Expressive Speech Characteristics in the Communication with Artificial Agents

Kerstin Fischer[⋆]
[⋆]University of Bremen
FB10: Sprach- und Literaturwissenschaften
Postfach 330440
28334 Bremen
tel: +49-421-218-9735
fax: +49-40-42883-2515
kerstinf@uni-bremen.de

Abstract

This paper deals with emotional speech characteristics in human-computer- and human-robot interaction. The focus is on the users' involuntary expression of emotion in reaction to system malfunction, which may cause severe problems for automatic speech recognition and processing. Investigating different user groups is shown to be a useful method for determining what makes speakers respond emotionally and for understanding the interpersonal differences that can be observed in reaction to system malfunction. Which aspects may be involved is illustrated by discussing the example of the personal relationship between user and system as evident from the different forms of address that can be found in the corpora. We shall draw on corpora of human-computer and human-robot communication involving children and adults from both sexes. However, it will be demonstrated that the major factor that determines the users' expressive behaviour is their conceptualisation of the artificial agent and the situation. Knowledge about this is then used to develop means for influencing the speakers' attitude towards the system, which can be shown to change their expressive speech characteristics in situations of system malfunction much more effectively than by assessing the linguistic behaviour directly. Thus, interestingly, what turns out to be most suitable for guiding the user into a linguistic behaviour that is understandable for an automatic speech processing system is employing aspects of expressive speech as well.

1 Introduction

Not all emotional expression in the communication with artificial agents may be welcome. Of course expressive speech is a natural characteristic of conversational behaviour in human-human situations. In conversation, topics and purposes are free to vary, speakers may report on emotionally engaging events, and co-participants may display their involvement in order to align with the speaker. Communication with artificial agents in contrast is usually task-oriented and domain-restricted. With the exception of *eliza*-like systems, human-computer interaction (HCI) and human-robot interaction (HRI) are generally carried out for a particular, usually pre-defined, purpose, and thus topics are not free to vary. Emotional topics may thus only arise if planned for by the system designers.

Emotional expression may then occur for three reasons: either it is designed to be part of the domain, or it is used as an accompanying feature to make also domain- and task-specific interaction more human-like, or it occurs as a not planned reaction by system users to particular behaviours of the system, such as system malfunction. In this paper, I will address on the one hand the expressive characteristics of speech humans direct to dialogue systems in situations of miscommunication, the problems these emotional characteristics cause for automatic speech processing systems, and the consequences of these findings for the design of dialogue systems, especially for dialogue management and personality modelling. Thus, it will be discussed how applications can be enriched so as to deal with the problems outlined. On the other hand, it will be discussed in how far, and under which conditions, expressive behaviours from the side of the artificial agent can be used to make the communication more human-like.

The paper thus aims at the descriptive adequacy of the linguistic phenomena involved, and it can be evaluated by a measurable increase or decrease of expressive properties in reaction to particular dialogue acts.

2 Data and Method

The data employed in this study are elicited on the basis of a particular methodology. The method consists in eliciting data of human-computer and human-robot interaction by keeping as many variables constant as possible and by systematically varying only particular aspects of the communicative situation. This means that the system's output, as much as the robot's behaviour, are controlled by means of predefined schemata. This method ensures the interpersonal comparability of the data and the identification of those contextual features that determine the users' linguistic behaviour. Another aspect of the method is the repeated use of system malfunction in order to get the users to reformulate their utterances and thus to uncover their hypotheses about what may have caused the communicative problem. A simple example would be that in case of a miscommunication if the speaker starts speaking very loudly, she displays her hypothesis about her communication partner as someone who needs to be talked to loudly. For details on the method, see Fischer (2003).

The data for this study stem from two sources. On the one hand, Wizard-of-Oz data have been elicited that provide an independent method for identifying emotional speech characteristics.[1] In particular, there are 64 German and 8 English dialogues of appointment scheduling with an average length of 18-33 minutes. There are about 248 turns per dialogue elicited in a Wizard-of-Oz scenario (simulated human-machine dialogues), and questionnaire results show that speakers have not doubted to be speaking to a real system. A particular feature of the corpus is the control for inter- and intrapersonal variation which is achieved by employing a fixed schema according to which 'system' output is generated. The system malfunctions simulated are misunderstanding, failed understanding, generation/synthesis errors, and extra long processing time. The sequences of system output are repeated at least three times throughout the dialogues so that we can compare how speakers react to a particular dialogue act, say, for the first, the third and the fifth time, and thus have an independent measure of changes in speaker attitude (since nothing else changes within these dialogues). Table 1 shows a short extract from the fixed schema of system utterances. The turn ID allows the identification of the occurrence of the turn within each dialogue, and as sequences 2101-2103 and 3101-3103 illustrate, the sequences of system output are repeated throughout the dialogue three to five times. All dialogues were annotated for prosodic, lexical, and conversational peculiarities (Fischer, 1999a).

On the other hand, in the framework of a larger project investigating human-robot interaction (SFB/TR8 'Spatial Cognition' at the Universities of Bremen and Freiburg), a constantly growing body of HRI dialogues are being elicited; by controlledly varying particular situational variables the goal is to find out which situational parameters influence the different ways of speaking to a robot.

