

Annotating State of Mind in Meeting Data

Dirk Heylen

Dennis Reidsma

Roeland Ordelman

University of Twente
Human Media Interaction
 {heylen,dennisr,ordelman}@ewi.utwente.nl

ABSTRACT

We discuss the annotation procedure for mental state and emotion that is under development for the AMI (Augmented Multiparty Interaction) corpus. The categories that were found to be most appropriate relate not only to emotions but also to (meta-)cognitive states and interpersonal variables. The history of the development of the annotation scheme is briefly described. The discussion centers around the presentation of the procedure.

Keywords

Annotation procedure, Emotion and Mental State, Meetings

INTRODUCTION

The AMI corpus (see www.amiproject.org) is a collection of multi-modal meeting recordings. The majority of meetings were elicited using a scenario whereby groups of four participants played different roles in a corporate design team. More than one hundred hours of meeting data have been collected. The goals of the AMI project are manifold. The prime goal is to develop several kinds of “meeting technologies”. The technology around which much of the research is centered in the first phase of the project is called the “Meeting Browser”. The Meeting Browser is a collection of programs that allows people to have access to the recordings that were made. It involves special techniques in multimedia indexing, multimedia retrieval and multimedia extraction. The major effort at the first stage of the project was to acquire the data. The second stage involved defining and testing annotation schemes for various relevant dimensions for the various applications that will make use of the data collection. Currently, the main work is to annotate the recorded meetings manually with all kinds of information that can be used as meta-data for the recordings or, more importantly, for use by machine learning techniques that will automatically extract features from the data or that will generate annotations automatically. The recordings are of interest not just to the signal processing researchers, or the researchers dealing with multimedia information retrieval/extraction but also for people interested in face-to-face conversation for its own sake: conversational analysts, linguists, social psychologists etcetera.

Procedures that can determine what people feel and think when they are engaged in conversation can help us to

retrieve information from recordings of people in a meeting. From the point of view of the relevance for meeting browsing and other techniques for building up memories of what happened in a meeting, it is obvious that what is relevant about what goes on in people’s minds is not only what they “felt” about what was being said in the emotional meaning of the word, but also whether they were surprised by the things that were said, certain, sceptical or how clear or confusing certain issues were presented.

In the following sections, we discuss in detail our approach to the annotation for mental state for the AMI corpus. The main discussion centers around the presentation of the procedure. This involves segmenting the video and audio streams first and then assigning a number of features and categories to each of these segments. The annotation labeling consists of assigning categorical labels as well as indications of dimensional parameters (intensity and valence). We will discuss the selection of the categories. In passing, the history of the development of the annotation scheme is briefly described.

ANNOTATION PROCEDURE

The job of the annotator is to watch videos of the AMI meetings and annotate them with information about what we will call the “mental state” of the participants. This means that the annotator continuously tries to answer the question: “What state of mind is this person in?” Is the person happy, surprised, interested, bored?

There are two important points to make at the start. First, the notion “mental state of a person” can be loosely interpreted as “the feeling of a person”. We want to stress that we consider feeling in a broad sense to include not only typical emotional categories such as “irritated” or “amused” but also so-called (meta-)cognitive states and processes such as “trying to remember”, “paying attention” “interest”, “distraction” or “puzzlement” and what could be termed propositional attitudes such as “scepticism” or “uncertainty”. Another kind of category that was used included terms such as “dominance”, “defensiveness” and “support” that provide a characterisation of the interpersonal variables.

A second point to note is that we tell annotators that we do not expect them to be able to completely mind-read what is going on with a participant. We discourage them from psychoanalysing a person in depth. Instead we ask them to give a broad description only of what they can reasonably

assume to be the mental state of the person, based on what they observe.

The instructions for the annotator and the whole procedure were developed after a number of sessions in which the developers of the scheme (i.e. the authors of this paper) watched and analysed several meetings together and individually; trying to achieve a consensus agreement on segmentation and labelling.

