Rocco: A RoboCup Soccer Commentator System

Dirk Voelz, Elisabeth André, Gerd Herzog, and Thomas Rist

DFKI GmbH, German Research Center for Artificial Intelligence
D-66123 Saarbrücken, Germany
{andre,herzog,rist,voelz}@dfki.de

Abstract. With the attempt to enable robots to play soccer games, the RoboCup challenge poses a demanding standard problem for AI and intelligent robotics research. The rich domain of robot soccer, however, provides a further option for the investigation of a second class of intelligent systems which are capable of understanding and describing complex time-varying scenes. Such automatic commentator systems offer an interesting research perspective for additional integration of natural language and intelligent multimedia technologies.
In this paper, first results concerning the realization of a fully automated RoboCup commentator will be presented. The system called Rocco is currently able to generate TV-style live reports for arbitrary matches of the RoboCup simulator league. Based upon our generic approach towards multimedia reporting systems, step-by-step even more advanced capabilities are to be added with future versions of the initial Rocco prototype.

1 Introduction

The Robot World-Cup Soccer (RoboCup) challenge [8, 7] poses a common standard problem for a broad spectrum of specialized subfields in Artificial Intelligence and intelligent robotics research. Obviously, advanced techniques related to autonomous agents, collaborative multi-agent systems, real-time reasoning, sensor-fusion, etc. are required to enable robots to play soccer games. The rich domain of robot soccer, however, provides a further option for the development of a second class of intelligent systems that are capable of understanding and describing complex time-varying scenes. Such automatic commentator systems offer an interesting perspective for the additional integration of natural language and intelligent multimedia technology into the research framework.

In general, the combination of techniques for scene analysis and intelligent multimedia generation has a high potential for many application contexts since it will open the door to an interesting new type of computer-based information system that provides highly flexible access to the visual world.

A first approach towards the automated generation of multimedia reports for time-varying scenes on the basis of visual data has been introduced in our own previous work [2]. For the experimental investigations reported in [2] short

M. Asada and H. Kitano (Eds.): RoboCup-98, LNAI 1604, pp. 50-60, 1999.

sections of video recordings from real soccer games had been chosen as domain of discourse. In [3] our initial approach has been further elaborated and carried forward to the RoboCup domain. The system ROCCO (RoboCupCommentator) presented here constitutes a first practical result of our recent work related to the domain of robot soccer. The current ROCCO prototype is able to generate TV-style live reports for arbitrary matches of the RoboCup simulator league.

The exceptional research potential of automatic commentator systems is well reflected by the fact that at least two more related research activities have been started within the the context of RoboCup. Similar to the initial ROCCO version, the system MIKE (Multi-agent Interactions Knowledgeably Explained) described in [12] is designed to produce simultaneous spoken commentary for matches from the RoboCup simulator league. A specific focus of the MIKE approach is an elaborated analysis of agent behavior. The system aims to explain and classify interactions between multiple agents in order to provide an evaluation of team play and to generate predictions concerning the short-term evolution of a given situation.

The system BYRNE [4] employs an animated talking head as a commentator for matches of the RoboCup simulator league. The main focus of this work is on the generation of appropriate affective speech and facial expressions, based on the character's personality, emontional state, and the state of the play. Currently, BYRNE does not connect to the RoboCup soccer simulator program directly but uses pre-analysed game transcipts as input.

2 The Rocco System

A practical goal of our current experimental investigations in the area of multimedia reporting is the development and continuous advancement of ROCCO, a system for the automatic generation of reports for RoboCup soccer games.

As a first step, we restrict ourselves to matches of the simulator league which involves software agents only. This allows us to disregard the intrinsically difficult task of automatic image analysis instead of using real image sequences and starting from raw video material like in our previous work on the SOCCER commentator [1].

The basic architecture of the ROCCO system is sketched in Fig. 1 (cf. [3]). ROCCO relies on the RoboCup simulator called SOCCER SERVER [10], which is a network-based graphic simulation environment for multiple autonomous mobile robots in a two-dimensional space. The system provides a virtual soccer field and allows client programs (i.e. software robots) to connect to the server and control a specific player within the simulation. ROCCO utilizes the real-time game log that the SOCCER SERVER is able to supply to monitoring programs. The continously updated information includes the absolute positions of all mobile objects (i.e players and ball) as well as additional data concerning game score and play modes (like for example a *goal-kick* or a *throw-in*). On the basis of this material, ROCCO aims to provide a running report for the scene under consideration. The graphical user interface for monitoring object movements