The two corpora of German and English HRI used here were elicited in the following two settings: The first set of dialogues was elicited in a joint attention scenario where the participants' task was to verbally instruct Sony's Aibo (see Figure 1) to move to a particular object pointed at by the leader of the experiment and through a sequence (a parcour) of such object localisation tasks.[2] Again the robot's behaviour was predetermined, involving some malfunctions in order to get users to present several different solutions to the same problems. The robot's behaviour was actually manipulated by a wizard according to a fixed schema. The method thus relies heavily on speakers' reformulations because those show what their hypotheses are about what could have gone wrong (Fischer, 2003) and thus their mental models of the system. We have furthermore access to dialogues recorded at the University of Erlangen (Batliner et al., 2004), for which the setting was coordinated with ours, yet which involves children, not adults, as in our data.

Table 1: Fixed Schema of Computer Output

ID	Dialogue Act	System Output
2101	Nonsense	What for date whatthehell bla.
2102	Nonsense	Bla rabartibla blurb.
2103	Request proposal	Please propose a date.
2201	Reject proposal	This time is already occupied.
2202	Misunderstanding	Vacation time is June 15 to July 20.
2203	Request proposal	Please propose a date.
2301	Misunderstanding	7th of February is a Sunday.
2302	No understanding	I did not understand.
2303	Misunderstanding	The weekend is already occupied.
2304	Misunderstanding	It is impossible to meet at 4am.
2305	Reject proposal	This time is already occupied.
2306	Misunderstanding	Friday suits me well.
2307	Misunderstanding	1rst of March is already taken.
2308	Accept proposal	I have noted the appointment.
3101	Nonsense	What for date whatthehell bla.
3102	Nonsense	Bla rabartibla blurb.
3103	Request proposal	Please propose a date.

The second set of HRI dialogues was elicited in a scenario in which the users' task was to instruct the robot, a pioneer 1, to measure distances between two objects from a set of seven pointed at by the leader of the experiment. In this case, the users typed their instructions into a notebook. The robot's output was also predetermined, consisting for the most part either of error messages or of messages naming the distances between the objects to be measured. 21 German speakers participated in this experiments, about half of which were computer scientists.

[1] The transcription conventions are <P> for pause, <B> for breathing, <L> for syllable lengthening, and interpunctuation (,.?) for level, falling and rising intonation contours respectively.

[2] The transcription conventions are: -, –, and (1sec) for the pauses, (at=loud) (/a) for attributes of speech, such as loudness, and interpunctuation (,.?) for level, falling and rising intonation contours respectively.

Figure 1: Sony’s Aibo Robot

3 Emotional Speech in the Communication with Artificial Agents

The analyses of emotional peculiarities in HCI allow us to present a typology of the expressive characteristics of speech in the communication with artificial agents that are due to system malfunction. Such an account is non-trivial since most research on expressive properties of speech has been carried out on data produced by actors (cf. Tischer, 1993; Batliner et al., 2003), and very little research has been conducted on real situations of system use (but see Oviatt, 1995; Oviatt et al., 1998a,b; Levow, 1999). Using actors’ speech has the methodological advantage that it is clear which strings of speech are supposed to be emotional. In actual situations it may be very difficult to lay one’s finger at the moment in which a speaker gets emotionally involved (Batliner et al., 2000). Oviatt and collaborators solve this problem by focussing on particular speech acts, such as repetitions. In the current study, we have the possibility, because of the particular design of the corpus, to identify the emotional properties of speech on independent grounds, namely on the basis of changes within the linguistic behaviour in reaction to the same system behaviour at different points in time.

In the following example, the same speaker, e058, is confronted with the same system utterances three times. While in the earlier phase of the dialogue the speaker uses metalanguage to discuss the misunderstanding, he uses an ironic comment [*yes, great*] some minutes later, and towards the end of the dialogue he uses many prosodic peculiarities and an impolite evaluation [*you should be fired*]:

(1) s0582201: dieser Termin ist schon belegt. [*this date is already occupied*]

e0582201: FÜNFter erster. [*5th of January*]

s0582202: die Urlaubszeit ist fünfzehnten Juni bis zwanzigsten Juli. [*vacation time is from 15th of June to 20th of Juli*]

e0582202: ja, das hat ja auch nicht viel damit zu tun, da wir uns im Januar befinden, ne? [*well this has not much to do with the fact that we are in January, right?*]

s0582203: bitte machen Sie einen Vorschlag. [*please make a proposal*]

e0582203: fünfter Januar. [*5th of January*]

(2) s0584101: dieser Termin ist schon belegt. [*this date is already occupied*]

e0584101: ACHTZEHN bis zweiundzwanzig Uhr. [*6 to 10pm*]

s0584102: die Urlaubszeit ist fünfzehnten Juni bis zwanzigsten Juli. [*vacation time is from 15th of June to 20th of Juli*]

e0584102: <B> ja, klasse. <P> Dienstag, zwölfter erster, ACHTzehn bis zweiundzwanzig Uhr. [*yes, great. Tuesday, 12th of January, 6 to 10pm*]

s0584103: bitte machen Sie einen Vorschlag. [*please make a proposal*]

e0584103: Dienstag, zwlfter erster, achtzehn bis zweiundzwanzig Uhr. [*Tuesday, 12th of January, 6 to 10pm*]

