Towards investigating effective affective dialogue strategies

Gregor Bertrand,Florian Nothdurft,Wolfgang Minker,Steffen Walter,Andreas Scheck,Henrik Kessler

University of Ulm - Institute for Information Technology, Clinic for psychosomatic medicine and psychotherapy
Albert-Einstein-Allee 43, Am Hochsträß 8
89081 Ulm
Germany
gregor.bertrand@uni-ulm.de, florian.nothdurft@uni-ulm.de, wolfgang.minker@uni-ulm.de,
steffen.walter@uni-ulm.de, andreas.scheck@uni-ulm.de, henrik.kessler@uni-ulm.de

Abstract

We describe an experimental Wizard-of-Oz-setup for the integration of emotional strategies into spoken dialogue management. With this setup we seek to evaluate different approaches to emotional dialogue strategies in human computer interaction with a spoken dialogue system. The study aims to analyse what kinds of emotional strategies work best in spoken dialogue management especially facing the problem that users may not be honest about their emotions. Therefore as well direct (user is asked about his state) as indirect (measurements of psychophysiological features) evidence is considered for the evaluation of our strategies.

1. Introduction

Automatic recognition of emotions is still being a hot topic in research on adaptive human-computer interfaces. However, in emotion-aware Spoken Language Dialogue Systems (SLDS) not only accurate recognition of user emotion plays an important role, but also effective strategies for coping with different emotional user states have to be developed. Otherwise frustrated users might just break up the conversation. In this area very little research has been done yet.

Our approach to the integration of emotional strategies into spoken dialogue systems is to integrate them into the dialogue manager component. If we consider the common architecture of a spoken dialogue system, dialogue management is the central component dealing with input and output from and towards the user and also with communication with the application that the user would like to access via speech. In figure 1 our approach is depicted.

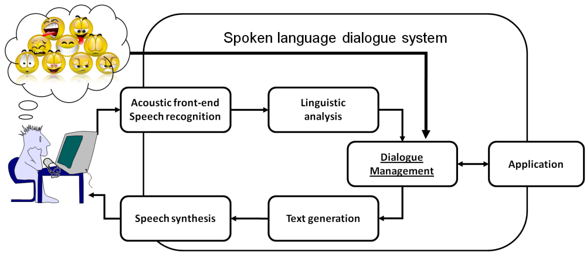

Figure 1: Integration of emotional strategies into the common spoken dialogue system architecture.

In this paper we present an experimental Wizard-of-Oz setting (Experiments where parts of the system are simulated by a human person) with the purpose to investigate different strategies for influencing the user's emotional state. Wizard-of-Oz (WoZ) experiments developped as a means to facilitate usability focused software design and were presented by (Kelley, 1984) as an iterative method to design user-friendly applications. We use these experiments in order to collect user data that may later be used to improve the usability of spoken dialogue systems. The concept of WoZ-experiments is widely used for software development and prototyping in the area of human computer interation and interface design (Akers, 2006; Höysniemi et al., 2004; Molin, 2004; Bernsen et al., 1994).

Especially speech communication is a process that is extremely dependent on emotion. That is why it is important to take into account how the user's current emotional state can be categorized. It's not only important to keep track of the user's emotional state but also to react in an appropriate way, if the user is feeling uncomfortable or even angry. Otherwise there is a risk that the user aborts communication with the SLDS. Therefore the focus of this work is to develop dialogue strategies for SLDS which are appropriate for regaining a cooperative user attitude in case of frustration or dissatisfaction.

On the one hand we investigate how users differ in their reactions towards the system and on the other hand, how effective different types of strategies are.

2. Related Work

There have been several studies concerning the issue of emotion recognition. Therefore emotion seems to be a field of high interest in human computer interaction. For example in (Pittermann, 2008) the focus was on integrating emotion recognition into spoken language dialogue systems. In (Cowie et al., 2001) the authors concentrated on bimodal emotion recognition from facial and voice features. In the experimental setup in (Anttonen and Surakka, 2005) the heart rate of the participants was used as an indicator for emotional stimulation. In (Branco et al., 2005) was shown that there is a correlation between facial muscle activity and task difficulty. An extensive summary of scientific work treating emotion recognition can also be found in (Pittermann, 2008).

