Designing Multiparty Interaction Support in Elva, an Embodied Tour Guide

Jun ZHENG
School of Computing
National University of Singapore
3 Science Drive 2
Singapore 117543
65-97698795
zhengjun@comp.nus.edu.sg

Xiang YUAN
School of Computing
National University of Singapore
3 Science Drive 2
Singapore 117543
65-97698795
g0202623@nus.edu.sg

Yam San CHEE
National Institute of Education
Nanyang Technological University
1 Nanyang Walk
Singapore 637616
65-67903261
yschee@nie.edu.sg

ABSTRACT

Although social research into group interaction has flourished since the 20th century, the technology of embodied conversational agents for handling multiparty interaction is still immature. In view of this, we aim at designing agent systems to support group interaction in virtual environments. We intensively researched group formation processes and leadership in terms of turn taking management and nonverbal behavior in facilitated casual social groups. Towards achieving agent autonomy and believability, we employ group interaction mechanisms in an integrated four-layer agent architecture. This approach takes into account group modeling, turn taking moderation and multimodal coordination mechanism to ensure appropriate coupling of the agent's perception and actions at the group level. We also present an application of an embodied tour guide, named Elva, in an interactive art gallery. The application demonstrates how an agent can act as a leader in coordinating and facilitating multiparty interaction in a facilitated casual social group, in particular, a tour group context.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**], I.2.1 [**Applications and Expert Systems**], I.2.4 [**Knowledge Representation Formalisms and Methods**] and I.2.11 [**Distributed Artificial Intelligence**].

General Terms

Human Factors, Design.

Keywords

Virtual Reality, Multi-User Environment, Virtual Environments, Embodied Conversational Agent, Speech Act, and Virtual Art Gallery.

1. INTRODUCTION

Immersive virtual worlds offer enormous potential and opportunity to human users for the interactive experience and realistic three-dimensional graphical environment that they provide. Advanced agent technologies allow Embodied Conversational Agents (ECAs) to be built, in order to create more customized and user-oriented virtual experiences. To date, most of the agents effectively operate at one-to-one level. Meanwhile, Collaborative Virtual Environments (CVE) technology is becoming more popular. It aims to transform today's computer networks into the networks that support collaborative work and social play. CVE represents a natural extension of current single-user technology to support multiple-users. Furthermore, CVE can also support some aspects of social interaction which are not readily accommodated by single-user ECA technology.

In previous research, we built an embodied tour guide, Elva [1], that is able to engage conversationally with a system user about gallery exhibits, and capable of behaving according to social norms in terms of gestures and facial expressions (see Figure 1). Its layered architectural design has ensured appropriate coupling between the agent's perception and action. We utilized the notion of "schema" to support structured and coherent verbal behavior. Elva's nonverbal behavior is dynamically generated from communicative goals to mimic a face-to-face conversation.

Figure 1: Elva and Virtual Art Gallery

While the previous research has focused on Elva's interaction with a single user, multiparty interaction support is a desirable feature for the further development of this project. A multiparty scenario not only extends the questions from a two-party scenario, such as modeling of interaction and coordination of multimodal behaviors, but also presents entirely new challenges posed by the large number of users. How to embed relevant social rules in a multiparty system, how to exhibit leadership, and how to establish appropriate interaction style impose great challenges to the designer of the virtual agent. The successfully implemented multi-party interaction support aims to improve agent autonomy and believability and to achieve higher user satisfaction.

In this paper, we present an integrated agent design with multiparty interaction support. Section 2 reviews agent technology and social theory development. Section 3 discusses the specific issues in a facilitated casual social group. Section 4 focuses on system design and architecture. Section 5 presents a demonstration and evaluation, and section 6 concludes the paper.

2. RESEARCH BACKGROUND

In this section, we review the agent technology and social theory development in the group behavior domain.

2.1 Agent Technology Development

To date, researchers have heavily explored the effective use of ECAs in one-to-one interaction application context. Some research findings address the social effectiveness of one-to-one ECA. ECAs exhibit natural facial expression, eye gaze and head movement. In Elva [1], a schema-based framework is utilized to model the dialogues based on the narrative intention and communicative goal, and also to perform discourse planning and tour planning. Speech act theory was used to further enhance agent's reasoning ability. On the other hand, multiagent systems also attract a great deal of research attention. Many theories and technologies have been developed in order to design and specify multiagent systems. Gaia methodology [2] is proposed for the analysis and design of a multiagent system. It exploits organizational abstraction to provide clear guidelines for the design. Jung and Tambe [3] also exploit the performance model of coordination or conflict resolution strategies for a large group of cooperative agents. A number of teamwork models have been deployed successfully in complex dynamic multi-agent applications, such as Steam [4] and Alliance [5]. Teamwork in Steam [4] is based on agents' building up a hierarchy of joint intentions. Agents exploit such models to autonomously reason about coordination and communication. In Alliance [5], robots decide which tasks to perform in a behavior-based fashion. Thalmann worked on agent crowd research. In [6], the possibility of generating simulations with various realism levels and the hierarchical structure used to model crowds are presented though it is not yet real-time application.