(3) s0586206: dieser Termin ist schon belegt. [*this date is already occupied*]

e0586206: sechster <P> Januar, <P> zwanzig bis zweiundzwanzig Uhr. [*6th of January, 8 to 10pm*]

s0587102: die Urlaubszeit ist fünfzehnten Juni bis zwanzigsten Juli. [*vacation time is from 15th of June to 20th of Juli*]

e0587102: dich sollte man feuern. <B> sechster <P> Januar , <P> zwanzig bis zweiundzwanzig Uhr. [*you should be fired. 6th of January, 8 to 10pm*]

s0587103: bitte machen Sie einen Vorschlag. [*please make a proposal*]

e0587103: se<L>chster Ja<L>nua<L>r, <P> <;<zwa<L>nzig bi<L>s zweiundzwa<L>nzig Uhr> ;with very low voice>. [*6th of January, 8 to 10pm*]

Thus, using the method proposed, the prosodic peculiarities, lexical means and conversational strategies, such as metalanguaging or the use of repetitions can be correlated with particular staes of the dialogues and thus with changes in speaker attitude.

3.1 Prosodic Peculiarities in Reaction to System Malfunction

The phonetic and prosodic peculiarities identifiable in our German corpus (cf. also Pirker and Loderer, 1999) are very similar to those identified in a number of studies for English (Levow, 1998, 1999; Oviatt, 1995). Thus, speakers were found to employ the following prosodic strategies:

- hyper-articulation
- syllable lengthening (e.g. Mon<L>day)
- pauses (between words and syllables, e.g. on <P> Thurs<P>day)
- stress variation
- variation of loudness
- variation of intonation contours
- laughter or sighing

In order to design automatic speech processing systems that can deal with really occurring speech properties, it is necessary to know what can be expected.

For instance, the example shows that the speaker regards slow speed, syllable lengthening (<L>), pausing (<P>), and strong emphasis (capital letters) as something that makes her speech easier to process for her communication partner:

(4) e4077101a: you didn't even let me finish. how do you know it's if it's occupied or not? <B> Monday <P> the eLEVenth <P> of JANuary <P> at twelve pm.

s4077102: vacation time is from the tenth of June till the fifteenth of July.

e4077102: no<L>, no <P> <<;slow> JAN<L>uary.> <P> <Swallow> <B> JANuary the <:<B> elEVenth:> <P> <B> at TWELve pm.

s4077103: will you please make a suggestion for an appointment?

e4077103: <Swallow> okay. <Swallow> let's try JANuARy <B> the <:<B> e<L>LEVenth:> <P> <B> at <P> TWELve pm.

If recognisers and dialogue managers are not designed to take into account the peculiarities that arise in real situations of miscommunication, since situations of communicative problems often trigger even more peculiarities, a vicious circle may arise that may lead to interruptions of the communication and, in the worst case, to the loss of a customer. That is, if malfunction occurs, and the speaker employs speech peculiarities either to increase the understandability of her utterances or to express her anger, the system, not being trained on such features, will understand even less Levow (1998).

Consequently, speech recognisers need to be adapted to the actually occurring, for instance, by being trained on data that include the phenomena arising. Alternatively, speech recognisers trained particularly on emotional data can be employed that are used as soon as emotional language is detected. This presupposes means to identify the moment at which the speaker get emotionally engaged (Batliner et al., 2003, see). Another possibility is to try to use dialogue management to calm down the angry user and to find other ways to prevent those emotional characteristics that are problematic for automatic speech processing systems from occurring; this is the perspective taken in this study.

3.2 Conversational Peculiarities in Reaction to System Malfunction

In emotional speech phonetic and prosodic characteristics are only one aspect of the expressive behaviour; speakers are furthermore also found to employ a number of conversational strategies that are peculiar to the situation of communicative problems, such as the repetitions investigated by Levow (1998, 1999); Oviatt (1995). Moreover, these behaviours may be easier to identify than increases of prosodic peculiarities (see Glockemann, 2003), in case emotional involvement in the speaker is to be monitored (Batliner et al., 2003). Here, what can be found are reformulations, additional specifications, meta-linguistic statements, new proposals without any relevant relationship to the previous utterances, thematic breaks, repetitions, and evaluations.

Reformulation

e4032301: the fifth of January, Tuesday <P> an appointment for five hours.

e403s4032302: I did not understand.

e403e4032302: an appointment on Tuesday January fifth <P> for five hours.

Metalanguage

e4022306a: Tuesday, January fifth, from eight o'clock until one o'clock.

s4022307: the first week of March is already occupied.

e4022307: I mean January fifth.

Additional Information

s4015104: please make a proposal.

e4015104: okay. <P> twelfth of January ninety-nine?

s4015201: I did not understand.

e4015201: the twelfth of January nineteen-ninety-nine?

New Proposal without Relevant Sequential Relation

e4022201: then the<L> twenty-second? <P> at <P> eight in the morning? <P> until two in the afternoon?

s4022202: vacation time is from the tenth of June till the fifteenth of July.

e4022202: <B> <P> uhm <P> on January fifth, at eight o'clock?

Thematic Breaks

e4072302: <Swallow> <B> <P> I have time on Thursday the twenty-first of January <P> <B> at two pm.

s4072303: the weekend is already occupied.

e4072303: <B> <Smack> okay. <P> <B> let's try <P> <B> okay, I have a another suggestion. how about Monday, <P> the eighteenth of January <P> <B> at <B> twelve pm?

