Dialogue & Discourse 13(1) (2022) 1–40 doi: 10.5210/dad.2022.101

When to Say What and How: Adapting the Elaborateness and Indirectness of Spoken Dialogue Systems

Juliana Miehle JULIANA.MIEHLE@UNI-ULM.DE
Institute of Communications Engineering
Ulm University

Wolfgang Minker WOLFGANG.MINKER@UNI-ULM.DE
Institute of Communications Engineering
Ulm University

Stefan Ultes STEFAN.ULTES@DAIMLER.COM
Mercedes-Benz AG Research & Development
Sindelfingen

Editor: Amanda Stent and Barbara Di Eugenio

Submitted 01/2021; Accepted 03/2022; Published online 04/2022

Abstract

With the aim of designing a spoken dialogue system which has the ability to adapt to the user's communication idiosyncrasies, we investigate whether it is possible to carry over insights from the usage of communication styles in human-human interaction to human-computer interaction. In an extensive literature review, it is demonstrated that communication styles play an important role in human communication. Using a multi-lingual data set, we show that there is a significant correlation between the communication style of the system and the preceding communication style of the user. This is why two components that extend the standard architecture of spoken dialogue systems are presented: 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. We consider the communication styles *elaborateness* and *indirectness* as it has been shown that they influence the user's satisfaction and the user's perception of a dialogue. We present a neural classification approach based on supervised learning for each task. Neural networks are trained and evaluated with features that can be automatically derived during an ongoing interaction in every spoken dialogue system. It is shown that both components yield solid results and outperform the baseline in form of a majority-class classifier.

Keywords: Communication Styles, Dialogue Management, Interactive Adaptation, Supervised Learning, Classification, Neural Approach

1. Introduction

Even though intelligent assistants like Amazon Alexa, Apple Siri, Google Assistant or Microsoft Cortana are becoming increasingly popular, they do not consider different communication styles to adapt their behaviour. Current systems focus on content (*what* is said) rather than formulation (*how* is it said). However, it has been shown that people adapt their interaction styles to one another across many levels of utterance production when communicating.

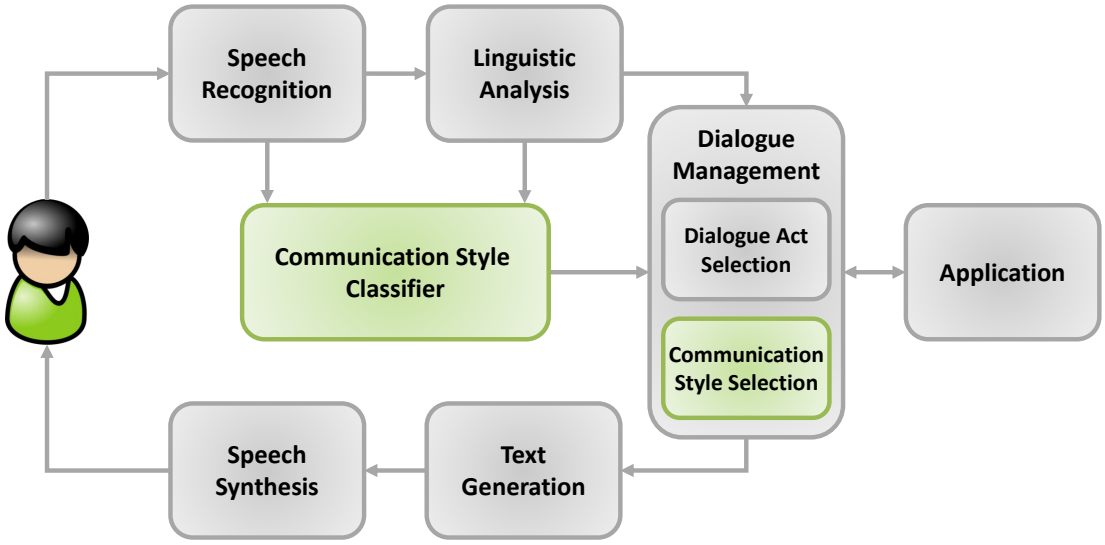

Figure 1: The standard architecture of spoken dialogue systems is extended by two components: 1) a communication style classifier that automatically identifies the user communication style and 2) a communication style selection module that selects an appropriate system communication style.

The goal of this article is to investigate if it is possible to carry over insights from the usage of communication styles in human-human interaction to human-computer interaction. Building upon a long history of communication research for human-human interaction, we investigate if the used communication styles of the user and the system influence each other in human-computer interaction. To demonstrate the principal usage of communication style adaptation within a spoken dialogue system, the problem of identifying the user's communication style and the problem of selecting the system's communication style are framed as classification problems (see Figure 1). Thus, the main contributions of this article are as follows:

1. Comprehensive overview over the general field of communication styles literature for human-human interaction

2. Introduction to interactive adaptation for human-human and human-computer interaction

3. Analysing the correlation between user and system communication style in human-computer interaction on a multi-lingual data set

4. A communication style classifier that automatically identifies the user communication style using supervised learning

5. A communication style selection module that selects an appropriate communication style of the system response using supervised learning

Various studies suggest that adapting the communication styles of spoken dialogue systems to the individual users in a similar way to what humans do will lead to more natural interactions (Stenchikova and Stent, 2007; Reitter et al., 2006; Mairesse and Walker, 2010). The work described in this paper builds upon and extends work published in (Miehle et al., 2020) and considers the communication styles *elaborateness* and *indirectness*. Pragst et al. (2019) have shown that both styles influence the user's perception of a dialogue and are therefore valuable candidates for adaptive dialogue management. Miehle et al. (2018b) have shown that varying the *elaborateness* and *indirectness* of a spoken user interface influences the user's satisfaction and the user's perception of the dialogue. The *elaborateness* thereby 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.

The structure of the paper is as follows: In Section 2, we introduce communication styles and interactive adaptation in human-human and human-computer interaction. Related work on the adaptation and recognition of communication styles in human-computer interaction is discussed in Section 3. In Section 4, the corpus used in this work is described and the correlation between user and system communication style is investigated. We present the user communication style classifier in Section 5 and the system communication style selection in Section 6, before concluding in Section 7.

2. Communication Styles and Interactive Adaptation

In this section, we introduce communication styles and interactive adaptation in human-human and human-computer interaction. Showing that these aspects play an important role in human communication, we provide the background for our work on communication style adaptation in spoken dialogue systems. A summary of the references is provided in Table 1.

