
Emotion Analysis and Emotion-Handling Subdialogues

Michael Streit¹, Anton Batliner², and Thomas Portele³

¹ DFKI GmbH, Saarbrücken, Germany
streit@dfki.de

² Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
batliner@informatik.uni-erlangen.de

³ Philips Research Laboratories GmbH, Aachen, Germany
Thomas.Portele@philips.com

Summary. The chapter presents the cognitive model-based approach of abductive interpretation of emotions that is used in the multimodal dialogue system SMARTKOM. The approach is based on Ortony, Clore and Collins' (OCC) model of emotions, which explains emotions by matches or mismatches of the attitudes of an agent with the state of affairs in the relevant situation. We explain how eliciting conditions, i.e., abstract schemata for the explanation of emotions, can be instantiated with general or abstract concepts for attitudes and actions, and further enhanced with conditions and operators for generating reactions, which allow for abductive inference of explanations of emotional states and determination of reactions. During this process concepts that are initially abstract are made concrete. Emotions may work as a self-contained dialogue move. They show a complex relation to explicit communication. Additionally, we present our approach of evaluating indicators of emotions and user states that come from different sources.

1 Introduction

For a long period, the concept of rational agents, which exchange rational arguments, was the predominant paradigm for research on dialogue systems. In the last decade the scientific community became aware of the fact that emotions, moods and other attitudes play an important role in natural communication. This insight in human behavior resulted in research on affective interfaces, development of believable agents, and in research that aims at supporting the recognition of problematic dialogue situations by analyzing the affective state of the user.

While there are considerable advancements in generating affective artificial agents that display believable emotions in appropriate situations (Picard, 1997), the recognition and interpretation of human emotions in dialogue systems is still in its infancy. In SMARTKOM we explore the handling of emotions by focusing on exemplary cases, which show how emotions work as dialogue moves or and how the emotional color of an utterance may change the literal meaning into the converse.

The term *emotion* normally aims at pronounced, clear forms of human states marked by strong feelings such as, e.g., anger, fear, sadness, joy, etc. — the so-called “full-blown, big” *n* (*n* typically ranging between 4 and some 20) — emotions. Under a close look, however, almost nothing is that clear-cut: the underlying (bodily and cognitive) processes are not yet fully understood, emotions do often occur in mixed, not in pure forms, their marking can be overtly suppressed due to social constraints and rules (Cornelius, 2000; Batliner et al., 2003a,c,b). Finally there is no full agreement as for a catalogue of emotions, and of pivotal characteristics, telling emotions apart from other states such as attitudes, mood, etc.

The OCC Model of Emotions

Research concerned with the generation of an affective and believable behavior of artificial agents is often based on the so-called OCC model of emotions (Ortony et al., 1988), which explains emotion by cognitive processes relating the user’s goals, standards, likes and dislikes to the actions of other agents and the state of the world that results from these actions. This approach starts from clearly distinguished emotions, which result from a cognitively comprehensible *appraisal* of the situation by the agent. How intensive emotions are, or if they are displayed at all is determined by the intensity of the attitudes of the agent and by social display rules.

Though mixing or suppressing emotions is a problem for the recognition of emotions as well as for the fine-tuning of the artificial generation of emotional behavior, the OCC model provides a systematic account for relating a certain situation of an interaction to emotional states that fit to this situation. The logical structure of the situation that causes a certain emotion is not affected by the question of how intensive an emotion is or if it is displayed at all. This is not only attractive for the generation of affective behavior, but also for inferring appropriate reactions on emotions that are observed in a certain situation.

Emotions and Other Affective States

For research, which is concerned with the detection of problematic situations in communication by analyzing the user’s behavior, not only emotions are relevant. This is the case, independently from the question of whether the catalogue of emotions is completely defined or not. For instance, if the user is hesitant, she may need help, or if she is tired this may affect the presentation of information, or the system may even request the user to stop some activity that needs high attention.

We use the term “(emotional) user states” to encompass all nonneutral, somehow marked behavior of the user within a human–machine communication. Thus, user states as bored, stressed, irritated, tired, etc., can and have to be addressed as well, irrespective of whether they belong to the one or the other psychological or physiological category. In contrast, the psychological or physiological category of a state is relevant for its interpretation. The SMARTKOM system will react to this match or mismatch, e.g., by trying to repair the mismatch or by promoting further matches.