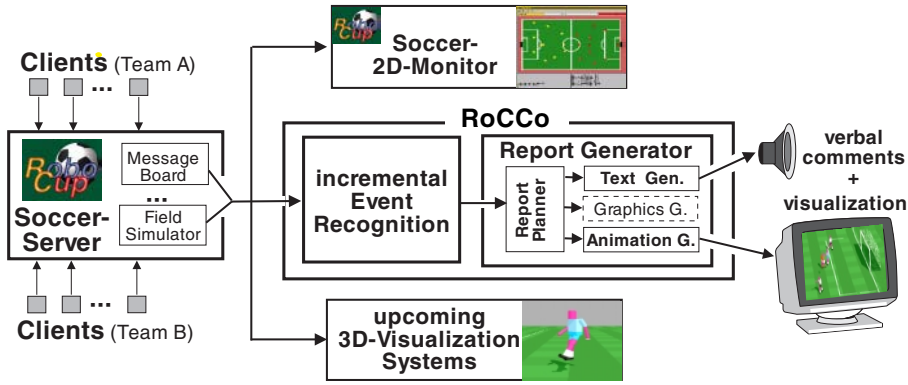

Fig. 1. Architecture of the ROCCO system

included within the SOCCER SERVER environment can be exploited as a simple visualization component. More advanced graphical displays become possible with already announced components for three-dimensional visualization (cf. [8]).

The automatic description of a time-varying scene can be thought of as a complex transformation process. Starting from a continuous flow of raw data, which basically describe simple geometric aspects of a scene, a more advanced analysis of the evolving scene is required in order to recognize relevant occurences that can be communicated to the user. ROCCO includes a dedicated component for the incremental recognition of interesting events in the time-varying scene. The resulting higher-level conceptual units form the input of the core report generator which exploits them for the construction of a multimedia report employing verbal comments and accompanying visualizations.

The initial Java-based ROCCO prototype, as it is shown in Fig. 2, is able to generate TV-style live reports for arbitrary matches of the RoboCup simulator league. The screenshot was taken during a typical session with the ROCCO system. In this case, the SOCCER SERVER logplayer is used to playback a previously recorded match. The window on the left contains a transcript of the spoken messages that have been generated for this example scene. Using the TRUETALK text-to-speech software, ROCCO is able to control intonation in order to generate a more lively commentary. The testbed character of the ROCCO system makes it easy to experiment with generation parameters to further explore the various possibilities for report generation.

3 High-level Scene Analysis

The continuously updated real-time game log from the RoboCup SOCCER SERVER forms the input for the ROCCO commentator system. This kind of *geometrical scene description* (cf. [9]) contains information concerning visible objects and

