Autonomic gaze control of avatars using voice information in virtual space voice chat system

Kinya Fujita, Toshimitsu Miyajima and Takashi Shimoji

Tokyo University of Agriculture and Technology
2-24-16 Nakacho, Koganei , 184-8588, Tokyo, Japan
kfujita@ cc.tuat.ac.jp

Abstract

Avatars play an important role for the embodiment of the users in virtual space communication systems. However, the conventional systems to realize the gaze have required camera-based precise eye tracking and the numbers of the users have been restricted. This paper proposes a simple substitution for the gaze control in a virtual space voice chat system. The method is to control the gaze target of the avatar based on the Appeal Point that is calculated from the voice levels of the other users. The subjective evaluation experiment demonstrated the effectiveness of the method on the naturalness of the virtual space communication.

1 Introduction

Rapid growing of Internet and three-dimensional computer graphics technology made multi-user communication systems employing shared virtual space inexpensive and popular personal application software. DIVE (Carlsson 1993) is one of the early-developed systems that provide shared virtual space for distributed users. After that, numbers of distributed multi-user virtual space systems, that provide text or voice chat functions, have been developed, such as Massive (Greenhalgh 1995), FreeWalk (Nakanishi 1996), Community Place (Lea 1997) and so on. Some of these systems are mainly focused on to provide casual chat function, not formal meeting, for distributed users. The authors also developed a virtual space communication system (Fujita 2003), which enables the users walk around in the space using walk-in-place locomotion interface devices (Fujita 2004) and casual voice chat. In the virtual space communication systems, the embodiment of the users (Bowers 1996) is an important issue for natural communication. Avatars that represent the remote users have been employed for the embodiment of the users in the distributed multi-user virtual space systems (Cassell 1999, Wray 1999).

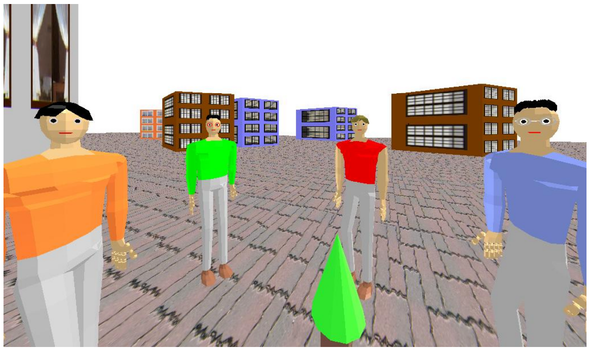

Figure 1: An example of the avatars looking at the user in the developed virtual space voice chat system.

However, the avatars for the embodiment of the users need to act adequately to provide the natural sense to the other users. In the real world, the nonverbal information such as gesture, facial expression and intonation is utilized for smooth human communication in addition to the verbal information. Gaze, as one of the most important nonverbal information, has essential functions to regulate the flow of the conversation, to provide the reaction and to send social signals (Argyle 1976). Several methods have been proposed to realize the gaze function in the multi-user videoconference systems. MAJIC attained the mutual gaze by placing the video projectors and cameras behind the screen (Okada 1994). Hydra (Sellen 1995) and GAZE (Vertegaal 1999) detected the user's actual gazes using a camera-based tracking system and the video images of the users were located in virtual space. These systems also enabled the users look each other. However, these systems require the cameras, the precise gaze point detection technology and broad bandwidth for real-time video streaming. The numbers of the users are also restricted. Therefore, we propose a simple substitution in multi-user virtual space voice chat system without additional devices to control the avatar gaze by a method based on the voice information and spatial relationship of the users in virtual space.

2 Method

The gaze control problem is divided into two problems, the physiological eye movement and the gaze target (target avatar) selection. For the former problem, the human eye movement is being statistically analysed in order to synthesize the physiologically adequate avatar eye movement (Lee 2002, Bilvi 2003). This study is focused on the later problem. When we observe the conversation among several users in the real world, the audience tends to look at the speaker. That is the natural feedback of listening. Moreover, if another person starts to speak while someone is speaking, the person who starts speaking later tends to attract the attention of other persons. In this study, the former is called speaker effect and the latter is called starting effect. These effects were defined as Appeal Point (AP) as an index of attention attraction, which is computed from each user’s voice level and duration. The gaze target avatar was chosen as the avatar that has the highest appeal point.