The spirit of the approach, namely to consider what type of conditions elicit the affective state of the agent, extends to some nonemotional states, but not to all. For

instance, the user may be hesitant, because she has a goal, but does not know how she can achieve it, or she does not know how to decide between certain alternatives, etc. Such considerations add up to a very similar kind of cognitively comprehensible appraisal of a situation, as elicit emotions in the OCC model. In contrast, the state of tiredness (in the literal meaning) cannot be accounted for by such an approach that analyses the *cause* of the user state. Rather, the system should consider the possible consequences of the user's state.

Preconditions for a Working System

Some important conditions have to be met, however, if one wants to deal with user states in an automatic system:

- It must be possible to classify the state correctly up to a satisfying extent.
- Thus there has to be a sufficiently large training sample.
- The respective user state can be processed within the whole system, not only at the classification stage.

The first condition means that we should start with user states that are clearly marked. This rules out such states as "slightly irritated," even if they might occur quite often and have a strong impact on the felicity of communication. We are thus left with those pure emotions like anger or joy, which do not, alas, occur often in real or in Wizard-of-Oz human-machine communications (Batliner et al., 2003c,a,b; Ang et al., 2002). Thus we decided to concentrate on some few user states and to collect acted data for our training sample.

Overview

The focus of the paper is to present the cognitive model-based approach of abductive interpretation of emotions as it is used in the SMARTKOM system. Additionally, we present our approach on the collection and evaluation of indicators of emotions, user states, problematic situations and other states of the user or the interaction. Recognition of emotions from prosody and facial expressions is described in Zeißler et al. (2006) and in Frank et al. (2006).

We start in the second section with a brief description of the architecture of emotion analysis in SMARTKOM. In the third section we introduce the type of interaction we want to focus on. In the fourth section we describe the method for evaluating indicators from different sources. At the time, we use only a little part of the information provided there, but much more would be useful. The remaining sections are dedicated to the interpretation of emotions and user states and the generation of reactions to these states. First, we introduce the OCC model of emotions, then we say a little bit about abduction and the problems that occur if we use the OCC model for an abductive interpretation. In the next step we show how we generate explanatory hypotheses by introducing abstract concepts as fillers of eliciting conditions. We introduce conditions for explanation selection and recovering operators for generating initiate reactions of the system. Conditions and generation of reactions are interwoven: usually conditions provide parameters that are used by the recovering operators.

2 Emotion Processing in the SmartKom System

SMARTKOM (<http://www.smartkom.org>) is a multimodal dialogue system that provides access to multiple applications (Wahlster, 2003). In addition to input-modalities that are used for intentional communication, the system accounts for the emotional state of the user as it is displayed by facial expression or by prosody. The processing of emotions and user states consists of three stages.

- At the first stage the emotional state of the user is recognized from facial expression and prosody.
- At the second stage indications of problematic situations and the emotional state of the user are collected from several sources and collectively evaluated. The component also analyzes the dialogue in respect to the style of interaction and the task and paradigm knowledge of the user (Portele, 2003).
- The interpretation of emotions and user states, and the generation of reactions to these states build the third stage. It is realized by so-called dynamic help. This component is dedicated to manage subdialogues, and to provide presentation specification and intention analysis in problematic situations that are not handled by the standard dialogue component of SMARTKOM.

3 The Use Cases

To demonstrate the added value of user state classification and its subsequent processing in the SMARTKOM system, we designed so-called **use cases**. The first use case is intended to show how a merely emotional reaction, without explicit communication, can work as a self-contained dialogue move. In this case, joy or anger are interpreted as positive or negative feedback, respectively. In the second use case emotion works as a semantic operation that turns a positive feedback into a negative one, which is considered as a form of sarcasm. In both use cases, the system suspects that the emotional reaction may be caused by a like or a dislike concerning the properties of the presented objects. If reasonable candidates of such likes or dislikes can be identified that are not already known by the system, it starts a preference update dialogue.

If the system knows positive or negative preferences, it first presents objects that contain a preferred feature. Objects that show a disliked feature will be shown last.¹

user: *What's on TV tomorrow?*

system: Shows talk show at the top of the display, in the middle popular music, and crime at the bottom.

user: *And what's in the evening, in the first program?*

system: Shows a science fiction movie.