A study concerning the issue if negative emotional states, in particular frustration, could be alleviated by reacting via speech to the user's facial expressions was conducted by

(Jaksic et al., 2006). The results show that affective responses, i.e. responses showing emotion, of a virtual social agent can in fact influence the user's emotional state. In case of a moderately frustrated user, the results indicated that the frustration can be reduced by the use of affective responses. In contrast, a very frustrated user was rather annoyed by the social agent, which led to a slightly increased frustration level. Furthermore the number of affective responses correlated with the reduction of frustration.

The results of this study lead to the conclusion that responses can in fact contribute to reducing frustration. Late or infrequent intervention, however, will make it difficult or maybe even impossible to influence the emotional user state in a positive way. Therefore reacting appropriately to early stages of negative emotional states of the user becomes a very important issue.

Nevertheless this study lacks discrimination of different affective responses and is limited to identifying the emotional user state by questioning the user. So the user state that was identified is depending on the user telling the truth about his emotional state which may not always be the case.

However, affective response doesn't need to be given via speech by a social agent to reduce frustration. As seen in (Klein et al., 2002), text-only responses can be sufficient to help the user recover from frustration and other negative emotional states. The system built in (Klein et al., 2002) tried to relieve user frustration by giving pro-active emotional support. In order to achieve these goals support strategies from different fields of psychological studies were used. These were shown to have an effect on emotional states. The results showed that giving affective text-only responses along the communication with an agent can reduce user frustration and bring him back to a more positive emotional state. This included communicating a sense of sympathy and empathy, among other things. However, what's lacking in this study and in others as well, is the attempt to differentiate between the support strategies.

There have been several studies concerning the issue why communication may fail like (Aberdeen et al., 2001; Aberdeen and Ferro, 2003). Their main focus was on identifying situations and conditions leading to errors in understanding and communication.

Other studies focused on categorizing errors in communication (Paek, 2003) so that different levels of misunderstanding can be described.

Apart from these scientific works which are centered around the description of how and why errors in communication occur there are also efforts to find ways to remedy communication failure and develop strategies for handling errors.

In (McTear et al., 2005) research was primarily focused on guaranteeing a common ground of understanding by the notion of grounding (Traum, 1999).

In (Skantze, 2005) strategies for coping with communication errors have been investigated. This study suggests a non-problem oriented approach for error recovery based on the analysis of human error recovery strategies. This means that a spoken dialogue system should avoid to signal misunderstanding because the user interprets misunderstanding as decreased task success. This seems to imply for emotional strategies, that they should not be problem-oriented either but focus on the strengths of the user and the system in order to convey a positive image of the dialogue.

In (Boyer et al., 2008) was investigated how affective feedback should be balanced with cognitive feedback in a learning environment. Considering tutorial or learning there have been further studies like (Rebolledo-Mendez et al., 2006; Tan and Biswas, 2006), limited on the task how to motivate a learner; they could, however, not provide any information in a broader context on how to motivate a common user of a spoken dialogue system to continue interaction. Nevertheless these studies suggested that in the field of learning it is possible to increase motivation through carefully designed strategies applied in an appropriate context. This means that motivational feedback is only useful given the learner is unmotivated.

In short there have been studies dealing with emotion recognition and also with analysis of communication failure. Furthermore there have been investigations of how to respond to frustration and of models how to improve understanding. Further studies focused on increasing motivation in the field of learning. However, only little research has been done in the area of strategies which may influence the user in an emotionally positive way. Emotionally positive in this context means that the user is more willing to interact with the system and is interested in completing his work rather than aborting the communication out of frustration or dissatisfaction.