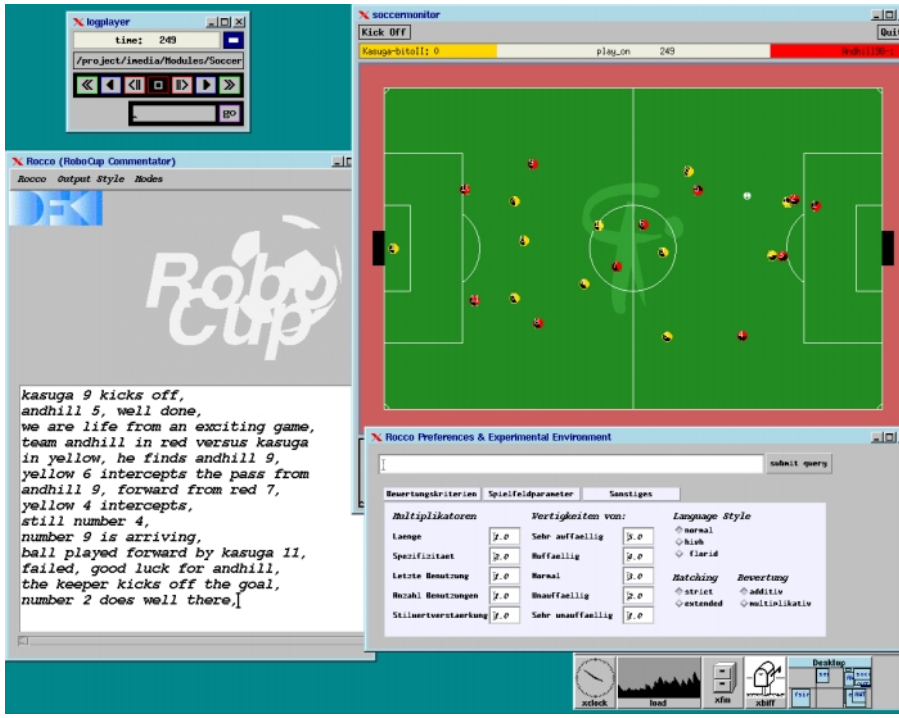

Fig. 2. The basic windows of the initial ROCCO system

their locations over time, together with additional world knowledge about the objects. High-level scene analysis provides an interpretation of this basicly visual information and aims at recognizing conceptual units at a higher level of abstraction. The information structures resulting from such an analysis encode a deeper understanding of the time-varying scene to be described. In general, they include spatial relations for the explicit characterization of spatial arrangements of objects, representations of recognized object movements, and further higher-level concepts such as representations of behaviour and interaction patterns of the agents observed (see [6]).

Within the ROCCO prototype we restrict ourselves to the incremental recognition of relevant events, i.e. interesting object movements involving players and the ball. The representation and automatic recognition of events is inspired by the generic approach described in [6], which has been adopted in ROCCO according to the practical needs of the application context. Declarative event concepts represent a priori knowledge about typical occurrences in a scene. These event concepts are organized into an abstraction hierarchy, grounded on specialization and temporal decomposition. Simple recognition automata each of which corresponds to a specific concept definition are used to instantiate events on the basis of the underlying scene data.

A peculiarity of the simulator league setting is the fact that the SOCCER SERVER already provides some elementary events related to selected actions of a single player (like *'kick'* or *'catch'*). Hence, ROCCO only needs to translate this data into a suitable representation format instead of doing a full event recognition. Elementary event and state predicates, which are oriented towards a single timepoint, can be utilized to define higher-level event concepts. Specific examples for these kind of elementary predicates are the following:

- *(Type:HasBall Time:time Agent:X)*
  Judging which agent actually is in possession of the ball, is an important part of event recognition in the soccer domain. As robots and softbots sometimes do not even perceive the ball when standing next to it, this kind of decision is not always an easy task in the RoboCup environment. For ROCCO, an agent X is considered to have the ball, if (1) the distance between them is less than a predefined constant value, (2) there is no one else closer to the ball, and (3) the agent is or was at least recently facing the ball.
- *(Type:Speed Time:time Object:X Speed:speed)*
  The speed of an object X (player or ball) will be interpreted as no movement if it is less than some very small constant value.
- *(Type:Angle Time:time Object:X Angle:angle)*
  The absolute angle associated with an object X depends on its current orientation or direction of movement.
- *(Type:Kick Time:time Agent:X)*
  The fact that agent X kicks the ball is recorded in the data from the simulator.
- *(Type:Region Time:time Object:X Region:region)*
  The playground is divided into several regions, like for example the penalty area. This predicate signifies that a certain object, i.e. the ball or a relevant player, enters a specific area of high interest.
- *(Type:Playmode Time:time Mode:mode)*
  A playmode change is to be interpreted as a game command by the referee (e.g. *'goal'*, *'upside'*) which also counts as an event that may be communicated.

From current and previous elementary predicates higher order events are calculated. Consider a *ball-transfer* event into the penalty area as an example:

(Type:BallTransfer
StartTime:178 EndTime:185
Spec:EnterPenaltyArea Agent:red9 Recipient:red7)

The recognition of a ball-transfer event depends on the recognition of all of its sub-events as listed below:

(Type:HasBall Time:start-time Agent:agent)
(Type:Kick Time:start-time Agent:agent)
(Type:Speed Time:time Object:Ball Speed:speed) with
$speed > 0$ for all $start - time \leq time \leq end - time$.

(Type:Angle Time:time Object:Ball Angle:angle) with
$angle$ is constant for all $start - time \leq time \leq end - time$.
(Type:Region Time:start-time Object:Ball Region:NotPenaltyArea)
(Type:Region Time:end-time Object:Ball Region:PenaltyArea)
(Type:HasBall Time:end-time Agent:recipient)
As a further condition, there must be no event of the form
(Type:HasBall Time:time Agent:agent) with
$start - time < time < end - time$.

The other event concepts in the knowledge base of the ROCCO system are defined in a similar way. The initial ROCCO system includes about 15 event definitions covering essential soccer event concepts.