¹ It is possible that an object has both liked and disliked attributes, e.g., there may be a movie with a preferred genre, in which a disliked actor plays.

First constellation: emotion only**user:** Displays joy via facial gestures.**system:** *Do you like science fiction? Shall I account for that in future presentations?***Second constellation: emotionally marked verbal communication****user:** *That's really a beautiful program!* (She produces this sentence with an angry prosody. The positive feedback is analyzed as being sarcastic.)**system:** *You don't like science fiction? Shall I account for that in future presentations?***user:** *Yes./No.***system:** *OK. I'll take care of that!*

(Suppose the user's answer was *yes*: In the first constellation **science fiction** will be presented at the beginning of a presentation, in the second constellation at the end.)

user: *Please, again tomorrow's program!***system:** Shows **science fiction** at the beginning (at the end) of a presentation.

Instead of answering *no* or *yes*, the user may also correct the supposed like or dislike, e.g., by saying *No, I like crime movies*, or she may just ignore the question, by moving to a different topic. In such cases, the system will simply not rearrange the order of presentation.

4 Indications of User States and Problematic States of Interactions

We introduced in SMARTKOM a component, the *interaction module*, that collects and evaluates indications of emotions, problematic situations and other aspects of the interaction. This approach is used, on the one hand, to estimate recognition results by taking other indicators into account that may support or devalue the result. On the other hand, the component introduces its own results concerning states of the user or characteristics of the interaction. It operates by analyzing a set of possible *indicators*.

Indicators can have values between 0 and 1 and these values may change in time. For example, the mimic analysis module delivers information about the user's current emotional state as a set of four probabilities (likelihood of anger, likelihood of joy, likelihood of dilatoriness, neutral). The set of indicators is fairly fixed and is defined by the capabilities of the modules in the system (Table 1).

The interaction module provides a set of *models* as output. Each model value is also in the range between 0 and 1. Several models support the recognition of emotions and try to detect problematic situations during a dialogue.

4.1 Indicators

The indicator values are mapped to the models by means of a matrix multiplication. One element of the matrix denotes the influence of one indicator value to one model. This design is motivated by the observation that most indicators can contribute to different models, and that the combination of simple indicators to complex models may be optimized by machine-learning algorithms. Furthermore, new models on demand from other modules can be constructed easily by combining the indicator set with a different weighting scheme.

Table 1. List of indicators

Source	Description
Mimic recognizer	Mimically conveyed anger
Prosody recognizer	Prosodically conveyed anger
Mimic recognizer	Mimically conveyed joy
Prosody recognizer	Prosodically conveyed joy
Mimic recognizer	Mimically conveyed dilatoriness
Prosody recognizer	Prosodically conveyed dilatoriness
Speech recognition	Linguistically conveyed anger
Speech understanding	Ratio of unanalyzable words
Intention analysis	Overall score of the best hypothesis
Intention analysis	Difference in score between first and second best hypotheses
Intention analysis	Number of possible hypotheses (depth of lattice)
Speech recognition	Score of the speech recognizer
Gesture recognition	Score of the gesture analyzer
Speech understanding	Score of the language analyzer
Media integration	Score of multimodal integration
Discourse history	Score of the discourse module
Domain model	Score of the domain module
Intention analysis	Final score of the intention module
Intention analysis	Number of elements in the user input
Discourse history	Number of new (not previously mentioned) elements
Speech understanding	Number of elements addressed by speech
Gesture analysis	Number of elements addressed by gesture
Media integration	Number of elements addressed by speech and gesture
Intention analysis	Importance of speech recognition score for overall score
Intention analysis	Importance of gesture analysis score for overall score
Intention analysis	Importance of domain model score for overall score
Intention analysis	Importance of language understanding score for overall score
Intention analysis	Importance of discourse model score for overall score
Speech understanding	Relative number of sentence-like units in one turn
Speech understanding	Relative number of words in one turn
Speech understanding	Relative frequency of pronouns
Speech understanding	Relative frequency of verbs
Speech understanding	Relative frequency of adverbs
Speech understanding	Relative frequency of nouns
Speech understanding	Relative frequency of content words
Speech understanding, language generation	Relative frequency of content words appearing in the system output
Speech understanding, language generation	Relative frequency of content words not appearing in the system output