However, less work has been done to investigate how best to create ECAs in group contexts. Therefore, multiparty support system research is still an immature field and warrants further investigation. Particularly, greater research effort is required in incorporating social norms into the system to create natural group behavior.

2.2 Social Theory Development

Social research into group interaction first became a distinguishable field in the early part of the 20th century. It flourished in related social and behavioral science disciplines such as speech communication, political science, organizational behavior, social work, and educational psychology. Group interaction is very complicated process, and it involves a large number of factors that affect the interaction simultaneously. Tubbs [7] identifies six basic types of groups. They are 1) primary group, 2) Casual and social group, 3) Educational groups, 4) Work groups, 5) Problem solving groups, and 6) Computer assisted groups. Casual and social groups, by definition, include neighborhood groups and fraternities in which the relationships may be relatively short-lived. Sometimes, casual and social groups have similar goals rather than common-shared goals, and in such a case there is less interdependence. A tour group is one kind of casual and social group. In tour groups, group visitors come and go at their own discretion. Group visitors also possess different backgrounds, thus the requirement and preference from each visitor is rarely the same.

Group size affects both the communication among participants and the quality and style of interaction. James's study [8] shows that a small group tends to have more discussion and interaction. However, interaction among members of a large group is more towards narrative style where fewer people are actively involved in conversation. Furthermore, Tubbs also shows that leadership styles differ in the degree and location of control.

There are three distinct styles, namely, authoritarian, laissez-faire, and democratic. Authoritarian and laissez-faire are two extremes, whereas democratic style represents an attempt to find a reasonable compromise between the extremes.

Multimodal behavior style in group context has been intensively studied. Cassell [9] discussed the use of gaze, gesture, and body posture which play an essential role in proper execution of many conversational behaviors, such as conversation initiation and termination. Goodwin [10] also investigated some particular aspects of the interaction of speaker and hearer in the construction of the turn at talk in natural conversation, in particular about the negotiation of an appropriate state of mutual gaze within a speaker-hearer relationship and the engagement display through a collaborative process of interaction.

3. FACILITATED CASUAL SOCIAL GROUP

Our research has focused on the casual social group domain, in particular, a tour group in a gallery. In this section, we discuss a casual social group's formation, leadership and nonverbal behavior style.

3.1 Casual Social Group's Formation

In the tour group context, group boundary needs to be identified very clearly. There are many ways to define group boundary. We use physical proximity as an indicator of group boundary because visitors have more flexibility to go to any place. Thus, only people who are physically close to the tour guide are interested in what the tour guide is saying. At the same time, due to the very nature of informal grouping, there are times that new visitors want to join the tour group, causing the group boundary to change over time. Moreland and Levine [11] argue that entry into a group is often marked by some ceremony or ritual. In order to further promote group dynamics, a great level of importance is attached to this ceremony. The initiation ceremony could take different forms, such as a warm welcome or pleasant eye contact. The ceremony serves a symbolic function both for the newcomer and for the group itself. For the newcomer, it helps in the process of identity transition. At the same time, the group also needs symbols to define its boundaries. However, the ceremony should not disturb the existing members of the group.

3.2 Leadership in Turn Taking Management

In earlier discussion, three leadership styles were discussed. Of the three, the democratic style is most relevant to this application. In a democratic style, the leader attempts to provide direction and to perform both task and social leadership functions, but at the same time he or she tries to avoid dominating the group with one person's view. Leadership is reflected in coordinating turn taking and choosing a suitable interaction style.

Turn-taking coordination is about managing the shift of turns among members in an appropriate manner. Turn-taking actions include take turn, request turn, release turn, assign turn, and hold turn.

Interaction happens continuously in tour group. From empirical study, two dominant interaction styles are evident in the conversation of tour group: narration and discussion. Each style exhibits different characteristics in terms of information flow, virtual agent role, and turn taking techniques.