4 Report Generation

The initial ROCCO prototype is designed as a robust TV-style live commentator system for the RoboCup simulator league. The system generates simple multimedia reports which combine a 2-dimensional graphical animation of the soccer scene with a spoken commentary. ROCCO employs the monitor program provided with the RoboCup simulation environment for the graphical display. Currently, the central task of the reporting module is the generation of a verbal description of the scene under consideration.

One of the characteristics of a live report is the fact that temporal constraints makes it impossible to comment every detail within the time-varying scene. Given the large number of simultaneous occurances within a soccer scene, the verbalisation process continuously needs to carefully select those events that should be communicated. Selected events are passed on to the realisation component which uses a template-based approach to generate corresponding natural language utterances. For each event instance, an appropriate template is selected and instantiated. Then the templates are annotated with intonational information, and finally text-to-speech software is used to generate audio output.

4.1 Discourse Planning

Report generation in a system that is continuously provided with new data calls for a flexible content selection mechanism. At each point in time, the system has to decide which events should be verbalized. To solve this problem, ROCCO relies on an incremental discourse planning mechanism, paying special attention to timing constraints. Before generating a new comment, ROCCO performs the following steps:

Threshold calculation: A minimum importance value for the next comment is calculated first. It consists of a *basic threshold* which decreases during periods of inactivity of the commentator system, then allowing less important events to be commented. Furthermore, a longer time of inactivity will also increase the chance to state some background comment instead of describing the current activities.

Calculation of topicality: All new recognized events are evaluated for their topicality. Topicality is determined by the salience of an event and the time that has passed since its occurrence.

Content selection: The event with the highest topicality is selected for verbalization if the topicality value exceeds the current threshold for the minimum importance value. If no such event exists, either a comment containing background information is randomly selected, or the systems waits until more important events have been recognized or the threshold has lowered enough to select a less important event. Once an event is selected, the minimum importance value will be set back to the basic threshold.

Human soccer commentators often report an event while it is taking place, which leads to common phrases like: *"Miller passes in ··· to Meier"*. To meet the requirements of simultaneous scene description, the recognition component also provides information concerning only partly-recognized events. As a consequence, the verbalization component cannot always start from completely worked-out conceptual contents. Starting with the verbalization of *"Miller passes"* before complete information about the whole event is available, limits the possibilities of discourse planning and sometimes urges even human reporters to use gramatically incorrect sentences. The verbalization component in ROCCO contains various templates for most combinations of partial event instantiations. In the example given, ROCCO achieves to comment the complete event of Miller passing to Meier in the penalty area with a phrase well known from soccer life-reports: *"Miller passes ··· now in the penalty area ··· Meier gets it"*.

4.2 Generation of Verbal Comments

One objective of the ROCCO system is to provide a large variety of natural-language expressions that are typical of soccer live reports. A more flexible template-based approach is being employed since it is rather difficult to generate syntactic specifications for soccer slang expressions so that may be processed by a fully-fledged natural-language generator. The template knowledge base consists of about 300 text-fragments transcribed from about 12 hours of soccer matches during the WorldCup qualifications 1997. Each event specification is assigned to a list of templates, that might be applicable depending on further contextual constraints. A template consists of fixed strings and variables which are instantiated when a template is applied to a specific event instance. A separate *nominal phrase generator* is utilized to describe agents, teams, etc. with appropriate nominal phrases. To improve the variety of expressions some templates can be modified at random. For example the word *there* can be added to the phrase *well done* with a chance of 30 percent. Table 1 contains a list of some templates for the *ball-transfer* event. For the appropriate choice of a template for the verbalization of the selected event several aspects are taken into account:

Applicability Constraints: Constraints specify whether a certain template can be employed in the current context. For instance, a phrase like *"Now*

Template	Constraints	Vb	Floridity	Sp	Formality	Salience
(?x passes the ball to ?y)	None	8	dry	4	formal	normal
(?x playes the ball towards ?y)	None	8	dry	4	formal	normal
(?x towards ?y)	None	5	dry	3	slang	normal
(?x combines with ?y)	None	6	normal	3	slang	lower
(?x, now ?y)	(NotTop ?y)	5	dry	2	slang	low
(ball played towards ?y)	None	5	dry	3	colloquial	normal
(the ball came from ?x)	(NotTop ?x)	6	dry	3	colloquial	higher
(shot)	None	1	dry	1	colloquial	normal
(?x)	(NotTop ?x)	2	dry	1	slang	low
(?y was there {'again':WasTopic(?y)})	None	4	dry	1	slang	normal
(well done {'there':Random(30%)})	None	2	dry	0	colloquial	normal
(a lovely ball)	None	3	flowery	1	colloquial	higher

Table 1. Some verbalization templates for the event *ball-transfer*

Miller" should not be uttered if Miller is already topicalized. To avoid close similarities of templates in the knowledge base, constraints further control some details of instantiation. Consider the template *(?y was there {'again':WasTopic(?y)})*. If, for example, Miller has been mentioned not long ago, the template will be instantiated as *"Miller was there again"* instead of *"Miller was there"* otherwise.

