natural Language Generation Conference, pp. 129–137.
- Kubat, M. (2017). An introduction to machine learning. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-63913-0.
- Levelt, W. J. M. and Kelter, S. (1982). "Surface form and memory in question answering". In: Cognitive psychology 14.1, pp. 78–106.
- Lewis, J. R. (1995). "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: International Journal of Human-Computer Interaction 7.1, pp. 57–78.
- Lewis, R. D. (2010). When Cultures Collide: Leading Across Cultures. Brealey.
- Liscombe, J., Hirschberg, J., and Venditti, J. J. (2005). "Detecting certainness in spoken tutorial dialogues". In: Proceedings of the Ninth European Conference on Speech Communication and Technology.
- Lison, P. (2015). "A hybrid approach to dialogue management based on probabilistic rules". In: Computer Speech & Language 34.1, pp. 232–255. ISSN: 0885-2308.
- Lison, P. and Kennington, C. (2016). "OpenDial: A Toolkit for Developing Spoken Dialogue Systems with Probabilistic Rules". In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Demonstrations). Berlin, Germany: Association for Computational Linguistics, pp. 67–72.
- Litman, D. J. and Pan, S. (2002). "Designing and evaluating an adaptive spoken dialogue system". In: User Modeling and User-Adapted Interaction 12.2-3, pp. 111–137.
- Lupton, D. (2013). "Quantifying the body: monitoring and measuring health in the age of mHealth technologies". In: Critical public health 23.4, pp. 393–403.
- Madaio, M., Cassell, J., and Ogan, A. (2017). "The impact of peer tutors' use of indirect feedback and instructions". In: Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL). Philadelphia, PA: International Society of the Learning Sciences.
- Mairesse, F. and Walker, M. A. (2010). "Towards personality-based user adaptation: psychologically informed stylistic language generation". In: User Modeling and User-Adapted Interaction 20.3, pp. 227–278.
- Mairesse, F. and Walker, M. A. (2011). "Controlling User Perceptions of Linguistic Style: Trainable Generation of Personality Traits". In: *Computational Linguistics* 37.3, pp. 455–488.
- Matsuda, Y., Arakawa, Y., and Yasumoto, K. (2016). "Design and Evaluation of Participatory Mobile Sensing Platform for Diverse Sensing and Gamification Scenarios". In: Proceedings of the 14th Annual International Conference on Mobile Systems, Applica-

tions, and Services Companion (MobiSys '16 Companion), p. 57. DOI: https://doi.org/10.1145/2938559.2948836.

- McTear, M. (2004). Spoken dialogue technology: toward the conversational user interface. Springer Science & Business Media.
- Miehle, J., Bagci, I., Minker, W., and Ultes, S. (2019a). "A Social Companion and Conversational Partner for the Elderly". In: Advanced Social Interaction with Agents. Vol. 510. Lecture Notes in Electrical Engineering. Springer International Publishing, pp. 103–109. DOI: https://doi.org/10.1007/978-3-319-92108-2_12.
- Miehle, J., Feustel, I., Hornauer, J., Minker, W., and Ultes, S. (2020). "Estimating User Communication Styles for Spoken Dialogue Systems". In: Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020). Marseille, France: European Language Resources Association (ELRA), pp. 533–541.
- Miehle, J., Feustel, I., Minker, W., and Ultes, S. (2021a). "A Script Knowledge Based Dialogue System for Indoor Navigation". In: *Conversational Dialogue Systems for the Next Decade*. Vol. 704. Lecture Notes in Electrical Engineering. Springer Singapore, pp. 379–385. DOI: https://doi.org/10.1007/978-981-15-8395-7_29.
- Miehle, J., Gerstenlauer, N., Ostler, D., Feußner, H., Minker, W., and Ultes, S. (2017a). "An Intelligent Digital Assistant for Clinical Operating Rooms". In: Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL). Saarbrücken, Germany, pp. 164–165.
- Miehle, J., Gerstenlauer, N., Ostler, D., Feußner, H., Minker, W., and Ultes, S. (2018a). "Expert Evaluation of a Spoken Dialogue System in a Clinical Operating Room". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA).
- Miehle, J., Minker, W., and Ultes, S. (2018b). "Exploring the Impact of Elaborateness and Indirectness on User Satisfaction in a Spoken Dialogue System". In: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP). Singapore: ACM, pp. 165–172.
- Miehle, J., Minker, W., and Ultes, S. (2018c). "What Causes the Differences in Communication Styles? A Multicultural Study on Directness and Elaborateness". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA).
- Miehle, J., Minker, W., and Ultes, S. (2022). "When to Say What and How: Adapting the Elaborateness and Indirectness of Spoken Dialogue Systems". In: *Dialogue & Discourse* 13(1), pp. 1–40. DOI: https://doi.org/10.5210/dad.2022.101. This publication has been under review at the submission date of the dissertation.
- Miehle, J., Ostler, D., Gerstenlauer, N., and Minker, W. (2017b). "The next step: intelligent digital assistance for clinical operating rooms". In: *Innovative Surgical Sciences* 2.3, pp. 159–161. DOI: https://doi.org/10.1515/iss-2017-0034.
- Miehle, J., Wagner, N., Minker, W., and Ultes, S. (2021c). "Culture-Aware Dialogue Management for Conversational Assistants". In: *Conversational Dialogue Systems for the Next Decade*. Vol. 704. Lecture Notes in Electrical Engineering. Springer Singapore, pp. 103–115. DOI: https://doi.org/10.1007/978-981-15-8395-7_8.

