A Realistic Chat Environment for Virtual Avatars in Cyber Space

Soo-Hyun Park, Seung-Hyun Ji, Dong-Sung Ryu and Hwan-Gue Cho
Dept. of Computer Science and Engineering
Pusan National University, Republic of Korea
{shpark, shji, dsryu}@pearl.cs.pusan.ac.kr, hgcho@pusan.ac.kr

Abstract

Chat is a crucial function in current virtual worlds. Since virtual worlds are becoming increasingly popular, we need a realistic and efficient communication framework for multi-agents participating in a virtual world. This paper proposes a unified communication framework for multiple avatars in a virtual world, (e.g., Second-Life). Current chat systems in virtual worlds provide a set of special communication protocols such as pairwise, group and secret chat, independently. Therefore, modern virtual communication systems have adapted artificial chat techniques such as ‘One-to-One Chat’ and ‘Group Chat’ to a virtual 3-D space. Our framework exploits spatial relationships (distance, and viewing vector) between several agents, without including a specialized chat protocol layer in communication. The main contribution of our work is threefold. First, we propose a realistic and unified communication framework which enables ‘Complete Chat’ and ‘Partial Chat’ in terms of spatial relationships between avatar agents without including an additional communication protocol. Second, we propose a single greedy algorithm to find an optimal position in 3-D virtual space which optimizes chat availability. Third, our system reconstructs a dialogue graph which maintains all transcripts in the form of directed graphs with temporal(dialogue sequences) and spatial information(physical positions) about communicating agents.

1. Motivation and Aims

Computing technologies are extending the means by which people communicate. As the popularity of Internet services grows, another form of communication - electronic chat rooms - has been introduced [6]. Although these electronic chat programs originally supported only text-based communication, fast modems and network connectivity techniques have enabled developers of these chat programs to create a richer user experience via the inclusion of multimedia features including graphics, sound and gestures. Agents in these graphical chat rooms communicate not only by typing text, but also by changing gestures or facial expressions of an avatar [1,3,9]. Especially, virtual worlds have recently become successful, due to rapidly improving Information Technologies (ITs). So, recent Internet technology enables us to participate in virtual worlds, Second Life is representative of successful virtual worlds [7].

Today, the success and momentum of virtual worlds is undeniable. The market for Massively Multiplayer Online Games(MMOGs) is estimated to be valued in excess of one billion US dollars [14]. However, these virtual world systems have several problems and limitations of communication in their virtual space. They use only 2-D dialog boxes and word balloons, even though they provide a 3-D virtual space. To overcome these problems, modern virtual world systems support specialized chat techniques(Internet protocols) such as ‘One-to-One Chat’ and ‘Group Chat’. However we consider that these kinds of communication frameworks are quite different from realistic chat.

We propose a realistic smart virtual communication system for avatar agents, using spatial relationships such as distance and direction between agents. Our system is based on a very simple and robust framework, including information about spatial relationships of agents in virtual worlds. We do not provide any specialized communication protocol or Internet layer similar to those used in common chat tools. In our system, if agent, $A_p$, wants to join a small talk group, then he/she must approach the group by himself/herself. As $A_p$ approaches the chat group, he/she hears the chat dialogues produced by small group more clearly.

The current chat system in virtual space generally provides a 2-dimensional flat dialog box, including chat text. Or, in order to construct a chat for two particular people, another communication protocol must be applied, in the form of an Internet application layer such as an Instant Messenger(e.g. MSN). These kinds of chat layers increase In-
ternet computational load, which makes virtual communication systems difficult to construct and very error-prone. We consider that current communication systems in virtual worlds are quite different to real-world communication. In the real-world, we can perceive, but do not hear in detail, all dialogues of agents in a wide hall. This kind of “Partial Chat” is not possible in the current communication systems in virtual worlds.

Figure 1 shows a snapshot of our system. Eight agents, $A_i$, are in chat in a conference hall. We can see two groups. In the group in the foreground (comprising three agents), my agent (wearing a green shirt) can hear the two agents opposite clearly, since they are close by and within hearing distance. We recognize that five agents in the background are chat, but we cannot completely distinguish the dialogue, since they are distant from my avatar.

![Figure 1. System snapshot. In this environment, My agent (wearing a green shirt) can clearly hear two agents opposite, but can only recognize that a distant group (Five agents) are in chat, and cannot completely distinguish the dialogue, due to distance constraints.](image)

In our system, participating agents can hear clearly the dialogue of agents nearby. The extent to which dialogue is heard is dependent on the distance between my agent and the corresponding agents, which is quite realistic and natural in real-world space. This kind of natural constraint in virtual communication enables us to provide a simple and unified framework. So, the potential capacity of chat is only dependent on spatial relationships of multiple agents, instead of the system application layer. If someone wants to chat to more people, then that person must join a chat group more completely. If someone wants to chat to more people, then that person must speak more loudly. Our system transforms "speaking loudly" into a larger word balloon with a higher (vertical) position, which increases the visibility of word balloons to agents. So the core component of our system is limited to developing algorithms for transforming word balloons, according to an agent's chat dialogue (volume) and the means of placement of the constructed word balloon in the virtual space.