4.2 Models

The module delivers four sets of models. The distribution of the models to the sets is somewhat arbitrary and is mainly governed by the intended use within the SMART-KOM system. Therefore the models for task and paradigm knowledge use many in-

Table 2. List of models

Set	Description
Problem	Likelihood of a problem
Problem	Likelihood of an analysis problem
Problem	Discourse progress rate
Problem	Likelihood of the user being angry
Problem	Likelihood of the user being happy
UserKnowledge	Estimation of user familiarity with task
UserKnowledge	Estimation of user familiarity with system
Modality	Ratio of spoken input content
Modality	Ratio of gestural input content
Modality	Ratio of multimodal input content
ModalityContrastive	Ratio of contrastive usage of multimodal input
ModalityRedundant	Ratio of redundant usage of multimodal input
Linguistic	Adaptivity of user's lexical choices to former system output
Linguistic	Likelihood of long turns
Linguistic	Likelihood of long sentences
Linguistic	Ratio of pronoun usage
Linguistic	Ratio of verb usage
Linguistic	Ratio of adverb usage
Linguistic	Ratio of noun and verb usage

dicators that also contribute to the problem detection models. This overlap reflects the design principle of the module.

- One set of model values reflects the assumed task and paradigm knowledge of the user. The task knowledge describes the user's knowledge of the current task (e.g., programming a VCR), while the paradigm knowledge indicates how well the user is accustomed to dealing with multimodal dialogue systems, and, especially, with SMARTKOM. These models can be employed by the dynamic help and the presentation/language generation to deliver an appropriate amount of feedback and assistance.
- A second set of models describes the linguistic behavior of the user regarding the number of, e.g., referential expressions, usage of complete sentences and average length of input. Some of these features can, in principle, be used by language models or parsers (although this is not the case in the current SMARTKOM system). All of them may help to adapt the language generation in order to reflect the user's style—based on the assumption that this is beneficial. Furthermore, the adaptivity of a user's lexical choices to former system output is estimated, which can help adapting dynamic language models used in SMARTKOM and language generation in order to maximize this value as a measure of the common vocabulary.
- The third set compares the use of different modalities by the user (for instance, a preference for gestures or spoken input). The presentation and the behavior of the animated character can be adapted toward the user's preferred distribution of the different modalities — users who prefer pointing gestures could be supplied with an interaction display with more possibilities and details, while users interacting mainly through speech should get the most important system feedback also by spoken output.
- Problematic situations and user state information are expressed by three models in the fourth set of the module:

- One model describes the likelihood that the user is angry by combining scores from mimic analysis, emotion extraction from prosody and use of certain words. These indicators are either computed directly from other modules or are obtained by counting appearances of specific words (a related model assesses the user's satisfaction by looking for joyous emotional expressions; this and the anger model are used for setting content-based user preferences).
- A second model combines confidence values from recognizers and similar scores from speech analysis, domain model, discourse history and intention recognition as well as differences in the distribution of these values among concurring hypotheses; this model is supposed to indicate problems in the analysis part of the system.
- A third model estimates the dialogue progress. Here, the ratio of new information items to total information items (after completion by overlay in the discourse history) is employed as one indicator, another one being the ratio of overlaid (i.e., changed, interpretable as corrected) information to unaltered information. A further indicator is the overall number of information items in the user input. Because of the system design, all other indicators can contribute to the model values as well (and the indicators used here are also used for other models, e.g., about user experience with the task or the interaction paradigm), but the indicators named above are assumed to be the most important ones for the respective models.

These model values are part of the input of the dynamic help module.

5 Cognitive Model–Based Interpretation of Emotions

Our approach to the analysis of emotions is based on the OCC model of emotions developed by Ortony, Clore and Collins. Following the OCC model, emotions are characterized by their eliciting conditions. These conditions consist of a certain combination of

- the goals of the agent in this situation
- her attitudes to certain events (mainly likes and dislikes)
- the standards that she uses to (morally) judge an event
- the facts that hold in a certain situation
- the actions (of other agents) that caused these facts