Repetition

e4024101: January fourteenth, <P> from six until ten at night?

s4024102: vacation time is from the tenth of June till the fifteenth of July.

e4024102: January fourteenth, from six until ten at night?

Evaluation

e4075206: <Swallow> okay. <B> you're very busy for a computer. <P> <B> how about <B> Sunday the seventeenth of January <B> at <B> ten am?

The features described are likely to be due to emotional arousal; one reason may be that almost all speakers answered in the questionnaire that they filled out after the dialogues that they have been emotionally involved.

Second, there are systematic changes in the course of the dialogues, regarding their prosodic, conversational, and lexical properties. As shown in Figure 3, for instance, the conversational strategies are correlated with different phases of the dialogues, such that reformulations etc. occur more in earlier phases of the dialogue, associated with co-operative linguistic behaviour, whereas repetitions, rejections, and evaluations are located at the other end of the spectrum, occurring typically towards the end of the dialogues. Thus, when repetitions occur (see Levow, 1998, 1999), then this fact points to some dissatisfaction of the user already Fischer (1999b,c). The same holds for the prosodic and the lexical properties of the users' speech in the current corpus. For instance, regarding lexical material used, expressions with the German equivalent of *shit* are twice as frequent in the second half of the dialogues. Expressions with *Gott* [*god*] occur five times as frequently in the second half of the dialogues.

Third, if the lexical, conversational and prosodic strategies described were only due to the speakers' attempts to make themselves more understandable, there would be no prosodic peculiarities if the speakers feel that understanding is not at issue. For instance, rejections of proposals are understood as relevant answers:

(5) s0323202: dieser Termin ist schon belegt. [*this date is already occupied.*]

e0323202: aha. <B> wenigstens eine korrekte Antwort. [*uhuh. <B> at least a correct answer.*]

Although the speakers feel understood when the system rejects their proposals, the prosodic peculiarities of turns following rejections also increase throughout the dialogues (Fischer and Batliner, 2000). The changes observable can therefore not be solely attributed to different strategies to increase one's understandability, and thus emotional arousal must be involved as well.

Like speech recognizers, dialogue managers need to be designed to deal with the conversational characteristics of speech in the context of communicative problems, both regarding the recognition of the respective dialogue acts, and the capability to respond appropriately. Very few studies address the prevention or resolution of communicative problems in dialogue systems (cf. Batliner et al., 2003). Two aspects are relevant here: on the one hand the employment of expressive speech by the artificial agent, which can help calm down the user significantly, on the other hand the development of personality modelling on the basis of interpersonal differences observable in the data.

4 User Groups

If we now ask which measures can be taken to calm down an angry user and to prevent the situation from escalating, many possible behaviours of the dialogue manager could be used to guide the conversation even in the case of miscommunication. We may proceed by investigating whether all users behave in the same way and if not, what we can learn from those who display fewer problematic characteristics.

Sociolinguistic variables that have been found to often influence speech are age, gender, and social class. Unfortunately our data do not allow conclusions about social class, but about age and gender. As we shall see in the discussion of these two variables, it is more the attitude the speakers display towards their communication partner and their perception of the situation than any extralinguistic speaker characteristics that may be relevant for user modelling. Thus, the following three subsections address age, gender, and speaker attitude respectively.

4.1 Age

In the use of expressive characteristics in HRI the speakers' age may be relevant; it is intuitively plausible that children may approach a robot in a much more playful way than an adult may. Accordingly, in experiments with children carried out in a similar set up like ours at the University of Erlangen, many more expressive speech characteristics can be found, compared to our experiments with adults:

(6) Ohm_21.062: Aibo steh auf [*Aibo get up*]

Ohm_21.063: brav so ist es brav lauf lauf geradeaus [*nice this is nice run run straight ahead*]

Ohm_21.064: lauf geradeaus Aibo hopp geradeaus laufen [*run straight ahead Aibo hopp straight ahead*]

Ohm_21.065: lauf [*run*]

Ohm_21.066: lauf Aibo Aibo lauf geradeaus [*run Aibo Aibo run straight ahead*]

Ohm_21.067: Aibo [*Aibo*]

Ohm_21.068: hörst du nicht geradeaus laufen Aibo [*don't you listen run straight ahead aibo*]

Ohm_21.069: brav so ist es brav [*nice this is nice*]

Ohm_21.070: lauf weiter [*go on*]

Ohm_21.071: Aibo steh auf [*Aibo get up*]

Ohm_21.072: Aibo [*Aibo*]

Ohm_21.073: steh auf [*get up*]

Ohm_21.074: steh auf [*get up*]

Ohm_21.075: b"oser Hund [*bad dog*]

Ohm_21.076: lauf [*run*]

Ohm_21.077: so ist es brav lauf [*this is nice run*]

Ohm_21.078: lauf weiter laufen [*run go on*]

Ohm_21.079: Aibo lauf lauf [*Aibo run run*]

Ohm_21.080: lauf Aibo [*run Aibo*]

Ohm_21.081: lauf [*run*]

Ohm_21.082: Aibo lauf [*Aibo run*]

Ohm_21.083: sitz Aibo Aibo sitz [*sit down Aibo Aibo sit down*]

Ohm_21.084: sitz [*sit down*]

Ohm_21.085: Aibo sitz [*Aibo sit down*]

Ohm_21.086: sitz Aibo [*sit down Aibo*]

In this example, the child uses the robot's name numerous times to get the robot's attention and to make it attend to his instructions. Furthermore, we find several evaluations of the robot's behaviour, such as *good dog* or *bad dog*. We also find interjections *na*, *oh Gott*, *ach herrje* or, as in this example, the secondary interjections *hopp* and *komm*.

