On the use of the multimodal clues in observed human behavior for the modeling of agent cooperative behavior

Jean-Claude MARTIN

(1) LIMSI-CNRS, BP 133, 91403 Orsay, FRANCE
(2) LINC, IUT de Montreuil, Univ. Paris 8, 140 rue de la Nouvelle France, 93100 Montreuil, FRANCE
martin@limsi.fr http://www.limsi.fr/Individu/martin/

Abstract

We introduce TYCOON a framework we are developing for the analysis of human verbal and non-verbal behavior. This framework includes a typology made of six primitive types of cooperation between communicative modalities: equivalence, specialization, transfer, redundancy, complementarity and concurrency. We have used this typology when annotating videotaped multimodal human-computer interaction and human-human communication. We have defined some metrics for the analysis of such multimodal behavior of observed subjects. The computed values of such metrics are being used for the specification of the multimodal behavior of an embodied conversational agent.

We will present how these low-level specifications of combination between modalities might be used in the future as building blocks for bridging the gap between experimental human behavior analysis and the modeling of cooperation-related multimodal attitudes related for instance to autonomy, delegation and control.

Introduction

Recent techniques have been developed for the observation of human multimodal behavior (Maybury and Martin 2002). We claim that such corpora and tools which have been mainly developed for the analysis of human verbal and non-verbal behavior can also be useful for studying how people make their communication modalities cooperate in order for them to reach their communicative goal, but also how a group of people cooperate for a collaborative task such as a collective oral presentation.

We will first describe our work on experimental human behavior analysis. We will then define the related framework for studying and monitoring cooperations between not only user interface modality agents, but also between software agents.

Experimental analysis of human cooperative multimodal behavior

When communicating with someone else, we use several modalities such as speech and gestures in a cooperative way. The mechanisms that underlie this multimodality of human communication are not completely identified nor understood. Similarly we do not know completely the behavior a subject might have when facing a system, which allows her to use these different modalities of communication.

The study of this multimodality of human communication is currently driving federative research grouping several domains such as Linguistics and Computer Sciences. It includes several dimensions such as the need to collect data on existing corpora but also the definition of coding schemes, the development of annotation tools and of algorithms computing behavioral metrics.

Following previous work on manual annotation of human multimodal behavior, we have developed tools making easier the annotation and the computation of behavioral metrics. We have defined a XML DTD grammar for specifying the annotations. According to this grammar, such annotations are composed of several sections. A first section describes the objects the subject is referring to in the corpus (i.e. Drawings on a blackboard). Each of the following sections of an annotation contains a multimodal segment, itself composed of several sub-section (one for each modality). A Java software has been developed in order to parse such annotations and compute behavioral metrics. These tools have been applied to 40 samples taken in several corpora. One example is given in Figure 1.

For defining these behavioral metrics, we use TYCOON a framework we are developing for the analysis of human verbal and non-verbal behavior (Martin et al. 2001). The framework includes a typology made of six primitive types of cooperation between communicative modalities: equivalence, specialization, transfer, redundancy, complementarity and concurrency. We have used this typology when annotating videotaped multimodal human-computer interaction and human-human communication.