For each annotation assignment, annotators watch the video recordings of a meeting. They have a choice as to which viewpoints they want to watch. In particular, there are close-up recordings available for every single participant along with several overview videos that allow one to see the behaviour of the participant in context of the other participants. There are videos that show the pairs of meeting participants sitting at the same side of the tables and a video track, shot from the side that shows all participants. The close-up recording is generally used for the annotation of mental state, often accompanied by an overview video that provides more information about the context. The annotators can choose which videos they prefer to have open for inspection.

The annotation task consists of two parts: first, defining “cuts” (segmentation points) in the video of a person at places where a *distinctive change* in the mental state of this person occurs, and second, to fill in a form that describes each segment that is thus created. In brief, the instructions to the annotator read as follows:

- Start watching the video and try to imagine what the mental state is of the person you are observing.
- As soon as you notice a distinctive change in the mental state you press a key that will mark a segment boundary. The video stops playing. Note that this

boundary will in fact be an “end-point”. The start of this particular segment is the end-point of the previous segment (or in case it is the first segment, the start will obviously be the start of the video).

- You fill in a form that describes the mental state of the participant in the segment. This will include specifying the intensity, and the quality or evaluation of the mental state (whether it is a positive or a negative one). This will be explained in more detail below. You also have to choose one or more relevant category words from a predefined set that fits the mental state of the participant in the segment best.
- You press the “continue” key. The video resumes playing at the beginning of the next segment and the annotation process reiterates.

The video and audio controls are shown in Figure 1. The controls for the annotation form are shown in Figure 2.

CHANGES IN MENTAL STATE

The basis of the annotation process is marking up changes in the mental state of the participant in a video. These define the segments for which labels are defined. There are two types of change that we want to be annotated and that consequently should lead to the creation of a segment boundary: a change in mental state type and a clear change in the intensity of the mental state. Being amused, annoyed, angry, happy or relaxed are examples of mental state types. We explicitly defined one special “mental state”: the neutral state. Evidently, a neutral mental state is a construct that does not really exist, but we define it here as an observed mental state that does not have a distinctive type with a particular intensity level. Such a segment will be denoted as a neutral segment with intensity 0.

A change in mental state type will be observed for example when somebody has a neutral look for a while, and one can observe that this person starts to look amused. Or if somebody has been looking very annoyed for a while, and then relaxes again to a neutral state. Typically the changes in mental states have longer or shorter fade-ins and fade-outs. A look of surprise may arise suddenly and disappear quickly, whereas amusement might start with a slight pulling up of the corners of the mouth (almost unnoticeable) that gradually builds into a complete smile and then slowly dissipates again. A clear change in the intensity of a mental state can be observed for example when somebody has been looking vaguely annoyed for a while, and suddenly the person starts to look extremely annoyed and frustrated. In this example, the annotator is asked to place a segment boundary when the intensity level starts to increase. The segment to the left of that boundary receives a type description “annoyed” with a relatively low intensity, and the segment to the right of the boundary a type description “annoyed” with the highest intensity value. Another typical example of an intensity change occurs when someone first looks amused and then starts laughing. The segment boundary should be placed just before the laughing starts. There are many cases where it is difficult to