This paper is organized as follows. In Section 2, we survey and compare some previously released systems enabling virtual chat in 2-D or 3-D space. The preliminary section introduces basic definitions. Section 3 explains the means of constructing a 3-D word balloon by considering spatial information about agents and the means of placing them in a virtual space to optimize the chat environment.

2. Previous Work

Initially, Internet chat services only supported text-based communication. But, fast modems and network connectivity techniques have enabled developers of these chat programs to create a richer user experience via the inclusion of graphics or simple images. Today, advanced chat software enables us to participate in voice/image chat by utilizing web-cameras. This chat software and component tools are rapidly evolving to adapt to the complicated and user-friendly environments demanded by users (personal agents) in virtual worlds.

Text-based chat is the simplest means of communication, via text messages between people in the same chat room [11]. The oldest form of chat systems are mainly based on this text-based chat system. Freelancin' Roundtable was a modem-based text chat system circa 1980s [12]. Major IT companies such as Microsoft and Yahoo freely provide these chat services for both text and voice, simultaneously. These text-based chat systems are simple and convenient, as they can be used anywhere and anytime, for anything from ‘One-to-One chat’ to ‘Group chat’.

Fast modems and network connectivity techniques have enabled developers of chat programs to create a richer user experience via the inclusion of graphics or simple images. Especially, ‘Comic Chat’ is a stylized system that automatically represents on-line communications in the form of comics. ‘Comic Chat’ provides the means to generate numerous aspects of comics, including word balloons, characters, gestures and semantic panels. It has several advantages over other graphical chat programs, including a graphical history and a dynamic graphical presentation. Figure 2 shows the results of communication using ‘Comic Chat’.

The means of placement of the text body (word balloon) on the chat screen is an issue in these graphic-template based chat systems. Simple placement does not work, since all comic books have their own reading order in texts. So, ‘Comic Chat’ adopted a greedy algorithm for placement of bodies of word balloons, since word balloons are critical in ‘Comic Chat’. This algorithm is a kind of constructor, since once the algorithm selects the placements of word balloons, it never changes them. However, the word balloon tail placement is not greedy. Instead of a greedy algorithm, the algorithm delays placing the tails until all word balloon
Comic Chat automatically represents communication in the form of comics, including word balloons, characters, gestures and semantic panels. Three different balloon types, including speech balloons, thought balloons and whisper balloons, have been placed. Therefore ‘Comic Chat’ provides a number of template characters and their gestures. When an agent types some text, ‘Comic Chat’ uses this text to determine a default gesture and expression. Even though ‘Comic Chat’ is no longer available, its aims and ideas were significant, since it provided a graphical form of a dialogue sequence such as Question & Answer.

Some chat systems for virtual space utilize a 3-D graphical representation of avatar agents that can freely move and act in a virtual space, in a manner similar to a real human. These virtual systems dominate communication, education and medical research etc. Especially, virtual space in multiplayer online games is quite realistic. We note the great success of virtual world ‘Second Life’ and ‘IMVU’. Figure 3 shows a snapshot of the chat environment of ‘Second Life’ and ‘IMVU’. All word balloons are orthogonally visible to a viewer. All dialogue texts can be read by any avatar agent without regard to spatial constraints(distance). A special feature of ‘Second Life’ and ‘IMVU’ is a ‘Shout Word Balloon’. If a user applies the ‘Shout Word Balloon’ feature, messages are sent farther and are visible to more people. In ‘IMVU’ if a sentence contains an exclamation mark, a ‘Shout Word Balloon’ is automatically generated.

The main goal of this paper is to provide a means of constructing a realistic chat environment in virtual space, by enabling ‘Partial Chat’ with novel dialogue tracking algorithms. For example, in [4], we can hardly track a sequence of interactive dialogue(Question & Answer). In our system, a dialogue(Question & Answer) among avatars is traced and stored in an elegant directed graph structure, by exploiting spatial information about each avatar (the physical location and viewing vector) with temporal information about transcripts.

3. Preliminary

In the following sections, we consider a virtual hall consisting of two small sub-spaces separated by a door. Seven virtual agents (denoted as $A, B, \ldots, G$) are strolling in the hall. Figure 4 shows a top view, including seven avatars. Group $G_1$ consists of Agent $A$ and $B$. Group $G_2$ consists of Agent $D, E, F$ and $G$. Chat Agent $C$ at the passage gate is looking at Group $G_2$ and Group $G_1$ in Figure 4(b) and (d), respectively. Figure 4 shows a visibility graph of the chat room and the resulting scene.