2.1 Speaker effect

Basically, the strength of gaze attraction seems to be affected by the loudness and the length of the speaking duration, because a frequent speaker has higher probability to speak again. Therefore, the Appeal Point generated by the speaker effect: APc was defined as the sixty seconds integral of the logarithmic voice level. The voice level was exponentially weighted by the past time to give the current voice level higher priority than the past. The parameter values in the equation were experimentally decided.

$$AP_C = \int_{-60}^{0} \log(v(t)) \exp(\frac{t}{60} t) dt \qquad (1)$$

2.2 Starting effect

Although APc realizes the gaze at the speaker, the integral in the calculation potentially raises a problem that the current speaking user may not attract the gaze. Moreover, the higher priority of the overriding speaker is also to be realized. Therefore, the starting effect Appeal Point APs was defined. APs was given a constant value at the onset of the speaking and decreased linearly in five seconds, in order to give instantaneous character. In this AP calculation, additional restriction was applied that APs will not be generated if the silent duration is less than 5 seconds, in order to avoid the misjudgement of the natural intermittence of the voice as the speaking start.

$$AP_S = \frac{5-(t-t_s)}{5} \qquad (2)$$

The gaze at the avatar who has highest AP, which is the summation of APc and APs, provides a gaze control function, however, the speaker is continuously gazed until another speaker starts speaking. The gaze target user was randomly changed once in about 30 seconds in order to avoid the unnaturalness or unconsciousness by continuous gaze..

$$AP = aAP_C + bAP_S \qquad (3)$$

Figure 2 represents the conceptual diagram of the speaker effect Appeal Point APc function. The integral calculation smoothly interpolates the intermittence of the voice as duration I and attains the continuous gaze at the last speaker as duration II. However, the integral calculation delays the speaker change as duration III. The starting effect compensates this delay in addition to the later starting speaker’s Appeal Point enhancing effect.

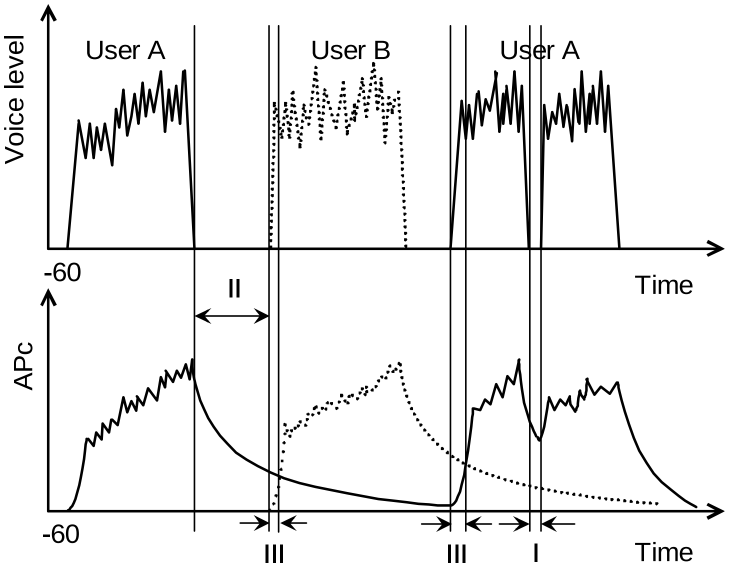

Figure 2: The conceptual diagram of the speaking effect APc.

The gaze at the user who has highest AP, which is the summation of APc and APs, provides a gaze control function, however, the speaker is continuously gazed until another speaker starts speaking. The gaze target user was randomly changed once in about 30 seconds in order to avoid the unnaturalness or unconsciousness by continuous gaze.

2.3 Priority enhancement of local user

In general, being gazed is expected to enhance the feeling of being listened. The enhancement of the head turning probability of the avatars to the local user may give more impression being listened. Therefore, the higher priority was given to the local user by changing the generation restriction condition of the starting effect APs from 30s silent to 5s silent only for the local user. By applying this local user priority enhancement, the avatars of the distant users more easily turns their heads to the local user.

3 Experimental evaluation

The subjective evaluation of the gaze control was performed in order to examine the effectiveness of the speaker effect, the starting effect and the randomness on smooth and natural communication. Ten university students were divided into two groups and one group of subjects participated in the experiment at a time. Each avatar of the experimental subjects was located a place, where allows the user look at other 4 avatars, in the virtual space. The subjects were requested to talk with the other users about a theme given by the experimenter for 5 minutes about a daily-life subjects, such as sport, culture, study and so on. The gaze control conditions are the combinations of the speaker effect, starting effect and random gaze, as shown in table 1.

Table 1: Gaze control conditions in the subjective evaluation experiment.

No control (fixed gaze)
Random
Speaker
Speaker + Starting
Speaker + Random
Speaker + Starting + Random

As seen in figure 3, the subjective scores of six conditions in ordinary scale were 0, 18, 37, 39, 24, 32 respectively. The result shows that all conditions with gaze control, including random condition, gave more natural sense than the fixed gaze. The score of the random condition was the second lowest in the experiment. It appears that the gaze control provided the avatar autonomic action and the autonomy caused the user more natural impression, even if the change of the gaze is random. The four conditions with gaze control using AP were obviously more natural than the other two conditions. It was demonstrated that the avatar gaze control based on the speaking state makes users feel more natural in conversation as expected.

