

Linguistic Expression of Emotion in Human-Machine Interaction: The NIMITEK Corpus as a Research Tool

Milan Gnjatović¹, Manuela Kunze¹, Xian Zhang¹, Jörg Frommer² and Dietmar Rösner¹

Abstract. Since the end of 2005, the NIMITEK consortium investigates issues in spoken human-machine interaction (HMI). We employ the multimodal NIMITEK corpus of affected behaviour in HMI as a tool that provides an empirical foundation for the development of integrated emotion detection from both signal based as well as content based emotion recognition. This paper primarily discusses various linguistic features that may carry affect information (e.g., key words and phrases, lexical cohesive agencies, dialogue act sequences, etc.). Finally, we report on a first prototype of an automatic annotator for recognition and tracking of the user's emotional state from linguistic information.

1 INTRODUCTION

Since the end of 2005, the NIMITEK consortium investigates issues in spoken HMI. In the first phase of this project, specially designed wizard-of-oz experiments (WOZ) have been carried out in order to collect realistic material from simulated human-machine conversations with *everyday* users. Our point of departure was that subjects have to be motivated to accomplish a given task in order that a successful accomplishment or a failure to accomplish could induce an emotional state. Therefore, the scenario was the following: we instructed subjects to undergo a test of both intelligence and communicational abilities, supported by a spoken dialogue system. For this purpose, they were confronted with a set of graphical tasks (e.g., Tower-of-Hanoi puzzle, Tangram, etc.) and the subjects could only verbally address the system. Instructions accepted by the system were not predefined. The result is the multimodal NIMITEK corpus of affected behaviour in HMI [1] that comprises 15h of annotated video and audio recordings of WOZ sessions with German speaking subjects, including genuine emotional expressions elicited in the laboratory settings. The corpus was annotated with respect to its emotional content, and all dialogues were also transcribed.

The corpus had an important role in the development of the NIMITEK prototype system [2]. The NIMITEK prototype system is an adaptive spoken dialogue system for supporting users while they solve problems in a graphics system. The central feature of the system is adaptive dialogue management. The underlying idea is that the dialogue manager dynamically adapts a dialogue strategy according to the current state of the interaction, including the emotional state of the user.

¹ Department of Knowledge Processing and Language Engineering, Otto-von-Guericke-University Magdeburg. Email: {gnjatovic, makunze, xzhang, roesner}@iws.cs.uni-magdeburg.de.

² Psychosomatische Medizin und Psychotherapie, Otto-von-Guericke-University Magdeburg. Email: joerg.frommer@med.ovgu.de.

In the NIMITEK system, information about the emotional state of the user is provided to the dialogue management module by the emotion recognition module. In the actual implementation, this module combines two knowledge resources within an integrated classification: prosodic cues and cues from facial expressions. In this paper, we discuss the possible integration of an additional emotional classifier in the emotion recognition module. We focus on the recognition and tracking of the emotional state of the user from lexical information and other linguistic features.



Figure 1. Interaction with the NIMITEK system.

Previous research experiences show that various sources of information (e.g., non-verbal information, interruption in fluency, parts-of-speech, n-grams, etc.) may be used to improve automatic recognition of emotion that is based on prosodic features. In the following, we use the transcribed version of the NIMITEK corpus to investigate various linguistic features that may carry affect information (e.g., key words and phrases, lexical cohesive agencies, dialogue act sequences, etc.).

2 SIGNALLING EMOTIONS

Other projects of the NIMITEK consortium work on detecting emotion from the speech signal (especially from prosody) and from an analysis of the facial expressions. In the following, we will report about analyses of emotional content from the linguistic structures in the transcribed conversations.

We start our discussion showing a sequence of commands taken from the NIMITEK corpus. The discourse was produced by the subject while she was solving a graphical puzzle:

I take the parallelogram ... Yes, move slowly to the right ... More ... stop ... Please move slowly up ... stop ... Please move slowly to the right.

Figure 2. A sequence of commands produced by the subject solving a graphical task.