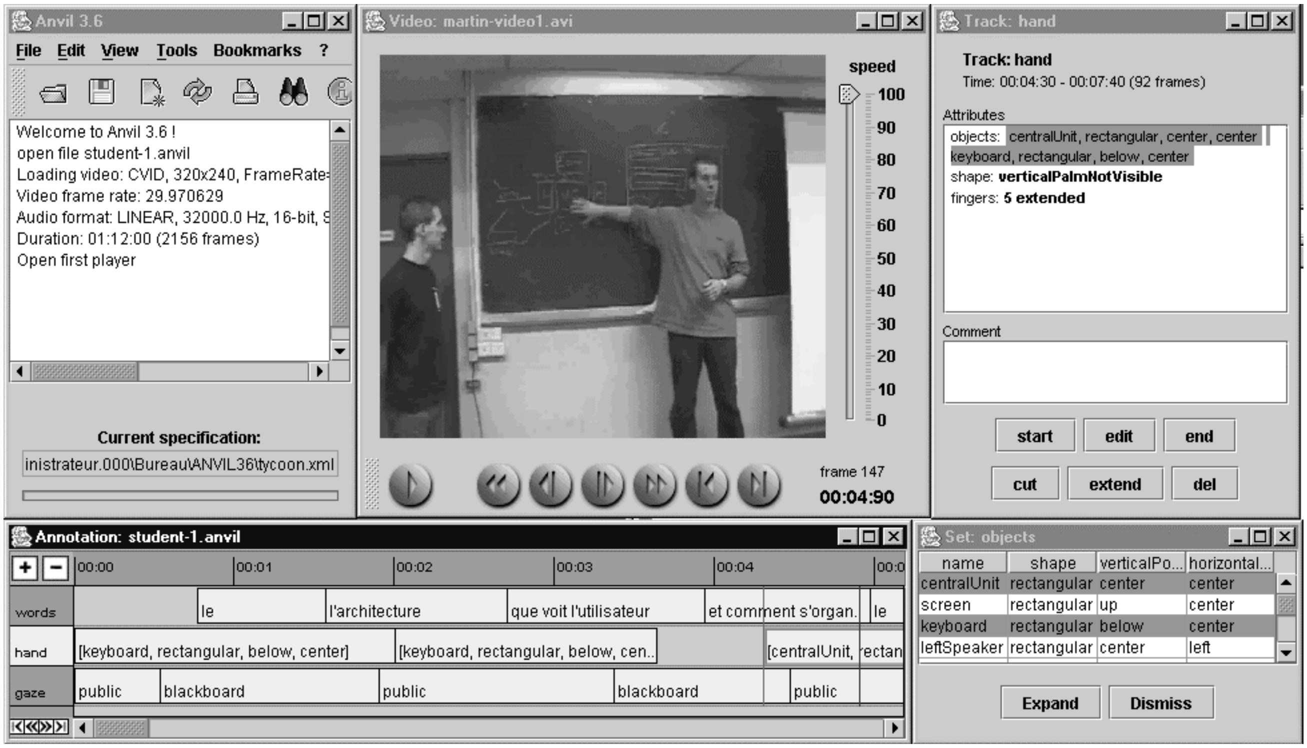

Figure 1: Screenshot of an annotated example featuring two students cooperating during a collective oral presentation. The annotation scheme (lower left window) contains 3 tracks (spoken words, hand gestures, gaze). A new Anvil (Kipp 2001) feature enables the annotation of objects referred in gesture and speech (lower right window) as described in (Martin & Kipp 2002).

We have defined some metrics for the analysis of such multimodal behavior of observed subjects. The computed values of such metrics are being used for the specification of the multimodal behavior of an embodied conversational agent (Figure 2).

Framework for studying and monitoring cooperation between agents

Following these measures of human multimodal behavior, we propose a tentative framework for studying and monitoring cooperations between agents, would these agents be the communicative modalities that one human use in order to reach his/her communicative goal, or would it be several human or software agents cooperating in order to reach a collective communicative presentation goal.

Cooperative environment. A cooperative environment is composed of: an application domain, a set of referenceable objects, a set of agents and a set of types of possible cooperation between these agents.

Application domain. An application domain is defined by a set of message templates (including command names and associated parameters in the case of Human-Computer Interaction) and a set of object.

Referenceable object. A referenceable object embeds an object of the application with knowledge on how to refer to this object (with linguistic or non-linguistic means).

Information chunk. An information chunk is represented by a set of features. A feature provides the value of only one attribute of an information chunk (i.e. the date, at which it was detected, or a word that was recognized, or a request that was understood). A name, a content and a confidence factor define a feature.

Agent. An agent is a computational process represented by: its name, a set of input information chunks it may process, a set of output information chunks it may produce and a confidence factor associated with this process. An agent may be human or software.

Cooperation. A cooperation requires the exchange of information in order to achieve a common goal. In Tycoon, we have distinguished six possible types cooperation between agents.

Equivalence. A cooperation by equivalence is defined by a set of agents, a set of chunks of information, which can be produced by either of the agents and a criterion, which is used to select one of the agents. When several agents cooperate by equivalence, this means that a chunk of information may be produced as an alternative, by either of them.

Figure 2: Screendump of the LEA multimodal agent. A XML language is used to specify each modality configuration (buisine et al. 2002).

Transfer. Cooperation by transfer is defined by two agents and a function mapping the output of the first agent into the input of the second agent. When several agents cooperate by transfer, this means that a chunk of information produced by one agent is used as input by another agent.

Specialization. A cooperation by specialization is defined by an agent, a set of agents A and a set of chunks of information this agent is specialized in when compared to the agents of the set A. When agents cooperate by specialization, this means that the same agent always produces a specific kind of information.