Such strategies to influence the user's emotional state are of vital importance because even in human-human interaction understanding is not always guaranteed. When communicating with a dialogue system, misunderstanding and other errors are even more relevant. Misunderstanding can cause dissatisfaction and frustration with the dialogue system. However (Klein et al., 2002) and (Jaksic et al., 2006) have shown that negative emotional states can be reduced by corresponding system reactions.

The focus of our study is on the reaction of the SLDS.

3. Description of the Experiment

Hence, we are conducting an experiment, which compares how effective different support strategies are. Each strategy is based on one of the following components: empathy, motivation and thankfulness. The strategies do not combine more than one component and are therefore disjoint among each other. For control purposes we are including a neutral strategy.

Concerning the setup it is important to elicit user emotions in a manner that is strong enough to be relevant. For this reason the user has to be involved to a large degree in the experiment.

The following desiderata guided the scenario development:

- The user should interact with the system via spoken language in a relatively natural way.

- The system should ask the user to complete tasks that are cognitively demanding and call for attention to and engagement with the system.

- At various points the system should offer help to test the efficiency of presumably helpful dialogue strategies of the system.

- The scenario should include the induction of user emotions that are important for dialogue characteristics and occur in real-life human-computer-interactions.

We decided to implement a scenario that mainly consists of a cognitive task, where the user has to memorize visually presented items under time constraints and with increasing difficulty. All commands and answers of the user can be given using spoken language. The system responds with the use of a text-to-speech engine to encourage dialogues.

As a test-bed for our dialogue management system we created a Wizard-of-Oz (WOZ) scenario. The setup of the scenario is shown in figure 2.

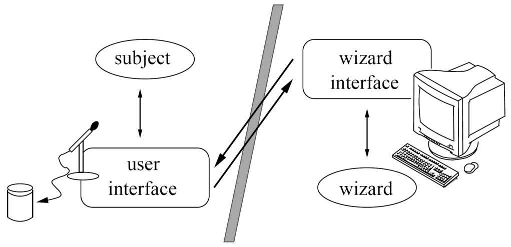

Figure 2: Wizard-of-Oz setup (Bernsen et al., 1994).

In order to induce emotions in a quasi-experimental way, the human operator of the system ('Wizard') can control the type of feedback the system gives (e.g. friendly or impolite), the difficulty of the task (e.g. adding countdowns to increase stress levels), the reaction time (e.g. delayed answers to cause frustration) and the quality of help (e.g. being very helpful to please the user).

With this scenario we assume to comply with all four components mentioned above and allow the emergence of quasi-natural dialogues that form the basis of our modelling of dialogue strategies. To collect additional data and accurately assess the user's emotional state we use several psychobiological devices.

The model that we apply in order to measure the user's emotional state is the Pleasure-Arousal-Dominance (PAD) space (Mehrabian, 1996). The PAD-space is a simple 3-dimensional space where the user can be categorized with respect to being in a positive or negative mood (pleasure domain), being agitated or tired (arousal domain) and feeling in control of the situation or not (dominance domain). We chose this form of categorization because of its simplicity and because it can be swiftly handled, e.g. using 3 variables $-1 \leq x_i \leq 1$ with $i \in \{1, 2, 3\}$.

A 9-channel electroencephalography (EEG) device measures brain activity over the central cortex in real-time to search for indicators of shifts in attention and emotion. Peripheral data (blood volume pulse, heart rate, skin conductance level and breathing) complement this assessment. To measure the user's reactions (positive or negative) in the pleasure domain (Mehrabian, 1996) electromyography (EMG) is applied over the eye and mouth region. Additionally a frontal camera films facial reactions that can later be assessed via automated algorithms in order to estimate the emotional content.

The experiment itself has an approximate duration of two hours which includes preparation time for the measurement devices. For each proband we get about 40 minutes of video material, EEG data and so on. At present 14 participants have been recorded. In the future, 6 more test persons will be recorded. There should be 10 male participants and 10 female participants of different ages.

