Gefördert vom

bmb+f

Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil

Verbundvorhaben

Annotating Emotional Language Data

Kerstin Fischer

Universität Hamburg

Vm

Report 236
December 1999

December 1999

Kerstin Fischer
AB NatS
Fachbereich Informatik
Universität Hamburg
Vogt-Kölln-Str. 30
22527 Hamburg
Tel.: (040) 42883 - 2516
Fax: (040) 42883 - 2515
e-mail: `fischer@nats.informatik.uni-hamburg.de`

Gehört zum Antragsabschnitt: 4.8, Dialog und Prosodie

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 701 F7 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor. (ISSN 0949-6084)

1 Background

The phenomena addressed in this study are the linguistic characteristics of emotional utterances. The background of this research is constituted by the observation that at human-computer speech interfaces irritations caused by system malfunctions cannot be completely avoided and that these may cause linguistic behaviour by the speakers which may be problematic for automatic speech processing systems; reasons are that speakers may be emotionally involved or that they try to increase the understandability of their utterances by means of linguistic strategies which are even more difficult for such systems to process. For instance, the acoustic characteristics of the users' repetitions or reformulations of previous utterances have been found to be very different (Levow, 1998; Fischer, 1999a); that is, if the system claims not to have understood a contribution by the speaker, the speaker may repeat her utterance, however, possibly with a different stress pattern, different phrasal intonation, with a strong emphasis on exact pronunciation or even hyper-articulation,[1] and short pauses between the words or even syllables. Some of these properties may cause severe problems for current automatic speech processing systems; for instance, Levow (1998) finds that the error rate in speech recognition rises from 16% to 44% for repetitions.

Besides these linguistic properties which are caused by speakers' local strategies to make themselves understood, as displayed in repetitions and reformulations, speakers may also become emotionally involved when they are repeatedly confronted with errors by the automatic speech processing system, such that the system's malfunctions may provoke emotional responses in the user; thus, the speakers' attitude towards the system may change over time. This change in attitude may have global consequences for the prosodic, lexical, and conversational properties of the speakers' utterances. For instance, the average pitch may rise, the local properties as the above may occur also when no irritation directly precedes the current utterance, people may start talking to themselves, and words (e.g. four-letter words) may be used the system has not been trained for. Like the local changes observed in direct reaction to system malfunctions, these linguistic properties (prosodic, lexical, and conversational) may constitute great problems for current automatic speech processing systems which need to be addressed if human-computer speech interfaces are to be successful. What needs to be prevented is thus a vicious circle:[2]

[1] The term *hyper-articulation*, in contrast, for instance, to Oviatt *et al.* (1998b), is taken here to refer to phonological changes only.

[2] The adjective *deviant* is taken to refer to those aspects of verbal behaviour for which automatic speech processing systems are generally not trained.

system malfunction
causes
deviant speaker behaviour
causes
even greater system malfunction

Any method to intervene in this vicious circle presupposes that it is recognised when the speaker reacts emotionally. Within the Verbmobil project, this problem is approached by means of an automatic classifier which distinguishes two classes of utterances: emotional and neutral. As described in Huber *et al.* (1998), researchers from the University of Erlangen have already successfully trained such an automatic classifier to classify non-authentic data, in which the speakers were asked to behave emotionally. Within the current project, carried out at the University of Hamburg, authentic data are elicited; in order to identify emotional turns for the training of this classifier, which may indicate when the vicious circle described above needs to be interrupted, emotional and neutral training material needs to be selected. The criteria for this selection are not obvious, and there are a number of methodological problems which have to be taken into account.

In this paper, it will be described how emotional language dialogues are elicited and how they can be annotated for the selection as training material; that is, it will be determined which irritations can be found in reaction to system malfunction and how these can be related to a supposed emotional state in the hearer, such that it can be identified which utterances are emotional utterances, and how the speakers' attitude towards the system is realized linguistically.

2 Methodology

There are a number of different possibilities for determining the emotionality of utterances, and there is already a large body of literature on what the linguistic correlates of emotions are; likewise, there are results regarding the local properties of human-computer interaction in the context of error resolution. These two research areas provide us with possible candidates for linguistic properties which can be relevant for distinguishing neutral and emotional utterances. Thus, the properties of human-to-machine communication Oviatt *et al.* (1998a,b); Levow (1998) have found as local strategies of speakers in reaction to instances of failed oder misunderstanding by the system are increased pause interjection and pause elongation, fewer disfluencies, speech elongation, the increased use of falling intonation contours in utterance-final positions, an increase in hyper-clear phonology, and a lower pitch minimum and average (Oviatt *et al.*, 1998b, p. 102). Oviatt *et al.* (1998a) show how some of these features increase during error resolution cycles, that is, they show how speakers adapt their linguistic behaviour according to these features while they are trying to solve one particular problem; the authors have not looked at whether speakers are more likely to employ such features in earlier or later phases of the dialogues and whether they are thus influenced by the speakers' attitude towards the system, or whether they are only local properties as part of error resultion strategies. Oviatt *et al.* (1998b) argue that the properties determined are used to increase the understandability of utterances since in human-to-human communication these features indeed facilitate understanding. Also Giles & Williams (1992) discuss hyperarticulation (the term comprising here all kinds of linguistic features), but they consider only communication between human speakers; the motivations for hyperarticulation they detect are "speakers' *intense* desires for social approval, interpersonal affiliation or group identification" on the one hand and the wish "to increase evaluations of competence, culture or control, and to emphasize social distance" on the other (1992, p. 351, emphasis original). Whether in human-computer interaction local linguistic strategies such as hyperarticulation are also caused by emotional reactions, is still an open question.

Regarding the dimensions of emotional behaviour, the linguistic parameters which are suspected to be involved in the expression of emotionality are mainly prosodic, conversational, lexical and non-verbal means. These however constitute only a small part of possible dimensions of emotional reaction which altogether constitute the multidimensional response pattern 'emotion' (Battacchi *et al.*, 1997): physiological, tonic posture-related, instrumental and expressive motorical (the paralinguistic properties are situated here) reactions, expressive linguistic reactions and the subjective experiential component (1997, p. 21-22). Similar to Battacchi *et al.* (1997)'s distinction between expressive linguistic and motorical reactions, Fiehler (1990) distinguishes talking about and expressing emotions, then between spontaneous unvoluntary expression of emotion and the 'production' of such an expression (1990, p. 102). Accordingly, Kehrein (1998) distinguishes explicit communication of emotions by lexical means from morphological, syntactical, phonetic-phonological and non-verbal implicit strategies. As a lexical class, Drescher (1997) investigates interjections, as well as conversational procedures such as reduplications and discourse organisation, as strategies of the expression of emotionality. Similarly, Battacchi *et al.* (1997) focus on those strategies in which emotions or emotional experiences are the topic of the conversation and on the lexical inventory of the language as a prerequisite for this kind of conversation.

A concrete list of prosodic phenomena which are relevant for the description of emotional expression can be found in Scherer (1986), for instance; he regards

F0 pertubation, mean, range, variability, contour; F1 and F2 mean formant bandwidth and precision; intensity mean, range, and variability; frequency range; high-frequency energy; spectral noise and the speech rate as the major acoustic parameters (1986, p. 149) for the vocal expression of emotion. In Scherer & Wallbott (1990), furthermore facial and motoric expressions of emotion are considered. Similarly, Schmidt-Atzert (1993) investigates mimics, gesture, body movement and posture, voice and language, in particular mean F0 frequency, variation in F0 frequency, energy, speech rate, pauses, and disfluencies (1986, p. 25). Finally he considers writing rate and gaze. However, methodologically many of these studies (for instance, Williams & Stevens, 1982; Tischer, 1993; Pirker & Loderer, 1999) are problematic such that they rely a) on intuition, or b) on material which is not authentic, or c) they use a descriptive inventory which is too vague or unreliable. Tischer (1993), for instance, has determined a number of linguistic parameters on the basis of the same sentences read by an actor who attempts to reproduce a particular emotional expression. It is questionable whether these results are similar to what can be found in authentic data; for instance, Scherer (1982) argues:

> Selbst wenn professionelle Schauspieler zur Darstellung von emotionalen Inhalten in Szenarios, die es den Schauspielern erlauben, bestimmte emotionale Erlebnisse aus dem täglichen Leben zu reproduzieren, verwandt werden (...), gibt es noch eindeutige Unterschiede zu natürlich vorkommenden spontanen Emotionsausdrücken (...). Darüber hinaus kann man starke individuelle Unterschiede in der Darstellungsfähigkeit bei naiven Versuchspersonen gegenüber Schauspielschülern oder professionellen Schauspielern erwarten. Außerdem werden oft nur ein oder zwei Schauspieler in dieser Art Forschung verwandt und die Variabilität der Enkodierer wird nur sehr selten systematisch untersucht. Weiterhin wird oft nicht ausgeschlossen, daß der individuelle Enkodierer verschiedene Arten der gewünschten Emotionen darstellen kann, etwa 'unterdrückten Ärger' gegenüber 'explosivem Ärger' (Scherer, 1982, p. 298).