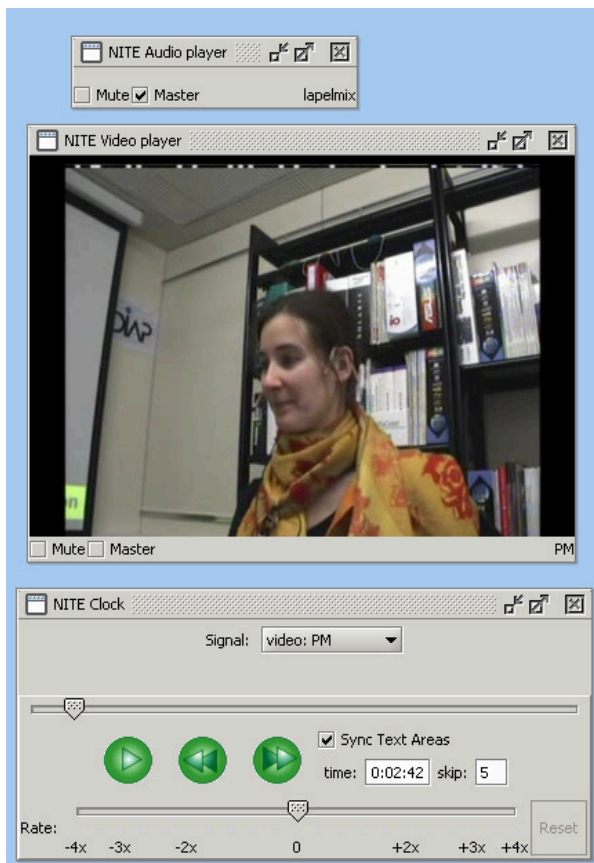


Figure 1 The video and audio controls

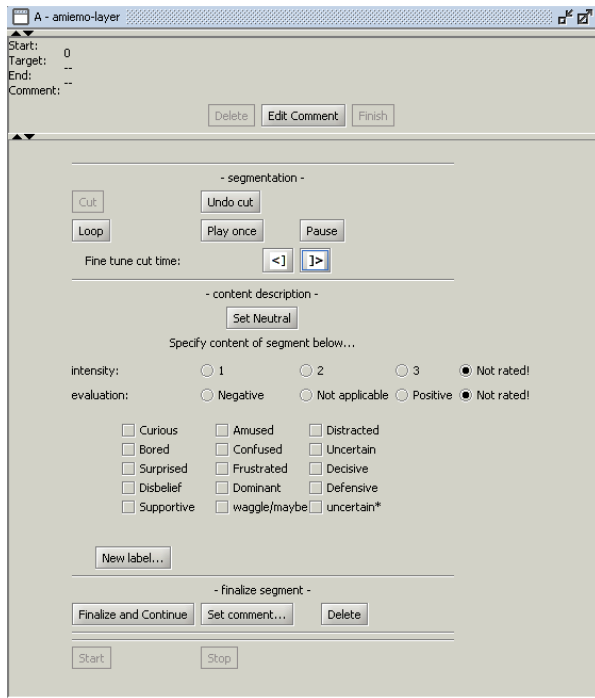


Figure 2 The emotion annotation interface

decide whether a type or intensity change should be marked as a segment boundary or not. In the example of the person that becomes amused (a smile develops, a smile remains for a while and gradually disappears) the choice will usually be to interpret this event as a single segment or “episode” which covers all three phases, instead of marking this as three separate segments differing in intensity. However, when the smile turns into a roaring laughter, the smile and the laughter should definitely be marked as two separate segments with different intensity values. In order to get familiar with these kinds of decisions, the instruction demos that are used for training the annotators contain segments with displays of the kinds of fluctuations in the intensity of certain mental states that we consider to be minor differences that do not deserve further segmentation. The instruction demos were created by the developers of the scheme.

Annotation dimensions

After having created a segment boundary, the mental state in the segment of focus has to be described. Forms¹ were designed for the annotator to make it easy to rate the intensity and the valence of the mental state and to choose a category label from a list of predefined labels that the annotator thinks best fits the observed mental state of the meeting participant in the segment of focus.

¹ The modules and tools were developed using the Nite XML Toolkit, an open source toolkit downloadable from <http://www.sourceforge.net/projects/nite>.

Intensity: Intensity is rated on a three-point scale: (1) low intensity, (2) medium intensity and (3) high intensity. A neutral mental state will automatically receive the intensity level 0.

Evaluation: People constantly evaluate the events, their own and others actions in many ways. This can be judgements about they are good or bad in a moral, ethical sense; whether they are good or bad for the goals they are pursuing; whether what is being said is true or false, believable or unbelievable, etcetera. Annotators are asked to mark the evaluation of a segment ‘negative’ in case the negative aspects of evaluation seem predominant for the participant, and positive otherwise. Sometimes positive and negative are not applicable, in which case one can choose the “not applicable” option. ‘Positive’ might for example be appropriate when someone is laughing; ‘negative’ for example when someone is very angry.

Category labels: Finally, a label that best describes the mental state of the participant in the segment has to be chosen from a predefined list. More than one label can be chosen, if the annotator feels that the segment fits into multiple categories. However, the annotators are told to do this only if there is not possible to assign one dominant state or if multiple mental states are clearly marked and deserve to be labelled as such.