The task itself is subdivided into several rounds which are increasing in difficulty and time constraints. Our aim is to make the cognitive strain experienced by the user rise in a similar way.

During the test, the wizard can make use of predefined system utterances that are then read to the user. These utterances are designed to increase or decrease the stress level felt by the user and to make him experience different emotions.

In the first round the user should at the start experience positive valence, low arousal and high dominance. At the end of the first round positive valence, high arousal and high dominance should be achieved. In the second round he should experience positive valence, high arousal and high dominance. The design of the third round should evoke positive valence, low arousal and high dominance in the beginning and positive valence, high arousal and low dominance in the end. At the start of the fourth round the user should feel negative valence, low arousal and high dominance and then negative valence high arousal and low dominance. The fifth round is designed in a way that should create negative valence, high arousal and low dominance in the beginning and finally negative valence, high arousal and low dominance should even be more intense. In the sixth round the user should feel positive valence, low arousal and high dominance again. The emotional states that we aim to evoke in the user are listed in table 1.

Round	Pleasure	Arousal	Dominance
1	high high	low high	high high
2	high	high	high
3	high high	low high	high low
4	low low	low high	high low
5	low low	high high	low low
6	high	low	high

Table 1: Emotional states for the experiment

These emotional states should be achieved by the means of providing slow feedback and false feedback of the application which is a realistic scenario in spoken dialogue systems because of the sometimes still poor speech recognition. Executing additional pressure on the user via speech feedback, e.g. urging him to hurry up, asking him to abort or to speak more clearly is another means for evoking relevant emotions. Timing constraints and help that is provided or denied are other features for influencing the user's mood. During these six rounds the user should have experienced certain emotions which serve on the one hand as some kind of calibration for the measurement devices to the user. One can detect how strong the feedback of certain devices is regarding the user and the supposed emotional reaction. On the other hand the emotions measured are an indicator how easily the user can be influenced emotionally.

After the user has completed the task, a speech dialogue is initiated by the agent which requests the user to self-reflect on his performance. Due to the fact that certain steps of the experiment are designed to frustrate the user, we assume that his self-reflection is not purely positive. In response to the user's answer we present different dialogue strategies, as in figure 3, in order to find out to what extent the emotional state of the user can be influenced in a positive way.

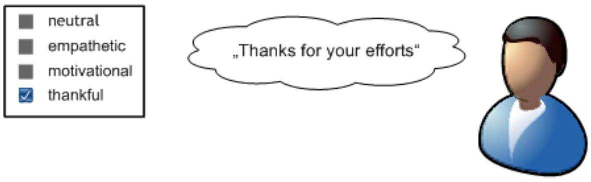

Figure 3: Strategy selection

'Positive' in this case means that the pleasure level can be increased describing the user state according to (Mehrabian, 1996). These are the different strategies that we want to compare regarding their effect on the user:

- motivational
- praising
- thankful
- neutral

One of these strategies is randomly presented to the user by a text to speech engine simulating the SLDS. Following the user's response he will be interrogated how he is feeling currently.

The data acquired during the experiment will be reviewed together with the user concerning his emotional state and furthermore the data will be classified according to emotional categories with the help of emotion recognition by speech, automatic EEG-analysis, skin conductivity and the analysis of the user's facial expression.

Important for the categorization is not so much his statements, but the way of presenting them. We analyze his statement via emotion recognition by speech and via other available information to investigate whether his emotional state has changed. This means in addition, that we are able to compare the user's experienced emotional state to the detected user's emotional state.

One major advantage will be that the recorded emotion is real. It is not played by actors but authentic emotion, that is spontaneous and occurs in human computer interaction.

4. Conclusion and Future Directions

With this study we try to fill an important gap in current research. Affective dialogue strategies are known to be a major factor reducing user frustration and relieving negative emotional user states, but previous work lacked a differentiation and comparison between the individual factors of affective dialogue strategies.