- Miehle, J., Wieluch, S., Minker, W., and Ultes, S. (2021d). "Decide or Delegate: How Script Knowledge Based Conversational Assistants Should Act in Inconclusive Situations". In: Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking (ICDCN 2021). Nara, Japan: Association for Computing Machinery, pp. 69–73. DOI: https://doi.org/10.1145/3427477.3428185.
- Miehle, J., Yoshino, K., Pragst, L., Ultes, S., Nakamura, S., and Minker, W. (2016). "Cultural Communication Idiosyncrasies in Human-Computer Interaction". In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). Los Angeles, USA: Association for Computational Linguistics, pp. 74–79.
- Moon, Y. and Nass, C. (1996). "How "real" are computer personalities? Psychological responses to personality types in human-computer interaction". In: *Communication research* 23.6, pp. 651–674.
- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, C. (1995). "Can computer personalities be human personalities?" In: *Conference companion on Human factors in computing systems*, pp. 228–229.
- Nenkova, A., Gravano, A., and Hirschberg, J. (2008). "High frequency word entrainment in spoken dialogue". In: Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies: Short papers. Association for Computational Linguistics, pp. 169–172.
- Neuliep, J. W. (2018). Intercultural communication: a contextual approach. Seventh edition, international student edition. SAGE.
- Nguyen, D. Q., Nguyen, D. Q., Pham, D. D., and Pham, S. B. (2014). "RDRPOSTagger: A ripple down rules-based part-of-speech tagger". In: Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp. 17–20.
- Niederhoffer, K. G. and Pennebaker, J. W. (2002). "Linguistic style matching in social interaction". In: Journal of Language and Social Psychology 21.4, pp. 337–360.
- Nothdurft, F., Honold, F., and Kurzok, P. (2012). "Using explanations for runtime dialogue adaptation". In: Proceedings of the 14th ACM international conference on Multimodal interaction, pp. 63–64.
- Oraby, S., Reed, L., Tandon, S., Sharath, T. S., Lukin, S., and Walker, M. A. (2018). "Controlling Personality-Based Stylistic Variation with Neural Natural Language Generators". In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 180–190.
- Oviatt, S., Darves, C., and Coulston, R. (2004). "Toward adaptive conversational interfaces: Modeling speech convergence with animated personas". In: ACM Transactions on Computer-Human Interaction (TOCHI) 11.3, pp. 300–328.
- Pardo, J. S. (2006). "On phonetic convergence during conversational interaction". In: The Journal of the Acoustical Society of America 119.4, pp. 2382–2393.
- Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., and Nass, C. I. (2006). "Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice". In: *Proceedings of the SIGCHI conference on Human Factors in computing* systems, pp. 1177–1180.