The human operator that plays the role of the system in the WOZ settings performs the instructed command properly, so the interaction between the subject and “the system” unfolds without problems in communication. In contrast to such an “unproblematic” dialogue fragment, we recognized that there are different styles how subjects approach the system in a case when a problem occurs. One could be termed *pedagogical* or *teacherese* and is characterized by trying to teach the computer how it should behave properly. The following dialogue fragment (translated from German) illustrates this:

SUB: The smallest ring from the one to the three ... Stop! Back ... Where should the smallest ring go? ... On the three ... three! ... What are you doing there?
WOZ: I am doing just what you are asking.
SUB: Incorrect! ... Down ... these rings stay down! ... Understood?
WOZ: What rings?
SUB: The middle ring and the large ring,... now put...
WOZ: I don't understand you.
SUB: The smallest ring! ... On the three please, ... on the three ... No! On the three ... Where is the three? ... On the three! ... Where should the smallest ring be placed?
WOZ: On the three.
SUB: Please, do it!

Figure 3. A dialogue fragment illustrating *pedagogical* or *teacherese* talk style.

Another style is characterized by open signals of despair and helplessness, when problems pile up:

SUB: No! No! No! ... Execution not correct!
WOZ: You still haven't solved the task.
SUB: I know that.
WOZ: Do you need help?
SUB: The communication is not working. You don't understand my instructions and you don't do what I say.
WOZ: I am doing only what you are saying.
SUB: No, it is not true.

Figure 4. A dialogue fragment illustrating open signals of helpness.

We are investigating a typology of users' utterances and sequences of users' utterances that signal emotional state of the user. In the next sections, we discuss insights from the NIMITEK corpus relating to different linguistic features that may carry affect information.

2.1 SPECIFIC KEY WORDS AND PHRASES

One way to recognize an emotional state is to detect key words and phrases in users' utterances. Here are some examples of key words and phrases that relate to certain emotion-related states and attitudes (translated; German original in parentheses):

Annoyed: Sh*t (Sche*ße), stupid (blöd), Do what I say (Tu was ich sage), I've had enough of it (Es reicht mir), It is mean (Das ist gemein). Oh ... something like this I hate just like the plague. (oohh... so was hasse ich doch wie die Pest.)
Retiring: I don't understand it (Ich versteh' das nicht), It's not working at all (Das geht doch gar nicht), I don't understand the task (Ich versteh'

die Aufgabe nicht).
Indisposed: I am going now (ich geh' gleich), Oh man (Oh man), God (Gott), I don't feel like doing any more. (Ich hab' kein' Bock mehr.)
Offending: You think, doll. (Denkst du, Puppe)
Satisfied: Super (Super), yeah! (yeah!), awesome (geil), I am good, am I not? (Bin gut, was?)
Polite: Please (Bitte), I would like (Ich hätt' gern).
Friendly: Dear computer, ... (Lieber Computer, ...)

Figure 5. Examples of key words and phrases that relate to various emotional state.

However, expressions of emotions are not limited to a single dialogue act, but they map over a range of mutually related dialogue acts. Therefore, we consider also lexical cohesive agencies [3, p. 309-334] (that relate dialogue acts in a structure) and dialogue acts sequences, cf. [4], in order to detect signals of emotion-related states. In the next sections, we are primarily focused on recognizing signals of negative emotional states.

2.2 LEXICAL COHESIVE AGENCIES

Ellipsis-substitutions: Ellipsis-substitution is a form of anaphoric cohesion in a discourse, *where we presuppose something by means of what is left out* [3, p.316]. For example, in:

Why are you **moving** it on peg 2? Why? Why are you **doing** this step? (Warum **fährst** du auf Säule 2? Warum? Warum **machst** du diesen Schritt?)

Figure 6. An example of a dialogue sequence containing question with ellipsis-substitution.

the subject replaces the verb *move* (*fahren*) with the general verb *do* (*machen*). It is important to note, that in *ellipsis-substitutions the typical meaning is not one of co-reference. There is always some significant difference between the second instance and the first* [3, p. 322]. To illustrate this, let us observe a typical example for an ellipsis-substitution in the NIMITEK corpus: Please do it! (Bitte tu das!). This utterance does not explicitly carry information what is the system expected to do. It contains an elliptical-substitution (*do*), a reference (*it*) that relates to context and an element of politeness (*please*). Here the ellipsis-substitution is used to signal that the action that the system performed is not the same as the action instructed by the user. Thus, ellipsis-substitutions may carry a signal of a potential problem in the communication. Such a problem may be related to the given task or to the interface language.