For triggering an emotion, it is important to know how facts are related to the goals and the likes and dislikes of the user. In particular, it is interesting if they coincide or not. Standards are important for emotions such as anger or gratitude that contain criticism or praise of another agent based on her actions. Eliciting conditions can be viewed as expressing the cause of an emotion by providing a cognitively comprehensible explanation of an emotion. The following eliciting condition for anger is taken from Prendinger et al. (2002):


```

anger(Agent1,Agent2,State,Sit) if
  holds(did(Agent2,Action),Sit),
  causes(Action,State,Sit0),
  wants(Agent1,non_State,Sit),
  blameworthy(Agent1,Action) ,
  (Sit0 < Sit)

```

This condition means that the agent is angry if she believes that another agent caused some state of affairs that contradicts her goals by performing an action that is not acceptable according to the user's standards (expressed by the *blameworthy* predicate). By the situation variables *Sit*, *Sit0*, one can express how the elements of the conditions are connected with respect to the sequence of situations that occur (subsequently we will omit situation variables).

Recognizing the intensity of emotions could provide additional valuable information, e.g., slight anger may occur at the beginning of a problem, while strong anger may indicate an enduring problem. But the recognition of the situation that caused the emotion and the generation of appropriate reaction is basically the same whether emotions are displayed slightly or strongly.

5.1 Abductive Interpretation of Eliciting Conditions

The OCC model is mainly used for the generation of the behavior of an animated agent. In this case, one can deliberately define the agent's likes, dislikes and standards in advance. If we want to interpret emotions that are displayed by an agent, we have to find out which combination of facts, attitudes and standards may have caused the emotion. Our approach is to achieve this by analyzing eliciting conditions in an abductive manner. Abduction as a form of practical inference is introduced by Peirce (1995). Abduction is often characterized as inference to the best explanation: Suppose we observe some fact *A*, which is surprising for us. If we know the rule

$$B,C \rightarrow A$$

(i.e., *A* is true if *B* and *C* are true), then we may suspect that also *B* and *C* are true, because this would plausibly explain *A*. If we know that there is another rule

$$D \rightarrow A$$

then *D* is another candidate for explaining *A*. Hence we need a criterion to decide which explanation is better. The quality of an explanation depends on two factors: Do we know all relevant rules (i.e., explanations)? Do we possess criterions to choose from explanations? With eliciting conditions we have the advantage of possessing schemata that claim to characterize all possible explanations of an emotion.

5.2 Problems with Abductive Interpretation

Eliciting conditions are abstract schemata that cannot be used directly to infer possible causes of emotions. To perform abductive reasoning on eliciting conditions, we have to identify concepts that could be filled into the schemata. Seemingly, we are in a problematic situation. The system has no information about the user's standards,

likes and dislikes in advance. It can get information about her goals from the user's input. But, on the one hand, this information may be based on misunderstanding, and, on the other hand, the user may have goals which cannot be recognized from her utterances. Similar problems occur with the actions of the system. The concrete actions that are based on misunderstanding are not relevant for the analysis of the user's emotion.²

5.3 Abstract Goals and Actions for Emotion Interpretation

To overcome the problems mentioned in the last paragraph, we introduce metagoals concerning general principles of communication and abstract goals concerning user needs that (to some extent) depend on the application. For every metagoal or abstract goal we introduce an abstract action that satisfies the goal.

For instance, to account for misunderstandings, we introduce *understanding* as an action on the metalevel and *to be understood* as a goal on the metalevel. To account for user preferences, we introduce the concept that *a presentation accounts for the user's preferences* as an abstract action of the system — let it be called *present-ByPreferences* — and accordingly the possible abstract fact or user goal *isPresented-ByPreferences*.³ This goal is abstract and underspecified because we do not know the concrete preferences of the user. Further, the relevant types of preferences depend on the type of the application.

Reasonable goals (facts, actions, likes, standards) have to be identified by careful analysis of general principles of communication and the needs of the user with respect to the type of applications with which she is working. This needs empirical validation, which could not be provided within the scope of the SMARTKOM project. Which set of concepts is chosen also depends on practical decisions: which goals will the system support at all, will the system possibly recognize goals that it is not able to handle, will the system react on any recognized emotion in some way (e.g., by regretting as a default in case of anger), or will it only react to emotions to which it can provide a repair or other meaningful cooperative reaction? We demonstrate the approach by the example of **anger**.

General Concepts

We first look for actions or facts that may contradict the user's wishes, likes, dislikes or standards on a general level. Important candidates for abstract actions that contradict the user's wishes are **misunderstanding**, **slow processing** and **requests with a negative or disliked outcome**. Accordingly, we stipulate abstract or general goals, e.g., the goal to be understood properly.