While it can be argued that actors produce prototypical emotional behaviour, it is unclear whether what speakers produce in the communication with an automatic speech processing system is similarly prototypical emotional behaviour as well. When actors are asked to display an emotion, they may do just this, yet what we do not know is whether speakers in natural situations display their emotions at all, and, if so, whether they do so with the same linguistic means. Speakers may furthermore react to particular situations from their real lives differently so that they may display different kinds of emotions when prompted by a particular scenario; thus, as Scherer (1986) writes, "the same situation and the same stimulation can produce different affective states in different individuals depending on the nature of their cognitive appraisal of the situation" (1986, p. 146). In the data elicited in the current project, some speakers report, for instance, to be angry, others to be amused, although the scenario is absolutely identical. Interindividual differences regarding the display of emotional properties have also been found in the literature, for instance, in Williams & Stevens (1982). Fiehler (1990), for instance, has argued for a sociological approach to emotions such that the display of emotionality is a sign which, like all signs, is displayed for the communication partner to consider (Clark, 1996, p. 160). Thus it is expectable that there are differences between the properties of emotional utterances in the speech of actors and in natural situations as well as between different speakers since the display of emotionality fulfils a particular communicative function (see also Drescher, 1997). Furthermore, whether speakers display their anger to a computer at all, is still in question. Thus, the relationship between the prototypes produced by actors and the observable properties of human-computer interaction still needs to be determined.

An alternative to eliciting non-authentic data from actors is to look at utterances from corpora from which particular emotional data are selected, and thus to analyse the linguistic properties of this emotionally suspicious linguistic material (for instance, Pirker & Loderer, 1999); suspicious in this respect are in German, for instance, judgements such as *schade, der tillt gerade, Scheißteil*, the interjection *hm*, or statements about ongoing mental processes: "Ich werde noch wahnsinnig." However, which linguistic properties are thus considered has often been determined intuitively, yet it would be desirable if they could be selected on independent grounds. To rely on intuition in such a decision means that an analysis of emotionality in interaction is preceded by an *a priori* conception of what the properties of emotional language are. This procedure is essentially circular, and by employing it, one runs the risk of ignoring potentially useful and relevant information.

There are a number of ways to constrain one's intuitions; for instance, there is the possibility of using other dimensions of the expression of emotion as a basis for such a selection; for instance, Picard (1997) takes physiological and motorically-expressive means into account (see also Buck, 1994; Picard, 1999). However, the question here is which physiological properties of speakers correspond to the experience of emotion such that these physiological properties have some kind of experiential relevance to the speakers themselves and, more importantly here, that they have consequences for the speakers' linguistic behaviour. There is no reason to believe that there is an *a priori* status of the connection between observable physiological properties and the experience and expression of emotion. While, for instance, Scherer (1986) argues that emotion is to be conceptualized "as a 'syndrome' with various components (e.g. physiological arousal *and* expression *and* feeling) in response to an evaluation of significant events in the environment" (1986, p. 146), Lang (1988); Schmidt-Atzert (1981); Battacchi *et al.* (1997) argue that, for methodological reasons, as long as the relationship between subjective experience, physiological, and behavioural reactions is not clear, they should be treated as independent phenomena.

For the experience of emotions eventually the same may hold as for physiological reactions: Their relationship to the speakers' observable linguistic behaviour may not be clear. Information on the speakers' emotional experiences are perhaps easiest to get by asking the participants after the interaction with their communication partners how they have felt. So far, all speakers in the experiments described below have claimed to have reacted emotionally. Most of them are angry, some find the interaction funny. However, many of them also said that they have hidden their anger. Their verbal statements about their involvement often differ significantly from what they put down in the questionnaire they fill out after the interaction with the simulated automatic speech processing system; for instance, they may write "leicht genervt" but may hand the questionnaire back to the experimenter saying "mein Gott, das Ding kann einen ja wirklich zur Weißglut bringen." Consequently, it is dubious whether asking the speakers themselves may yield reliable results, and results which are of influence for the description of the linguistic properties of their utterances observable.

Asking not the participants but having other speakers evaluate the emotionality of particular utterances (Gumperz, 1982) may be an alternative method, but it is costly and relies on intuitive judgements, too. Most problematic is here the descriptive inventory; while the procedure may be uninformative regarding the criteria employed, if natural language emotion terms are being used, these descriptive categories are furthermore methodologically dubious for other reasons; as Scherer (1982), for instance, argues regarding actors in the quote presented above, there may be interindividual differences between different speakers such that they may understand particular emotion terms differently; similarly, Scherer (1986) holds that "there may be large differences in the connotations of emotion labels and in labeling strategies" (1986, p. 146). In Cowie *et al.* (1999), for instance, a rating system was developed in which emotion terms are ordered within a two-dimensional space; raters have to place every utterance within this space. However, as Wierzbicka (1992a,b, 1995) has shown, emotion terms are language specific and highly complex concepts and therefore not suited as a descriptive metalanguage. Correspondingly, Cowie *et al.* (1999) do not present intercoder reliability ratings; in fact, by the time of the presentation of their paper, the order of terms had already been rearranged, and it is dubious whether a consistent order and consistency in raters' judgements can be found. A way to distinguish discrete emotions may therefore be a controlled componential analysis of emotions, as proposed, for instance, in Wierzbicka (1992a,b) or in Scherer (1986).

Finally, emotional utterances from corpora can be selected for analysis by exploiting the speakers' own treatment of utterances in conversational interaction; if, for instance, a speaker reacts to her partner's turn by saying "Du brauchst nicht gleich sauer zu werden," then this is an indicator that she has understood her partner's utterance as displaying the mental state of anger. In Conversation Analysis, this method is referred to as the "next-turn proof procedure" (Sacks *et al.*, 1974). However, in human-computer interaction this method is not easily applicable since artificial communicators such as automatic speech processing systems do not ususally display or even discuss their interpretation of their communication partner's emotional states; still, in the overnext turn the speaker may provide an interpretation of his previous turn himself, as in the following example in which turn e0023205 displays that the speakers has meant his previous turn to be a request for a proposal and that he does not consider the system to have met his request:

(1) e0023204: machen SIE einen Vorschlag. (*make a proposal.*)

s0023205: Donnerstag von 8-10 Uhr ist schon belegt. (*Thursday from 8 to 10am is already occupied.*)

e0023205: das ist kein Vorschlag. (*that's not a proposal.*)

However, while the speakers may provide an interpretation of their own previous turns, they have not been found to comment on their own emotionality; the reason may be that they do not believe the system to understand such signals or to be able to discuss aspects of emotionality anyway. Likewise, instances in which the system displays an interpretation of the speaker's emotionality to which the speaker may react are rare; automatic speech processing systems, at least currently, are rarely designed to be capable of making emotions the topic of human-to-computer communication. Thus, while this method provides a suitable analytical tool in natural conversation, for our purposes, to determine the emotionality of particular turns or even words in human-computer interaction, other methods have to be found which allow to classify turns as input for an automatic classifier and which are not based on a dubious classification in emotion terms, which rely on authentic data and which are not dependent on intuitive decisions about the emotionality of utterances.

The method proposed here[3] builds on the analysis of intrapersonal variation. The idea is that if speakers are exposed to identical system behaviour several

[3] The method and scenario used here have been described in more detail in Fischer (1998).

times, their changes in linguistic behaviour can indicate their changing speaker attitude towards the system; these globally changing linguistic properties can then be taken as indicators of emotionality in human-computer interaction. The method is thus to elicit dialogues in which the system does not function properly, and in which these malfunctions occur in stable contexts repeatedly throughout each dialogue, so that the speakers' reactions to the malfunctions can be compared through time. Therefore, Wizard-of-Oz dialogues have been recorded, i.e. dialogues in which speakers believe to be talking to a machine while the system output is actually manipulated by a human 'wizard' (see Fraser & Gilbert, 1991). In these experiments, different kinds of system malfunction are being simulated, for instance, the system repeatedly misinterprets the speakers' utterances, simulating speech recognition errors by providing irrelvant responses. Further 'malfunctions' are the complete failure to understand, long pauses (30 secs.), and 'wrongly synthesised,' not understandable, utterances.

As a methodology for controlling inter- and intrapersonal variation, a fixed dialogue schema has been created which determines the utterances made by the system; thus certain sequences of system output have been defined which are combined in a fixed order, all phases appearing at least twice. These recursively recurring dialog phases make it possible to analyse the same sequences of utterances in different phases of the dialogue. The system output is thereby completely independent of the users' utterances. For instance, the system may ask the user to make a proposal for a day when to meet. Irrespective of the contents of the user's reaction, the system will then utter that the first of January is a holiday, simulating a speech recognition error. After the next speaker utterance, the system will assert that it is impossible to meet at four o'clock in the morning. This sequence may occur four times in each dialogue. The speakers' reactions to these utterances may change systematically during time.