Lexical cohesion: The choice of lexical items to create cohesion in the discourse can also signal an emotion-related state, both on the lexical level (e.g., repetitions), as well as on the semantic level (e.g., reformulations). For example:

Simple repetition: It just cannot be. It just ... It just cannot be. (Das kann doch nicht sein. Das ist doch ... das kann doch nicht sein.)
Repetition and remark: What is the problem? What is the problem? (Was ist das Problem? Was ist das Problem?)
 Left up. Left up. Left up. **I said** left up. (Links oben. Links oben. Links oben. **Ich habe gesagt** links oben.)

Reformulation: Not **true** at all. That's **definitely wrong**.
(Gar nicht **wahr**. Das **stimmt** gar nicht.)

Figure 7. Examples illustrating how the choice of lexical items to create lexical cohesion can relate to negative user states.

Dialogue act sequences. The type of dialogue acts in a sequence may also carry affect information. For example, a sequence of questions may signal potential problems in communication. An illustration of such a sequence containing a question with ellipsis substitution is given in Figure 6.

2.3 QUESTIONS

Subjects in the NIMITEK corpus used also questions to signal their frustration or uncertainty. Common for these questions is a relatively high level of abstraction, as we illustrate below. We differentiate several groups of such questions.

The first group of questions signals the frustration of the user. It contains probe questions that start with 'what' or 'why', usually are characterized by short structure and contain ellipsis substitutions, e.g.: What are you doing? (Was tust du?) What's the point of that? (Was soll das?) Why are you doing this? (Warum machst du das?) Question from the second group relate to a concrete action that was or should be performed by the system, e.g.: Why don't you move the 8 to left? (Warum schiebst du die 8 nicht nach links?) The third group signals that the user is confused or retiring. These questions usually contain a reference to a previous utterance, such as: But there is one more, isn't there? (Aber es gibt noch eins, oder?) Finally, the fourth group contains rhetorical questions, e.g.: Did I say disk downward? No! (Hab ich gesagt Scheibe nach unten? Nein!)

2.4 NEGATION

Negation, combined with other functional elements (e.g. modal particles) may also signal a potential problem in communication. For example:

Negation: Right up. **No**, right up.
(Rechts oben. **Nein**, rechts oben.)

Negation with enhancement: **No**, it is not right, It is **simply** not right.
(**Nein**, das stimmt nicht. Das stimmt **einfach** nicht.)
I don't understand it. I **really** don't understand it. (Ich verstehs nicht. Ich verstehs **echt** nicht.)

Negation with confirmation: Do I mean it? **No**, I don't mean it, **do I?**
(Meinte ich das? **Nee**, das meinte ich nicht, **oder**)

Negation with generalization: **No**, that will come to nothing. **No**, that will all come to nothing. (**Nee**, das wird **alles** nichts mehr. **Nee**, das wird nichts mehr.)

Figure 8. Example illustrating how negation can relate to negative user states.

3 EVALUATION OF THE NIMITEK CORPUS

The evaluation of the emotional content of the NIMITEK corpus was performed in two phases. Since our research focuses on examining aspects of affected speech rather than of facial gestures, the evaluators were allowed only to hear audio recordings from the corpus. In the first phase of the evaluation process, we defined a data-driven model [2, 5] of user states.

Evaluators (German speakers and non-German speakers) performed the perception test assigning one or more labels to each evaluation unit. We selected a dialogue turn or a group of several successive short dialogue turns as an evaluation unit. The selection of (these sometimes rather long) evaluation units was motivated by the intention to demonstrate the naturalness of the collected recordings, i.e., to demonstrate that emotional expressions are extended in time (cf. [1]). Only utterances produced by the subjects were evaluated, while wizard's expressions were ignored. An example of an evaluation unit is given in Figure 9.

SUB: Repeat the task. ... O... I'm too stupid for this, I don't understand the task.

WOZ: You can manage it!

SUB: I don't understand this. ... How can I move this?

WOZ: Do you need an advice?

SUB: Yes.

Figure 9. An example of an evaluation unit.