- Pesch, R., Bouncken, R. B., and Kraus, S. (2015). "Effects of communication style and age diversity in innovation teams". In: *International Journal of Innovation and Technology Management* 12.06, p. 1550029.
- Pickering, M. J. and Garrod, S. (2004). "Toward a mechanistic psychology of dialogue". In: *Behavioral and brain sciences* 27.2, pp. 169–190.
- Porayska-Pomsta, K. and Mellish, C. (2004). "Modelling politeness in natural language generation". In: International Conference on Natural Language Generation. Sprin-ger, pp. 141–150.
- Pragst, L., Miehle, J., Minker, W., and Ultes, S. (2017a). "Challenges for adaptive dialogue management in the KRISTINA project". In: ISIAA 2017: Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents. Glasgow, UK: ACM, pp. 11–14. DOI: https://doi.org/10.1145/3139491. 3139508.
- Pragst, L., Minker, W., and Ultes, S. (2017b). "Exploring the Applicability of Elaborateness and Indirectness in Dialogue Management". In: Proceedings of the 8th International Workshop On Spoken Dialogue Systems (IWSDS). Farmington, USA.
- Pragst, L., Minker, W., and Ultes, S. (2020). "Comparative Study of Sentence Embeddings for Contextual Paraphrasing". In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 6841–6851.
- Prokofieva, A. and Hirschberg, J. (2014). "Hedging and speaker commitment". In: Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland, pp. 10–13.
- Qingxue, L. (2003). "Understanding different cultural patterns or orientations between East and West". In: *Investigationes Linguisticae* 9, pp. 21–30.
- Rach, N., Minker, W., and Ultes, S. (2017). "Interaction Quality Estimation Using Long Short-Term Memories". In: Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). Saarbrücken, Germany: Association for Computational Linguistics, pp. 164–169.
- Rau, L. F., Jacobs, P. S., and Zernik, U. (1989). "Information extraction and text summarization using linguistic knowledge acquisition". In: Information Processing & Management 25.4, pp. 419–428.
- Raveh, E., Steiner, I., Gessinger, I., and Möbius, B. (2018). "Studying Mutual Phonetic Influence with a Web-Based Spoken Dialogue System". In: International Conference on Speech and Computer. Springer, pp. 552–562.
- Reitter, D., Keller, F., and Moore, J. D. (2006). "Computational modelling of structural priming in dialogue". In: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. Association for Computational Linguistics, pp. 121–124.
- Rosenblatt, F. (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

- Samuel, A. L. (1959). "Some Studies in Machine Learning Using the Game of Checkers". In: *IBM Journal of Research and Development* 3.3, pp. 210–229. DOI: https://doi.org/10.1147/rd.33.0210.
- Schank, R. C. and Abelson, R. P. (1977). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Lawrence Erlbaum Associates.
- Schober, M. F. (1993). "Spatial perspective-taking in conversation". In: Cognition 47.1, pp. 1–24.
- Searle, J. R. (1975). "Indirect Speech Acts". In: Speech acts. Brill, pp. 59-82.
- Shrout, P. E. and Fleiss, J. L. (1979). "Intraclass correlations: uses in assessing rater reliability". In: *Psychological bulletin* 86.2, p. 420.
- Sigala, M., Christou, E., and Gretzel, U. (2012). Social Media in Travel, Tourism and Hospitality; Theory, Practice and Cases. Ashgate Publishing, Ltd.
- Smestad, T. L. and Volden, F. (2019). "Chatbot Personalities Matters". In: Internet Science. Springer International Publishing, pp. 170–181.
- Spearman, C. (1904). "The proof and measurement of association between two things". In: *The American journal of psychology* 15.1, pp. 72–101.
- Stenchikova, S. and Stent, A. (2007). "Measuring adaptation between dialogs". In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue.
- Sutton, R. S. and Barto, A. G. (1998). Reinforcement learning: An introduction. 1st. Cambridge, MA, USA: MIT Press.
- Suzuki, N. and Katagiri, Y. (2007). "Prosodic alignment in human-computer interaction". In: Connection Science 19.2, pp. 131–141.
- Swanson, R. and Gordon, A. S. (2008). "Say anything: A massively collaborative open domain story writing companion". In: *Joint International Conference on Interactive Digital Storytelling*. Springer, pp. 32–40.
- Tapus, A. and Mataric, M. J. (2008). "Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance." In: AAAI spring symposium: emotion, personality, and social behavior, pp. 133–140.
- Tondello, G. F., Mora, A., Marczewski, A., and Nacke, L. E. (2019). "Empirical validation of the Gamification User Types Hexad scale in English and Spanish". In: *International Journal of Human-Computer Studies* 127, pp. 95–111. ISSN: 1071-5819. DOI: https: //doi.org/10.1016/j.ijhcs.2018.10.002.
- Ulinski, M., Benjamin, S., and Hirschberg, J. (2018). "Using hedge detection to improve committed belief tagging". In: Proceedings of the Workshop on Computational Semantics beyond Events and Roles, pp. 1–5.
- Ultes, S., Budzianowski, P., Casanueva, I., Mrksic, N., Rojas-Barahona, L. M., Su, P.-H., Wen, T.-H., Gasic, M., and Young, S. J. (2017). "Domain-Independent User Satisfaction Reward Estimation for Dialogue Policy Learning". In: *INTERSPEECH*, pp. 1721–1725.
- Ultes, S. and Minker, W. (2014). "Managing adaptive spoken dialogue for Intelligent Environments". In: Journal of Ambient Intelligence and Smart Environments 6.5, pp. 523–539.
- Van Dolen, W. M., Dabholkar, P. A., and De Ruyter, K. (2007). "Satisfaction with online commercial group chat: the influence of perceived technology attributes, chat group

