Walkie-Talkie MIKE

Ian Frank[1], Kumiko Tanaka-Ishii[2], Hitoshi Matsubara[1], and Eiichi Osawa[1]

[1] Future University, Hakodate, Japan, {ianf,matsubar,osawa}@fun.ac.jp
[2] Tokyo University, Japan, kumiko@ipl.t.u-tokyo.ac.jp

Abstract. We address the problem of information flow in disaster relief scenarios by presenting an architecture for generating natural language dialogues between large numbers of agents. This architecture is the first step towards real-time support systems for relief workers and their controllers. Our work demonstrates how techniques from the MIKE commentary system for RoboCup soccer can be carried over to the domain of RoboCup Rescue. Thanks to this background, the initial product of our research is a system that explains a RoboCup Rescue simulation not to the agents in the domain themselves but to a watching *audience*. This "commentary" is produced by recreating the actual dialogues most likely be occurring in the domain: walkie-talkie conversations.

1 Introduction

In a disaster relief situation, *information* is crucial. The disaster scene itself will be confusing, complex and quickly-changing. Yet the different groups of agents in the domain (relief workers, ground controllers, and civilians) will each need fast access to different kinds of knowledge.

Our theme in this paper is that thinking only of *information acquisition* in such domains is not enough. Rather, we suggest that it is the *flow* of information between groups of agents that makes the difference between "disaster" and "disaster relief". As a concrete first step towards understanding this information flow, we investigate probably the single most important communication method in the domain: *walkie-talkie conversations*.

Specifically, we demonstrate how walkie-talkie conversations can be generated for the RoboCup-Rescue simulator [1]. That is, we take the current (silent) simulator and, based on the events in the simulated world, produce a "soundtrack" of the exchanges you might expect to hear from *real* relief workers and *real* control centres if the simulation being played out was itself also real.

In effect, the output of our system is a kind of "commentary", and this is no accident, since it builds on the research that produced the MIKE commentator system for the RoboCup simulator league [2] and small-size robot league [3]. We show in this paper that many of the lessons learned in domains such as soccer commentary carry over remarkably well into natural language generation for a disaster relief scenario. Most notably, we demonstrate how the large number of voices in the walkie-talkie conversation is actually an extreme version of decomposing an explanation task among several agents [4]. For such dialogues, and in

the face of the impracticality of constructing prior plans for the discourse, the agents benefit from using different discourse strategies. Most obviously in the disaster relief domain, the controller needs to concentrate on high-level analysis of the overall situation, whereas relief workers on the ground need to pay more attention to the low-level events happening around them.

In its role as a "commentator", we call our system Walkie-Talkie MIKE, and we intend to demonstrate it at RoboCup events, to add atmosphere to RoboCup-Rescue demonstrations in much the same way that MIKE itself adds atmosphere to soccer games. The main motivation for our research, however, is not this "entertainment" but instead the investigation of how best to handle information flow in disaster relief situations. The use of natural language is critical in this respect, as it is "resource sensitive"; it presents information in a way that leaves recipients free to use their hands and eyes for other important tasks. Yet, to our knowledge, this is the first time that research on walkie-talkie dialogues has been carried out. We anticipate that our work will be a first step towards a deeper understanding of natural language dialogues between large numbers of agents in rapidly changing domains.

2   The Basic Walkie-Talkie Architecture

The RoboCup-Rescue simulator is graphically impressive but presents a large amount of information in a small screen. Helping viewers understand this simulation was one of the initial motivations for considering the Walkie-Talkie MIKE project. To a significant extent, we succeeded in processing the information generated by the server with an architecture similar to that of the original MIKE soccer commentator [2]. Space limitations prevent us from giving more than a sketch of our implementation, but we highlight the main similarities and differences below, before giving an example commentary in §3.