2.1 Communication Styles

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:

- 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.

Topic	References	
Communication styles	Bultman and Svarstad (2000) Grice (1975) Holtgraves (1986) Kroeger (2019) Madaio et al. (2017) Miehle et al. (2018a) Miehle et al. (2018b)	Miehle et al. (2016) Neuliep (2018) Pesch et al. (2015) Pragst et al. (2019) Searle (1975) Van Dolen et al. (2007)
Cultural models	Elliott et al. (2016) Hofstede (2009)	Kaplan (1966) Lewis (2010)
Interactive adaptation in human-human interaction	Branigan et al. (2000) Brennan and Clark (1996) Burgoon et al. (1995) Garrod and Anderson (1987) Jungers et al. (2002) Levelt and Kelter (1982)	Nenkova et al. (2008) Niederhoffer and Pennebaker (2002) Pardo (2006) Pickering and Garrod (2004) Reitter et al. (2006) Schober (1993)
Interactive adaptation in human-computer interaction	Bell et al. (2003) Bergmann et al. (2015) Branigan and Pearson (2006) Branigan et al. (2010) Branigan et al. (2003) Brennan (1991) Brennan (1996) Brennan and Ohaeri (1994)	Coulston et al. (2002) Darves and Oviatt (2002) Doran et al. (2003) Koulouri et al. (2016) Oviatt et al. (2004) Pearson et al. (2006) Suzuki and Katagiri (2007)

Table 1: Summary of references on communication styles and interactive adaptation.

- Manner: Be perspicuous.
    1. Avoid obscurity of expression.
    2. Avoid ambiguity.
    3. Be brief (avoid unnecessary prolixity).
    4. Be orderly.

The listener, for his/her part, naturally assumes that an utterance follows the cooperative principle, i.e. he/she 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 from 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 and form the basis for our work on communication style adaptation.

One special 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 by 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/she 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/she 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/she 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/she 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/she can infer that speaker B cannot do both in one evening. As he/she is not able to perform the proposed act, he/she 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; Lewis, 2010). 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/she interacts, what is the goal of the interaction and which effect he/she wants to achieve. Miehle et al. (2016) investigated cultural differences in communication style preferences between the Germans and the Japanese. The results revealed that communication idiosyncrasies in human-human interaction may also be observed during human-computer interaction in a spoken dialogue system context. Moreover, Miehle et al. (2018a) presented another study examining five European cultures whose communication styles are much more alike than the German and Japanese communication idiosyncrasies. The study explored not only the influence of the user's culture but also of the gender, the frequency of use of speech based assistants as well as the system's role. The results showed that the system's role significantly influences the user's preference in the system's communication style whereas the frequency of use of speech based assistants has no influence. Moreover, the findings showed differences among the cultures and, depending on the culture, there are gender differences with respect to the user's preference in the system's communication style.

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 showed 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 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 extent of sometimes ignoring the task at hand, making jokes, showing understanding, using emoticons and rewarding the input of the customers). The results showed 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 showed that a collaborative communication style enhances the clients' knowledge and thus positively influences the treatment outcomes. It was not required that the given information was exhaustive, but it was required that the physician clearly communicated essential details (i.e. 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 was that the physician communication style varied between the initial visit and follow-up visits, even with the same patient. The perception of direct and indirect speech was investigated by Holtgraves (1986). The results indicated 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 showed that, in comparison with friend tutors, stranger tutors provided more positive feedback and used more indirect instructions. Moreover, tutees attempted and solved more problems if the stranger tutor used 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.

Pragst et al. (2019) investigated the applicability of *elaborateness* and *indirectness* as possibilities for adaptation in spoken dialogue systems. In order to do so, they compared four conditions: a high level of *elaborateness* and *indirectness* and an involved user, a low level of *elaborateness* and *indirectness* and an involved user, a high level of *elaborateness* and *indirectness* and a distracted user, and a low level of *elaborateness* and *indirectness* and a distracted user. The results showed multiple significant differences between the two levels of *elaborateness* and *indirectness* and that the assessment changes depending on the situation of the user. Hence, it is concluded that *elaborateness* and *indirectness* influence the user's perception of a dialogue and are therefore valuable candidates for adaptive dialogue management. Miehle et al. (2018b) addressed the issues of 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*. The results showed that the system's communication style influences the user's satisfaction and the user's perception of the dialogue and that there is no general preference in the system's communication style, i.e. not every participant preferred the same communication style. Based on that, we consider the *elaborateness* and *indirectness* to be highly suitable for our adaptation approach.

2.2 Interactive Adaptation

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) reviewed a broad range of interaction adapta-

tion theories and models and presented 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, was outlined in (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.

2.2.1 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 showed that a question's surface form can affect the format of the answer given in the way that answers tend to match 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). It was shown that speakers adopted similar forms of descriptions, suggesting that interlocutors adapt their description styles to one another. 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 suggested 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 offered 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 was the lack of a relationship between the perceived interaction quality and the degree of linguistic style matching. Nenkova et al. (2008) presented a corpus study examining entrainment in the use of high frequency words (i.e. the most common words in the corpus). The results showed 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.

Syntactic adaptation has been investigated by Branigan et al. (2000). It was examined whether speakers in a dialogue tend to coordinate the syntactic structures of their contributions, irrespective of lexical and semantic content. The results revealed 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

dialogues for syntactic repetitions. Positive effects were found in both corpora, both for within-speaker and between-speaker repetitions. However, the comparison of both corpora indicated that spontaneous conversation shows significantly less repetitions 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 sentence than a fast sentence, and significantly shorter following a fast sentence than a slow sentence, but the speakers were also influenced by their own preferred production rate. Therefore, the authors concluded that both the preferred rate and the rate of the previously heard sentence 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 showed 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.