The choice of labels was data driven—the evaluators were allowed to introduce labels according to their own perception. They introduced 9 emotion labels, 10 subject's state labels and 3 talk style labels. The evaluators performed the perception test independently from each other and some of the introduced labels represent different but closely related emotions or emotion-related states. In order to define a usable data-driven model of user states, we grouped labels that relate to similar or mixed emotions or emotion-related states. Using clarifications provided by the evaluators, we mapped these labels onto six classes that form the ARISEN model of user states: **Annoyed**, **Retiring**, **Indisposed**, **Satisfied**, **Engaged**, **Neutral**. In the second phase of the evaluation, the experimental sessions were re-evaluated by a new group of German speaking evaluators. These evaluators could then use only labels from the ARISEN model. Each evaluation unit was evaluated by four or five evaluators. We used majority voting in order to attribute labels to evaluation units. If at least three evaluators agreed upon a label, it was attributed to the evaluation unit.

4 EXPERIMENTS IN AUTOMATIC ANNOTATION

We used the UIMA framework, cf. [6], to implement a first prototype of an automatic annotator for recognition of the user emotional state from linguistic information. In addition, we developed interfaces³ for graphical representation of the annotator's recognition results and for graphical comparison of the human evaluators' results and the annotator's results. The implementation of the annotator is based on the observations discussed above. For emotion recognition in transcribed dialogues, we used regular expressions. These expressions described specific key words and phrases as well as additional features from the transcriptions of the dialogues from the corpus. Although it was not required to mark prosodic features in the transcribed text, some of the transcribers occasionally and voluntarily did it. For example, they inserted vowels, e.g., *Oh*,

³ Screenshots of these interfaces can be found at: http://wdok.cs.uni-magdeburg.de/forschung/projekte/uima-workbench/projects/uima-based-analyses/analyses_dialogue_data/

Go_o_od!!! (Oh Gooott!!!), or only used upper case letters within a word, e.g., *Next FIELD (Nächstes FELD)*. These corpus specific features are also used to recognize emotions in the transcribed dialogue data. All these patterns are assigned to a concrete class of the ARISEN model. This assignment is based only on an interpretation of linguistic features. The selection and the assignment of words and phrases were performed independently from the voting of the human evaluators described in previous section. Only the transcriptions of observed experimental sessions were used, not video recordings.

The process of the automatic annotation was performed in the same manner as those performed by the human evaluators. The automatic annotator attributed – in accordance with detected linguistic features – zero, one or more labels from the ARISEN model to each subject’s utterance. There were two reasons to implement the annotator in this manner. First, subjects’ expressions in the NIMITEK corpus contain often mixed emotion. Second, we wanted to get results from the automatic annotator that can be compared with the results of the human evaluators, both individually and collectively. The process of evaluating the NIMITEK corpus and the automatic annotator is illustrated in Figure 10.

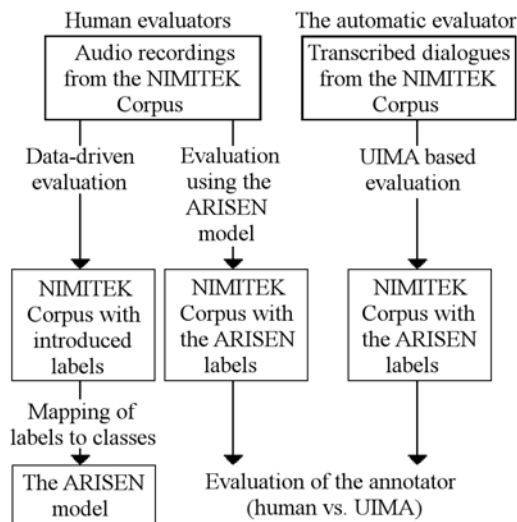


Figure 10. The process of evaluating the NIMITEK corpus and the automatic annotator.

For the purpose of evaluating the automatic annotator, we selected a part of a WOZ experimental session from the NIMITEK corpus, with a duration of 45 min. that was divided into 55 evaluation units. We compare the results of the automatic annotator with the results of the majority voting of the German group of evaluators. Tables 1 and 2 provide evaluation results for two evaluation units in order to illustrate how we compared the obtained results. The first example is given in Table 1. The human evaluators – after we used the majority voting to summarize their individual evaluations – attributed labels **Retiring** and **Annoyed** to the given evaluation unit. In Table 1, it is shortly written as R, A. The automatic annotator attributed labels **Annoyed**, **Indisposed**, **Satisfied** and **Engaged** to the same evaluation unit, i.e., A(2),I,S,E(2). A number in brackets following a label – we refer to it as *the weight of the attributed label* – has the following meaning:

- For human evaluators, it relates to the intensity of the emotion represented by the given label as the evaluators perceived it. For example, the evaluation unit given in Figure 9 was attributed with the label **Retiring** (intensity 1) and the label **Indisposed** (intensity 1.25.)
- For our automatic annotator, it relates to the frequency of occurrence of the given label in the annotation of the observed evaluation unit. E.g., in the evaluation unit given in Figure 9, the annotator detected a phrase that relates to the label **Annoyed**, a phrase that relates to the label **Indisposed**, and two phrases that relate to the label **Retiring**.