Verbosity: The verbosity of a template depends on the number of words it contains. While instantiated slots correspond to exactly one word, non-instantiated slots have to be forwarded to the nominal phrase generator which decides what form of nominal phrase is most appropriate. Since the length is not known at the time of template selection a default word number is assumed in this case.

Floridity: We distinguish between dry, normal and flowery language. Flowery language is composed of unusual ad hoc coinages, such as *"a lovely ball"*. Templates marked as normal may contain metaphors, such as *(finds the gap)*, while templates marked as dry, such as *(playes the ball towards ?y)* just convey the plain facts.

Specifity: The specifity of a template depends on the number of verbalized deep cases and the specifity of the natural-language expression chosen for the action type. For example, the specifity of *(?x looses the ball to ?y)* is 4 since 3 deep cases are verbalized and the specifity of the natural-language expression referring to the action type is 1. The specifity of *(misdone)* is 0 since none of the deep cases occurs in the template and the action type is not further specified.

Formality: This attribute can take on the values *slang*, *colloquial* and *normal*. Templates marked as formal are grammatically correct sentences which are more common in newspaper reports. Colloquial templates, such as *"ball played towards Meier"*, are simple phrases characteristic of informal con-

versation. Slang templates are colloquial templates peculiar to the soccer domain, such as *"Miller squeezes it through"*.

Salience: To generate more interesting reports, unusual templates, such as *"Miller squeezes it through"* should not be used too frequently.

To select a template, ROCCO first determines all templates associated with the given event concept and checks whether the constraints are satisfied. For each template, a situation-specific preference value will be calculated and the template with the highest preference value is finally selected. If the system is under time pressure, a template will be punished for its length, but rewarded for its specifity. In phases with little activity, all templates will get a reward both for their length and specifity. In all other situations, only the specifity of a template will be rewarded. Templates that have recently or frequently been used are punished and unusual templates, i.e. templates that seldom occur in a soccer report, as well as very flashy comments get an additional punishment while inconspicuous phrases get a bonus.

To generate descriptions for a player or a team, the nominal phrase generator is applied which takes into account the discourse context to determine whether to generate a pronoun or a noun phrase (with modifiers). In some cases, the template structure allows to leave out the object description if the object is already topicalized. An example is: *"Here comes Miller $\cdots$ (he) combines with Meier"*.

4.3 Using Intonation to Convey Emotions

Intonation should not only correlate with the syntactical and semantical structure of an utterance, but also reflect the speaker's intentions. Furthermore, intonation is an effective means of conveying the speaker's emotions.

To generate affective speech, ROCCO first identifies the emotions of the assumed speaker. For instance, starting with a neutral observer, succesful actions of a team or player will lead to excitement, failed actions to disappointment. Excitement will increase with decreasing distance to the goal and find its peak when a goal is scored or, in the case of disappointment, when a shot misses the goal. As excitement and disappointment are main emotions occurring in a soccer report, these were selected as a starting point for the generation of affective speech in the ROCCO prototype.

Trying to map emotions onto instructions for the speech synthesizer, we realized that already the variation of only two parameters could lead to amazing results. In particular these parameters are *pitch accent* and *speed*. A pitch accent is applied to a word to convey sentential stress. In ROCCO we currently use two out of six pitch accents introduced in [11]:

1. *H** is a *'high'* accent which, roughly speaking, stresses the indicated syllable with a high tone.
2. *L** is a *'low'* accent which, is realized as a tone that occurs low in the speaker's pitch range.

Speed is another effective means of conveying emotions. ROCCO increases its talking speed to express excitement, and slightly slows down to express disappointment.

These parameters are modeled similar to the emotions *glad/indignant* and *sad/distraught* as described in [5] where a highly elaborated system of synthesizer instructions for the production of different emotions is introduced. The approach presented in [5] proposes a slightly faster speech rate and a high pitch range for gladness, and a very slow speech rate and a negative pitch for sadness.