2.2.2 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 showed both differences and similarities between a simulated computer partner and a human partner. There were significantly more acknowledgements, first-person 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, regardless of whether they believed they were talking to a computer or another person. 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 showed 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 was shown that people adopted the terms of their computer partners during text-based 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 showed 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 revealed that the children's response latencies differed when they conversed with an animated character that spoke with the extrovert versus 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. In (Coulston et al., 2002), the amplitude convergence in the children's conversational speech with animated personas was investigated. It was 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 was 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 was no perfect match, the children were not doing mimicry. Bell et al. (2003) investigated whether people adapt their speaking rate while interacting with an animated character. The results confirmed that the users adapted to the speaking rate of the system, even if the subjects afterwards stated that they had 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 participants 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 demonstrated syntactic alignment for both conditions and suggested 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) showed that the users' lexical alignment is influenced by their expectations about a system. When users believed the system to be unsophisticated and restricted in capability, they adapted their language to match the system's language more than when they believed the system to be sophisticated and capable. This tendency was unaffected by the actual behaviour that the system exhibited. In (Branigan and Pearson, 2006), the findings of the studies were summarised and it was 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 were discussed in (Branigan et al., 2010). Bergmann et al. (2015) explored lexical and gestural alignment with real and virtual humans. It was shown that adaptation takes place regarding communicative features (lexical alignment) as well as features without obvious communicative function (handedness alignment).

2.3 Summary

Communication styles play an important role in human communication. We have introduced the theoretical background and the definitions of communication styles in general and for the *elaborateness* and *indirectness* in particular. These definitions are used throughout this work for annotations and classifications. Furthermore, we have provided a broad review of studies investigating the phenomenon of 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: 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, in the following we concentrate on this aspect. Our aim is to recognise the user's *elaborateness* and *indirectness* and adapt the system communication style accordingly.

3. Adaptation and Recognition of Communication Styles

In this section, related work on the adaptation and recognition of communication styles in human-computer interaction will be discussed. A summary of the references is provided in Table 2.

3.1 Adaptation of Communication Styles in Human-Computer Interaction

Various studies suggest to adapt spoken dialogue systems to the users in a way similar to how 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. We elaborate on this 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

Topic	References	
Development of computer personalities	Aly and Tapus (2016) André et al. (2000) Irfan et al. (2020) Isbister and Nass (2000) Mairesse and Walker (2010) Mairesse and Walker (2011)	Moon and Nass (1996) Nass et al. (1995) Oraby et al. (2018) Smestad and Volden (2019) Tapus and Mataric (2008)
Style variation	De Jong et al. (2008) Gupta et al. (2007) Hofs et al. (2010) Johnson et al. (2004) Kruijff-Korbayová et al. (2008)	Porayska-Pomsta and Mellish (2004) Wang et al. (2005) Whittaker et al. (2003) Wilkie et al. (2005)
Style adaptation	Ball and Breese (2000) Brockmann et al. (2005) Buschmeier et al. (2009) Hoegen et al. (2019)	Hu et al. (2018) Stenchikova and Stent (2007) Walker et al. (2007)
Elaborateness recognition	Di Buccio et al. (2014)	Gharouit and Nfaoui (2017)
Indirectness recognition	Adel and Schütze (2017) Aubakirova and Bansal (2016) Danescu-Niculescu-Mizil et al. (2013) Dral et al. (2011) Forbes-Riley and Litman (2011)	Goel et al. (2019) Liscombe et al. (2005) Prokofieva and Hirschberg (2014) Ulinski et al. (2018)

Table 2: Summary of references on the adaptation and recognition of communication styles in human-computer interaction.

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 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 in to its body, and did not ever 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 except 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 was 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 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 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 Wizard-of-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 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 classification 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. (2019) 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 2.1 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 2.1) 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 a 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 detection (Liscombe et al., 2005; Dral et al., 2011; Forbes-Riley and Litman, 2011; Adel and Schütze, 2017).

3.3 Summary

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 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. Therefore, we argue that the decision about 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. Regarding the recognition of *elaborateness* and *indirectness* in spoken dialogue systems, only little previous work has been done. For *elaborateness*, only queries submitted to search engines have been examined, and for *indirectness*, merely different categories have been explored. In contrast, in

this work, *indirectness* is classified in a more global way and not based on fixed structures or key words, and *elaborateness* is classified in spoken language.

4. Investigating the Correlation between User and System Communication Style

In order to analyse whether the communication style of the system is correlated to the communication style of the user, we have created a corpus[1] with annotations for *elaborateness* and *indirectness* for each user and system dialogue act. For the communication style annotations, we have used the definitions presented in Section 2.1.

Our 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. For scenarios where medical knowledge is required, the participant who takes the role of the system is legally qualified to practise a medical profession, i.e. is a trained medical doctor or nurse. 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 3. Along with the dialogue acts, the respective utterances are also added to the data set. An example dialogue is shown in Table 4. Overall, the corpus covers 258 dialogues containing 2,880 turns and 7,930 annotated dialogue actions. The dialogues are in four different languages: German, Polish, Spanish and Turkish. The language distribution is shown in Table 5. 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 were identical for the different languages. 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 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 *indirectness* and *elaborateness* would be: "It will rain." The higher the rating for *indirectness*, the more hidden are the intentions of the speaker (2 = slightly indirect, 5 = extremely indirect). The higher the rating for *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. An example dialogue with annotated *elaborateness* (E) and *indirectness* (I) scores is shown in Table 4.

Each dialogue act was annotated by three different raters. They were instructed with annotated sample dialogues. Moreover, uncertainties were discussed in a weekly meeting where the raters talked about concrete annotation sentences and clarified how to understand a sentence in the specific context. However, in these meetings no decision was made regarding which value to annotate, only

[1] Unfortunately, it is not possible to publish the corpus due to privacy reasons.

Dialogue acts	
Accept	PersonalApologise
Acknowledge	PersonalGreet
Advise	PersonalBye
AfternoonGreet	PersonalThank
AfternoonBye	ReadNewspaper
AnswerThank	Reject
AskMood	RepeatPreviousUtterance
AskPlans	RephrasePreviousUtterance
AskTask	Request
AskWellBeing	RequestAdditionalInfo
CheerUp	RequestMissingInfo
Console	RequestNewspaper
Declare	RequestReasonForEmotion
EveningGreet	RequestRepeat
EveningBye	RequestRephrase
ExplicitlyConfirmRecognisedInput	RequestWeather
ImplicitlyConfirmRecognisedInput	ShareJoy
IndividualisticallyOrientedMotivate	ShowWeather
MeetAgainBye	SimpleApologise
MorningGreet	SimpleGreet
MorningBye	SimpleMotivate
Obligate	SimpleBye
Order	SimpleThank
PersonalAnswerThank	

Table 3: List of predefined dialogue acts.