According to Brown and Gilman (1962), address forms reveal aspects of the relationship between communication partners along the dimensions of power and solidarity. In particular, the informal T-forms can be distinguished from the more formal V-forms, the former expressing more equal and more solidary relationships, the latter being used in hierarchical and less solidary relationships. German distinguishes two forms of address, the informal *Du* and the formal *Sie*. The child in the above example, as all the other children as well, employs the T-form, expressing solidarity and an equal relationship, and very frequently the robot's (first) name.

The use of the robot's name is very typical for all the dialogues with the children. It may be an indicator that the children are building up a much stronger personal relationship with the robot than the adults in our corpus in a very similar setting did. Such a strong personal relationship becomes apparent in the next example in which the speaker, another boy, points out this relationship to the robot as a motivation to follow his instructions:

(7) Ohm_27.174: Aibo tanz [*dance*]

Ohm_27.175: mach's für mich bitte [*do it for me please*]

Ohm_27.176: oh wie lieb [*oh how kind*]

Actually, there is no child among the 26 children recorded who would not use the robot's name. There are two children who use it only in situations of 'disbehaviour', and all others use it consistently throughout the dialogues. Correspondingly, Batliner et al. (2004) report the name *Aibo* to be the most frequent word in the German child-aibo data.

Furthermore, many features of human-to-human dialogues occur in the dialogues with the children that are very rare in adult human-computer interaction. One such example are discourse particles, among them interjections, and modal particles. Thus, Batliner et al. (2004) report the modal particle *mal* and the secondary interjection *komm* to be even among the ten most frequent words. This is quite surprising since the numbers of discourse and modal particles usually decrease in human-computer interaction (Hitzenberger and Womser-Hacker, 1995). The function of modal particles is to relate the current utterance to an assumed common ground, whereas the function of discourse particles is to mark an utterance as non-initial (Fischer, 2000a). The modal particles used by another child in the following example are *mal* and *schon*:

(8) Ohm_25.006: okay Aibo jetzt lauf mal [*okay Aibo now start running*]

Ohm_25.007: komm schon Aibo lauf [*come on Aibo run*]

Ohm_25.008: Aibo lauf [*Aibo run*]

In contrast, in our English and German adult-aibo dialogues, very few adults addressed the robot. For instance, the following speaker employs it after a successfully carried out task:

(9) turn right. –

turn right ? (2secs)

(at=loud)move forward(/at). -

move forward. (2secs)

(at=slow)ok(/at). –

(at=quiet)good robot(/at)

Furthermore, discourse particles and, in the German data, modal particles are extremely rare. A typical example from a German adult-aibo dialogue is the following:

(10) VP: rechts [*right*]

R: noncomply: straight ahead

VP: rechts, rechts, rechts [*right, right, right*]

R: comply: right

VP: vorwärts [*straight ahead*]

R: comply: straight ahead

VP: vorwärts [*straight ahead*]

R: comply: straight ahead

VP: vorwärts [*straight ahead*]

R: comply: straight ahead

VP: links [*left*]

R: comply: left

VP: vorwärts [*straight ahead*]

R: comply: straight ahead

VP: vorwärts [*straight ahead*]

However, it is possible that the differences between adults and children observable in these data are not exclusively due to speakers' age. In the adult-aibo scenario, the users were confronted with an experimental situation in which they had to give verbal instructions while there were three people, the leader of the experiment and two students operating the camera and taking notes, were present. In contrast, in the child-aibo data, the children had had the chance to get 'familiar' with the robot, and the robot was explicitely introduced to them by its name. In English child-aibo data using the same scenario, the British children displayed a very different linguistic behaviour (Batliner et al., 2004), the data being more similar to our adult-aibo dialogues. Part of the story may thus be that the children in the German dialogues with aibo experienced the experimental situation as very playful, which is supported by the fact that they all reported afterwards that they had had fun (even though the robot was as 'malfunctioning' as the robot in the other scenarios) and that they were made acquainted with the robot before the experiments.

Looking at another set of HRI data provides further evidence that variation between speakers cannot be simply traced back to extralinguistic factors, such as speaker age. In the second set of HRI data used here, adult users typed instructions into a notebook to instruct the robot and thus were much more private in their interaction with their artificial communication partner than in the previous scenario. Here it turns out that many users are much more playful. Some users address the robot like the children do, for instance:

(11) VP17-1: hallo roboter [*hello robot*]

sys:ERROR

VP17-2: hallo roboter [*hello robot*]

sys:ERROR

VP17-3: Die Aufgabe ist, den Abstand zu zwei Tassen zu messen. [*the task is to measure the distance to two cups.*]

sys:ERROR 652-a: input is invalid.