Redundancy. Several agents, a set of chunks of information and three functions define cooperation by redundancy. The first function checks that there are some common attributes in chunks produced by the agents, the second function computes a new chunk out of them, and the third function is used as a fusion criterion. If agents cooperate by redundancy, this means that these agents partly produce the same information.

Complementarity. Cooperation by complementarity is defined similarly as cooperation by redundancy except that there are several non-common attributes between the chunks produced by the two processes. The common value of some attributes might be used to drive the fusion process. When modalities cooperate by complementarity, different chunks of information are produced by each agent and have to be merged.

Concurrency. A cooperation by concurrency means that several agents produce independent chunks of information at the same time. These chunks must not be merged.

Goals of cooperation. Several agents may exchange information and cooperate for several reasons such as enabling a fast interaction between agents or improving mutual understanding of the agents.

Application to monitoring cooperation between software agents

Our framework is focusing on the way an agent integrates messages coming from several other agents. Several agents A_j may indeed cooperate according to several types of cooperation to process a message m received by a "classical" central "facilitator" agent (Figure 3):

- *equivalence*: each agent A_j can process the message m but with different response time or confidence which will lead the facilitator to send the message to only one of these agents,

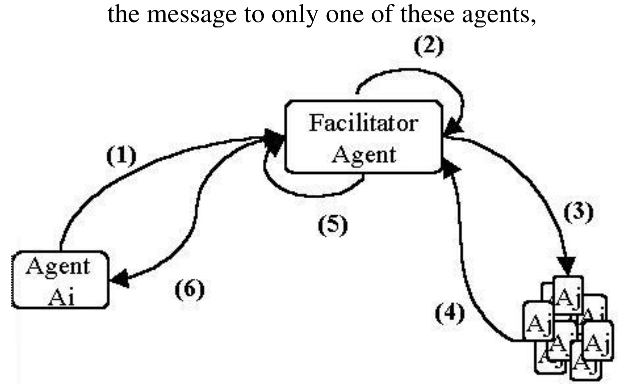

Figure 3: The types of cooperation can be used by a facilitator agent: 1) An agent Ai sends a message to the facilitator agent, 2) Considering the services declared by a set of agents {Aj}, the facilitator selects one or several agents Aj as well as the type of their cooperation, 3) The facilitator builds some messages and sends them to the selected agents Aj, 4) One or several agents report to the facilitator, 5) In the case of redundancy and complementarity these messages are integrated, 6) The facilitator sends a reply to the agent Ai.

- *redundancy*: each agent A_j can process the message m but with different response time or confidence which will lead the facilitator to send the same message m to all the agents A_j, to wait for the results and to merge them,
- *complementarity*: each agent A_j can process only part of the message m which will lead the facilitator to send parts of the message m to all agents, to wait for the results and to merge them,
- *specialization*: the facilitator will send the message m to the only agent who can process it.

A preliminary version of the program has been tested in the case of a very simple multi-agent system where agents share the knowledge about the values of different variables. Each agent may either know the value of a variable, or how this value depends on the value of other variables. We consider the agents displayed in Figure 4.

- Agent A1 knows that x = 10.
- Agent A2 knows that y = 2*x + 3.
- Agent A3 knows that z = x + y * 2.
- The first message is from agent A4 that asks to the facilitator the value of variable z.

We consider two ways of computing rates of use of each type of cooperation.

- "Statically": the allowed types of cooperation as a function of the initial knowledge of the agents (such as agent A1cooperates by specialization with the other agents when considering the value of variable X).
- "Dynamically": the types of cooperation as observed from the history of messages collected by the facilitator (agents A2 and A3 have been observed to cooperate by complementarity for the computation of the value of variable Z during the processing of the request sent by A4).

In order for an agent to decide whether it should act autonomously, by delegation or by control, it should have some information on the type of cooperation that it might expect from other agents.

Autonomy could be used when the agent is specialized in a specific chunk of information or service. Delegation may require the agent to know the possible cooperation between other agents.

Conclusion: cooperative attitudes related to autonomy, delegation and control

We believe that our TYCOON low-level specifications of cooperation between agents might be used in the future as building blocks for bridging the gap between experimental human behavior analysis and the modeling of higher-level cooperation-related attitudes related to autonomy, delegation and control.

Potential future questions include: How the knowledge of possible types of cooperation between available agents can be used in solving conflicts between agents? What effects does it have on task performance and communication performance (such as the number of exchanged messages)? How such cooperation principles can be formalized? How much is this typology/metrics a commonality across research camps involved in the workshop? Can it be applied to the monitoring of cooperation between real and complex autonomous agents? Between human and software agents? Can the knowledge of such types of cooperation be useful to select initiative strategies? How much tycoon metrics can be indicators of agent cooperation efficiency? Could tycoon metrics be used to monitor team creation or inter-team cooperation?