Role	Utterance	Annotation
User	Hello Kristina.	PersonalGreet, E: 1, I: 1
System	Hi, nice to meet you.	SimpleGreet, E: 3, I: 1
User	I have got a problem.	Declare, E: 1, I: 1
System	You can ask me everything.	Declare, E: 1, I: 1
User	Okay.	Acknowledge, E: 1, I: 1
	I have heard that old people should drink a lot and I have the impression that my father does not drink enough.	Declare, E: 3, I: 2
System	How much does he drink on average every day?	RequestMissingInfo, E: 1, I: 1
User	In the morning he likes a cup of coffee, at 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. Yes, and in the evening, sometimes a glass of beer, but not much more.	Declare, E: 4, I: 2
System	Do you have the impression that he has very dry skin?	RequestMissingInfo, E: 2, I: 1
User	Not really.	Reject, E: 2, I: 2
	Well, I always put cream on his hands and face and also on his head, so I couldn't say that now, but I would have to see how his arms and legs are.	Declare, E: 3, I: 2
	Would that be a sign that he is not drinking enough?	Request, E: 1, I: 2
System	It is a sign that he is not drinking enough.	Accept, E: 4, I: 1
User	Okay.	Acknowledge, E: 1, I: 1
System	Is he very dry in the mouth?	RequestMissingInfo, E: 1, I: 1
User	I've never paid attention to that before.	Declare, E: 2, I: 2
System	It is said that old people should drink about 1.3 litres a day.	Declare, E: 2, I: 2
User	I don't think he drinks that much.	Declare, E: 1, I: 1
System	Try to give him something he enjoys drinking.	Advise, E: 2, I: 1
User	Okay, I will try that.	Acknowledge, E: 3, I: 1
System	It might also be nice if you drink together. For example, you could have a cup of tea together in the afternoon.	Advise, E: 4, I: 3
User	That's a great idea!	Acknowledge, E: 3, I: 2
	Thank you very much!	SimpleThank, E: 2, I: 1
System	You're welcome.	AnswerThank, E: 1, I: 1
User	See you next time!	MeetAgainBye, E: 2, I: 1
System	Bye.	SimpleBye, E: 1, I: 1

Table 4: Example dialogue with annotated dialogue acts and *elaborateness/indirectness* scores.

	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 5: Language distribution of the dialogues in the annotated corpus, whereby D is the number of dialogues and DA is the number of dialogue acts.

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

Table 6: Class distribution of the annotated *elaborateness* and *indirectness* scores (median of the three ratings).

the meaning of ambiguous sentences was clarified so that the annotators could rate them afterwards on the basis of a common understanding. The class distribution of the annotated *elaborateness* and *indirectness* scores (median of the three ratings) is shown in Table 6. It can be seen that the 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 combined the classes 3, 4, and 5 to one new class, reducing the corpus to three classes. For *indirectness*, the annotation showed 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 into one indirect class for binary classification.

Elaborateness (5 classes)				
	R1/R2	R1/R3	R2/R3	Av.
κ	0.560	0.515	0.516	**0.530**
ρ	0.848	0.813	0.799	**0.820**
ICC				**0.934**
Elaborateness (3 classes)				
	R1/R2	R1/R3	R2/R3	Av.
κ	0.670	0.612	0.608	**0.630**
ρ	0.826	0.794	0.767	**0.796**
ICC				**0.916**
Indirectness (5 classes)				
	R1/R2	R1/R3	R2/R3	Av.
κ	0.315	0.423	0.368	**0.369**
ρ	0.387	0.504	0.442	**0.444**
ICC				**0.686**
Indirectness (3 classes)				
	R1/R2	R1/R3	R2/R3	Av.
κ	0.335	0.439	0.382	**0.385**
ρ	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**
ρ	0.377	0.500	0.440	**0.439**
ICC				**0.701**

Table 7: 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.

	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*

Table 8: 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.

We have analysed the quality of the annotated scores by use of the following measures: Cohen's Kappa κ, Spearman's rank correlation coefficient Rho ρ and the Intraclass Correlation Coefficient ICC. The results can be seen in Table 7. 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 we reduce the classes to three or two (in case of *indirectness*), we obtain a higher agreement while the correlation and the inter-rater reliability do not change significantly. Overall, we have a good inter-rater reliability for both communication styles given the difficulty of the annotation task.

In order to use the communication style annotations as target for our classification tasks, we need a final score to be calculated from the three ratings. As we have applied an ordinal scale for the ratings, we have used the median of the three ratings. According to Stevens (1946), this is the appropriate measure for ordinal scales.

Using this corpus, we analysed whether the communication style of the speaker who assumed the role of the system (hereafter referred to as system) is correlated to the communication style of the 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. In Section 2, it was already shown that humans adapt their communication styles during an interaction. However, we want to show that this is also the case in the current setting. In order to do so, we extracted 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. One exchange contains up to five consecutive user actions U and up to four consecutive system actions S. Therefore, we analysed 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}. The results in terms of Spearman's rank correlation coefficient Rho ρ for both *elaborateness* and *indirectness* in five, three and two classes can be seen in Table 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 9.

	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 9: 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.

Parameter	Grid
#Nodes	3, 6, 12, 24, 48, 96, 144, 192
#Epochs	50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000
Optimiser	adadelta, adam, nadam, adagrad, sgd, rmsprop
Output function	sigmoid, softmax
Loss function	categorical crossentropy (CC), mean squared error (MSE)

Table 10: Grid of parameter values for the user communication style classifier.

It can be seen that there is a significant correlation for both *elaborateness* and *indirectness* for all four languages. However, the effect size for *elaborateness* varies between 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 speakers were identical for the different languages, we conclude that the discrepancy is due to differences in the languages/cultures.