characteristics, and advisor communication style". In: *Journal of retailing* 83.3, pp. 339–358.

- Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., and André, E. (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.
- Walker, M. A., Stent, A., Mairesse, F., and Prasad, R. (2007). "Individual and domain adaptation in sentence planning for dialogue". In: *Journal of Artificial Intelligence Re*search 30, pp. 413–456.
- Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., and Collins, H. (2005). "The Politeness Effect: Pedagogical Agents and Learning Gains." In: *AIED*, pp. 686–693.
- Whittaker, S., Walker, M. A., and Maloor, P. (2003). "Should I tell all?: An experiment on conciseness in spoken dialogue". In: *Eighth European Conference on Speech Communication and Technology.*
- Wilkie, J., Jack, M. A., and Littlewood, P. J. (2005). "System-initiated digressive proposals in automated human-computer telephone dialogues: the use of contrasting politeness strategies". In: *International Journal of Human-Computer Studies* 62.1, pp. 41–71.
- Zhang, Z., Miehle, J., Matsuda, Y., Fujimoto, M., Arakawa, Y., Yasumoto, K., and Minker, W. (2021). "Exploring the Impacts of Elaborateness and Indirectness in a Behavior Change Support System". In: *IEEE Access* 9, pp. 74778–74788. DOI: https://doi. org/10.1109/ACCESS.2021.3079473.

List of Contributing Publications

- Kawanaka, S., Miehle, J., Matsuda, Y., Suwa, H., Yasumoto, K., and Minker, W. (2020). "Design and Evaluation on Task Allocation Interfaces in Gamified Participatory Sensing for Tourism". In: *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services.* New York, NY, USA: Association for Computing Machinery, pp. 458–463. DOI: https://doi.org/10.1145/ 3448891.3448957.
- Kawanaka, S., Miehle, J., Matsuda, Y., Suwa, H., Yasumoto, K., and Minker, W. (2021).
 "Task allocation interface design and personalization in gamified participatory sensing for tourism". In: arXiv:2105.12032. URL: https://arxiv.org/abs/2105.12032.
- Miehle, J., Bagci, I., Minker, W., and Ultes, S. (2019a). "A Social Companion and Conversational Partner for the Elderly". In: Advanced Social Interaction with Agents. Vol. 510. Lecture Notes in Electrical Engineering. Springer International Publishing, pp. 103–109. DOI: https://doi.org/10.1007/978-3-319-92108-2_12.
- Miehle, J., Feustel, I., Hornauer, J., Minker, W., and Ultes, S. (2020). "Estimating User Communication Styles for Spoken Dialogue Systems". In: Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020). Marseille, France: European Language Resources Association (ELRA), pp. 533–541.
- Miehle, J., Feustel, I., Minker, W., and Ultes, S. (2021a). "A Script Knowledge Based Dialogue System for Indoor Navigation". In: *Conversational Dialogue Systems for the Next Decade*. Vol. 704. Lecture Notes in Electrical Engineering. Springer Singapore, pp. 379–385. DOI: https://doi.org/10.1007/978-981-15-8395-7_29.
- Miehle, J., Gerstenlauer, N., Ostler, D., Feußner, H., Minker, W., and Ultes, S. (2017a). "An Intelligent Digital Assistant for Clinical Operating Rooms". In: Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL). Saarbrücken, Germany, pp. 164–165.
- Miehle, J., Gerstenlauer, N., Ostler, D., Feußner, H., Minker, W., and Ultes, S. (2018a). "Expert Evaluation of a Spoken Dialogue System in a Clinical Operating Room". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA).
- Miehle, J., Minker, W., and Ultes, S. (2018b). "Exploring the Impact of Elaborateness and Indirectness on User Satisfaction in a Spoken Dialogue System". In: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP). Singapore: ACM, pp. 165–172.
- Miehle, J., Minker, W., and Ultes, S. (2018c). "What Causes the Differences in Communication Styles? A Multicultural Study on Directness and Elaborateness". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA).