Human evaluators: Majority voting	Automatic annotator results	Hit (%)	Miss (%)	False hit (%)
R, A	A(2), I, S, E(2)	33.3%		66.7%

Table 1. Example of comparing results from human evaluators and the automatic annotator.

If no number is specified, the weight of the given label is considered to be 1. The weight of a label is used to define absolute and relative weights of the annotator’s results. For example, for the evaluation unit given in the Table 1:

- The absolute weight of the annotator’s result is the sum of the weights of labels attributed by the annotator, i.e., 2+1+1+2=6.
- We refer to the part of the annotator’s result that matches with the majority voting of the human evaluators as *Hit*. The *Hit* in the annotator’s result is A(2).
- The absolute weight of the *Hit* in the annotator’s result is 2.
- The relative weight of the *Hit* in the annotator’s result is a ratio between absolute weight of the *Hit* and absolute weight of the whole result, i.e., 2/6 (33.3%).
- We refer to the part of the annotator result that represents incorrectly recognized emotional state of the subject as the *False hit*. The *False hit* in annotator’s result is: I, S, E(2).
- The absolute weight of the *False hit* in the annotator’s result is 1+1+2=4.
- The relative weight of the *False hit* in the annotator’s result is 4/6 (66.7%).

Human evaluators: Majority voting	Automatic annotator Results	Hit (%)	Miss (%)	False hit (%)
A, E, R	R, E(2)	66.7%	33.3%	

Table 2. Another example of comparing results from human evaluators and the automatic annotator.

Another example, when the annotator’s recognition of the emotional state is correct but incomplete is given in Table 2, i.e., when the set of labels attributed by the annotator is a strict subset of the set of labels attributed by the human evaluators. Then, the comparison of the results is performed similarly as in the previous case: The relative weight of the *Hit* in the human evaluators’ result is 2/3 (66.7%). We refer to the part of the human evaluators’ result that was not detected by the automatic

annotator as the *Miss*. The *Miss* in the human evaluators' result is: A. Its relative weight is 1/3 (33.3%).

Using these rules, we compared – for each evaluation unit independently – evaluation results of the selected part of a WOZ experimental session from the NIMITEK corpus provided by human evaluators and the automatic annotator. Then we summed *Hit*, *Miss* and *False hit* rates for all evaluation units. In 6 evaluation units (out of 55 in total) there is no agreement among human evaluators with respect to the label. These evaluation units are not considered. The summarized results of the evaluation of the automatic annotator are given in Table 3.

<i>Hit (%)</i>	<i>Miss (%)</i>	<i>False hit (%)</i>
31.70%	34.35%	33.92%

Table 3. Evaluation using the ARISEN model.

For the given 6-classes emotional model ARISEN (**A**nnoyed, **R**etiring, **I**ndisposed, **S**atisfied, **E**ngaged, **N**eutral), the annotator showed the following performance:

- 31.70% of subject emotional states were correctly recognized as they were detected by the human evaluators,
- 34.35% of subject emotional states were not recognized,
- 33.92% of subject emotional states were incorrectly recognized.

Interpreting these results that show differences in evaluation between the human evaluators and the automatic annotator, we make two remarks.

The first remark relates to the process of annotation. The annotator doesn't use additional prosodic information –available for the human evaluators - that can improve recognition rate.

The second remark relates to the NIMITEK corpus. Emotional states of the subjects in the NIMITEK corpus can be often characterized as mixed and extended in time. During the first phase of the evaluation of the NIMITEK corpus, German evaluators attributed more than one label to 39.86% of all evaluation units. Non-German evaluators attributed more than one label to 52.12% of evaluation units. Assigning more than one label to an evaluation units means that there was a change in the expressed emotion within the given evaluation unit (to recall, the smallest evaluation units in the first phase of evaluation is a dialogue turn). This fact makes emotion recognition (even) more challenging. Further evaluation details are given in [1].