5. User Communication Style Classifier

As we have shown that there is a significant correlation between the user and the system communication style, in (Miehle et al., 2020) we presented a classification approach to automatically estimate the user's communication style during an ongoing dialogue. The estimated communication style can then be used in the dialogue management to adapt the system behaviour to the user, as depicted in Figure 1. We utilise a supervised learning approach with a multi-layer perceptron (MLP) classifier with one hidden layer. A comparison with a support vector machine (SVM) classifier and a recurrent neural network (RNN) classifier consisting of two long short-term memory (LSTM) layers can be found in (Miehle et al., 2020). To mitigate overfitting, the neural net is trained and evaluated with a 10-fold cross-validation setting on the German part of the corpus described in Section 4. 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 10. To account for the imbalanced data during the grid search optimisation, the Unweighted Average Recall (UAR) was used, which is the arithmetic average of all class-wise recalls. The best parameter values that were chosen as final setting for the models are shown in Ta-

Features	#Nodes	#Epochs	Optimiser	Output function	Loss function
DA	48	250	nadam	sigmoid	CC
DA+G	48	250	nadam	softmax	MSE
U	48	50	adadelta	softmax	CC
U+DA	48	50	adadelta	softmax	CC
U+DA+G	48	50	adadelta	softmax	CC
UB	48	50	adadelta	sigmoid	CC
UB+DA	48	50	adadelta	sigmoid	CC
UB+DA+G	48	50	adadelta	softmax	CC
WE	144	250	nadam	sigmoid	CC
WE+DA	48	250	nadam	sigmoid	CC
WE+DA+G	144	250	nadam	sigmoid	CC
DA	48	250	nadam	softmax	MSE
DA+G	48	250	nadam	sigmoid	MSE
U	48	50	adagrad	sigmoid	CC
U+DA	48	50	adagrad	sigmoid	CC
U+DA+G	48	50	adagrad	sigmoid	CC
UB	48	50	adadelta	softmax	CC
UB+DA	48	50	adadelta	softmax	CC
UB+DA+G	48	50	adadelta	softmax	CC
WE	144	1000	nadam	sigmoid	CC
WE+DA	48	350	nadam	sigmoid	CC
WE+DA+G	48	350	nadam	sigmoid	CC
DA	48	250	nadam	softmax	CC
DA+G	48	250	nadam	softmax	CC
U	48	50	adagrad	sigmoid	CC
U+DA	48	50	adagrad	sigmoid	CC
U+DA+G	48	50	adagrad	sigmoid	CC
UB	48	50	adagrad	sigmoid	CC
UB+DA	48	50	adagrad	sigmoid	CC
UB+DA+G	48	50	adagrad	sigmoid	CC
WE	48	350	adadelta	sigmoid	MSE
WE+DA	48	350	adadelta	sigmoid	CC
WE+DA+G	48	350	adadelta	sigmoid	CC

Table 11: The best parameter values for the user communication style classifier (top: 3-class *elaborateness*, middle: 3-class *indirectness*, bottom: 2-class *indirectness*).

		Elaborateness (3 classes)	*Indirectness* (3 classes)	*Indirectness* (2 classes)
	UAR	0.840	0.555	0.753
	ACC	0.838	0.832	0.848
DA	F1	0.838	0.582	0.761
	κ	0.749	0.467	0.527
	ρ	0.862	0.523	0.541

Table 12: Classification results using dialogue act features (DA) in terms of Unweighted Average Recall (UAR), Accuracy (ACC), F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ.

ble 11. We define a dialogue act feature set and investigate how grammatical and linguistic features influence the performance.

5.1 The Dialogue Act Features

As a first approach, the MLP was trained using only *dialogue act features (DA)* that can directly be derived from the data[2]. These features contain the dialogue act and the amount of words in the utterance of the corresponding dialogue act. 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 1). 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. The results are shown in Table 12.

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 75%. The results for *indirectness* clearly show the difficulty of the task, as was already shown by the corpus creation. There, it was quite hard for the annotators to distinguish between different levels of *indirectness* so that the class distribution of *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.

5.2 The Contribution of Grammatical and Linguistic Features

To address the question of whether *grammatical features (G)* 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, 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

[2]During our experiments, we also tested 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), but this led to worse results.

		Elaborateness (3 classes)	*Indirectness* (3 classes)	*Indirectness* (2 classes)
DA+G	UAR	0.841	0.558	0.753
	ACC	0.840	0.834	0.848
	F1	0.839	0.588	0.761
	κ	0.753	0.470	0.526
	ρ	0.864	0.521	0.540

Table 13: Classification results using dialogue act features as well as grammatical features (DA+G) in terms of Unweighted Average Recall (UAR), Accuracy (ACC), F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ.

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 13. 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 greatly contribute to the overall classification performance. In order to encode linguistic features, a Bag-of-Words (BoW) approach was used in combination with unigrams (U), unigrams and bigrams (UB) and word embeddings (WE). Using BoW and the corpus presented in Section 4, two distinct vocabularies were created:

- The BoW-U vocabulary contains every word occurring in the database.
- 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 led 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)[3]. 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 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.

[3]During our experiments, we also tested self-trained word vectors, but this led to worse results.

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

Table 14: Classification results using linguistic features encoded as unigrams (U) or unigrams and bigrams (UB) (separately and in combination with dialogue act features and grammatical features) in terms of Unweighted Average Recall (UAR), Accuracy (ACC), F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ.

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

Table 15: Classification results using linguistic features encoded as word embeddings (WE) (separately and in combination with dialogue act features and grammatical features) in terms of Unweighted Average Recall (UAR), Accuracy (ACC), F1-Score, Cohen's Kappa κ and Spearman's rank correlation coefficient Rho ρ.

In addition to using these linguistic feature sets individually, we used them in combination with dialogue act features (DA) and grammatical features (G). The results with the U and UB feature sets are shown in Table 14, the results with the WE feature set can be found in Table 15.

For *elaborateness*, the best results are achieved with the dialogue act feature set. Grammatical and linguistic features do not seem to have any effect on the classification performance. This leads to the conclusion that for *elaborateness*, analysing the utterance length dependent on the dialogue act seems to contain enough information to achieve good classification performance. For *indirectness*, the overall performance improves by using linguistic information encoded as word embeddings. This in combination with grammatical and dialogue act features (WE+DA+G) led to UARs of 59% and 76% for the estimation of the *indirectness* using three classes and two classes, respectively. Using the BoW approach in combination with unigrams and bigrams did not improve the classification performance.

To sum up, linguistic features are beneficial for the estimation of *indirectness*, but not for the estimation of *elaborateness*. For the latter, the dialogue act features (i.e. the dialogue act and the amount of words in the utterance) seem to be sufficient. All features can be automatically recognised during an ongoing interaction in any spoken dialogue system, without any prior annotation. Hence, this user communication style classifier can be used as additional component for any spoken dialogue system.

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

Table 16: Class distribution of the annotated *elaborateness* and *indirectness* scores for the 2,880 dialogue turns.

6. System Communication Style Selection

In this section, the task of automatically selecting the system communication style during an ongoing interaction with a spoken dialogue system is addressed. As depicted in Figure 1, this is part of the dialogue management so that it not only decides *what* is said next, but also *how*. We suggest 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).