Speakers are instructed to schedule ten appointments with the system. Before speakers are confronted with the (simulated) malfunctioning system, they are involved in a 'test phase' of which they are told that it is necessary so that the system can adjust to the quality of their voices. In this phase the wizard is, contrary to the 'real' dialogues, cooperative. Utterances from this 'test phase' serve as the background for the further analyses since speakers are not expected to be emotionally engaged if their communication partner is cooperative. Utterances from this phase are therefore good candidates for the selection as 'neutral' for the training of the automatic classifier. After this phase, the speakers are confronted with the fixed dialogue schema. Each of the recordings is ended by a sequence of system output 'I did not understand' and is then interrupted by the experimenter with the comment that the machine has obviously hung itself up. The speakers are then asked to answer some questions about their satisfaction with the system,

Table 1: Schema for the Macro-Structural Annotation of Turns

Speaker	Filenumber	Subdialogue	Phase	Turn
e.g. *e* or *s*	002	2	1	03

whether they believe to have been emotionally engaged and whether they have believed to be talking to a computer. Afterwards they are informed about the real purpose of the recording.

The data are currently 58 dialogues of between 18 to 33 minutes length of which 40 are transcribed.[4] They consist of 248 turns on the average,[5] 124 of which are uttered by the speaker. The relationship between female and male speakers is about equal; participants are between 17 and 61 years old and all native speakers of German. None of them has reported that s/he has realized that the system output was created by a human 'wizard'.

3 Analysis and Annotation of Data

Since intrapersonal variation is taken to be the indicator of changing speaker attitude, all external aspects of the situation being equal, the speakers' behaviour has to be analysed through time. Therefore, the dialogues have been annotated regarding the macro-structural position of each turn, such that for each turn the position with respect to the dialogue phase is immediately identifiable. Thus, e002, for instance, is the identification number of the speaker, while the other four digits represent the position of the respective utterance in the macro-structure of the dialogue. This annotation schema is summarized in Table 1.

The first digit describes the subdialogue which is marked by the system's acceptance of a date suggested by the speaker; altogether there are eight subdialogues while the speakers' task is to schedule ten appointments with the system, which means that it is impossible for them to fulfil the task. The next digit refers to a particular sequence of (simulated) system output which is repeated several times throughout the dialogue. The last two digits count the utterances within such a phase. Table 2 shows an extract from the fixed schema which determines the order of (simulated) system output; phase 2101-2103, for instance, which is also instantiated in example (2), occurs five times per dialogue, one further occurrence being 3101-3103, only the phrasing having been altered slightly. Likewise, the

[4] Of these, 14 had been prosodically, conversationally and lexically annotated by the time of some of the statistical analyses presented in section 3.

[5] The variation, in spite of the fixed dialogue schema, is due to the fact that speakers sometimes wait until "the system has recovered", and thus the 'wizard' has to initiate a nonscheduled request to propose a date in order to continue the dialogue.

sequence of utterances by the system in phase 2301-2308 occurs four times per dialogue. The speakers' reactions to these sequences can therefore be compared over time.

Example (2) shows how the speaker's linguistic behaviour in reaction to the same sequence, here the utterance of incomprehensible speech by the system, may change during time. The system utterances are, depending on the macro-structural position, either a sequence of 'bla irgendwas Termin wotsoewer?' and 'was soll Termin blablurb wot sehell bla' or 'bla irgendwas Termin wotsoewer?' and 'bla rapartibla Termin blurb.' The verbal reactions by, for instance, speaker e002 are the following:

(2) a. e0022101: ich hab' nicht verstanden. (*I didn't understand.*)

e0022102: das kann ich nicht verstehen. (*this I can't understand.*)

b. e0023101: ich hab' nix verstanden.(*I didn't understand nothing.*)

e0023102: das<L> ist Kauderwelsch. (*this is gibberish.*)

c. e0025101: <B> das kann ich nicht verstehen. (*<B> this I can't understand.*)

e0025102: das kann ich auch nicht verstehen. (*this I can't understand either.*)

d. e0026101: <Smack> damit kann ich nichts anfangen. (*<Smack> this I can't handle.*)

e0026102: so ein Unsinn. (*nonsense.*)

e. e0028101: <B> <Smack>

e0028102: <B> Donnerstag, siebter Januar, achtzehn bis zwanzig Uhr. (*<B> Thursday, 7th of January, 6 to 8 pm.*)

While speaker e002 attempts to solve the communicative problem in the first pair of incomprehensible utterances, he blames the system for the misunderstanding in the next two occurrences, accusing it in turn e0026102 of talking nonsense. In the last pair, he does not react to the system's utterance any more at all, remaining silent after the first utterance and simply repeating his previous proposal in the second. These changes in speaker behaviour, from cooperative behaviour to

Table 2: Part of the Fixed Dialogue Scheme

	dialogue act	actual utterance
...	...	...
2101	nonsense	bla irgendwas Termin wotsoewer?
2102	nonsense	was soll Termin blablurb wot sehell bla.
2103	request proposal	bitte machen Sie einen Vorschlag. *'please make a proposal'*
2201	reject proposal	dieser Termin ist schon belegt. *'this date is already occupied'*
2202	misunderstanding	die Urlaubszeit ist vom zehnten sechsten bis fünfzehnten siebten. *'vacation time is from 10.6.-15.7.'*
2203	request proposal	bitte machen Sie einen Vorschlag. *'please make a proposal'*
2301	misunderstanding	der erste erste ist ein Feiertag. *'the first of January is a holiday'*
2302	failed understanding	ich habe Sie nicht verstanden. *'I did not understand'*
2303	misunderstanding	der Wochenendtermin ist schon belegt. *'the weekend is already occupied'*
2304	misunderstanding	ein Termin um vier Uhr morgens ist nicht möglich. *'an appointment at 4am is not possible'*
2305	reject date	dieser Termin ist schon belegt. *'this date is already occupied'*
2306	misunderstanding	Freitag, der achte erste paßt sehr gut. *'Friday 8.1. suits me well'*
2307	misunderstanding	die erste Märzwoche ist schon belegt. *'the first week in March is already occupied'*
2308	accept proposal	ich habe den Termin für Sie notiert. *'I have noted the appointment'*
3101	nonsense	bla irgendwas Termin wotsoewer?
3102	nonsense	bla rapartibla Termin blurb.
3103	request proposal	bitte machen Sie einen Vorschlag. *'please make a proposal'*
...	...	...

uncooperative simple repetition, can be found in reaction to other system irritations as well, for instance in the following irrelevant system reactions which simulate speech recognition errors. Here the speaker has made a proposal and is told that either after 10pm or before 4am appointments are not possible (s4304/7304: 'nach zweiundzwanzig Uhr ist ein Termin nicht möglich.' s5304: 'ein Termin um vier Uhr morgens ist nicht möglich.'):

(3) e0144304: nein, vor zweiundzwanzig Uhr. am Freitag, den zweiundzwanzigsten Januar, von acht bis zwölf Uhr. haben Sie diesen Termin eingetragen? (*no, before 10 pm. on Friday the 22nd of January, from 8 to 12am. have you noted down the appointment?*)

e0145304: aber ich möchte auch gar keinen Termin um vier Uhr morgens, sondern ein' Termin am vierzehnten Januar, von zehn bis vierzehn Uhr. (*but I don't want an appointment at 4am but on January 14th, from 10am to 2pm.*)

e0147304: von zwanzig bis zweiundzwanzig Uhr. (*from 8 to 10pm.*)

Similarly, in the next example in which speakers e002 and e009 are confronted with the information that the seventh of February is a Sunday (s2301/ 4301/ 5301/ 7301: 'der siebte Februar ist ein Sonntag.'), they use metalanguage to try to solve the misunderstanding in reaction to the first instances of this system utterance; in the fourth occurrence, the speakers only repeat their utterances, speaker e009 uses furthermore a signal of contradiction:

(4) a. e0022301: dies<L> hat nichts mit meinem Vorschlag zu tun. ich schlage vor Donnerstag, vierzehnter erster, achtzehn bis zwanzig Uhr. (*this has nothing to do with my proposal. I propose Thursday the 14th, 6 t0 8pm.*)

e0024301: das ist mir <P> total egal. (*I don't care at all.*)

e0025301: <B> das<L> ist mir ganz egal. <Smack> ich schlage nochmals vor, Samstag, der sechzehnte erste, sechzehnter Januar, <B> zwölf bi/ (*<B> this doesn't matter to me at all. <Smack> I propose again Saturday the 16th, 16th of January.*)

e0027301: mein Vorschlag, Sonntag, vierundzwanzigster Januar, neunzehnhundertneunundneunzig. (*my proposal, Sunday, 24th of January, 1999.*)

b. e0092301: <Cough> nee, das ist doch der neunte erste, ein Samstag. (*<Cough> no, but this is the 9th, a Saturday.*)

e0094301: <Smack> ich hab' doch <P> was hab' ich gesagt? Dienstag. Dienstag hab' ich gesagt. Dienstag, den fünften ersten, neunundneunzig. (*<Smack> but I have <P> what have I said? Tuesday. Tuesday was what I said. Tuesday the 5th, '99.*)

e0095301: <Smack> nee, Sonntag kann ich auch nicht, also <B> Mittwoch, am zwanzigsten ersten, ist da noch was frei? (*<Smack> no, Sunday I can't either, so <B> Wednesday, on the 20th, is there something free?*)

e0097301: nein, am einundzwanzigsten ersten, um achtzehn Uhr. (*no, on the 21rst, at 6pm.*)

In the following example, speaker e009 reacts to the system's information regarding the vacation time (s2202/4102/7102: 'die Urlaubszeit ist vom fünfzehnten Juni bis zwanzigsten Juli.') which is quite irrelvant since the speakers are supposed to schedule appointments in January:

(5) e0092202: nee, ich möchte ein/ ähm nur am Wochenende, keine Ferien, also, es soll nur vom Samstag, den neunten ersten, bis Sonntag, den zehnten ersten gehen. (*no, I want a uhm only on the weekend, no vacation, so, it is only supposed to take from Saturday, the 9th, to Sunday, the 10th.*)

e0094102: und am Freitag, am zweiundzwanzigsten ersten? (*and on Friday the 22nd?*)

e0097102: <Smack> ja, aber im Januar, am siebten ersten. (*<Smack> yes, but in January, on the 7th.*)

While speaker e009 reacts by means of explaining what she had proposed when she is firstly confronted with the irrelevant information about the vacation time, she only repeats her utterance the second time, while in the third utterance, her repetition is combined with a rejection.

The examples discussed so far have shown that the speakers' reactions do change throughout the dialogues, even in reaction to absolutely identical sequences of utterances, while also all contextual variables remain constant. For a particular speaker it will now be shown more extensively how her linguistic strategies develop during the recording; the example taken is the claim by the system not to have understood at all. While it occurs occasionally within different dialogue phases, every dialogue is ended by a sequence of these utterances by the system. The recording is then 'interrupted' by the experimenter by saying that the system has obviously crashed.

(6) a. e0105203: Samstag, der sechzehnte Januar, von acht bis zwölf Uhr. (*Saturday the 16th of January, from 8 to 12am.*)

s0105201: ich habe Sie nicht verstanden. (*I did not understand*)

e0105201: Sonnabend, <P> der sechzehnte Januar, <P> neunzehnhundertneunundneunzig, von acht bis zwölf Uhr. (*Saturday, <P> the 16th of January, <P> 1999, from 8 to 12am.*)

b. e0107301: Freitag, der zweiundzwanzigste Januar, von sechzehn bis achtzehn Uhr. <P> wird es Ihnen dort passen? (*Friday the 22nd of January, from 4 to 6pm. <P> will it suit you there?*)

s0107302: ich habe Sie nicht verstanden. (*I did not understand*)

e0107302: am Freitag, den zweiundzwanzigsten Januar, von sechzehn bis achtzehn Uhr. (*Friday the 22nd of January, from 4 to 6pm.*)

c. e0108301: <B> <P> Freitag, der achte erste, von acht bis zehn Uhr? (*<B> <P> Friday the 8th from 8 to 10am?*)

s0108302: ich habe Sie nicht verstanden. (*I did not understand*)

e0108302: <Smack> Freitag, <P> der achte erste, von acht bis zehn Uhr? (*<smack> Friday <P> the 8th from 8 to 10am?*)

s0108303: ich habe Sie nicht verstanden. (*I did not understand*)

e0108303: ist der Termin, am Freitag, den achten Januar noch frei? (*is Friday the 8th still free?*)

s0108304: ich habe Sie nicht verstanden. (*I did not understand*)

e0108304: <B> Gottes Willen. <;gemurmelt> <P> Donnerstag, der siebte Januar? (*<B> good god. <;mumbled> <P> Thursday the 7th of January?*)

s0108305: ich habe Sie nicht verstanden. (*I did not understand*) e0108305: ich würde gerne ein' Termin, am Freitag, dem achten Januar. <P> haben Sie mich verstanden? (*I would like to a date on Friday the 8th. <P> did you understand?*)

s0108306: ich habe Sie nicht verstanden. (*I did not understand*)

e0108306: <Smack> hm, Montag, der elfte Januar? (*<smack> hm, Monday the 11th of January?*)

s0108307: ich habe Sie nicht verstanden. (*I did not understand*)

e0108307: der *tillt gerade. <;leise> (*it's gone crazy <;leise>*)

s0108308: ich habe Sie nicht verstanden. (*I did not understand*)

e0108308: <B> ich weiß nicht. <;leise> (*<B> I don't know. <;quiet>*)

In sequence (6a.), the speaker reformulates her utterance, for instance, by using another term for Saturday, by using pauses and by providing additional information (the year). The second time (6b.), she repeats her proposal only, leaving out her partner-oriented question.

The last (c.) sequence of example (6) is characterized by several attempts by the speaker to reformulate her utterance, by inclusion of pauses (e0108302) and by using a different sentence structure. In turn e0108304, the speaker comments on the situation by means of *Gottes Willen* ('*good god*'), then tries another date, then goes back to the previous proposal. When all these strategies turn out to be unsuccessful, the speaker tries one more date, using an interjection which expresses beginning divergence (8306), and then she mutters an assessment of the system before finally giving up.

Analysing systematic changes in speaker behaviour throughout the dialogues can now point to those linguistic procedures which are characteristic of angry, annoyed or even desparate users. The macro-structural position of the turn is thus of central importance for analysing the speakers' intrapersonal variation and as a measure for their changing speaker attitude; it can also serve as the link between the local strategies observable in the context of error resolution, such as hyperarticulation or syllable lengthening, and the changing linguistic behaviour due to changes in the speakers' attitude towards the system, i.e. emotional changes.

In order to determine for each turn whether it is emotional or not, particular linguistic properties can now be determined which vary according to their distribution within the dialogues, i.e. with respect to their macro-structural position. These properties can be concluded to be of relevance for determining the emotionality of particular utterances. Candidates for such linguistic parameters according to which speakers vary their linguistic behaviour are, according to the examples discussed, prosodic means such as the inclusion of pauses, conversational strategies such as reformulations and repetitions, as well as lexical means such as the interjection *hm* or assessments like *Gottes Willen*.

3.1 Lexical Peculiarities

One obvious property which changes during the dialogues are the lexical means by means of which speakers evaluate the system's behaviour. In example (6), for instance, the speaker uses *Gottes Willen, hm*, and *der tillt gerade* in a later phase of the dialogue. Lexical items like these cannot be found in earlier phases of the dialogues.

Closer analyses of the distribution of, for instance, extremely positive assessments shows that *wunderbar* ('*wonderful*') occurs much more often in the first, cooperative, phase than in any other; this indicates that *wunderbar* is not a good indicator of increasing anger, for instance. However, Figure 1 shows that other positive evaluations, such as *hervorragend, großartig* ('*great*') and *super* ('*super*'), as *wunderbar* does to a certain extent as well, also occur in later phases of the dialogues, then, however, in an ironical interpretation. Thus, while their lexical form can be found throughout the dialogues, their interpretation changes considerably. While for automatic analyses these lexemes are not well suited, there are ironical assessments which only occur in later phases of the dialogues and which are therefore suitable indicators of changes in attitude towards the system, for instance, *interessant* ('*interesting*'). There are no instances of it in the cooperative phase 0 but it is used consistently in later phases of the dialogues. Similarly, *Scheiße* ('*shit*') and its combinations, as well as instances of idioms including *Gott* ('*lord*'), cannot be found in earlier phases of the dialogues but increasingly in later ones. These are therefore strong indicators of a changing speaker attitude.

Table 3 shows the lexical peculiarities as they develop throughout the dialogues. Their ranking is supposed to mirror their occurrence in earlier or later phases of the dialogues.

3.2 Conversational Peculiarities

Likewise, the classification of conversational peculiarities can be carried out with respect to the particular structure of the dialogues. As the examples above have already shown, the conversational strategies speakers employ vary throughout the dialogues.