- Narration: In narration interaction mode (see Figure 2), the virtual tour guide takes the initiative and *takes turns* to provide information about the gallery and its exhibits to visitors. In this mode, the tour guide mainly dominates the conversation to "broadcast" relevant information to all the visitors in the group. The tour guide also *holds turn* when moving to the next exhibit and waiting for the majority of group members to come closer. However, during the narration, group visitors can interrupt the tour guide at any time they want. In this case, the virtual tour guide releases the turn and attends to group members' questions or comments. Thus the interaction goes to Question and Answer (Q&A) sub-mode.

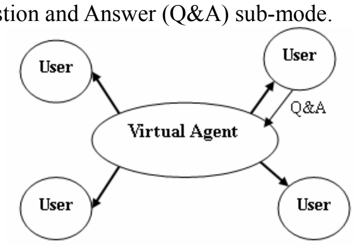

Figure 2: Narration Conversational Mode

- Discussion: In discussion mode (see Figure 3), the tour guide deliberately asks expansive questions to form a basis for discussion. In the discussion, the virtual tour guide mainly acts as facilitator at the side instead of being a sage on the stage. In order to promote group dynamics and group collaboration, the tour guide *releases turn* from most outspoken person and *assigns turn* to less outspoken visitors.

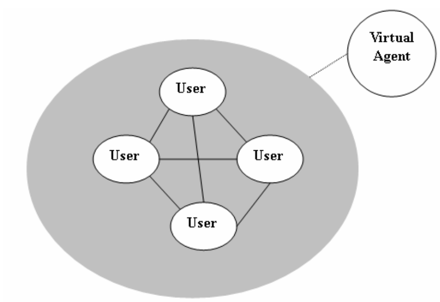

Figure 3: Discussion Conversation Mode

3.3 Nonverbal Behavior Style

Nonverbal behavior is instantiated as multimodal signals that distribute among different channels such as eye gaze, gesture, and so on. In previous research [1], we focused on designing interactional behaviors in turn taking and feedback. However, when designing non-verbal behavior in a group context, those behaviors should be modified according to social norms since a human tour guide handles a single visitor and a group of visitors differently. Through empirical study, we find that the social rules can be grouped into two main categories, namely boundary situation and turn holding situation.

Boundary situation includes the moment before speech, before taking turn, the moment after speech, after releasing turn, newcomer's arrival, and group member's exit and so on. But, turn holding situation represents the period when a certain user is holding the turn. Social rules group around these two situations. Appropriate nonverbal behavior needs to be generated according to each individual case.

Another area of non-verbal behavior is locomotion or movement. In a tour group, a tour can be divided into a sequence of locations. Through planning of the itinerary, a guide is able to intentionally direct the user from one artifact to another. At the same time, when the user navigates independently, the guide needs to track the visitor's movement to take appropriate action, such as following visitors and waiting for visitors. Sometimes, the guide does not need to respond to movement of the minority of visitors because it may disturb the majority of the visitors' experience.

4. DESIGN AND IMPLEMENTATION

In our design, the agent interfaces with the virtual world via a perception module and an actuation module (see Figure 4). The perception module provides the agent with high-level sensory information. The actuation module enables the agent to walk, gesture, and perform facial expressions. The gap between perception and action is bridged by mental processing with a four layered architecture of the agent's mind, a knowledge-based inference engine. The knowledge base at the heart of the system consists of knowledge entries that encode the domain knowledge as condition-action pairs. The reflexive layer implements a series of reflex behaviors that are quick and inflexible (Q&I). The reactive layer handles user's utterances as they arise. The deliberative layer provides planning and decision-making mechanisms. For details, refer to [1]. Above the three layers, the group interaction layer provides mechanisms to support multiparty interaction and multimodal generation in a group context. It takes into account the group modeling, turn taking moderation and multimodal coordination to ensure appropriate coupling of the agent's perception and actions at group level. Group Interactional Goal (GIG) (such as take turn and invite turn) is generated in the turn-taking moderation process. It is further utilized by the planner for discourse and schedule planning. Multimodal moderator generates the multimodal modifier to fine tune the multimodal behavior according to the communicative goal and group information. The following discussion focuses on the group interaction layer to show how the multiparty interaction is supported and how multimodal behavior style is moderated.