Analyser Modules

Walkie-Talkie MIKE uses "analysers" to post text fragments we call *propositions* into a shared memory. For example, fire brigade number 5 heading towards location id number 34 is represented as `(move fire 5 34)`, where `move` is the tag, and `fire`, `5` and `34` are the attributes.

The primary analysers used by Walkie-Talkie MIKE interpret the status of each type of agent in the domain: civilians, the police force, the ambulance and the fire brigade. These analysers have a rather different role to the analysers in soccer. Rather than trying to scrutinise the differences between two teams of agents with very similar abilities, they instead build up a picture of three types of information:

- Interpretation of the agents' actions (summary, analysis and prediction).
- Information that the agents will want to broadcast to other agents in the nearby area or further afield (different from the *actual* messages, if any,

broadcast by the agents themselves, since the protocol for inter-agent communication is not yet stable in the Rescue simulator, and any given set of clients may use semantics that are not known in advance).

– Information that the agents will need to know or be warned of. (All conversations — but especially walkie-talkie dialogues — are two-way processes. As well as simply listening and speaking, it is important to know what kind of information to *request* in order to perform at the best possible level.)

Additional Analysers

In addition to monitoring the actual agents, the RoboCup Rescue domain requires two further analysers. The first of these processes information about buildings and other infrastructure in the simulation. The second analyser is one that we use to represent a "control center".

In the original versions of the Rescue Simulator, there was no "control center". Our ControlCenter analyser therefore represented an imaginary control center of our own creation. However, the most recent version of the server allows for one control center for each type of relief worker (ambulance, police, fire brigade). The actual role of these control centres within the simulation is not yet clear (they are represented as a special type of "building" that can send messages), but we regard the notion of the control center as vital, since it is here where the lower-level inferences produced by the agent and the building analysers are most naturally built up into a higher-level picture of the overall domain. We have designed our ControlCenter analyser with two purposes in mind. Firstly, we require the ControlCenter to be a participant in the dialogue between the domain agents, so that we can experiment with the widest possible range of options for producing walkie-talkie dialogues. Secondly, we want to be able to use the ControlCenter as a real-time supporter and adviser for genuine, human ground controllers at a disaster scene. What we envisage here is something similar to the visualisation provided by the Statistics Proxy Server in RoboCup [5], but also including a natural language component.

Importance Scores

The Control Center also forms the basis for our use of importance scores: each proposition has a score that captures how important the event it describes should to a *ground controller* in the domain. The basic architecture of Walkie-Talkie MIKE then provides two mechanisms for selecting propositions to be uttered. The first, and simplest, is to allow the setting of an independent threshhold for each of the analysers. Any propositions with importance values higher than this value are then flagged and time-stamped. The second mechanism is more sophisticated, and requires the monitoring of *changes* in the values of importance scores over time. This flags and time-stamps any proposition whose change exceeds a configurable threshhold. Together, these mechanisms allow for maximum flexibility when applying the basic Walkie-Talkie MIKE architecture to different tasks (such as advising ground controllers, or generating a commentary).

To support the actual generation of speech, Walkie-Talkie MIKE associates a number of templates with each proposition, and selects randomly amongst these for the propositions that have been flagged for utterance. The actual generation of natural language depends on the application that the system is being used for. In this paper, we give the specific example of generating a "commentary" for the Rescue Simulator, as described below.

3   Walkie-Talkie Dialogues for the Rescue Simulator

Generating a realistic dialogue for the Rescue Simulator with the basic architecture of Walkie-Talkie MIKE requires the specification of a control algorithm for selecting among flagged propositions. We chose to implement this by concentrating on the high-level events identified by the ground controller (for the moment, we assume a single ground controller in our dialogues). The basic operation is thus for the controller to speak whilst there are important events to describe. In any gaps, individual domain events are selected in order of importance and an appropriate agent is chosen to speak a template describing the event.