When an avatar starts to chat when an agent types a dialogue text), the corresponding word balloon is created and placed above the agent, which is orthogonal to the viewing vector of the chat agent. That implies that each avatar always has his own word balloon above his head in the virtual space. We provide some basic notation and definitions.

Let $A_a$ and $A_b$ be virtual agents and $\mathbf{N}_x$ be the normal(viewing) vector of $A_x$. Also, let $W_a$ and $W_b$ be the
word balloons of $A_a$ and $A_b$, respectively, which are perpendicular to the viewing vector $\vec{N}_a$ and $\vec{N}_b$. And, let $l_{a,b}$ denote the distance between $A_a$ and $A_b$, which equals the distance between the centers of $W_a$ and $W_b$. And, $\theta_a(\theta_b)$ denotes the right-hand angle between $l_{a,b}$ and $W_a(W_b)$.

**Definition 3.1** For two chat agents $A_a$ and $A_b$, we define $VCB$ (Virtual Chat Bandwidth) as the communication capacity bandwidth, which is computed as follows. If $(\vec{N}_a \cdot \vec{N}_b \geq 0)$, then $VCB(A_a, A_b) = 0$. Otherwise let,

$$VCB(A_a, A_b) = C_1 \cdot \frac{(\sin \theta_a \cdot \sin \theta_b)k_1}{\text{dist}(A_a, A_b)} + C_2 k_2 + C_3$$

, where $C_1, C_2, k_1$ and $k_2$ are control constants. □

It is very clear that $VCB(A_a, A_b)$ is maximized if and only if $\theta_a, \theta_b = \pi/2$ and $VCB(A_a, A_b) = 0$ if and only if $\theta_a, \theta_b \geq \pi$. This implies that the higher $VCB(A_a, A_b)$, the greater the extent to which two agents $A_a$ chats with $A_b$ are in clear and complete chat, and visa versa.

**Definition 3.2** In a virtual chat, we know agent $A_i$ spoke some “MESSAGE” by time stamp $= t_j$. This condition is denoted as \[ \text{Talk}(A_i, t_j) = \text{“MESSAGE”} \] □

**Definition 3.3** For a given set of Talks, $\{\text{Talk}(A_i, t_j)\}$, A directed acyclic graph, $CFG$(Chat Flow Graph)$,(V,G)$ is defined as follows. $v_i \in CFG(V)$ corresponds to agent $A_i$. Let us consider two Talks, $\text{Talk}(A_i, t_k) \text{Talk}(A_j, t_{k+1})$. If $\vec{N}_a \cdot \vec{N}_b \geq 0$ and $VCB(A_i, A_j) > \delta_0$ and $|t_{k+1} - t_k| < \epsilon_0$, then we define an edge $(A_i, A_j) \in CFG(E)$. Otherwise we do not set an edge connection. □

4. Optimal Placement of 3D-Word Balloon

4.1 Evaluating Chat Visibility

In our system, it is crucial to simulate a ‘Partial Talk’ according to the $VCB$ between several word balloons. Previous chat systems utilized artificial chat techniques, such

![Figure 5](image1.png)

**Figure 5.** Avatars $A_a$ and $A_b$ are in chat. The solid line segment denotes the width of word balloon $W_a$ and $W_b$.

The text printed in each balloon in viewed differently according to viewing vectors and distances. We can see the chat text clearly at a distance of 0.04(a virtual space metric) in Figure-6(a), but we can not see clearly at a distance 0.38 in Figure-6(b). For agents who are not looking at each other, sentences are represented as ‘Partial Chat’. Figure-7 illustrates these clearly.

As shown in Figure-7, each agent has his/her own particular view point according to his/her viewing vectors and distances between agents. Figure-7(e) is a scene of view point $D$. Agent $D$ is connected to $E$, $F$, $G$ in Figure-4, so $D$ can see the word balloons of $E$, $F$, $G$. Figure-7(e), (d) are scenes from $C$’s view point. Agent $C$ is connected to $D$, $E$, $F$, $G$ at $time_t$ and is also connected to $A$, $B$ at $time_{t+1}$ in Figure-4. Thus $C$ can see chat dialogue of $D$, $E$, $F$, $G$ at $time_t$ and the word balloons of $A$, $B$ at $time_{t+1}$. But $C$ can partially see some balloons according to distances and viewing vectors. Also we note that $C$ can not ‘hear’ the chat dialogue of group $G_1$, $G_2$. $C$ only realizes that they are in chat.