The labels indicate quite general, diffuse categories. The names on the form have been chosen to indicate a category of labels rather than a specific state. For instance, the label “curious” should be used if one notices that the person observed shows a special interest in something. Other words that one might have used are “attentive” or “focused” or “interested”. If the annotator thinks another label or description is exactly to the point given a particular observed mental state, there is room to add this word or description. Also, if the annotator feels that the segment does not fit any of the categories indexed by the labels, he or she can propose a new label and add this. We encourage annotators to use the standard label set, that was defined especially for the meeting domain, as much as possible.

Labels

Each of the labels that are currently in use is presented below. These labels were introduced by looking at several fragments of the data by the designers of the schema and by other annotators during a number of trials of the software and the schema. As we just said, some of the labels that were nearly synonymous were collapsed into one category.

Neutral: nothing remarkable is happening.

Curious: (interested, attentive, focused) The participant shows special interest in a topic or issue. In many situations “paying attention” is the neutral state (people are listening to others, for instance). These cases are labeled as neutral.

Amused (cheerful, joking): The participant is clearly amused.

Distracted (inattentive): The participant is not paying attention to the central issue in the meeting. The participant can be distracted by specific other things, or the participants mind may be wandering.

Bored: The participant is clearly bored with the proceedings of the meeting.

Confused (puzzled): There is something that the participant does not understand or that the participant cannot work out.

Uncertain (hesitant): The participant is not certain about something. (“I don’t know”)

Surprised: Something occurs or is said that the participant had not expected.

Frustrated (annoyed): The participant appears frustrated or annoyed about something.

Decisive (certain, confident): The participant is decisive, or very confident and certain about something and shows this by being more assertive and resolute than normal. This may be about an issue in the meeting (e.g. giving an opinion of which the participant is very certain), or about the meeting process itself.

Disbelief (sceptis, doubt): The participant does not believe something, is e.g. sceptical whether an idea is good, a statement is true, a solution will work.

Dominant (challenging): The participant shows dominant behaviour with respect to someone else, e.g. the participant is commanding, controlling or persuading others.

Defensive (apologetic): The participant reacts defensively to e.g. protect own ideas, or authority.

Supportive (affirmative, agreeing, approval): The participant shows support for another participant, either with respect to a contributions to the meeting, or towards his or her presence in the meeting.

Discussion

As we noted in the introduction. The various labels describe a mixed group of phenomena: emotions as well as meta-cognitive states and interpersonal stances. It is also clear that by providing a simple list the systematic ways in which some labels are connected does not come to the fore. For instance, one could think of the labels “curious” (attentive) and “distracted” (inattentive) to be opposites. However, this does not pose a problem as such. In this case, one could say that the format is equivalent to a seven point scale for the “attention” dimension (3 positive, 3 negative, and the neutral label). Interesting though, is the fact that some labels have opposites that are not in the list because they have not been marked in the corpus so far.

There is a group of labels that “confused”, “surprised”, “uncertain”, “disbelief” that are not always easy to distinguish in context. Particularly because one thing may lead to another and cause and effect are, in general, hard to distinguish. Further trials should point out how confused annotators get applying these labels.

CONCLUSION

In this paper we have presented the annotation procedure for mental states of the AMI corpus. We are currently involved in extensive trials of the annotation schema. Results of reliability of the scheme, coverage, and the distribution of labels are presented in the main conference. The continuing trials should indicate the stability of the schema. Interesting to explore are the relations between these markers of mental states and other annotation levels,

particularly the dialogue act schema and the argumentation scheme. The signals that annotators perceive that make them choose a particular label are also signals that the other participants in the meeting can see and will interpret as signs for the mental state. It will be interesting to see how these shape and determine the way the meetings proceed.

ACKNOWLEDGMENTS

We thank the annotators for all their efforts and comments. This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication).

REFERENCES

1. Cowie, R. E. Douglas-Cowie, S. Savvidou, E. MacMahon, M. Sawey, M. Schröder (2000) ‘FeelTrace’: An instrument for recording perceived emotion in real time. ISCA Workshop on Speech and Emotion, p. 19-24
2. Reidsma, R. D. Heylen, R. Ordeman (2006) Annotating Emotion in meetings. Proceedings of LREC.