4.1 Group Interaction Layer

In the group interaction layer, the generation of group interactional behavior is achieved in three stages, namely appraisal stage, moderation stage, and expression stage. In the appraisal stage, the group modeler uses observed individual events from the perception module, processes the event, and accumulates the observed event in the state repository. It then passes a group level event to the moderators. In the moderation stage, the turn-taking moderator and multimodal moderator extract relevant information from the repository. Based on social rules, communicative goals and multimodal modifiers are

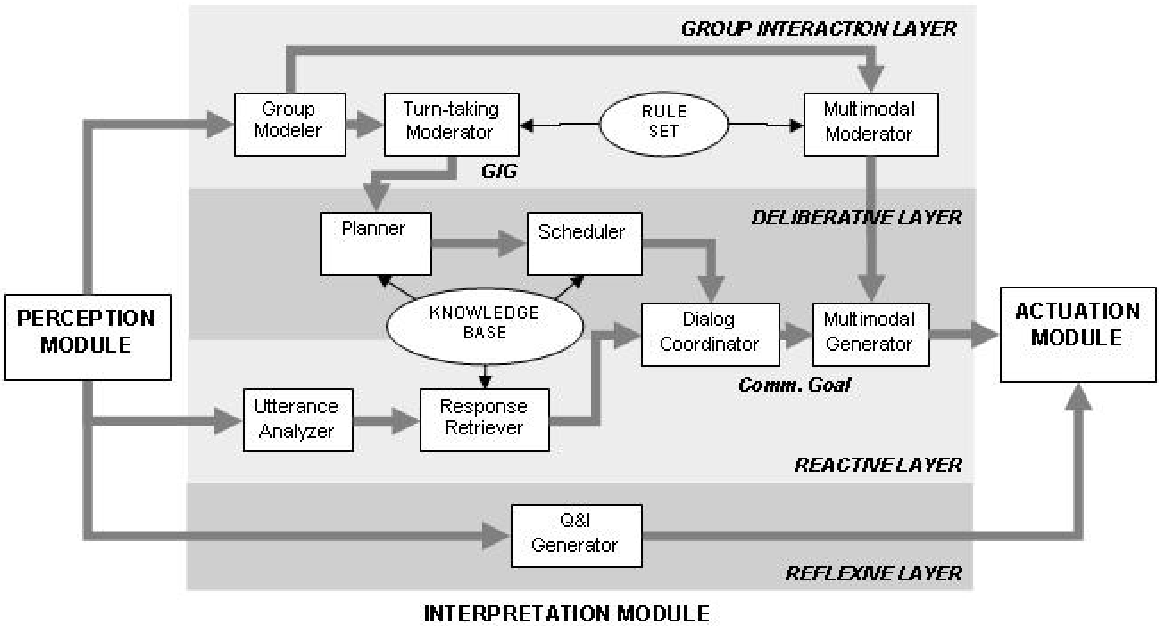

Figure 4: Four-layered Architecture for Agent's Mental Processing

generated and passed to planner and multimodal behavior generator in the expression stage. Finally, the planner and multimodal generator generate both verbal and non-verbal agent behavior.

In the group interaction layer (see Figure 5), the group modeler, turn-taking moderator, and multimodal moderator play an important role in supporting multiparty interaction and multimodal generation.

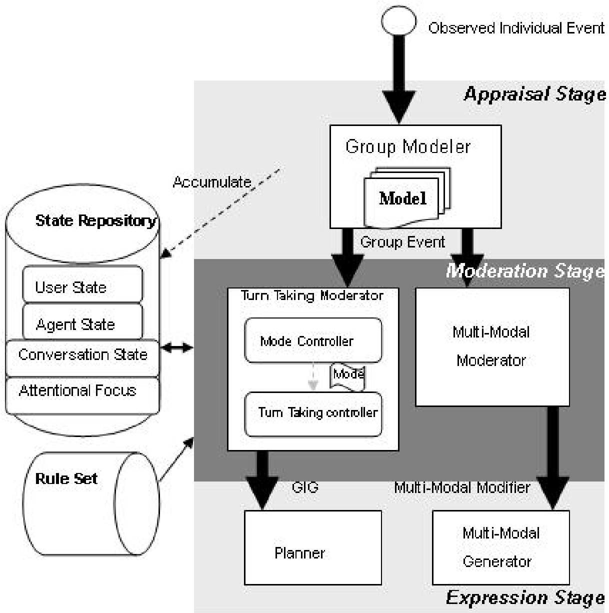

Figure 5: Three Stages in Group Interaction Layer

4.2 Group Modeler

The group modeler appraises observed events such as user's state and conversation state to generate group level events. Group level events are accumulated from individual events and are a combination and abstraction of many individual events. The virtual agent responds to group level events. For example, the agent observes user's navigation actions and models the actions as group navigation behavior such as NAVIGATION_IN_PROGRESS.