VP17-4: miß den Abstand zur Tasse genau vor Dir [*measure the distance to the cup right in front of you*]

sys:69,8 cm

(12) VP9-1: hallo roboter [*hello robot*]

sys:ERROR

VP9-2: wie kann ich entfernungen messen? [*how can I measure distances?*]

The following example stems from a speaker who uses not only the name of the robot and the T-form, but who also employs the modal particle *mal*:

(13) VP7-1: hallo roboter [*hello robot*]

sys:ERROR

VP7-2: miss mal die entfernung, roboter [*please measure the distance, robot*]

sys:ERROR

VP7-3: siehst du die tassen? [*do you see the cups?*]

However, in the following example, even though the user employs the T-form, the direct address is used as if it was an insult:

(14) VP4-28: Welchen Abstand haben die Tassen links hinten und hinten zueinander ? [*Which distance do the cups back left and back ?*]

sys:Unverstaendliche Eingabe. Bitte formulieren Sie neu. [*non-understandable input. Please reformulate.*]

VP4-29: Das ist nicht unverständlich, sondern sonnenklar, du Roboter. [*that's not non-understandable but completely clear, you robot.*]

sys:Bitte umformulieren! [*please reformulate!*]

VP4-30: Spinner. [INSULT]

However, not all users address the robot. As the following example shows, users sometimes believe that the robot does not even know where it is itself:

(15) VP10-1: das erste objekt ist das, das dir am nächsten ist [*the first object is the one which is closest to you*]

sys:ERROR

VP10-2: das erste objekt steht genau vor dir [*the first object is right in front of you*]

sys:ERROR

VP10-3: erstes Objekt: das am nächsten vor dem Roboter. [*first object: that is closest in front of the robot.*]

To sum up, in these dialogues users consistently use the informal T-form, and several users address the robot directly. Even though in later phases of the dialogues the robot uses 'natural language' and thus the formal V-form, users stick to the T-form.

In contrast, in the human-computer appointment scheduling dialogues, in which the computer's first utterance employs the V-form, the speakers consistently employ the V-form as well. Sometimes they switch to the T-form in the middle of the dialogues, explaining afterwards that they have thought about it and that they found that it does not make sense to use the formal form of address in the communication with an artificial agent. However, the following example shows that speakers may also use the different forms of address to mark on-stage versus off-stage talk:

(16) e0372301: nein, der siebte erste ist ein Donnerstag, du dumme Maschine. na, egal. machen wir den achten ersten als Termin ab. sind Sie damit einverstanden? [*no, the seventh of January is a Thursday, you (T-form) stupid machine. well, who cares. let's take the eighth of January. Do you (V-form) agree?*]

To conclude, it cannot be shown that a particular relationship with the robot, as evidenced by the forms of address used, can be directly related to particular user groups. That is, a relationship of solidarity and equality can not only be found in the child-aibo dialogues, but also in adult-computer and adult-pioneer interaction, where it may even be used consciously as a resource to mark an informal, or an off-topic, relationship. Although at first sight age seems thus to play an important role regarding the emotional involvement of the speakers, the perception of the situation and the artificial agent's output seem to be at least as relevant, as the different behaviours of the adults in the two human-robot andin the human-computer scenarios show. Thus, although we can see a bias of children towards approaching the human-robot interaction itself more playfully, it seems to be particularly important to predict how the users UNDERSTAND the situation.

4.2 Gender

A second often relevant extralinguistic variable is gender, and thus it may help to predict users' linguistic behaviour on the basis of their sex. However, if we investigate the use of conversational strategies between men and women in the human-computer dialogues, we find very few significant differences.

In the human-robot scenario in which users had to type instructions into a notebook in order to get the robot to measure distances between objects, different behaviours between males and females could be observed - yet, there was also the unfortunate coincidence that most of the male users were computer scientists, while most of the females were not. In those few cases in which a female was also a computer scientist, her language usage was much more similar to the male computer scientists than to the other females. Thus, the more important variable seems to be experience with artificial agents, but our data do not permit any reliable claims in this respect.

In contrast, our corpus of human-computer interaction allows us to make statistically valid assertions about the impact of the variable gender. For instance, in the use of reformulations (see Figure 2), women reformulate slightly longer than the males (the difference in phases 3 and 4 being significant), but display the same linguistic behaviour in the last phase of the dialogue. To reformulate one's utterances for the system is co-operative behaviour, and thus women seem to be slightly longer co-operative than males are.

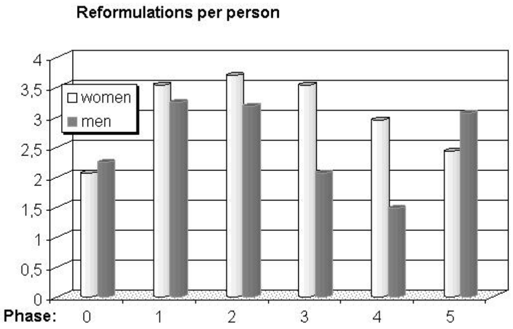

Figure 2: Reformulations by Women and Men

Conversely, issuing new proposals irrespective of the system's utterance is non-co-operative behaviour, and here women can be found to start a little later with this behaviour, as shown in Figure 3. Thus, for phase 3, there are significant differences, yet again women and men finish in the same way. Thus, the gender-specific behaviour just seems to be determined by a bit more patience with the system from the female side.