4.2 3D-Word Balloon Placement

In our virtual chat system, an important problem is the position of the word balloon representing $\{\text{Talk}(A_i, t_j)\}$. A simple placement of a balloon at a fixed position above an agent’s head reduces the readability of the balloon text, because one balloon can obscure another balloon. So, the balloon placement algorithm must maximize total

![Figure 6](image2.png)

**Figure 6.** Readability of chat dialogue. This is primarily dependent on distance.
5. Chat Flow Graph Construction

Maintaining transcripts is crucial in chat, since agents commonly interleave several tasks during chat. However, most chat programs provide only text transcripts without any explicit relation tags among them. For multi agent chat, previous systems only maintain text transcripts according to the temporal sequence. For example, based on the text maintained, we can not easily find a question and its corresponding answer. To address this problem, we propose a new Chat Flow Graph (CFG) which is automatically reconstructed in chat space. Algorithm 2 demonstrates the means of reconstructing a Chat Flow Graph (CFG) from the set of Talk($A_i, t_j$). Note that CFG must be a directed acyclic graph.

**Algorithm 1 Total VCB($A_i, A_j$) Calculation Algorithm**

For all $A_i, A_j \in AGENTS$, calculate $\sum_{A_i, A_j \in S} VCB(A_i, A_j)$

if $visible(A_i, A_j) = False$ then $VCB(A_i, A_j) = 0$

else

if $\theta_{a}, \theta_{b} \geq \pi$ then $VCB(A_i, A_j) = 0$

else $VCB(A_i, A_j) = C_1 \cdot \frac{(\sin(\theta_{a} - \sin \theta_{b})^{4}}{dist(A_i, A_j)^{2} + C_2}$

**Algorithm 2 Chat Flow Graph Construction Algorithm**

Procedure for constructing CFG from Talk($A_i, t_i$)

$V_i \leftarrow A_i$;
$N_i \leftarrow$ normal vector of $A_i$;
$T_i \leftarrow t_i$;
$E \leftarrow \{(v_i, v_j) | v_i, v_j \in V, i \neq j\}$;

for each edge $e \in E$ do

if $visible(v_i, v_j) = TRUE$ &&

$\alpha < T_j - T_i < \beta$ && $\gamma < N_i \cdot N_j < \delta$

then $E(v_i, v_j) = TRUE$

end procedure

Chat Flow Graph (CFG) edges are determined by differences between time stamps and the range of vision according to Time with parameters $\alpha, \beta$ and View with parameters $\gamma, \delta$. Table 1 shows an example of a Chat list and Figure 9 shows a Chat Flow Graph constructed using Table 1. We assume that no chats have the same time stamp.
Table 1. Talk List Construction.

<table>
<thead>
<tr>
<th>Time</th>
<th>Agent</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>&quot;Where will the conference be held?&quot;</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>&quot;It is Hangzhou.&quot;</td>
</tr>
<tr>
<td>8</td>
<td>D</td>
<td>&quot;It is Suzhou. isn’t it?&quot;</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>&quot;What did you have for lunch?&quot;</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>&quot;No! It is Hangzhou. Hey TJ! Are you ready to prepare about the conference?&quot;</td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>&quot;I ate Bulgogi for lunch.&quot;</td>
</tr>
<tr>
<td>17</td>
<td>D</td>
<td>&quot;So so.&quot;</td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>&quot;Was the food delicious?&quot;</td>
</tr>
<tr>
<td>22</td>
<td>B</td>
<td>&quot;Yes! very nice.&quot;</td>
</tr>
</tbody>
</table>

Figure 9. The corresponding CFG(Chat Flow Graph) of Chat shown in Table-3.

There is no ordering or sequencing of text transcripts in current chat systems. Most text transcripts in current chat systems merely consist of a list of talk according to time stamp. However, we maintain a Chat Flow Graph(CFG) in stead of text transcripts and analyze the order of chats and relationships between chats. This enables agents to understand the context of a chat.

6. Conclusion and Further Work

Previous chat systems utilized artificial chat techniques, such as One-to-One Chat and Group Chat using Complete Chat, that is either completely visible or invisible. In addition, they provide only text transcripts without any explicit relationships tags between them. Instead, We proposed a realistic and unified communication framework which enables Complete Chat and Partial Chat in terms of spatial relationships between agents without including an additional communication protocol. And, Our system reconstructs a dialogue graph which maintains all text transcripts in the form of a directed graph with temporal(dialogue sequences) and spatial information(physical positions) of communicating agents.

Currently this system has numerous limitations and drawbacks, which will be studied in future work.

- We need various types of word balloons which are common in comic books. That includes a ‘Shout Balloon’ and a ‘Thought Balloon’.

- It is important to handle sounds from sources beyond the range of view, such as calls from behind and sounds from speakers. This will improve the realism of the virtual world.

References