Individual event decomposition

Individual user events can be classified into two categories, movement events and speech utterance events. When either of these two events comes to the group interaction layer from the perception module, the group modeler gets user state information, conversation state, agent state, and attentional focus from the incoming event (see Table 1).

User locomotion state tracks user movement and keeps the states such as "NAVIGATION_IN_PROGRESS", "NAVIGATION_DONE" and "NAVIGATION_START". *Conversational state* keeps track of the speech event, such as "SPEECH_START" and "SPEECH_DONE". *User rating* reflects the user's activeness in terms of occupying air time. *Interaction mode* can be narration or discussion. *Conversation focus* and *KB focus* refer to the knowledge entry that constitutes the agent's current focus of attention. *Turn state* specifies the person who is holding the turn at that particular time.

Table 1: Individual event categories

User state:	User locomotion state
	User conversational state
	User Rating (active level)
	User location/orientation
Conversational state:	Interaction Mode
	Conversation focus/ KB focus
	Turn state
Agent state:	Agent location
	Agent orientation
	Agent conversational state

Group information accumulation

Group modeler decomposes an individual event into a set of meaningful fields and then stores them into the state repository. In the state repository, each user's state is stored separately and updated regularly whenever new individual events flow in. Conversation history is also recorded to be used as reference in future. Turn-taking moderator and multimodal moderator refer to state repository for relevant information to carry out appropriate group behavior.

Generation of group event

Group modeler generates a group level event based on the individual observed events. The group event is then passed to the moderation stage for further processing. Group level events include:

Group formation	Group termination
Group size change	Group navigation start
Group navigation done	Group idle
Group speech start	Group speech end

The group modeler employs the rule of physical proximity (using data like user-guide distance, user orientation relative to guide) to determine the boundary of group. When group boundary is redefined, group modeler fires a group formation event to the turn taking moderator and the multimodal moderator to generate appropriate "ceremonial behavior" such as smiling at the new-comer and saying "Welcome, Raymond."

4.3 Turn Taking Moderator

Turn taking moderator extracts data from the state repository. Through interaction mode control, the moderator suggests the shift of interaction mode in the current situation and then signals the planner. When appropriate, the planner searches in the knowledge base to find a suitable entry to realize the mode shift and turn assignment. Social rules are modeled inside the rule set to help generate the natural interactional behavior. Turn taking moderator has two inter-linked components: mode controller and turn taking manager.

Mode controller

In previous development, mode shift was readily specified in the planned discourse. With mode controller, the tour guide is now able to deliberately shift the mode, so as to cater for dynamic group size and duration of the mode.

First, the mode controller extracts information on group size and mode duration. Then, it decides whether a mode shift is desirable. We have developed a heuristic mode-shift-checking approach that is based on the characteristics of group interaction as discussed in Section 2.2. The basic assumption is that the possibility of mode shift depends on the group size and conversation mode duration. Given this assumption, the idea of discourse shift check is to use the group size and conversation mode as primary indicator to check when to actively change the mode. We then employ a checking function f to compute check score. f_1 is used in discussion mode, whereas f_2 is used in narration mode. n stands for group size and ***SignLevel(n)*** is a non-linearly incremental function on group size, which yields a corresponding significant level to measure the significance of the group size. One significant level can span a range of group sizes.

Furthermore, d is mode duration. β and γ are the coefficients. $\alpha 1$ and $\alpha 2$ are the powers of ***SignLevel(n)*** to denote the weight of ***SignLevel(n)*** to the function f. When the checking score exceeds a threshold, it triggers the mode change and updates the information repository. Finally, the turn taking manager signals a mode change request to the planner for further decision.

$$f_1(n, d) = (SignLevel(n))^{\alpha 1} + \beta \cdot d \quad (n=1,2,3\ldots)$$

$$f_2(n, d) = \frac{\gamma \cdot d}{(SignLevel(n))^{\alpha 2}} \quad (n=1,2,3\ldots)$$

The formulation of the checking function reveals our design rationale: first, turn taking mechanism is closely affected by group size. As mentioned, a small group tends to have more discussion and interaction. However, interaction among a large group is more towards narrative style where few people are actively involved in conversation. Second, mode duration is taken into consideration. The longer the duration in one mode, the more likely change of mode occurs. But group size has a greater effect than mode duration on mode shift since the group size determines the general pattern of the conversation, whereas mode duration only aims to fine tune the possibility of the mode change.