The following example shows how speakers' conversational strategies may change in contexts in which the system states complete recognition failures. While

Figure 1: Number of Instances of Particular Lexical Items (approx. 20 turns per phase, 40 dialogues)

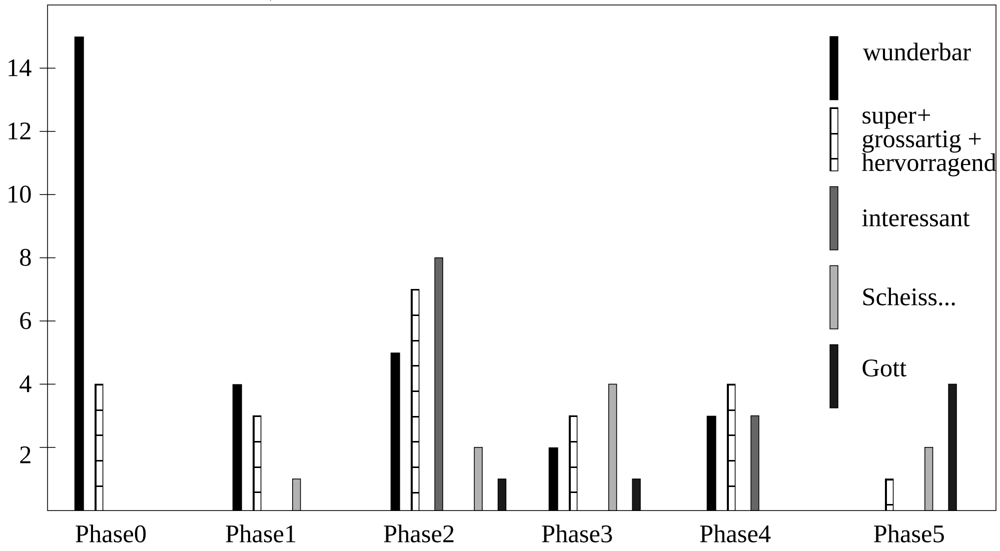


Table 3: Lexical Peculiarieties

	Lexical Peculiarities	Example
1	assessments with little involvement	*schade, Pech gehabt*
2	indirect assessments of dirscursive acts	*nein, nee, hm*
3	ironic, distanced assessments with little involvement	*interessant, das ist ja allerhand*
4	assessments with high involvement, not directed towards the communication partner	*Kauderwelsch, oh Gott, Unsinn, total egal*
5	assessments directed towards the communication partner	*Unverschämtheit, Scheißtyp*

in turn e0112302 speaker e011 reformulates her utterance to make it more precise, she makes the proposal more explicit in utterance e0114302. In turn e0115201, several dialogue phases later, her conversational strategy consists in repeating her original utterance with the inclusion of pauses between the individual phrasal constituents of her utterance. Finally, in turn e0118207, she does not even attempt to repeat her utterance; while her original utterance is realized by means of a number of particular prosodic properties such as syllable lengthening and pausing between the phrasal constituents, her strategy after the utterance by the system that it has not understood is to propose a new date. Thus, the use of repetitions, reformulations and new proposals has to be seen as strategies which speakers may use to solve particular problems arising in the interaction with the computer. While reformulations occur usually in earlier phases of the dialogues, it is more likely that repetitions occur towards the end of the dialogues rather than at the beginning.

(7) a. e0112301: <Laugh> <P> ich hätte gerne ein' Termin im Januar. <P> paßt es Ihnen am Dienstag, dem 19.1., von 8 bis 14 Uhr? (*<Laugh> <P> I'd like to have an appointment in January. <P> Does it suit you Tuesday the 19th from 8am to 2pm?*)

s0112302: ich habe Sie nicht verstanden. (*I did not understand*)

e0112302: ich hätte gerne einen Termin am Dienstag, den 19.1., <P> von 8 bis 14 Uhr. (*I'd like to have an appointment on Tuesday the 19th <P> from 8am to 2pm.*)

b. e0114301: <Swallow> Montag, 18.1., von 8 bis 12 Uhr. (*<Swallow> Monday 18th from 8 to 12 am*)

s0114302: ich habe Sie nicht verstanden. (*I did not understand*)

e0114302: ich hätte gerne ein' Termin, am Montag, dem 18.1., von 8 bis 12 Uhr. (*I'd like to have an appointment on Monday the 18th from 8 to 12 am*)

c. e0115103: <B> Freitag, der 22.1., von 8 bis 12 Uhr. (*<B> Friday the 22nd from 8 to 12 am*)

s0115201: ich habe Sie nicht verstanden. (*I did not understand*)

e0115201: <B> Freitag, der 22.1., <P> von 8 bis 12 Uhr. (*<B> Friday the 22nd <P> from 8 to 12am*)

d. e0118206: am Mo<L>ntag, <P> dem 4.1., <P> von 12 bis 14 Uhr. (*on Mo<L>nday <P> the 4th <P> from 12 to 2pm*)

s0118207: ich habe Sie nicht verstanden. (*I did not understand*)

e0118207: <B> am Dienstag, dem 5.1., von 12 bis 14 Uhr. (*<B> on Tuesday the 5th from 12 to 2pm*)

Especially in later phases of the dialogues, speakers may repeat their utterances irrespective of the speech act uttered by the system. A statistical analysis shows that 35% of all repetitions occur in the last 34 turns (of 124 turns total). Further instances of the use of this strategy were examples (4) and (5).

The findings about the employment of repetitions when all other, more cooperative, strategies have been unsuccessful may render repetitions as an indicator of changing speaker attitude. However, things are not that easy, that is, there is no simple rule which would render every repetition uncooperative; firstly, repetition is not like repetition; in the current dialogues, repeats are conditionally relevant, and therefore fully cooperative, when the system claims not to have understood. However, repetitions may also occur in positions in which they are not directly conditionally relevant, for instance, when the system utters something which is not understandable, or when it says something which is unrelated to the speaker's previous utterance, that is, in the context of simulated recognition errors. Thus, repeats have to be distinguished according to their local relevance in the sequential structure of the dialogues. The following example shows not only that the speaker herself asks the system for a repetition, but it also displays two instances of repetitions in which the first one is not conditionally relevant while the second one is since it occurs after a request to propose a date:

(8) s0222101: bla irgendwas Termin wotsoewer?

e0222101: nochmal. (*again*)

s0222102: was soll Termin blablurb wot sehell bla?

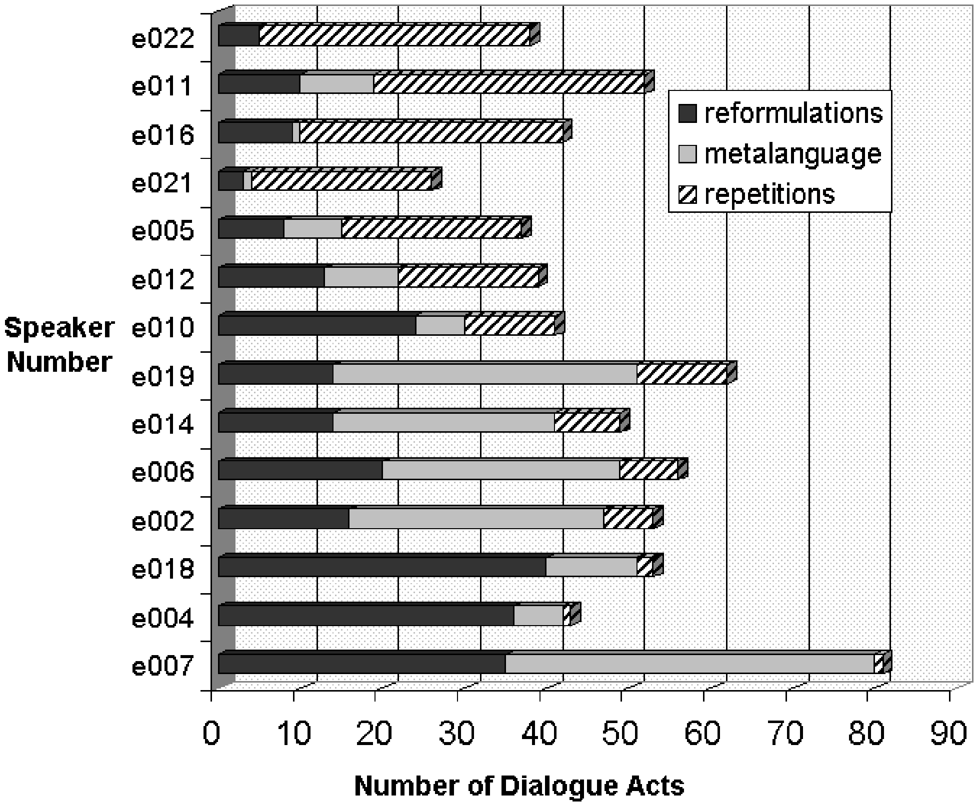
e0222102: Donnerstag, einundzwanzigster erster, zehn bis sechzehn Uhr. (*Thursday, 21rst of January, 10 am to 4 pm*)

s0222103: bitte machen Sie einen Vorschlag. (*please make a proposal*)

e0222103: Donnerstag, einundzwanzigster erster, zehn bis sechzehn Uhr. (*Thursday, 21rst of January, 10 am to 4 pm*)

Secondly, there are different speaker types, that is, speakers may react to the system output very differently. For instance, regarding repetitions, there are speakers (e004, e018) who use less than five repetitions throughout the dialogue while others use more than fifty (e016, e022). Figure 2 shows that speakers can be distinguished according to which communicative strategy they prefer: There are some who prefer to reformulate and to use metalanguage, and there are others who repeat their utterances very often instead. There is a negative correlation of -0.6 between the use of reformulations and metalanguage on the one hand and repetitions on the other. Thus, the number of repetitions occurring is dependent on who is speaking, and repetitions have to be seen as only one out of a number of strategies possible which are in linguistic opposition to the use of repeats in these dialogues.