As we obtained promising results for the classification of the user communication styles by using a supervised learning approach with a multi-layer perceptron (see Section 5), we think that this approach is also suitable for the task at hand. Hence, we utilise a MLP classifier with one hidden layer. To mitigate overfitting, 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 4. The class distribution for both communication styles is shown in Table 16. 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 17. To account for the imbalanced data during the grid search optimisation, the UAR is used. The best parameter values chosen as final setting for the models are shown in Table 18. For each of the 2,880 dialogue turns, we extracted the following features:

- 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 our experiments, we also tested part-of-speech tags and sentence embeddings (based on the respective utterances), though without improvement of the results. Note that all features can be automatically derived during an ongoing interaction in every 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 act selection in the dialogue manager and the user communication styles may be classified by use of the communication style classifier described in Section 5 (see Figure 1). However, in order to focus on the performance of the system communication style selection module

Parameter	Grid
#Nodes	10, 25, 50
#Epochs	10, 50, 100, 200, 500
Optimiser	adadelta, adam, nadam, adagrad
Output function	sigmoid, softmax
Loss function	categorical crossentropy (CC), mean squared error (MSE)

Table 17: Grid of parameter values for the system communication style selection.

Features	#Nodes	#Epochs	Optimiser	Output function	Loss function
S+L	50	200	adagrad	softmax	CC
W+U+CS+L	50	10	adagrad	sigmoid	CC
S+CS+L	25	100	adam	sigmoid	CC
S+W+U+CS+L	50	100	adadelta	sigmoid	CC
S+L	50	200	nadam	sigmoid	MSE
W+U+CS+L	50	100	adagrad	softmax	CC
S+CS+L	25	500	adam	softmax	CC
S+W+U+CS+L	50	100	adam	sigmoid	CC
S+L	25	10	nadam	softmax	CC
W+U+CS+L	50	100	adagrad	softmax	MSE
S+CS+L	10	10	nadam	softmax	CC
S+W+U+CS+L	25	10	nadam	softmax	CC

Table 18: The best parameter values for the system communication style selection (top: 3-class *elaborateness*, middle: 3-class *indirectness*, bottom: 2-class *indirectness*).

and avoid errors caused by the user communication style classification, the ground truth labels for the communication styles from Section 4 have been used for the following analysis. The number of classes of the user communication style has been adjusted to the number of classes of the system communication style (i.e. either three or two classes based on the task).

The results are shown in Table 19. 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 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 using 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.

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

Table 19: Classification results for the system communication style selection using different feature sets in terms of Unweighted Average Recall (UAR), Accuracy (ACC) and F1-Score.

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

Table 20: Classification results for the system communication style selection in terms of Unweighted Average Recall (UAR), Accuracy (ACC) and F1-Score of the overall test set and the individual languages.

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

Table 21: 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 Unweighted Average Recall (UAR), Accuracy (ACC) and F1-Score.

When dividing the test set based on the languages, we can see that the classification works differently for the individual languages (see Table 20). For the 3-class *elaborateness*, we achieve an UAR of 57% for German, 58% for Polish, 77% for Spanish and 51% for Turkish. 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 cultural differences, as already revealed by studies like (Miehle et al., 2016) and (Miehle et al., 2018a). For example, in the latter it has been shown that Spanish people like significantly more *elaborateness* than the other European cultures that have been investigated. Since the *elaborateness* is more dominant in the training data (see Table 16), the trained classifier better suits Spanish than the other cultures. However, this result might also be due to our 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 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 21. Hence, our trained system communication style selection module clearly outperforms a model which is just mimicking the user communication style at each turn.

7. Conclusion and Future Directions

In this work, we have introduced communication styles and interactive adaptation for human-human and human-computer interaction. In a broad literature review, we have demonstrated that those aspects play an important role in human communication. 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. Throughout this work, we

have focused on adaptation based on communication styles (i.e. how to formulate the utterance) due to the following reasons:

1. There is a strong theoretical background that allows to generalise the adaptation across multiple domains and applications.

2. The information required for this adaptation can be obtained during ongoing interactions.

3. There is a verified influence of communication style adaptation on user satisfaction (Miehle et al., 2018b).

Using a multi-lingual data set with *elaborateness* and *indirectness* annotations, we have shown that there exists a significant correlation between the communication style of the system and the preceding communication style of the user. Moreover, we have augmented the standard architecture of spoken dialogue systems with the ability to adapt to the user's communication idiosyncrasies. It 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 of the system response. We have presented a neural classification approach for each task.

For the user communication style classifier, a supervised learning approach has been utilised in order to estimate the user's *elaborateness* and *indirectness*. We have trained and evaluated a multi-layer perceptron with features that can be automatically derived during an ongoing interaction in every spoken dialogue system. We have tested different feature sets as input for our classifier and performed classification in two and three classes. 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. 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.

For the system communication style selection, we have used the same supervised learning approach. 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 and the user communication styles), we trained and evaluated a multi-layer perceptron. As for the first task, these features can be automatically recognised during an interaction in every spoken dialogue system. 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*.

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. So far, we have shown that both components (evaluated separately) yield solid results. In future work, we plan to conduct an evaluation of the overall system with real users. The user study presented in (Miehle et al., 2018b) showed that the system communication style selection has a direct influence on the user satisfaction. Based on these results, we will investigate whether a targeted increase in user satisfaction is achieved by our system. In order to do so, a spoken dialogue system that incorporates the user communication style classifier and the system communication style selection

will be compared with a system that does not contain these modules. Moreover, we will consider a reinforcement learning approach (instead of the herein presented supervised learning approach) as the system communication style selection in a spoken dialogue system might also depend on what the system and the user want to achieve in the long run.

Acknowledgements

This work is part of a project that has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 645012. We thank our colleagues from the University of Tübingen, the German Red Cross in Tübingen and semFYC in Barcelona for organising and carrying out the corpus recordings. Additionally, this work has received funding within the BMBF project "RobotKoop: Cooperative Interaction Strategies and Goal Negotiations with Learning Autonomous Robots" and the technology transfer project "Do it yourself, but not alone: Companion Technology for DIY support" of the Transregional Collaborative Research Centre SFB/TRR 62 "Companion Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG). We thank the anonymous reviewers and the editors for their constructive comments which helped us to improve the paper.

References

Heike Adel and Hinrich Schütze. 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*, pages 22–34, 2017.

Amir Aly and Adriana Tapus. Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction. *Autonomous Robots*, 40(2):193–209, 2016.