Turn taking manager

The design rationale of the turn taking manager is to create a more balanced group where everybody is encouraged to share comments and give feedback.

The turn taking manager controls the floor of conversation in the group. It extracts interaction mode, user rating, turn state and group navigation state from state repository. The turn taking manager decides who is to take the turn. From the agent perspective, the turn taking manager decides whether the agent should take turn, and to whom the agent shall assign a turn. As mentioned, there are five turn taking actions from the agent's perspective: take turn, request turn, release turn, hold turn, and assign turn.

In order to facilitate group interaction, the tour guide voluntarily releases the turn and opens the floor. Mode, user rating, conversation state and turn state are used to determine turn assignment result. The evaluation function g_{max} computes the maximum score of the multiplication of mode factor m (m for narration is 0.1; for discussion is 1) and speaker's user rating r (ranging from 0 to 10). If g_{max} exceeds threshold, meaning that there is a user who is too active, agent *releases turn* from conversation dominator. At the same time, g_{min} is calculated to identify the least outspoken person. Therefore the agent *assigns turn* to the least outspoken person.

$$g_{max} = \underset{i:1 \rightarrow n}{Max}\,(m \cdot r) \qquad g_{min} = \underset{i:1 \rightarrow n}{Min}\,(m \cdot r)$$

4.4 Multimodal Moderator

We focus our interest in designing interactional behaviors in the turn taking and the feedback which convey significant communicative information such as assigning a turn by gazing and raising eyebrows as suggested by Cassell [9]. Meanwhile, deictic behavior is also expressed together with interactional behavior to locate discourse entities in the virtual world, such as raising the hand to the addressee when assigning the turn.

Based on the Goodwin's [10] research, non-verbal behavior usually occurs in between two turns or during user turn. We adopt a goal-based model to generate nonverbal behavior. Communicative goals (such as take turn, give turn, seek turn, give feedback and request feedback) are mapped to a behavior script which corresponds to a nonverbal behavior.

The multimodal moderator serves to filter out unnecessary behaviors, expand and modify behaviors to fit in group context. Similar to the turn taking moderator, the multimodal moderator extracts information on group state, group boundary, group size and group location information from the state repository. The multimodal moderator adjusts nonverbal behavior based on the social norms (mainly gaze and locomotion), such as:

1) Agent scans the group with eye contact when talking.
2) Agent looks at the speaking member.
3) Agent moves closer to the group before talk.
4) Agent faces at the majority of group members.

Now we briefly describe the flow of generating and moderating a nonverbal behavior. The perception module perceives individual events, such as the user's movement, and sends to the group modeler. The group modeler updates the user's location/orientation in the state repository. Then the event is passed to the multimodal moderator and the turn taking moderator. The multimodal generator generates the behavior action such as "look at user". The multimodal moderator extracts the group location/orientation and the intentional focus to modify the action "look at user". The action is transformed to looking around at the crowd first, then focusing attention at the user who is currently the focus of conversation.

The multimodal moderator also fine-tunes the agent's navigational behavior. The virtual tour guide determines when to make a move to next sculpture, when to stop to wait for the visitors and when to stand in front of visitors before talking. In our computational approach, the virtual tour guide goes to next sculpture based on the existing planned tour itinerary, such as S1→ S2 → S3 → S4. In the case that the user navigates independently, the multimodal moderator decides if to filter out the event. Should the user navigation event be passed in, the virtual tour guide moves to the range of target sculptures first based on the update of the user's navigation information from state repository. Then the agent waits until it gets an update from the group modeler signaling that group navigation is done. The multimodal moderator extracts the updated group location and orientation information from the state repository and tunes/moderates the virtual tour guide's position before starting to talk.

5. DEMONSTRTATION AND EVALUATION

A Virtual art gallery has been designed using the design framework of the C-VISions, Collaborative Virtual Interactive Simulations system [12]. The C-VISions browser allows the user to interface with the virtual worlds: navigating and acting upon objects. A chatter's box allows the user to carry out conversation. The virtual guide, named Elva, appears in front of the user as an animated female character. Elva talks with system user through a speech synthesizer.

At present, the virtual gallery houses a virtual exhibition "Configuring the body: Form and Tenor" which utilizes existing content in the Ng Eng Teng Gallery of NUS Museums. We have tracked the following episode to illustrate the interaction between Elva and a group of system users as part of a virtual tour.

Elva is the virtual tour guide. Raymond, Alan and Joyce are the users' avatars in the system.

Elva is at the entrance of the gallery welcoming visitors.