Since we rely a template-based approach, it is possible to annotate natural-language utterances directly with intonational information. However, in some cases, ROCCO will override the default intonation to track the center of attention. Consider the phrase *"Miller got it"* as an example. If it has just been commented that he is trying to get the ball from Meier of the opposing team, and hence Miller is alread topicalized, ROCCO will accent the word "*got*" to indicate the new information. Alternatively if Miller is not topicalized, the player's name is to be emphasized. In this case, ROCCO will override the preset accent markers, and add a new *H** mark for the noun phrase "*Miller*". Depending on the current amount of excitement, ROCCO will then set speed and pitch range.

5 Conclusions and Further Work

The initial ROCCO prototype presented here constitutes a first promising step towards a multimedia reporting systems for RoboCup soccer matches. ROCCO is a robust TV-style live commentator system for the RoboCup simulator league. It combines emotional spoken descriptions of the running scene with the graphical display provided through the RoboCup simulation environment.

The approach described here is motivated by our vision of a novel type of computer-based information system that provides fully automated generation of multimedia reports for time-varying scenes on the basis of visual data. This work and related activities illustrate the high research potential for the integration of high-level scene analysis and intelligent multimedia presentation generation, especially in the context of the RoboCup challenge.

Our current experimental investigations concentrate on further improvements of the first ROCCO version. One interesting topic is the incorporation of a more elaborated model for the generation of affect in synthesized speech. A second line of research relates to the extension of multimedia presentation capabilities. We are currently working on our own 3D visualization component to enable intelligent control of the camera view as a novel feature of the multimedia presentation generator.

A third important aspect for us is the move from softbot games in the simulator league to RoboCup matches involving real robots. The main problem there is to obtain a suitable geometrical scene description that can be fed into our reporting system. Instead of relying on an external observer, the idea is to exploit the analyzed sensor information of the acting robots. Hence, no additional camera and no specific image analysis will be required. However, it can be expected

that the geometric scene description to be obtained from one of the robot soccer teams will be less exact and less complete than in the case of the RoboCup simulator. Our plans are to further investigate these issues in close cooperation with one of the leading robot soccer teams at Universität Freiburg.

References

1. E. André, G. Herzog, and T. Rist. On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: The System SOCCER. In *Proc. of the 8th ECAI*, pages 449–454, Munich, Germany, 1988.
2. E. André, G. Herzog, and T. Rist. Multimedia Presentation of Interpreted Visual Data. In P. Mc Kevitt, editor, *Proc. of AAAI-94 Workshop on "Integration of Natural Language and Vision Processing"*, pages 74–82, Seattle, WA, 1994. Also available as Report no. 103, SFB 314 – Project VITRA, Universität des Saarlandes, Saarbrücken, Germany.
3. E. André, T. Rist, and G. Herzog. Generating Multimedia Presentations for Robocup Soccer Games. In H. Kitano, editor, *RoboCup-97: Robot Soccer World Cup I*, pages 200–215. Springer, Berlin, Heidelberg, 1998.
4. K. Binsted. Character Design for Soccer Commentary. In *Proc. of the Second International Workshop on RoboCup*, pages 25–35, Paris, 1998.
5. J. E. Cahn. The Generation of Affect in Synthesized Speech. *Journal of the American Voice I/O Society*, 8:1–19, 1990.
6. G. Herzog and P. Wazinski. VIsual TRAnslator: Linking Perceptions and Natural Language Descriptions. *AI Review*, 8(2/3):175–187, 1994.
7. H. Kitano, editor. *RoboCup-97: Robot Soccer World Cup I.* Springer, Berlin, Heidelberg, 1998.
8. H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. RoboCup: A Challenge Problem for AI. *AI Magazine*, 18(1):73–85, 1997.
9. B. Neumann. Natural Language Description of Time-Varying Scenes. In D. L. Waltz, editor, *Semantic Structures: Advances in Natural Language Processing*, pages 167–207. Lawrence Erlbaum, Hillsdale, NJ, 1989.
10. I. Noda. Soccer Server Manual Rev. 2.00 (for Soccer Server Ver. 3.00 and later). Internal document, ETL Communication Intelligence Section, Tsukuba, Japan, 1998.
11. J. B. Pierrehumbert. Synthesizing Intonation. *Journal of the Acoustical Society of America*, 70:985–995, 1981.
12. K. Tanaka, I. Noda, I. Frank, H. Nakashima, K. Hasida, and H. Matsubara. MIKE: An Automatic Commentary System for Soccer. In *Proc. of ICMAS'98, Int. Conf. on Multi-agent Systems*, pages 285–292, Paris, France, 1998.