- Miehle, J., Minker, W., and Ultes, S. (2022). "When to Say What and How: Adapting the Elaborateness and Indirectness of Spoken Dialogue Systems". In: *Dialogue & Discourse* 13(1), pp. 1–40. DOI: https://doi.org/10.5210/dad.2022.101. This publication has been under review at the submission date of the dissertation.
- Miehle, J., Ostler, D., Gerstenlauer, N., and Minker, W. (2017b). "The next step: intelligent digital assistance for clinical operating rooms". In: *Innovative Surgical Sciences* 2.3, pp. 159–161. DOI: https://doi.org/10.1515/iss-2017-0034.
- Miehle, J., Wagner, N., Minker, W., and Ultes, S. (2021c). "Culture-Aware Dialogue Management for Conversational Assistants". In: *Conversational Dialogue Systems for the Next Decade*. Vol. 704. Lecture Notes in Electrical Engineering. Springer Singapore, pp. 103–115. DOI: https://doi.org/10.1007/978-981-15-8395-7_8.
- Miehle, J., Wieluch, S., Minker, W., and Ultes, S. (2021d). "Decide or Delegate: How Script Knowledge Based Conversational Assistants Should Act in Inconclusive Situations". In: Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking (ICDCN 2021). Nara, Japan: Association for Computing Machinery, pp. 69–73. DOI: https://doi.org/10.1145/3427477.3428185.
- Miehle, J., Yoshino, K., Pragst, L., Ultes, S., Nakamura, S., and Minker, W. (2016). "Cultural Communication Idiosyncrasies in Human-Computer Interaction". In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). Los Angeles, USA: Association for Computational Linguistics, pp. 74–79.
- Pragst, L., Miehle, J., Minker, W., and Ultes, S. (2017a). "Challenges for adaptive dialogue management in the KRISTINA project". In: ISIAA 2017: Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents. Glasgow, UK: ACM, pp. 11–14. DOI: https://doi.org/10.1145/3139491. 3139508.
- Zhang, Z., Miehle, J., Matsuda, Y., Fujimoto, M., Arakawa, Y., Yasumoto, K., and Minker, W. (2021). "Exploring the Impacts of Elaborateness and Indirectness in a Behavior Change Support System". In: *IEEE Access* 9, pp. 74778–74788. DOI: https://doi. org/10.1109/ACCESS.2021.3079473.

List of Additional Publications

- Callejas, Z., Lugrin, B., Martin, J.-C., McTear, M., and Miehle, J. (2021). "Adaptive Systems for Multicultural and Ageing Societies". In: *Multimodal Agents for Ageing and Multicultural Societies: Communications of NII Shonan Meetings*. Singapore: Springer Singapore, pp. 1–20. DOI: https://doi.org/10.1007/978-981-16-3476-5_1.
- Miehle, J., Minker, W., André, E., and Yoshino, K., eds. (2021b). Multimodal Agents for Ageing and Multicultural Societies: Communications of NII Shonan Meetings. Singapore: Springer Singapore. DOI: https://doi.org/10.1007/978-981-16-3476-5.
- Miehle, J., Pragst, L., Minker, W., and Ultes, S. (2019b). "Enabling Dialogue Management with Dynamically Created Dialogue Actions". In: arXiv:1907.00684. URL: http: //arxiv.org/abs/1907.00684.
- Pragst, L., Miehle, J., Ultes, S., and Minker, W. (2016). "Automatic Modification of Communication Style in Dialogue Management". In: Proceedings of the INLG 2016 Workshop on Computational Creativity in Natural Language Generation. Edinburgh, UK: Association for Computational Linguistics, pp. 36–40.
- Ultes, S., Miehle, J., and Minker, W. (2019). "On the Applicability of a User Satisfaction-Based Reward for Dialogue Policy Learning". In: Advanced Social Interaction with Agents. Vol. 510. Lecture Notes in Electrical Engineering. Springer International Publishing, pp. 211–217. DOI: https://doi.org/10.1007/978-3-319-92108-2_22.