5 ADAPTION OF THE ANNOTATOR

In the current implementation of the NIMITEK system, information about emotional state of the user is provided by the emotion recognition module. It integrates the prosodic classifier [7] and the facial expression classifier [8]. Our intention is to integrate our annotator for recognition and tracking of the user's emotional state from linguistic information.

The dialogue manager in the NIMITEK system differentiates between 3 emotional states (cf. [2]): positive, neutral and negative. Thus, we down-sample the ARISEN model to the 3-class problem: the states **A**nnoyed, **R**etiring and **I**ndisposed can be interpreted as negative; the states **S**atisfied and **E**ngaged as positive; and the state **N**egative remains. Results of the evaluation of the automatic annotator using this less fine-grained classification are given in Table 4. These results relate to the offline emotion recognition in the corpus. The annotator will be

also adapted to online emotion recognition in the NIMITEK system.

<i>Hit (%)</i>	<i>Miss (%)</i>	<i>False hit (%)</i>
51.20%	33.67%	17.26%

Table 4. Evaluation using a 3-class model: positive, neutral, negative.

6 DISCUSSION AND ONGOING WORK

This paper reports about work-in-progress: experiments in automatic annotation of emotional states. Further and more detailed investigations are planned. We discuss three lines of research to improve the performance of the automatic annotator.

The first, rather obvious, line of research is to integrate the annotator with other classifiers (e.g., prosodic classifier, facial expression classifier, etc.). A possible explanation for both incorrect recognition (*False hit*) and non-recognition (*Miss*) of emotional states could be that the annotator needs also prosodic information in order to recognize these emotional states. This is illustrated by the sequence of subject's utterances that is given in Figure 11. Due to its prosodic cues, this sequence was attributed by the human evaluators with the labels **A**nnoyed and **R**etiring. However, this sequence does not contain any "emotional" key word, so it was not recognized by the automatic annotator to carry emotional information.

Yes ... It is given this way. ... System, what is the solution of this task? ... Repeat the task. (Ja. ... Das steht da auch so drin. ... System, wie ist die Lösung dieser Aufgabe? ... Aufgabe wiederholen.)

Figure 11. A sequence of subject's utterance without emotional key words.

Another explanation relates to the linguistic features that the annotator takes into account. It leads us to the second line of research. The analysis of specific key words and phrases does not suffice for emotion recognition. Various problems occur when the emotion recognition is based only on detection of keywords and phrases (cf. [9]). Besides the above mentioned – e.g., the lack of prosodic information – we state some of them: (1) *Ambiguity in defining emotional keywords and phrases*. For example, the subject's exclamation "Oh" may express both surprise and disappointment in the NIMITEK corpus. (2) *Ambiguity in syntactic and semantic information*. For example, in the sequence of subject's utterances given in Figure 12, the system should resolve what the meaning of the second utterance is. It may be that the subject does not understand the given task, so she asks the system to repeat introducing instructions. Or it may be that a problem related to the interface language occurred, so the user, as a part of her dialogue strategy, asks the system to repeat the instructions previously uttered by the subject in order to control if the system understood them correctly. In the latter case, it is also a signal of a potential problem in communication.

Downward. ... System, repeat the instructions! (Nach unten. ... System, wiederhole die Anweisungen!)

Figure 12. An example of unclear meaning in a subject's utterance.

Incorrect recognition of emotional states may indicate that the annotator uses a set of key words and phrases that are ambiguous

with respect to the emotional state they can signal. The process of improvement of the annotator includes detection of such key words and phrases and investigation of possibilities to resolve them in a given dialogue context. On the other hand, non-recognition of emotional states may indicate that the set of linguistic features that the annotator takes into account should be extended in order to cover these emotional states. Therefore, additional linguistic features, such as information about structure of dialogue acts, context, lexical information, etc., should also be considered in the recognition process. Currently, we extend our approach to emotion recognition based on linguistic information – we integrate additional analyses of the above mentioned features.