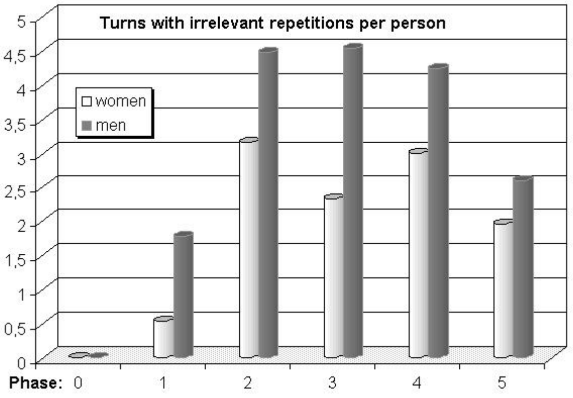

Figure 3: New Proposals by Women and Men

To conclude, gender differences, defined by the extralinguistic variable sex, do not seem to be particularly relevant in the communication with computers, besides the fact that the two sexes employ particular characteristics in different phases. This is partly in contrast with other findings on gender in human-computer interaction (see Fischer and Wrede, 1997). Nevertheless, much interpersonal variation in the dialogues can be observed, and so the question remains, if it is not age or gender, what this variation can be explained by.

4.3 Attitude

The last two sections have shown that extralinguistic factors may not be very helpful for predicting expressive speech characteristics. Instead, it may be useful to investigate the attitude that speakers employ towards the artificial agent. For instance, the occurrence of the prosodic peculiarities outlined above depends significantly on the speakers disposition. Thus, those who experience the situation as amusing will show significantly fewer prosodic peculiarities than those who experience it as annoying (Fischer, 2000b). Consequently, different user types can be identified on the basis of attitude.

I have developed a simple method for identifying different attitudes towards the system within the first utterance of the interaction; these attitudes can be shown to have direct consequences on the expressive characteristics of the speech directed towards the system. Thus, depending on the users' reaction to the question '*how do you do?*,' three user types can be distinguished: those who treat the computer as a tool and propose the first task instead of an answer, those who laugh and say '*fine*', and those who laugh, say '*fine*' and then ask the system '*and how do you do?*'. These three user groups can be significantly distinguished on the basis of their conversational behaviour on all linguistic levels.

Examples are the use of metalanguage and new proposals (that are not related to what has been previously discussed) in Figures 4 and 5. Here, we distinguished between players, those who pretend to have a normal conversation with the computer and thus ask it back politely about its well-being, and all others. In contrast to the two gender groups discussed above, here the differences between the two groups are significant for almost all phases of the dialogues.

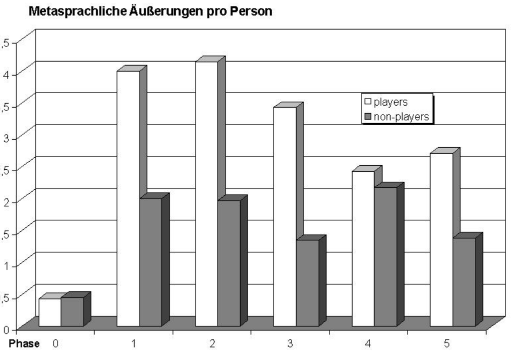

Figure 4: Metalinguistic Utterances by Players and Non-players

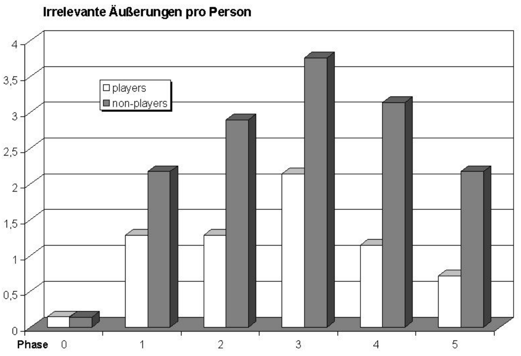

Figure 5: New Proposals by Players and Non-players

The way speakers conceptualise the situation and the artificial agent thus contributes significantly to the way they design their linguistic behaviour towards the system.

To sum up, in this section we have investigated some of the determinants of speakers' linguistic behaviour towards the system and in particular the role of extra-linguistic factors and the conceptualisation of the system as a communication partner. The results obtained, the fact that children are in general more likely to develop a playful and intimate relationship with a robot than adults are, that females tend to be more patient with artificial communication partners than males but otherwise do not differ very much in their linguistic behaviour, and that the speakers' attitude towards the system and the perception of the situation may be decisive for the users' design of their utterances and thus the expressive speech characteristics observable, can now be used to develop ways to guide the users' expressive speech behaviour.

5 Expressive Characteristics in System Output Design

In this section, two aspects will be discussed: on the one hand the role of expressive speech characteristics in system output in general, on the other the employment of system utterances to prevent those emotional features that are problematic for automatic speech processing.

Regarding the latter problem, in the HCI dialogues elicited, first steps towards influencing the users behaviour were taken. By experimenting with different system utterances, from directives like 'please speak more clearly, but not hyper-clearly' to excuses by the system, very different user reactions were obtained (cf. also Fischer and Batliner, 2000): while changes on the surface of the linguistic behaviour in reaction to directives were short-lived, and the peculiarities after the directive even increased, an excuse by the system for the communicative failures lead speakers to calm down immediately and for a longer time. Thus, approaching the users' linguistic behaviour on the level of the interpersonal relationship seems to be the most useful way of obtaining a linguistic behaviour that is unproblematic for automatic speech processing. This finding corresponds to the results obtained in this paper: Since the speakers attitude towards the system plays such a central role regarding the expressive properties of their speech, a useful way of guiding the users linguistic behaviour into something the system can deal with best may be located on the level of the attitude as well. Thus, using expressive characteristics in the robots or computers output may be one way to address the users emotional behaviour.