To increase the realism of the generated dialogue, we also allow the control center to acknowledge statements made by other agents in the domain. Further, before the control center describes an event, we may carry out some "scene-setting" by generating an extra utterance from an agent associated with the event. Finally, we also complement the Natural Language Generator with a small repertoire of sound effects. For instance, when a fire brigade agent is reporting a fire, we augment the audio with the sound of a fire burning.

To give an idea the effect of the resulting dialogues, we present in Figure 1 an example. Each phrase in this dialogue is generated by instantiating one of the NL templates contained in the basic Walkie-Talkie MIKE architecture. For the best overall effect, imagine hearing the exchanges as though they are being carried out over walkie-talkie communications, with bursts of radio static at the beginning and ends of each person's speech.

Already, the commentary of Figure 1 is of a type that rescue workers could simply have playing through an earpiece to pick out important pieces of information (e.g., the comments of the controller, or of other rescue workers they know are nearby). Note, though, that this example dialogue was not generated from an actual simulation, but was constructed by us manually to illustrate a number of key attractions of researching walkie-talkie dialogues in the disaster relief domain:

– **Appropriate Atmosphere**. As a commentary, the style of a walkie-talkie dialogue represents a very natural way of describing the situation to people watching the simulation. The example of Figure 1 could alternatively begin "This simulation is of an earthquake of size 7 on the Japanese scale. Such an earthquake causes structural damage and makes movement very hazardous... If you look here on the display you can see a fire brigade heading towards this fire in the North..." This kind of "explanation" may work in other domains, but it seems out of place in disaster simulation.

```
<Control> All units. The latest estimation of the size of the
earthquake is a seven on the Japanese scale. Expect structural damage
to roadside buildings. Proceed with extreme caution.
<Voice1> Control, Red brigade 3 heading towards fire in North.
<Control> Roger, fire spreading rapidly there. Red squares on our
monitors increasing.
<Voice2, background of flames roaring> Situation very bad
here. <background of victims screaming> people are trapped here.
<Control> Now have seven teams fighting fires in the North.
<Control> Cover in other areas is low. Don't over-commit, people.
<Voice4> We got three people out safely here.
<Voice3> Red 2 drawing back to the West.
<Control> Fire in the North quarter looks to be under control. Blue
squares our monitor there increasing.
```

Fig. 1. Example dialogue based on Walkie-Talkie MIKE templates

This is an issue of *appropriateness*. For instance, televised soccer almost always features announcers, so it is appropriate to simulate them in RoboCup. However, in disaster relief nobody really expects a commentary as such, so it is more appropriate to recreate the voices that would normally be present: the rescue workers themselves and their controllers.

– **High-level vs low-level**. The most recent versions of the MIKE soccer commentator system have demonstrated the benefits of separating the discourse strategies used to present high-level and low-level information. The same distinction also fits naturally into a walkie-talkie dialogue: we have low-level information about what each of the rescue workers or teams are doing, and also have high-level information based on the overall view of the disaster area. Thus, we are able to use the disaster relief domain to further investigate how to effectively combine dialogues generated by different strategies.

– **Continuity**. This is one of the issues that makes soccer commentary surprisingly hard. For instance, if a commentary contains the phrase "Manchester United are playing well today" the subsequent statements can be phrased "*they* are using space well" and "*their* pass success rate is high" rather than "Manchester United are using space well" and "Manchester United's pass success rate is high". Whilst overall continuity is difficult to achieve in RoboCup, a walkie-talkie dialogue in the RoboCup Rescue domain is more forgiving. Listeners hear the burst of static before each phrase and recognise that there may be a context shift coming. It is therefore less critical to maintain continuity.

Note that since each phrase gives a context to itself (emphasised by the burst of static at the beginning and end of each speaker's turn), we are not even limited to having all the voices speaking in the same *language*. The entire audience of the disaster simulation does not have to understand every phrase. It is enough to assume that every utterance is understood by one of the other agents in the simulation (or by the controller).