Finally, the third line of research is corpus specific. We recall that, according to the experimental settings, the wizard was intentionally trying to induce emotions in subjects. Stimuli used for an emotional response were e.g., intentional misunderstanding of subject's request and performing an incorrect operation, pretending not to understand subject's request and asking for a repetition, confronting subjects to unsolvable tasks, making provocative statements, etc. The third line of our research is to consider linguistic information from an additional source – the wizard. We state two reasons why we find this research line interesting. First, wizard's utterances were intended to induce an emotional state at the first place, and the inspection of the NIMITEK corpus shows that they influenced subjects. Second, in a realistic scenario including an implemented system, utterances produced by the system – as an equivalent to utterances produced by the wizard – are not subject of recognition – they are known to the dialogue manager. It should be mentioned that the efficiency of emotion recognition from linguistic information derived from the user depends to a large extent on the accuracy of automatic speech recognition (ASR). But, state-of-the-art ASR approaches still cannot deal with flexible, unrestricted users' language [10]. In contrast to this, utterances produced by the system contain information that is apriori known. We are aware that claiming that wizard's utterances necessarily influence subjects' states is too strong, and probably not true. However, we hypothesize that this additional information might be useful in two decision making processes related to emotion recognition: (1) resolving ambiguity of emotional keywords and phrases collected from the user, when prosodic information is not available, and (2) interpreting different emotion recognition results that are provided from different classifiers for the same user input. We plan to investigate this hypothesis.

7 CONCLUSION

The NIMITEK corpus is a unique research tool since it comprises genuine (i.e., non acted) emotions from subjects that can be judged as representatives for 'typical' non-trained, non-technical users of speech-based conversation systems. Since the subjects were not restricted by given predetermined linguistic constraints on the language to use for their utterances are as well indicative for the way in which such users probably like to converse with conversational agents. The current investigations of the recordings in the NIMITEK corpus are intended as an empirical foundation for the development of integrated emotion detection from both signal based as well as content based emotion recognition that in the long run should contribute to

make conversational agents more sensitive and responsive with respect to potential failures in human-machine conversations.

ACKNOWLEDGEMENTS

The presented study is performed as part of the NIMITEK project (<http://wdok.cs.uni-magdeburg.de/nimitek>), within the framework of the Excellence Program "Neurowissenschaften" of the federal state of Sachsen-Anhalt, Germany (FKZ: XN3621A/1005M). The responsibility for the content of this paper lies with the authors.

REFERENCES

- [1] M. Gnjatović and D. Rösner. The NIMITEK Corpus of Affected Behavior in Human-Machine Interaction. In: *Proceedings of the Second International Workshop on Corpora for Research on Emotion and Affect (satellite of LREC'08)*. Marrakech, Morocco, pages 5-8 (2008).
- [2] M. Gnjatović and D. Rösner. On the Role of the NIMITEK Corpus in Developing an Emotion Adaptive Spoken Dialogue System. In: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), Marrakech, Morocco, 8 pages, no pagination, (2008).
- [3] M.A.K. Halliday. *An introduction to functional grammar*, Second Edition, London, Edward Arnold (1994).
- [4] A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nöth. Desperately Seeking Emotions: Actors, Wizards, and Human beings. In: *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Newcastle, Northern Ireland, pages. 195-200 (2000).
- [5] Batliner, A., Steidl, S., Hacker, C. and Nöth, E. Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech In: *User Modelling and User-Adapted Interaction - The Journal of Personalization Research (umuai)* 18, pages 175-206 (2008).
- [6] M. Gnjatović, M. Kunze and D. Rösner. Processing Dialogue-Based Data in the UIMA Framework. In: *Proceedings of the UIMA Workshop at the GLDV 2007*. Web published, Tübingen, Germany (2007).
- [7] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll. Frame vs. Turn-Level: Emotion Recognition from Speech Considering Static and Dynamic Processing. In: *Proceedings of 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007)*, Lisbon, Portugal, pages 139-147 (2007).
- [8] R. Niese, A. Al-Hamadi, and B. Michaelis. A Novel Method for 3D Face Detection and Normalization. In: *Journal of Multimedia*, 2(5):1-12 (2007).
- [9] Wu, C.-H., Chuang, Z.-J., and Lin, Y. 2006. Emotion recognition from text using semantic labels and separable mixture models. In: *ACM Transactions on Asian Language Information Processing (TALIP)* Vol. 5, Num. 2, ACM, New York, NY, USA, pages 165-183 (2006).
- [10] C.-H. Lee, Fundamentals and Technical Challenges in Automatic Speech Recognition. In: *Proceedings of the XII International Conference "Speech and Computer" (SPECOM'2007)*, Moscow State Linguistic University, Moscow, Russia, pages 25-44 (2007).