Figure 2: Number of Reformulations, Metalanguage, and (conditionally non-relevant) Repetitions by Speaker (14 speakers)



On the whole, however, the occurrence of conditionally non-relevant repetitions increases in later phases of the dialogues; speakers are more ready to repeat an utterance when they have given up on other strategies. The point at which that may happen may differ, yet speakers are not found to begin with repetitions and to reformulate in later phases of the dialogue. Thus while this interpersonal variation makes it difficult to evaluate whether a particular conversational strategy is an indicator for a changing speaker attitude or not, independently of the speaker types, the likelihood that a speaker repeats her utterance, rather than, say, reformulates it, increases during the dialogues (see also Fischer, 1999a). Thus, when we consider the reactions to a particular utterance, for example, the system's statement that holidays will be in June and July (when the task is to find a date for an appointment in January), the likelihood that a speaker reacts by means of a repetition increases from 14% when this utterance occurs for the first time to 43% when it is uttered a third time towards the end of the dialogue. Likewise, if the system produces a sequence of incomprehensible utterances, the likeliness that the speakers will only repeat their utterances is five times higher when it occurs for the fifth time than when speakers are confronted with it for the first or even the second time. While in early phases of the dialogues speakers react directly to the system's output, that is, acknowledging what has been said and reacting relevantly, for instance, by means of reformulations and meta-communicational statements, they cease to try out different conversational strategies when they are more frustrated. Thus, the use of repeats after system utterances other than explicit statements of the failure to understand is a characteristic feature of later phases of the dialogue when the speakers are already emotionally engaged, and repetitions can therefore be regarded as an indicator of changing speaker attitude.

Repetitions, distinguished by means of their relevance, are therefore good candidates for indicators of a change of emotional state; the interpersonal variation however can eventually only be accounted for if several features are taken into account in determining the emotionality of an utterance automatically.

Table 5 shows the different conversational properties which have been found to change in the dialogues over time; the ranking is again supposed to account for the occurrence of these strategies in different phases of the dialogues.

3.3 Prosodic and Phonological Peculiarities

Like the lexical and the conversational peculiarities of the dialogues described so far, prosodic and phonological properties are identifiable whose occurrence changes during time within a particular dialogue.

The dimensions along which the prosodic features of the speakers' utterances change include the following:

- the variation of stress patterns;

Table 5: Conversational Peculiarities

	Conversational Strategies	Example
1	reformulation, failed understanding, specification, clarification question	*ich hab' Sie nicht verstanden, wie bitte?*
2	conditionally relevant repetition	–
3	metalanguage laughter, audible breathing	*ich meinte aber,* *das kann ich nicht verstehen*
4	conditionally not relevant utterance	–
5	conditionally not relevant repetition	–
6	objection	*das ist kein Vorschlag,* *das hat nichts zu tun*
7	thematic break, pause	*guten Tag*
8	assessments not directed at the communication partner	*der tilt wieder,* *das ist Kauderwelsch*
9	assessments not directed at the communication partner	*Scheißding,* *Unverschämtheit*

- the inclusion of pauses;
- the variation of speed;
- the variation of loudness;
- the variation of intonation contours;
- the duration of syllables and of particular consonants;
- the inclusion of audible breathing and laughter.

Furthermore, phonological changes can be observed which will be referred to as instances of hyperarticulation.

The phonological and prosodic properties observable are not essentially different form those found by, for instance, Pirker & Loderer (1999); Tischer (1993); Scherer (1986) as emotional properties and from those identified by, for instance, Oviatt *et al.* (1998b,a); Levow (1998) in contexts of error resolution; however, the aspects of variation listed above can be identified on independent grounds by means of the methodology proposed, and, furthermore, they can be shown to be dimensions of variation, rather than particular properties, such that they describe potential strategies which speakers can employ, with both increasing and decreasing values.

In the following, examples for changes along these dimensions will be given; the following examples show repeats in different structural contexts where phonological and prosodic properties of the repeated utterance differ from those in the original utterance:[6]

(9) e0021104: ich meine MONtag, den elften ERSten neunzehnhundertneunundneunzig [tsIC]. (*I mean Monday the eleventh of January, 1999.*)

s0021105: ich habe Sie nicht verstanden. (*I did not understand*)

e0021105: ich meine MonTAG, den elfTEN ersTEN neunzehnhundertneunundneunzig [tsIk]. (*I mean Monday the eleventh of January, 1999.*)

In the original utterance, the speaker emphasizes the major first syllables of the content words *Montag, elften* and *ersten*. The weak syllables of *elften* and *ersten* are reduced to nasal alveolars, which is the standard pronunciation for these syllables. Also according to German standard pronunciation, the final consonant in *-zig* is realized by a palatal fricative. In contrast, in the repetition the non-prominent syllables *-tag* of *Montag* and the *-en*-syllables of *elften* and *ersten* are emphasized. The latter are not reduced but realized by /-En/. The *-zig* ending is realized by means of /tsIk/ in the repetition. Thus, we find variation of emphasis and instances of hyperarticulation in the example.

Likewise, in the repetition of the next example, the speaker (hyper-) articulates the syllable *-ter* of *vierunzwanzigster*, whose pronunciation is /t6/ in the standard pronunciation, as /tEr/:

(10) e0027301: mein VorSCHLAG, SONNtag, vierundzwanzigster [st6] JANuar [janu'a:6], neunzehnhundertneunundneunzig. (*my proposal, Sunday, 24th of January, 1999.*)

s0027302: ich habe Sie nicht verstanden. (*I did not understand.*)

e0027302: <B> Sonntag, VIERundzwanzigster [stEr] JAnuar [“ja:nu?a:r], neunzehnhundertneunundneunzig. (*<B> Sunday, 24th of January, 1999.*)

Consequently, the speakers can be shown to use hyperarticulation as a linguistic strategy to make themselves understood. Furthermore, in the repetition

[6] The phonological transcription is according to Wells *et al.* (1992); emphasis on syllables is represented by capital letters.

a glottal stop is inserted between the two vowels of *Januar*, and the length of *Sonntag* increases by 50%.

In example (11), the repetition is more than a second longer than the original utterance, an increase of 17%. Furthermore, the duration of individual consonants, as for instance the duration of /S/ in *sechsstündig* which is more than doubled in duration from 120msec to 250msec in the repetition, is increased as well:

(11) e0198210: rei<L>cht nicht, wir suchen einen sechsstündigen Termin. DIENstag, neunzehnte erster, ACHT bis vierzehn Uhr. können Sie da? (*not enough, we are looking for an appointment of six hours. Tuesday, 19th, 8am to 2pm. does that suit you?*)

s0198301: ich habe Sie nicht verstanden. (*I did not understand*)

e0198301: einen SECHSs:tündigen Termin. DIENstag, NEUNzehnte erster, ACHT bis vierzehn Uhr. KÖNnen Sie da? (*an appointment of six hours. Tuesday, 19th, 8am to 2pm. does that suit you?*)

Duration is thus one of the linguistic properties to which speakers attend in the realization of repetitions.

In the previous examples, the utterances by the speakers were rejected by the simulated system as not understandable on the whole. In the following example, the system utters something which indicates that it has only misunderstood aspects of the proposal by the speaker. Thus in example (12), the speaker stresses the name of the day and the date very much in the repeat, i.e. those words which carry the main informational load of which she believes that they had been 'misunderstood' the first time. Thus, the speaker has hypotheses about what may have gone wrong and she tries to react accordingly, that is, she attempts to increase the understandability of those words which had not been previously understood.