<Elva waves her hand to users.>

1. Elva: How do you do, everyone?
2. Raymond: How do you do?
3. Joyce: Hi.

<Elva smiles at users.>

4. Elva: Welcome to the Ng Eng Teng Gallery! My name is Elva, and I am your guide.
5. Elva: In this tour, I will introduce the artist briefly. After that you can browse the artwork of your interest. I will be accompanying you along the tour to answer your queries.

Elva leads users to the sculpture "Oh, My Bump!" (Figure 6(a))

<Elva points to the sculpture using her right hand.>

6. Elva: Right here is this piece called "oh my bump".
7. Elva: Please rotate it to see the back.

New visitor, Vincent, comes closer to the group.

<Elva turns to Vincent and waves to him.>

8. Elva: Hi, Vincent. Welcome to join our group.
9. Now we are at the sculpture "oh my bump".

<Elva turns back to the rest of the group.>

10. Elva: What can you see from the sculpture?

Short Silence.

<Elva looks around at users and smiles>

11. Elva: Do not be shy. Feel free to share with us.
12. Raymond: I see a bump…
13. Joyce: Err. Is it a face?

<Elva looks at Joyce, and nods>

14. Elva: Good try. More?
15. Raymond: Face? Is it a body instead?
16. Elva: Yes. Please think along this line.
17. Raymond: I guess it is face and body.

<Elva turns to Alan.>

18. Elva: I notice that Alan is very quiet. Alan, what do you think about it?
19. Alan: Hmm. I see a torso as well as a face. (Figure 6(b))

Short Silence.

<Elva raises her hand.>

20. Elva: Well, thank you all for your guesses.

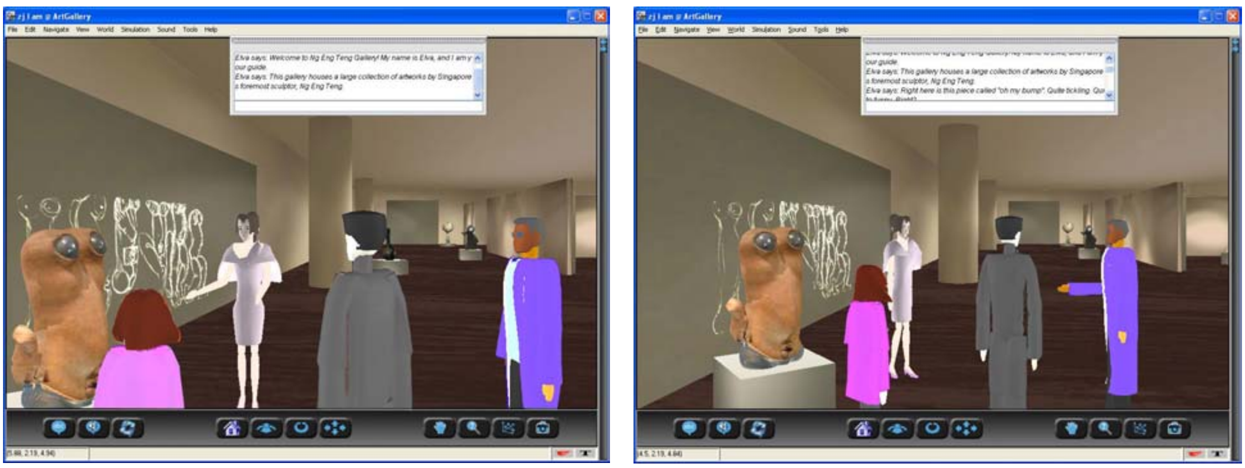

Figure 6: (a) Elva, second from left, is speaking to users. (b) User, first from right, is speaking.

21. Elva: To me, it is a human body with the head and limbs omitted, namely a "torso".
22. Elva: However, looking from the front, you can also see a mouth here, stating that this is a face. And we see the eyes and nose as well.
23. Elva: This is Eng Teng's creative concept of "torso-to-face".
24. Joyce: What is it made of then?

<Elva turns to look at Joyce.>

25. Elva: This sculpture is made of clay.

<Elva turns to look at other people.>

Short Silence.

Elva continues to describe the concept of "torso-to-face".