– **Explaining the Simulation**. In some of the controller's statements in our
example commentary, there are explanations of the graphical view of the
simulator itself. This is a convenient, indirect way to explain RoboCup-
Rescue to an audience. The implicit assumption is that the controller is using
the same viewer as we are. Note that even if a human is present to explain
the simulator (for example at a RoboCup-Rescue demonstration), Walkie-
Talkie MIKE can be used without significant modification as "background
colour". Since our goal is not to "explain" but to recreate a part of the
domain currently missing (the audio part), the walkie-talkie conversation
can be left running while the human talks about various specific features of
interest. Walkie-Talkie MIKE is also fairly robust to the possible changes that
may be made to the display (for example zooming, highlighting, panning)
for the benefit of the watching audience. The walkie -talkie phrase "Control,
Red brigade 3 heading towards fire in North" is acceptable even when Red
brigade 3 is not visible on the screen, unlike the corresponding "commentary"
of "Now you can see Red 3 heading North".

4   Future Work and Conclusions

We have described the basic architecture of Walkie-Talkie MIKE and shown how
to make use of this architecture to produce a "commentator" for the Rescue
Simulator. In doing so, we demonstrated the feasibility of extending established
natural language generation techniques to the disaster relief domain.

We will be demonstrating Walkie-Talkie MIKE as a live "commentator" at
the 2002 RoboCup. However, this represents just the first step of our work in this
area. In the long term, we expect our work on Walkie-Talkie MIKE to provide
new insights into multi-agent natural language generation. In the more immedi-
ate future, we plan to use our system to produce natural language systems for
summarising and passing high-level information from a control centre to individ-
ual (and teams of) relief workers. Building on this, we also plan on producing a
more complete system to assist ground controllers in real-time. Another possibil-
ity is the broadcasting of information to civilians trapped in a disaster area. For
instance, some parts of the world (including areas of Japan) have an infrastruc-
ture of public speakers that could be used to broadcast instructions directing
citizens to the safest routes away from the most dangerous areas.

When it comes to actually pushing RoboCup-Rescue techniques into practice
in the future, it will be necessary for robots to work with humans (and for
humans to work with robots). Through our work on Walkie Talkie MIKE, we
are starting to appreciate both the problems and the types of dialogues that
are important for such collaboration. Although the name "Walkie Talkie" MIKE
describes primarily the style of our system's current dialogue, we hope our work
will eventually justify a literal interpretation: allowing relief workers and robots
to walk and do their jobs whilst talking to them and for them about all the
important events that are unfolding in the domain.

References

1. Hiroaki Kitano, Satoshi Tadokoro, Itsuki Noda, Hitoshi Matsubara, Tomoichi Taka-hashi, Atsushi Shinjoh, and Susumu Shimada. RoboCup-Rescue: Search and rescue for large scale disasters as a domain for multi-agent research. In *Proceedings of IEEE Conference on Man, Systems, and Cybernetics (SMC-99)*, 1999.
2. K. Tanaka-Ishii, I. Noda, I. Frank, H. Nakashima, K. Hasida, and H. Matsubara. MIKE: An automatic commentary system for soccer. In *Proceedings of ICMAS-98*, pages 285–292, 1998. Also available from Electrotechnical Laboratory as Tech Report ETL-97-29.
3. I. Frank, K. Tanaka-Ishii, H. Okuno, Y. Nakagawa, K. Maeda, and H. Nakadai, K. Kitano. And the fans are going wild! SIG plus MIKE. In *Proceedings of the Fourth International Workshop on RoboCup*, Melbourne, Australia, 2000. To be published by Springer in LNCS/AI series.
4. K. Tanaka-Ishii and I. Frank. Multi-agent explanation strategies in real-time do-mains. In *Proceedings of ACL2000, the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, 2000.
5. I. Frank, K. Tanaka-Ishii, and K. Arai. The statistics proxy server. In *Proceedings of the Fourth International Workshop on RoboCup*, Melbourne, Australia, 2000. To be published by Springer in LNCS/AI series.