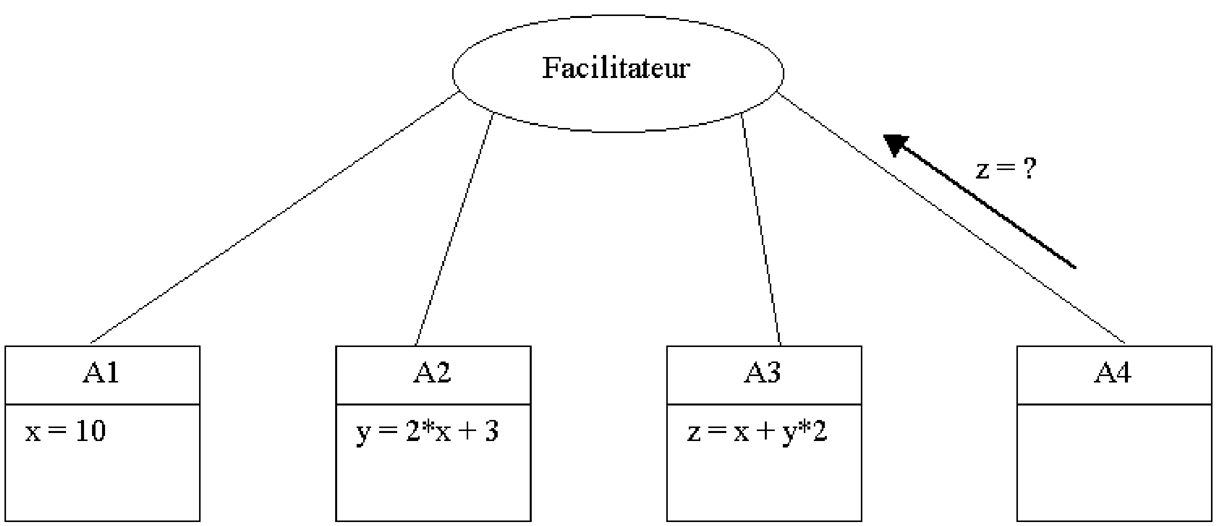

Figure 4: Example with 4 agents.

We will propose a tentative mapping between these cooperation-related and potential clues in multimodal behavior but also cooperation between these clues according to the TYCOON typology. The use of multimodal clues in observed behavior might indeed help to reach a better understanding and modeling of cooperative behavior.

Acknowledgments

The graphical design of the LEA agent has been achieved by Christophe RENDU who can be contacted at crendu@nexusanimation.com and +33.6.03.60.43.62

The author also wishes to thank Michael Kipp, Caroline Vignoles, Stéphanie Busine, and Sarkis Abrilian for their contribution to parts of the work described in this paper.

References

Buisine, S., Abrilian, S., Rendu, C., Martin, J.-C. (2002) Towards Experimental Specification and Evaluation of Lifelike Multimodal Behavior. Proceedings of the Workshop on "Embodied conversational agents - let's specify and evaluate them!" http://www.vhml.org/workshops/AAMAS in conjunction with The First International Joint Conference on Autonomous Agents & Multi-Agent Systems, 16 July, 2002, Bologna, Italy.

Kipp, M. 2001. “Anvil - A Generic Annotation Tool for Multimodal Dialogue”. *Proceedings of Eurospeech 2001*, pp. 1367-1370, Aalborg, September 2001.

Martin, J.-C. & Kipp, M. 2002. “Annotating and Measuring Multimodal Behaviour - Tycoon Metrics in the Anvil Tool”. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC’2002),* Las Palmas, Canary Islands, Spain, 29-31 may 2002.

Martin, J.-C., Grimard, S., & Alexandri, K. 2001. "On the annotation of the multimodal behavior and computation of cooperation between modalities". *Proceedings of the workshop on " Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents ",* May 29, 2001, Montreal, in conjunction with The Fifth International Conference on Autonomous Agents. pp 1-7. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-46/

Maybury, M., and Martin J.-C. 2002. *Proceedings of the workshop "Multimodal Resources and Multimodal Systems Evaluation". (Eds).* Third international conference on language resources and evaluation (LREC'2002), Las Palmas, Canary Islands, Spain, 1st June 2002 http://www.lrec-conf.org/lrec2002/lrec/wksh/Multimodality.html.