However, even if the system is embodied and employs expressive characteristics, it is not guaranteed that speakers will consider emotional expression as an appropriate part of the communication. In our HRI dialogues, some speakers took off their head sets which they wore for instructing the robot when they had to laugh about the robots behaviour. Thus, they regarded emotional expression to be off-topic. Another reason to be careful about implementing expressive behaviours in artificial agents may be that human-like properties of artificial agents may raise too high expectations. Thus, if human-like properties are being used, it has to be kept in mind that these must be functioning well - otherwise the opposite effect has been reported (see Bruce et al., 2001; Kanda et al., 2001).

6 Conclusion

Emotional speech, if it occurs unplanned in the context of system malfunction, may constitute a great problem for speech recognition and dialogue management. However, in order to prevent such problems, conversational behaviours by the artificial agent can be employed that may calm down an angry user. A number of such behaviours were proposed. Interestingly, the most effective behaviours are expressive behaviours as well. However, which measures should be taken depends essentially on the users' attitude towards the system, and thus personality modelling constitutes an important first step towards the users guidance by means of dialogue mana gement. How this can be done implicitly and online was exemplified in this paper as well.

Acknowledgements

Most of the research presented here was carried out in the framework of the SFB/TR8 'Spatial Cognition' at the University of Bremen, funded by the German Research Foundation. Many thanks furthermore go to Anton Batliner from the University of Erlangen for his co-operation in the corpus design and for his sharing of the data.

References

A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to find trouble in communication. *Speech Communication*, 40(1-2), 2003.

A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. Russel, and M. Wong. 'you stupid tin box' – children interacting with the aibo robot: A cross-linguistic emotional speech corpus. In *Proceedings of LREC 2004*, 2004.

A. Batliner, R. Huber, H. Niemann, E. Nöth, J. Spilker, and K. Fischer. The recognition of emotion. In W. Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin etc.: Springer, 2000.

R. Brown and A. Gilman. The pronouns of power and solidarity. *American Anthropologist*, 4(6):24–29, 1962.

A. Bruce, I. Nourbakhsh, and R. Simmons. The role of expressiveness and attention in human-robot interaction. In *Proceedings of 2001 AAAI Fall Symposium*, 2001.

K. Fischer. Annotating emotional language data. Technical Report 236, Verbmobil, 1999a.

K. Fischer. Discourse effects on the prosodic properties of repetetitions in human-computer interaction. In *Proceedings of the ESCA-Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands.*, pages 123–128, 1999b.

K. Fischer. Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In Hans-Jörg Bullinger and Jürgen Ziegler, editors, *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany.*, pages 560–565. Lawrence Erlbaum Ass., London, 1999c.

K. Fischer. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles*. Mouton de Gruyter: Berlin, New York, 2000a.

K. Fischer. What is a situation? *Gothenburg Papers in Computational Linguistics*, 00-05:85–92, 2000b.

K. Fischer. Linguistic methods for investigating concepts in use. In Thomas Stolz and Katja Kolbe, editors, *Methodologie in der Linguistik*. Frankfurt a.M.: Peter Lang, 2003.

K. Fischer and Anton Batliner. What makes speakers angry in human-computer conversation. In *Proceedings of the Third Workshop on Human-Computer Conversation, Bellagio, Italy*, 2000.

K. Fischer and B. Wrede. Discourse particles in female and male human-computer-interaction. In Milton Keynes De Montford University, editor, *Women into Computing*, pages 36–49, 1997.

M. Glockemann. Methoden aus dem bereich des information retrieval bei der erkennung und behandlung von kommunikationsstörungen in der natürlichsprachlichen mensch-maschine-interaktion. Master's thesis, University of Hamburg, 2003.

L. Hitzenberger and C. Womser-Hacker. Experimentelle Untersuchungen zu multimodalen natürlichsprachigen Dialogen in der Mensch-Computer- Interaktion. *Sprache und Datenverarbeitung*, 19(1):51–61, 1995.

T. Kanda, H. Ishiguro, and T. Ishida. Psychological analysis on human-robot interaction. In *IEEE International Conference on Robotics and Automation ICRA*, 2001.

G.-A. Levow. Characterizing and recognizing spoken corrections in human-computer dialogue. In *Proceedings of Coling/ACL '98*, 1998.

G. A. Levow. Understanding recognition failures in spoken corrections in human-computer dialogue. In *Proceedings of the ESCA-Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands.*, pages 123–128, 1999.

S. Oviatt. Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, (9):19–35, 1995.

S. Oviatt, J. Bernard, and G.-A. Levow. Linguistic adaptations during spoken and multimodal error resolution. *Langauge and Speech*, 41(3-4):419–442, 1998a.

S. Oviatt, M. MacEachern, and G.-A. Levow. Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, 24:87–110, 1998b.

H. Pirker and G. Loderer. I said ”two ti-ckets”: How to talk to a deaf wizard. In *Proceedings of the ESCA Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands*, pages 181–186, 1999.

B. Tischer. *Die vokale Kommunikation von Gefühlen*. Weinheim: Beltz, 1993.