Elisabeth André, Thomas Rist, Susanne Van Mulken, Martin Klesen, and Stefan Baldes. The automated design of believable dialogues for animated presentation teams. *Embodied Conversational Agents*, pages 220–255, 2000.

Malika Aubakirova and Mohit Bansal. Interpreting neural networks to improve politeness comprehension. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2035–2041, 2016.

Gene Ball and Jack Breese. Emotion and personality in a conversational agent. *Embodied Conversational Agents*, pages 189–219, 2000.

Linda Bell, Joakim Gustafson, and Mattias Heldner. Prosodic adaptation in human-computer interaction. In *Proceedings of ICPHS*, volume 3, pages 833–836. Citeseer, 2003.

Kirsten Bergmann, Holly P. Branigan, and Stefan Kopp. Exploring the alignment space–lexical and gestural alignment with real and virtual humans. *Frontiers in ICT*, 2:7, 2015.

Holly P. Branigan and Jamie Pearson. Alignment in human-computer interaction. *How People Talk to Computers, Robots, and Other Artificial Communication Partners*, pages 140–156, 2006.

Holly P. Branigan, Martin J. Pickering, and Alexandra A. Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25, 2000.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, Janet F. McLean, and Clifford I. Nass. 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*, pages 186–191. Lawrence Erlbaum Associates, 2003.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368, 2010.

Susan E. Brennan. Conversation with and through computers. *User Modeling and User-Adapted Interaction*, 1(1):67–86, 1991.

Susan E. Brennan. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96:41–44, 1996.

Susan E. Brennan and Herbert H. Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482, 1996.

Susan E. Brennan and Justina O. Ohaeri. Effects of message style on users' attributions toward agents. In *Conference Companion on Human Factors in Computing Systems*, pages 281–282, 1994.

Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. Modelling alignment for affective dialogue. In *Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling*, 2005.

Dara C. Bultman and Bonnie L. Svarstad. Effects of physician communication style on client medication beliefs and adherence with antidepressant treatment. *Patient Education and Counseling*, 40(2):173 – 185, 2000.

Judee K. Burgoon, Lesa A. Stern, and Leesa Dillman. *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press, 1995.

Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. An alignment-capable microplanner for natural language generation. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 82–89, 2009.

Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic Press, 1977.

Rachel Coulston, Sharon Oviatt, and Courtney Darves. Amplitude convergence in children's conversational speech with animated personas. In *Seventh International Conference on Spoken Language Processing*, pages 2689–2692, 2002.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 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)*, pages 250–259, Sofia, Bulgaria, 2013. Association for Computational Linguistics.

Courtney Darves and Sharon Oviatt. Adaptation of users' spoken dialogue patterns in a conversational interface. In *Proceedings of the 7th International Conference on Spoken Language Processing*, pages 561–564, 2002.

Markus De Jong, Mariët Theune, and Dennis Hofs. 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*, pages 207–214, 2008.

Emanuele Di Buccio, Massimo Melucci, and Federica Moro. Detecting verbose queries and improving information retrieval. *Information Processing & Management*, 50(2):342–360, 2014.

Christine Doran, John Aberdeen, Laurie Damianos, and Lynette Hirschman. Comparing several aspects of human-computer and human-human dialogues. In *Current and New Directions in Discourse and Dialogue*, pages 133–159. Springer, 2003.

Jeroen Dral, Dirk Heylen, and Rieks op den Akker. *Detecting Uncertainty in Spoken Dialogues: An Exploratory Research for the Automatic Detection of Speaker Uncertainty by Using Prosodic Markers*, pages 67–77. Springer Netherlands, 2011.

Candia Elliott, R. Jerry Adams, and Suganya Sockalingam. Multicultural toolkit: Toolkit for cross-cultural collaboration. Awesome Library. `http://www.awesomelibrary.org/multiculturaltoolkit.html`, 2016. Accessed: 2016-05-01.

Kate Forbes-Riley and Diane J. Litman. Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, 53(9):1115 – 1136, 2011.

Simon Garrod and Anthony Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218, 1987.

Kenza Gharouit and El Habib Nfaoui. A comparison of classification algorithms for verbose queries detection using babelnet. In *2017 Intelligent Systems and Computer Vision (ISCV)*, pages 1–5. IEEE, 2017.

Pranav Goel, Yoichi Matsuyama, Michael Madaio, and Justine Cassell. I think it might help if we multiply, and not add: Detecting indirectness in conversation. In *9th International Workshop on Spoken Dialogue System Technology*, pages 27–40. Springer Singapore, 2019.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning Word Vectors for 157 Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), 2018.

Herbert P. Grice. Logic and conversation. In *Speech Acts*, pages 41–58. Brill, 1975.

Swati Gupta, Marilyn A. Walker, and Daniela M. Romano. How rude are you?: Evaluating politeness and affect in interaction. In *International Conference on Affective Computing and Intelligent Interaction*, pages 203–217. Springer, 2007.

Rens Hoegen, Deepali Aneja, Daniel McDuff, and Mary Czerwinski. An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 111–118, 2019.

Dennis Hofs, Mariët Theune, and Rieks op den Akker. Natural interaction with a virtual guide in a virtual environment. *Journal on Multimodal User Interfaces*, 3(1-2):141–153, 2010.

Geert Hofstede. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Sage, 2009.

Thomas Holtgraves. Language structure in social interaction: Perceptions of direct and indirect speech acts and interactants who use them. *Journal of Personality and Social Psychology*, 51(2): 305, 1986.

Zhichao Hu, Jean E. Fox Tree, and Marilyn A. Walker. Modeling linguistic and personality adaptation for natural language generation. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 20–31, 2018.

Bahar Irfan, Anika Narayanan, and James Kennedy. Dynamic emotional language adaptation in multiparty interactions with agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8, 2020.

Katherine Isbister and Clifford Nass. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53(2):251–267, 2000.

W. Lewis Johnson, Paola Rizzo, Wauter Bosma, Sander Kole, Mattijs Ghijsen, and Herwin van Welbergen. Generating socially appropriate tutorial dialog. In *Affective Dialogue Systems*, pages 254–264. Springer Berlin Heidelberg, 2004.

Melissa K. Jungers, Caroline Palmer, and Shari R. Speer. Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, 1(2):21–35, 2002.

Robert B. Kaplan. Cultural thought patterns in inter-cultural education. *Language Learning*, 16 (1-2):1–20, 1966.