² Although the type of the action that the system wrongly performs may influence the intensity of the user's negative feelings.

³ For convenience we often identify the name of the fact and the name of the goal to make this fact true.

Application-Dependent Concepts: Problematic Results of Database Queries

According to our use cases we concentrate on requests with liked or disliked outcome as a source of negative or positive emotions. We identified four types of disliked results:

- The result is empty.
- The majority of retrieved objects show features that are not liked by the user.
- The objects are presented in a way that is contrary to the preferences of the user, e.g., by presenting disliked objects first.
- The user query resulted in a recall, which is too large. The user may need help for further specification possibilities.

We assume for this list of topics that the disliked or problematic results are not due to misunderstanding. Misunderstanding is taken as evoking its own class of constellations. If misunderstanding is involved, the result is not relevant for the analysis.

User-Specified Goals and System-Initiated Actions

As far as no misunderstanding is involved, the SMARTKOM system will usually simply follow the user's specification. If this works, no anger should arise with respect to the fact that the system tries to achieve this goal (but perhaps instead *joy*). In specific situations the system may initiate actions that are necessary from the point of view of the system, but that may be disliked or even considered blameworthy by the user. For instance, the system may require a biometric verification, which the user dislikes. Such actions are relevant for explaining negative emotions but are not considered in our implementation. Inappropriate or undesired reactions on emotions could also be a cause for anger (or for being bored). In fact, this is a subcase of disliked system-initiated actions.

6 Analyzing and Handling Pure Emotion

With the concepts introduced in the last section, we are able to build instantiations of eliciting conditions that allow us to infer combinations of goals, facts, actions, likes and dislikes that possibly explain the user's emotion. We call instantiations of eliciting condition schemata **eliciting constellations**. To get criteria for selecting the relevant constellation, we augment constellations with conditions and organize these conditions internally as a decision tree.⁴

Further, the system has to determine reactions that are appropriate for

- resolving the situation that caused the negative emotion
- avoiding negative emotions in future in similar situations
- promoting the occurrence of positive emotions in similar situations

⁴ As mentioned in the conclusion, we could perform testing only in a limited way. Thus no training of the decision tree was possible.

It is also desirable to include methods that provide abstract underspecified goals and actions with presumable values. Such values are not only used for determining concrete system reactions, they serve as a part of the constellation conditions.

According to our use cases, we have to consider database queries that retrieve disliked objects. The system offers as repair that it will regard the likes and dislikes of the user in its presentations.

A constellation for handling anger according to our use cases is given below (leaving out some minor details) in a Prolog-style notation. It applies to browsing television programs or cinema programs. For these applications preferences are actually taken into account for the presentation. These rules are basically processed in the following manner: First the conditions are tested (internally the conditions are processed in a decision tree-like order). Then the cause of the emotion, which is represented by the clauses above the conditions, is considered as a reasonable explanation, whereby the variables are filled by the result of the condition processing. Then the system action is performed.

```
anger(thisConstellation,user,system) if
  holds(did(system, non_presentByPreference(dislike(user,X))),
  causes(non_presentByPreference(dislike(user,X)),
    non_isPresentedByPreference(dislike(user,X))),
  wants(user, isPresentedByPreferences(dislike(user,X))),
  blameworthy(user, non_presentByPreferences(dislike(user,X))),
  conditions(thisConstellation,X),
  (proposed system action:) update(dislike(user,X)).
```

The constellation expresses that there is a concrete reading of the goal *present-ByPreferences* that may be a goal of the user, that this goal is not satisfied and that ignoring the goal is against the standards of the user. The constellation contains facts and actions that are not concretely specified. For instance, we do not know whether the presentation contains some possibly disliked feature, and we do not know which feature it is.

We test the salience of the constellation by establishing the following conditions. The predicate *presentationEntriesContainCommonFeature(X)* also delivers a concrete presumable instance of the user's dislike.

```
conditions(thisConstellation,X) if
  presentationEntriesContainCommonFeature(X),      (1)
  non_specified(user,X),                          (2)
  non_knows(system,like(user,X)),                  (3)
  non_knows(system,dislike(user,X)).              (4)
```

Condition 1 verifies if the user perceives *too many* objects with the supposed disliked feature. (It also excludes the case that there is no result at all, which would support a different explanation for anger.) It is important for the other tests that the predicate delivers a hypothesis for the disliked feature. Condition 2 excludes that the user is angry about the occurrence of features that she has specified in her request

(there is a possibility of misunderstanding). Condition 3 excludes, that the user is angry about a feature that she has already declared to like. Condition 4 excludes that the system in fact tried to present the disliked feature appropriately but just did not find other objects.