Future work, based on the results of this study concerning affective dialogue strategies, will include figuring out combinations of single influence factors to more effective dialogue strategies. These dialogue strategies will be no longer disjoint and more closely related to actual desirable dialogue strategies for SLDS.

Acknowledgements

This work is being made in the context of the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG).

5. References

J. Aberdeen and Lisa Ferro. 2003. Dialogue patterns and misunderstandings. In *Error Handling in Spoken Dialogue Systems*, pages 17–21, August.

John Aberdeen, Christine Doran, Laurie Damianos, Samuel Bayer, and Lynette Hirschman. 2001. Finding errors automatically in semantically tagged dialogues. In *HLT '01: Proceedings of the first international conference on Human language technology research*, pages 1–5, Morristown, NJ, USA. Association for Computational Linguistics.

David Akers. 2006. Wizard of oz for participatory design: inventing a gestural interface for 3d selection of neural pathway estimates. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 454–459, New York, NY, USA. ACM.

Jenni Anttonen and Veikko Surakka. 2005. Emotions and heart rate while sitting on a chair. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 491–499, New York, NY, USA. ACM.

Niels Ole Bernsen, Hans Dybkjaer, and Laila Dybkjaer. 1994. Wizard of oz prototyping: How and when? In *CCI Working Papers in Cognitive Science and HCI.*

Kristy Elizabeth Boyer, Robert Phillips, Michael Wallis, Mladen Vouk, and James Lester. 2008. Balancing cognitive and motivational scaffolding in tutorial dialogue. In *Intelligent Tutorin Systems*.

Pedro Branco, Peter Firth, L. Miguel Encarnaç ao, and Paolo Bonato. 2005. Faces of emotion in human-computer interaction. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1236–1239, New York, NY, USA. ACM.

R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. 2001. Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, 18(1):32–80.

Johanna Höysniemi, Perttu Hämäläinen, and Laura Turkki. 2004. Wizard of oz prototyping of computer vision based action games for children. In *IDC '04: Proceedings of the 2004 conference on Interaction design and children*, pages 27–34, New York, NY, USA. ACM.

Nada Jaksic, Pedro Branco, Peter Stephenson, and L. Miguel Encarnaç ao. 2006. The effectiveness of social agents in reducing user frustration. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 917–922, New York, NY, USA. ACM.

J. F. Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2(1):26–41.

J. Klein, Y. Moon, and R. Picard. 2002. This computer responds to user frustration:: Theory, design, and results. *Interacting with Computers*, 14(2):119–140, February.

Michael McTear, Ian O'Neill, Philip Hanna, and Xingkun Liu. 2005. Handling errors and determining confirmation strategies - an object-based approach. *Speech Communication*, pages 249–269.

Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, December.

Lennart Molin. 2004. Wizard-of-oz prototyping for co-operative interaction design of graphical user interfaces. In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, pages 425–428, New York, NY, USA. ACM.

Tim Paek. 2003. Toward a taxonomy of communication errors. In *Error Handling in Spoken Dialogue Systems*, pages 53–58, August.

Johannes Pittermann. 2008. *Speech-Emotion Recognition in Adaptive Dialogue Systems*. Ph.D. thesis, University of Ulm.

Genaro Rebolledo-Mendez, Benedict du Boulay, and Rosemary Luckin. 2006. Motivating the learner: An empirical evaluation. *Lecture Notes in Computer Science*, 4053:545–554.

Gabriel Skantze. 2005. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, 45(3):325 – 341. Special Issue on Error Handling in Spoken Dialogue Systems.

Jason Tan and Gautam Biswas. 2006. The role of feedback in preparation for future learning: A case study in learning by teaching environments. *Lecture Notes in Computer Science*, 4053:370–381.

David R. Traum. 1999. Computational models of grounding in collaborative systems. In Susan E. Brennan, Alain Giboin, and David Traum, editors, *Working Papers of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, pages 124–131, Menlo Park, California. American Association for Artificial Intelligence.