(12) e0118204: am MONtag, dem VIERten ersten, von ZWÖLF Uhr bis vierzehn Uhr. (*on Monday, the 4th of January, from 12am to 2pm.*)

s0118205: Mittwoch, der sechste erste, von acht bis zehn Uhr ist schon belegt. (*Wednesday, the 6th of January, from 8 to 10 am is already occupied.*)

e0118205: <B> <:<very loud> AM MONtag:>, dem VIERten ERSten, von ZWÖLF bis VIERzehn UHR. (*on Monday, the 4th of January, from 12am to 2pm.*)

s0118206: Donnerstag, von acht bis zehn Uhr ist schon belegt. (*Thursday, from 8 to 10am is already occupied.*)

e0118206: am Mo<L>nta<L>g, <P> dem VIERten ERSten, <P> von ZWÖLF bis VIERzehn UHR. (*on Mo<L>nda<L>y, <P> the 4th of January, <P> from 12am to 2pm.*)

Regarding prosodic peculiarities, this sequence is an example for how a speaker uses loudness and pauses between the individual chunks of information to increase understandability. Besides the varying loudness, in this example, the length of *Montag* ('Monday') increases from 950 msec to 1.15 sec to 1.3 sec in the second repetition. Furthermore, the pattern of emphasis changes between the first proposal of this date and the first repetition. Finally, pauses are introduced between each phrasal constituent in the second repetition.

In example (13), the repeated utterance, in particular the content word *Termin* (*appointment*), is distorted by laughter; however, unlike in the previous utterances, emphasis is reduced in the repeat, and the speaker utters an interjection at the beginning of the repetition which indicates that she is dissatisfied with the system's behaviour:

(13) e0052202: ich hätte gerne einen TerMIN <P> von ACHT bis zwölf UHR, am ZWEIundzwanzigsten JANuar. (*I'd like to meet <P> from 8 to 12am, on the 22nd of January.*)

s0052203: bitte machen Sie einen Vorschlag. (*please make a proposal*)

e0052203: hm <Lip Sound> ich hätte gerne ein' <:<Laugh> TerMIN:> am zweiundzwanzigsten Januar, von acht bis zwölf UHR. (*I'd like to meet on the 22nd of January from 8 to 12am.*)

Consequently, while emphasis in general is increased in local strategies to enhance understandability, speakers may also decrease emphasis if they regard it to be more helpful or the increase as useless. Therefore, it needs to be considered that speakers employ the features identified as variables according to which they vary their linguistic behaviour, rather than that they always increase or decrease these features. As figure 3 shows, the speaker of this utterance pair does not only vary the intonation contour when he repeats his utterance, the repetition is also much faster, less loud, it contains fewer pauses, and the duration of syllables and consonants is much shorter. Thus, again it needs to be considered that speakers are not all alike and that their use of particular strategies may be a trial and error towards an increase in understandability.

Methodologically, it may be problematic to distinguish those strategies which are meant to increase understandability, that is, which are local features, from those which are signs of a changing speaker attitude. However, for all of the properties described, it can be shown that statistically their occurrence is correlated

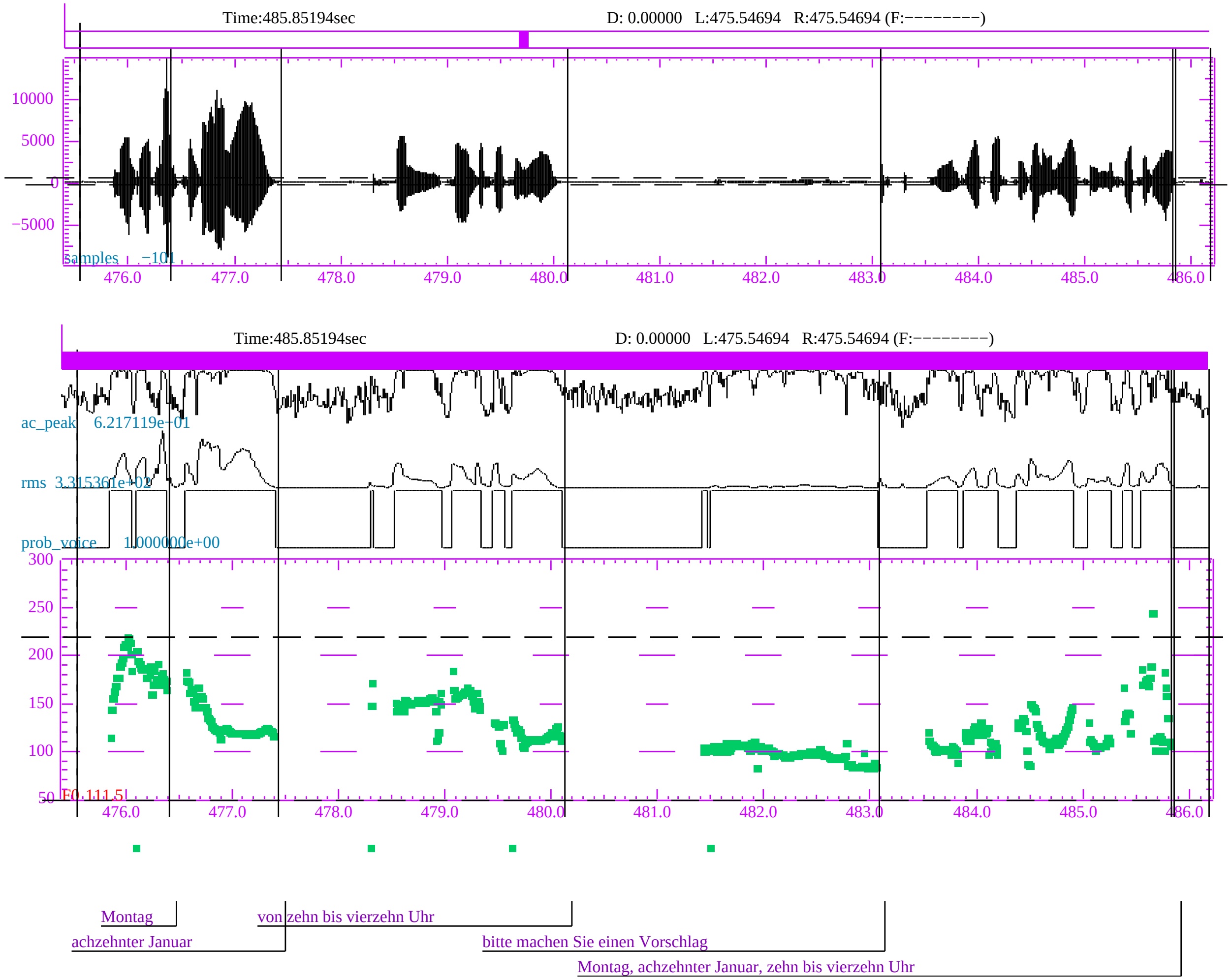


Figure 3: Variation of Intonation Contours in Repetitions

with different phases of the dialogues; the use of the strategies observed varies throughout the dialogues. Figure 4 shows a case in point: Hyperarticulation has been found to occur mainly in the later phases of the dialogues, such that 70% of all instances of hyperarticulation occur in the second half of the dialogues, and 30% can be found within the last 20 speaker turns (the total number being 124 turns per dialogue). Consequently, the prosodic properties of utterances as they occur in repetitions, for instance, have also to be regarded as global properties with respect to the macro-structure of the dialogues and therefore as indicators of emotional changes. While for the current task it is eventually irrelevant whether it is the one or the other since the occurrence of these properties is problematic for automatic speech processing systems in any case, the methodology proposed here allows to show that the use of the prosodic and phonological means which are meant to increase the understandability of utterances and which have been identified for English by Oviatt *et al.* (1998a,b); Levow (1998) and for German by Pirker & Loderer (1999) as local strategies in error resolution are also dependent on the speakers' attitude towards the system.

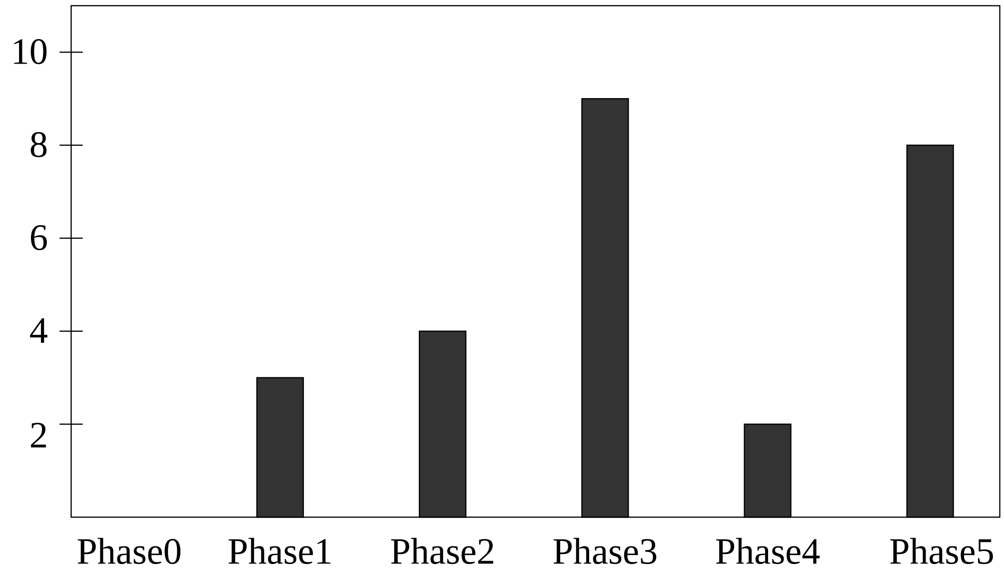


Figure 4: Instances of Hyperarticulation with Respect to the Different Dialogue Phases (approx. 20 turns per phase, 14 dialogues)

The prosodic properties which may change during the dialogues and which may therefore be indicators of changing speaker attitude are shown in Table 7.