For emotions displayed by facial expressions, we prove if the emotion emerges in a certain time interval after the presentation was displayed. With prosodically displayed emotion we prove if the verbally expressed content was compatible with the explanation of the emotion. It turned out that it is not sufficient to test if there are already stored preferences. It should additionally be proved, if a user has not agreed with storing a preference. This has to be remembered, otherwise the system may propose the same preference repeatedly.

The action *update(dislike,user,X)*, which is attached to the constellation, initiates a subdialogue that verifies if the user has the supposed dislike. It is not only a repair action, but takes part in the explanation process.

The conditions mentioned so far are not sufficient to **discriminate competing explanations**. Such competing explanations have to be modeled, even if no reaction is foreseen for these cases. We distinguished three main sources of anger: misunderstanding, slow processing and requests with a negative or disliked outcome. Evidence for problems in the analysis part is detected by the interaction module (Sect. 4). Slow processing is a possible explanation for anger, if anger occurs during the analysis. Also, the absolute duration of processing is a criterion. These dates are accessible via a module (the so-called watchdog) that monitors the processing state of the system.

Extending the Method — Beyond the Use Cases

We present the handling of an example not covered by the use cases in order to show how the method extends. By using the relaxation functionalities, which are provided by the dynamic help component, one could add constellations that handle anger concerning the factual result of a query. In this case, it is debatable if the observed anger must be directed to the system. It may also be directed to the television program. We stipulated as a possible source of anger (or, even more likely, as a source of disappointment) the fact that the recall on a query contains no objects that are liked by the user.

```

anger(user, system)
  (The relevant facts that hold in this constellation and
  the actions that caused the facts are omitted)
  if
    wants(user, getObjectAccordingToLikes)
    blameworthy(user, NONgetObjectAccordingToLikes)

  (conditions(X):)
  (
    knows(system,dislike(user,X)) ,
    NONspecified(user,X) ,

```

```

        ResultContainsProminentCommonFeature (X))
    )
    or
    (
        knows(system, like(user, X)) ,
        NONResultContains (X))
    )

    (proposed system action:)
    relaxRequest

```

Case: Anger on Empty Result

This is a simple case:

```

    anger(user, system)
    if
    wants(user, getNonEmptyResult)
    blameworthy(user, getNonEmptyResult)
    (conditions:) emptyResult
    (proposed system action:)
    relaxRequest

```

7 Emotions and Communicative Acts

Emotions that are signaled by facial expressions do not need to be accompanied by additional communication at all. Emotions expressed by voice are naturally related to some acoustic output. In the extreme, this output is only a container for the expressed emotion, but usually it contains a certain semantic content. The analysis of the relation between semantic content and underlying emotions is in its infancy, compared, e.g., with the relation between verbally communicated semantic content and pointing gestures. The latter is sufficiently known to build practical application. We distinguish in the following between communicative acts with semantic content, which are provided by speech and gestures, on the one hand, and emotions, on the other hand. The interpretation of pointing gestures and verbal utterances can be conceived as a fusion process that unifies pieces of information. Semantic contradictions between pointing gestures and verbally provided information are indications for errors. The relation between emotions and communicative acts is much more complicated. We give a presumably nonexhaustive classification of types of interaction between displayed emotion and communicated semantic content.

Redundancy

Semantic content redundantly expresses a simultaneously displayed emotion such as *that makes me angry* or *I'm glad about that*, or semantic content expresses an attitude that corresponds to the direction of the emotion (whether it is positive or negative) such as *great, bad*.

Contribution to the Explanation of the Emotion

Semantic content expresses a concrete attitude (like or dislike) that is involved in triggering the emotion such as *I don't like thrillers* or *great movies*, or semantic content addresses the facts and actions that caused the emotion such as *you didn't understand me* or *that takes too much time*, or simply by uttering *thrillers* accompanied by a positive or negative emotion.