Theodora Koulouri, Stanislao Lauria, and Robert D. Macredie. Do (and say) as i say: Linguistic adaptation in human–computer dialogs. *Human–Computer Interaction*, 31(1):59–95, 2016.

Paul R. Kroeger. *Analyzing meaning: An introduction to semantics and pragmatics*. Language Science Press, 2019.

Ivana Kruijff-Korbayová, Ciprian Kukina, Gerstenberger Olga, and Jan Schehl. Generation of output style variation in the sammie dialogue system. In *Proceedings of the Fifth International Natural Language Generation Conference*, pages 129–137, 2008.

Willem J. M. Levelt and Stephanie Kelter. Surface form and memory in question answering. *Cognitive Psychology*, 14(1):78–106, 1982.

Richard D. Lewis. *When Cultures Collide: Leading Across Cultures*. Brealey, 2010.

Jackson Liscombe, Julia Hirschberg, and Jennifer J. Venditti. Detecting certainness in spoken tutorial dialogues. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, pages 1837–1840, 2005.

Michael Madaio, Justine Cassell, and Amy Ogan. 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*. Philadelphia, PA: International Society of the Learning Sciences, 2017.

François Mairesse and Marilyn A. Walker. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, 20 (3):227–278, 2010.

François Mairesse and Marilyn A. Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37(3):455–488, 2011.

Juliana Miehle, Koichiro Yoshino, Louisa Pragst, Stefan Ultes, Satoshi Nakamura, and Wolfgang Minker. Cultural communication idiosyncrasies in human-computer interaction. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 74–79, Los Angeles, USA, September 2016. Association for Computational Linguistics.

Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 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)*. European Language Resources Association (ELRA), 2018a.

Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 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)*, pages 165–172. ACM, 2018b.

Juliana Miehle, Isabel Feustel, Julia Hornauer, Wolfgang Minker, and Stefan Ultes. Estimating user communication styles for spoken dialogue systems. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 533–541. European Language Resources Association (ELRA), May 2020.

Youngme Moon and Clifford Nass. How "real" are computer personalities? psychological responses to personality types in human-computer interaction. *Communication Research*, 23(6):651–674, 1996.

Clifford Nass, Youngme Moon, Brian J. Fogg, Byron Reeves, and Chris Dryer. Can computer personalities be human personalities? In *Conference Companion on Human Factors in Computing Systems*, pages 228–229, 1995.

Ani Nenkova, Agustin Gravano, and Julia Hirschberg. 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*, pages 169–172. Association for Computational Linguistics, 2008.

James W. Neuliep. *Intercultural Communication: A Contextual Approach*. SAGE, 2018.

Dat Quoc Nguyen, Dai Quoc Nguyen, Dang Duc Pham, and Son Bao Pham. 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*, pages 17–20, 2014.

Kate G. Niederhoffer and James W. Pennebaker. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360, 2002.

Shereen Oraby, Lena Reed, Shubhangi Tandon, T. S. Sharath, Stephanie Lukin, and Marilyn A. Walker. Controlling personality-based stylistic variation with neural natural language generators. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 180–190, 2018.

Sharon Oviatt, Courtney Darves, and Rachel Coulston. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(3):300–328, 2004.

Jennifer S. Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4):2382–2393, 2006.

Jamie Pearson, Jiang Hu, Holly P. Branigan, Martin J. Pickering, and Clifford I. Nass. 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*, pages 1177–1180, 2006.

Robin Pesch, Ricarda B. Bouncken, and Sascha Kraus. Effects of communication style and age diversity in innovation teams. *International Journal of Innovation and Technology Management*, 12(06):1–20, 2015.

Martin J. Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190, 2004.

Kaśka Porayska-Pomsta and Chris Mellish. Modelling politeness in natural language generation. In *International Conference on Natural Language Generation*, pages 141–150. Springer, 2004.

Louisa Pragst, Wolfgang Minker, and Stefan Ultes. Exploring the applicability of elaborateness and indirectness in dialogue management. In *Advanced Social Interaction with Agents : 8th International Workshop on Spoken Dialog Systems*, pages 189–198. Springer International Publishing, 2019.

Anna Prokofieva and Julia Hirschberg. Hedging and speaker commitment. In *Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland*, pages 10–13, 2014.

David Reitter, Frank Keller, and Johanna D. Moore. Computational modelling of structural priming in dialogue. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 121–124. Association for Computational Linguistics, 2006.

Michael F. Schober. Spatial perspective-taking in conversation. *Cognition*, 47(1):1–24, 1993.

John R. Searle. Indirect speech acts. In *Syntax and Semantics 3. Speech Acts*, pages 59–82. Academic Press, 1975.

Tuva Lunde Smestad and Frode Volden. Chatbot personalities matters. In *Internet Science*, pages 170–181. Springer International Publishing, 2019.

Svetlana Stenchikova and Amanda Stent. Measuring adaptation between dialogs. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 166–173, 2007.

Stanley S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946.

Noriko Suzuki and Yasuhiro Katagiri. Prosodic alignment in human–computer interaction. *Connection Science*, 19(2):131–141, 2007.

Adriana Tapus and Maja J. Mataric. Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, pages 133–140, 2008.

Morgan Ulinski, Seth Benjamin, and Julia Hirschberg. Using hedge detection to improve committed belief tagging. In *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*, pages 1–5, 2018.

Willemijn M. Van Dolen, Pratibha A. Dabholkar, and Ko De Ruyter. Satisfaction with online commercial group chat: the influence of perceived technology attributes, chat group characteristics, and advisor communication style. *Journal of Retailing*, 83(3):339–358, 2007.

Marilyn A. Walker, Amanda Stent, François Mairesse, and Rashmi Prasad. Individual and domain adaptation in sentence planning for dialogue. *Journal of Artificial Intelligence Research*, 30: 413–456, 2007.

Ning Wang, W. Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. The politeness effect: Pedagogical agents and learning gains. In *AIED*, pages 686–693, 2005.

Stephen Whittaker, Marilyn A. Walker, and Preetam Maloor. Should I tell all?: An experiment on conciseness in spoken dialogue. In *Eighth European Conference on Speech Communication and Technology*, pages 1685–1688, 2003.

Jenny Wilkie, Mervyn A. Jack, and Peter J. Littlewood. System-initiated digressive proposals in automated human–computer telephone dialogues: the use of contrasting politeness strategies. *International Journal of Human-Computer Studies*, 62(1):41–71, 2005.