This episode can be segmented into three parts. From the utterance 1 to 7, the *narration mode* is evident. Elva, the virtual agent, dominated the floor and gave out relevant information based on her expertise in the gallery domain. A mode shift was detected after the utterance 10. Therefore, the conversation went into *discussion mode* whereby Elva acted as a facilitator and floor was open to other users. User's actions and responses were recorded into state repository and each user's rating was generated based on his activeness. Like in the utterance 14, Elva provided some hint to guide the conversation flow. At the utterance 18, Elva identified that user Alan's user rating was quite low. Then Elva encouraged him to speak up by assigning a turn to him. At the utterance 20, conversation went back to the *narration mode*. At the utterance 24, a Q&A dialogue took place. After a short silence, Elva took initiative, and maintained the conversation in the *narration mode*. Furthermore, in the utterance 8 and 9, Elva identified a new visitor who is within the group boundary. Then Elva took the initiative to greet the user to a ceremony to welcome the visitor to join the group.

In the evaluation of Elva's capability in guiding a group, first, we focused on agent's ability to handle visitors with different degrees of activeness. Second, we studied the agent's performance when interacting with the groups with varying sizes. Ten groups of subjects participated in the empirical study. We collected subject's feedback via evaluation forms, which consisted of open questions on their overall experience as well as assessment on Elva's verbal and nonverbal behaviors using a 5-point scale. The feedback shows the majority of the subjects agreed that Elva's behavior was natural and they did not identify obvious violation of Elva's behaviors with social norms. The study also indicated that Elva has maintained a balance in discussion by inviting less outspoken party to join the conversation and discouraging "talkative" parties from dominating the discussion. In groups of larger size, the feedback shows that Elva played an effective role in controlling floor and facilitating the smooth conduct of the tour. Non-verbal behavior wise, the study revealed that Elva's eye contact and body language were appropriate and comprehensive. The majority of subjects agreed that they had a sense of involvement in the tour. Overall speaking, the evaluation indicates a relatively high level of agent autonomy and believability in group interaction.

6. CONCLUSION

The Integration of multiparty support into ECAs holds promise of creating natural and realistic group interaction in virtual environments. In this paper, we have described an integrated approach for modeling multiparty interaction of agent in a tour group context. In our layered agent architecture, group interaction layer enables the modeling of group processes through three stages, namely appraisal stage, moderation stage and expression stage. The group modeler utilizes the casual social group norms to determine the group boundary in virtual environment and formulate group data and event. Leadership of the agent is exhibited through turn taking moderation and multimodal moderation. The present development and evaluation of group interaction technology form the basis of our future research in this domain.

7. REFERENCES

[1] Yuan, X. & Chee, Y. S., 2003, Elva: an embodied tour guide in an interactive virtual art gallery, In Proceedings of *CyberWorld2003*, pp. 432-439.

[2] Zambonelli F., Jennings N.R., Wooldridge M., 2003, Developing multiagent systems: The Gaia methodology, *ACM Transactions on Software Engineering and Methodology*, Vol. 12, Issue 3, pp.317 – 370.

[3] Jung H., Tambe M., 2003, Performance models for large scale multiagent systems: using distributed POMDP building blocks, In Proceedings of *AAMAS2003,* pp. 297 – 304.

[4] Tambe, M. 1997. Towards Flexible Teamwork*, Journal of Artificial Intelligence Research*, Vol. 7, pp. 83-124.

[5] Parker, L. E., 1998. ALLIANCE: An architecture for fault tolerant multirobot cooperation. In *IEEE Transactions on Robotics and Automation*, 14(2).

[6] Musse, S. R. & Thalmann, D. 2001, IEEE Transactions on Visualization and Computer Graphics, Volume 7, Issue 2.

[7] Tubbs, Stewart L, 1992, A systems approach to small group interaction. 4th ed. McGraw-Hill Inc.

[8] James, John, 1951, A preliminary study of the size determinant in small group interaction, *American Sociological Review*, Vol. 16, No. 4, pp. 474-477.

[9] Cassell, J., Bickmore, T., Camphell, L., Vilhjalmsson, H., and Yan, H. 2000, *The human conversation as a system framework: designing embodied conversational agents*. In Cassell, J., Prevost, S., Sullivan, J., and Churchill, E. (ed), *Embodied Conversational Agents*. MIT Press, Cambridge, MA, pp. 29-63.

[10] Goodwin, C. 1981 *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

[11] Moreland, J.K. and Levine, J.M., 1982. Socialization in small groups: temporal changes in individual-group relations, in Berkowitz, L. (ed.) *Advances in Experimental Social Psychology*, Vol. 15. New York: Academic Press.

[12] Chee, Y.S. & Hooi, C.M., 2002, C-VISions: Socialized learning through collaborative, virtual, interactive simulations. In Proceedings of *CSCL2002*, pp. 687-696.