The *thriller* example contributes the concrete feature that may fill the abstract goal of being presented accordingly preferences. But this example does not necessarily express a like or dislike as *great movies*. With a negative emotion, the example may also belong to the topic *Semantic Content as Repair Action*.

Semantic Content as Repair Action

The semantic information is provided to repair the state of affairs that has caused the emotional state of the user. The example *thriller* works also here: *thriller* could be a correction of a misunderstanding of genre. There is no direct relation between the content of the utterance and the displayed emotion.

This is very common and important in human-machine dialogue as well as in human-human dialogue: The dialogue partner repeats or reformulates her request and concurrently displays a negative emotion. With overt anger, it could also be expected that the user cancels the interaction as a final form of repair.

Change of Semantic Content

The user displays a negative emotion and communicates verbally a positive attitude such as *marvelous*, *great movies*. The direction of the valenced attitude that is communicated verbally is changed by the direction of the displayed emotion. This is a simple form of sarcasm.

8 Conclusion

A complete implementation of the whole processing chain was available at the end of the project. There was no opportunity for systematic tests, which require high effort. For instance, the recognition of facial expression needs careful preparation of the environment in respect to lighting conditions in order to work. Our limited testing shows that, provided recognition is correct, the emotion interpretation generates the reactions that are requested by the use case specification.

We successfully implemented a cognitive model-based approach for analyzing emotions and other affective states of a user who participates in a multimodal human-machine dialogue. This is a success, but it will still take considerable effort to make it practically useful. The approach is based on an elaborated theory, which covers a broad range of phenomena. This is promising with respect to the extensibility of the approach. It is an important advantage of the approach that it generates conceivable explanations of emotions that allow for well-directed system reactions.

The approach is not restricted to handle classical emotions, but extends to other affective states. Also it is not restricted to states that are displayed nonverbally. Affective verbal feedback, such as *I like this*, can be explained along similar lines.

References

- J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog. In: *Proc. ICSLP-2002*, pp. 2037–2040, Denver, CO, 2002.
- A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to Find Trouble in Communication. *Speech Communication*, 40:117–143, 2003a.
- A. Batliner, C. Hacker, S. Steidl, E. Nöth, and J. Haas. User States, User Strategies, and System Performance: How to Match the One With the Other. In: *Proc. An ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, pp. 5–10, Chateau d’Oex, Switzerland, August 2003b.
- A. Batliner, V. Zeißler, C. Frank, J. Adelhardt, R.P. Shi, E. Nöth, and H. Niemann. We Are Not Amused – But How Do You Know? User States in a Multi-Modal Dialogue System. In: *Proc. EUROSPEECH-03*, vol. 1, pp. 733–736, Geneva, Switzerland, 2003c.
- R.R. Cornelius. Theoretical Approaches to Emotion. In: *Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, pp. 3–10, Newcastle, Northern Ireland, 2000.
- C. Frank, J. Adelhardt, A. Batliner, E. Nöth, R.P. Shi, V. Zeißler, and H. Niemann. The Facial Expression Module, 2006. In this volume.
- A. Ortony, G.L. Clore, and A. Collins (eds.). *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, UK, 1988.
- C.S. Peirce. Abduction and Induction. In: J. Buchler (ed.), *Philosophical Writings of Peirce*, pp. 150–156, Berlin Heidelberg New York, 1995. Springer.
- R.W. Picard (ed.). *Affective Computing*. MIT Press, Cambridge, MA, 1997.
- T. Portele. Interaction Modeling in the SmartKom system. In: *Proc. ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, Chateau d’Oex, Switzerland, August 2003. ISCA.
- H. Prendinger, S. Descamps, and M. Ishizuka. Scripting Affective Communication With Life-Like Characters in Web-Based Interaction Systems. *Speech Communication*, 16:519–553, 2002.
- W. Wahlster. SmartKom: Symmetric Multimodality in an Adaptive and Reusable Dialogue Shell. In: R. Krahl and D. Günther (eds.), *Proc. Human Computer Interaction Status Conference 2003*, pp. 47–62, Berlin, Germany, June 2003. DLR.
- V. Zeißler, J. Adelhardt, A. Batliner, C. Frank, E. Nöth, R.P. Shi, and H. Niemann. The Prosody Module, 2006. In this volume.