Table 7: Prosodic Peculiarities

	Prosodic Properties
1	pauses between phrases
2	particularly clear articulation
3	strong emphasis
4	pauses between words
5	particularly strong (contrastive) emphasis
6	pauses between syllables
7	syllable lengthening
8	hyperarticulation
9	messed up by laughter or audible breathing

4 Conclusion

To sum up, analysing the differences between the same sequences in different structural positions of the dialogue allows to determine those differences which are created by different attitudes towards the system such as increasing dissatifaction. Thus if the speaker reacts differently although the situation is constant and the sequences of system output are the same, the differences can be attributed to a change in attitude towards the system and to emotional involvement. By means of this method, lexical, conversational, and prosodic peculiarities could be identified which can be understood as indicators of emotionality in human-computer interaction. Especially their combination rather than the occurrence of single instances of these properties can be interpreted as indications of changes in speaker attitude. These properties are annotated in the dialogues such that, in addition to the macro-structural annotation shown in Table 2, the turns of each dialogue are marked regarding their lexical, conversational, and prosodic peculiarities which could be identified as those aspects which vary intrapersonally throughout the dialogues, i.e. which are variables according to which speakers adapt their linguistic behaviour to the special circumstances of non-successful human-computer interaction. The descriptive inventory identified contains therefore those features which are relevant as expressions of emotionality and of which speakers may believe that they facilitate understanding by the system. The method proposed here thus allows to determine the relevant properties on independent grounds instead of postulating them on the basis of intuition.

The annotion can now be used to select a turn or a word for the training of an automatic classifier which distinguishes emotional from neutral speech. Rather than deciding *a priori* or on the basis of intuition whether a turn is emotional or not, the annotation developed here provides three means for making this decision:

the lexical, conversational, and prosodic properties of the turn. How exactely these properties are taken to interact and on which bias to base the classification will have to be determined in several hypothesis-test cycles. Yet the method proposed provides a methodologically sound basis for identifying those properties which indicate the speaker's attitude toward the system. Being able to recognise automatically those utterances in which the speaker is becoming emotionally involved may allow system designers to introduce, for instance, aspects of system behaviour which may calm down the angry user (Fischer, 1999b).

References

Battacchi, M. W., Suslow, T., & Renna, M. 1997. *Emotion und Sprache*. 2nd edition edn. Peter Lang: Frankfurt a.M. et al.

Buck, R. 1994. The Neuropsychology of Communication: Spontaneous and Symbolic Aspects. *Journal of Pragmatics*, **22**, 265–278.

Clark, H. H. 1996. *Using Language*. Cambridge University Press.

Cowie, R., Douglas-Cowie, E., & Romano, A. 1999. Changing Emotional Tone in Dialogue and its Prosodic Correlates. *Pages 41–46 of: Proceedings of the ESCA Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands.*

Drescher, M. 1997. *Sprachliche Affektivität: Darstellung emotionaler Beteiligung am Beispiel von Gesprächen aus dem Französischen.* Habilitationsschrift, Universität Bielefeld.

Fiehler, R. 1990. *Kommunikation und Emotion. Theoretische und empirische Untersuchungen zur Rolle von Emotionen in der verbalen Interaktion.* De Gryuter, Berlin, New York.

Fischer, K. 1998. *Szenariodesign: Elizitieren von emotionalen Äußerungen.* Memo 140. Verbmobil, University of Hamburg.

Fischer, K. 1999a. Discourse Effects on the Prosodic Properties of Repetetitions in Human-Computer Interaction. *Pages 123–128 of: Proceedings of the ESCA-Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands.*

Fischer, K. 1999b. Repeats, Reformulations, and Emotional Speech: Evidence for the Design of Human-Computer Speech Interfaces. *Pages 560–565 of:* Bullinger, Hans-Jörg, & Ziegler, Jürgen (eds), *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany.* Lawrence Erlbaum Ass., London.

Fraser, N., & Gilbert, G.N. 1991. Simulating Speech Systems. *Computer Speech and Language*, **5**, 81–99.

Giles, H., & Williams, A. 1992. Accomodating Hypercorrection: A Communication Model. *Language and Communication*, **12**(3/4), 343–356.

Gumperz, J. 1982. *Discourse Strategies.* Studies in Interactional Sociolinguistics, no. 1. Cambridge University Press.

Huber, R., Nöth, E., Batliner, A., Buckow, J., Warncke, V., & Niemann, H. 1998. You BEEP Machine - Emotion in Automatic Speech Understanding Systems. *Pages 223–228 of:* Sojka, Petr, Matousek, Václav, Pala, Karel, & Kopecek, Ivan (eds), *Proceedings of the First Workshop on Text, Speech, Dialogue-TSD'98, Brno, Czech Republic, September 1998.* Masaryk University Press.

Kehrein, R. 1998. Linguistik der Emotionen - Suprasegmentalia und Bewertung. *In:* Hartung, Martin, & Brock, Alexander (eds), *Neuere Entwicklungen der Gesprächsforschung: Vorträge der 3. Arbeitstagung des Pragmatischen Kolloquiums Freiburg.* Tübingen: Narr.

Lang, P.J. 1988. What are the data of emotion? *In:* Hamilton, V., Bower, H.H., & Frijda, N.H. (eds), *Cognitive Perspectives on Emotion and Motivation.* Dordrecht: Kluwer Academic.

Levow, G.-A. 1998. Characterizing and Recognizing Spoken Corrections in Human-Computer Dialogue. *In: Proceedings of Coling/ACL '98.*

Oviatt, S., Bernard, J., & Levow, G.-A. 1998a. Linguistic Adaptations during Spoken and Multimodal Error Resolution. *Langauge and Speech*, **41**(3-4), 419–442.

Oviatt, S., MacEachern, M., & Levow, G.-A. 1998b. Predicting Hyperarticulate Speech During Human-Computer Error Resolution. *Speech Communication*, **24**, 87–110.

Picard, R. 1997. *Affective Computing.* The MIT Press, Cambridge, Mass.

Picard, R.W. 1999. Affective Computing for HCI. *Pages 829–833 of:* Bullinger, Hans-Jörg, & Ziegler, Jürgen (eds), *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany.* Lawrence Erlbaum Ass., London.

Pirker, H., & Loderer, G. 1999. I said "TWO TI-CKETS": How to talk to a deaf wizard. *Pages 181–186 of: Proceedings of the ESCA Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands.*

Sacks, H., Schegloff, E. A., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**(4), 696–735.

Scherer, K. R. 1982. Die vokale Kommunikation emotionaler Erregung. *Pages 287–306 of:* Scherer, Klaus R. (ed), *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens.* Weinheim & Basel:Belz.

Scherer, K. R. 1986. Affect Expression: A Review and a Model for Future Research. *Psychological Bulletin*, **99**, 143–165.

Scherer, K. R., & Wallbott, H. G. 1990. Ausdruck von Emotionen. *In:* Scherer, Klaus R. (ed), *Psychologie der Emotion. Enzyklopädie der Psychologie*, vol. C IV 3. Verlag für Psychologie: Göttingen, Toronto, Zürich.

Schmidt-Atzert, L. 1981. *Emotionspsychologie.* Stuttgart: Kohlhammer.

Schmidt-Atzert, L. 1993. *Die Entstehung von Gefühlen. Vom Auslöser zur Mitteilung.* Lehr und Forschungstexte Psychologie, no. 47. Springer: Heidelberg et al.

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

Wells, J., Barry, W., Grice, M., Fourcin, A., & Gibbon, D. 1992. *Standard Computercompatible Transcription.* Tech. rept. Phonetics and Linguistics Department, UCL.

Wierzbicka, A. 1992a. Defining Emotion Concepts. *Cognitive Science*, **16**, 539–581.

Wierzbicka, A. 1992b. Talking about Emotions: Semantics, Culture, and Cognition. *Cognition and Emotion*, **6**(3/4), 285–319.

Wierzbicka, A. 1995. Lexicon as a Key to History, Culture and Society. *Pages 103–155 of:* Dirven, René, & Vanparys, Johan (eds), *Current Approaches to the Lexicon.* Duisburg Papers on Research in Language and Culture, no. 24. Frankfurt a.M.: Lang.

Williams, C. E., & Stevens, K. N. 1982. Akustische Korrelate diskreter Emotionen. *Pages 307–325 of:* Scherer, Klaus R. (ed), *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens.* Weinheim & Basel: Belz.