The scores of the two conditions that have randomness in addition to the gaze control with AP were lower than the conditions without randomness. This is mainly attributed to the unexpected change of gaze target while the local user is speaking. On the other hand, two conditions with the starting effect provided more natural impression than those without the starting effect. The higher subjective score is attributed to the cognitive assistance effect in recognition of a speaker, because the starting effect reduces the gaze control latency.

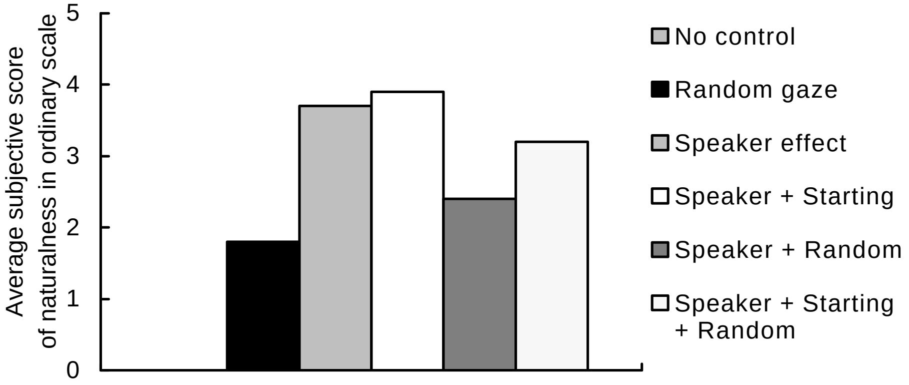

Figure 3: The effect of the voice-based gaze control with the combinations of the various effects.

Similar experiment to the previous one was performed to verify the effect of the priority enhancement of the local user. The experimental conditions and the subjective score of each condition are shown in table2 and figure 4.

Table 2: Local user priority control conditions in avatar gaze control.

No starting effect
Even priority (30s silent for all) in starting effect
Even priority (5s silent for all) in starting effect
Local user priority enhanced (5s silent condition only for local user)

As seen in figure 4, the scores of the three conditions with starting effect were higher than it of the condition without starting effect as observed in the previous experiment. The score of the 5s silent condition was lower than it of 30s silent condition. It appears that the increase of the head turning probability of the avatars from the local user to others affected in 5s silent condition. The score of the local users priority enhancing condition was slightly higher than the both even priority conditions. It appears that the enhancement of the local user priority enhanced the feeling of being listened.

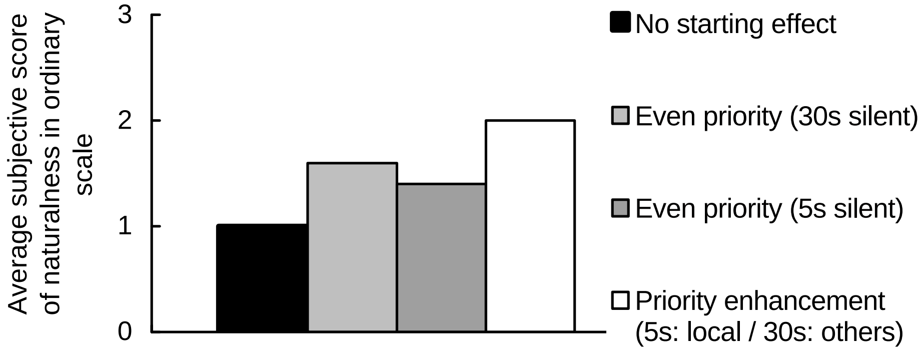

Figure 4: The effect of the priority control in the voice-based gaze control.

4 Conclusions

An avatar gaze control algorithm using Appeal Point, which is calculated using the voice information, was proposed for shared virtual space voice chat system. The effectiveness of the gaze control with the speaking and the staring effects for natural communication was experimentally demonstrated.

Acknowledgement

This work was partially supported by the program "The R&D support scheme for funding selected IT proposals" of the Ministry of Public Management, HomeAffairs, Posts and Telecommunications.

References

Carlsson, C. & Hagsand, O. (1993). DIVE - A platform for multi user virtual environments. Computer and Graphics, 17(6), 663-669.

Greenhalgh, C. & Benford, S. (1995). Massive: A Collaborative Virtual Environment for Teleconferencing, ACM Trans. on Computer-Human Interaction, 2(3), 239-261.

Nakanishi, H., Yoshida, C., Nishimura, T. & T. Ishida, (1996). FreeWalk: Supporting Casual Meetings in a Network, in Proc. CSCW’96, 308-314.

Lea, R., Honda, Y., Matsuda, K. & Matsuda, S. (1997). Community Place: Architecture and Performance, in Proc. VRML’97, 41-50.

Fujita, K. & Shimoji, T. (2003). Walkable shared virtual space with avatar animation for remote communication, in Proc. HCI International 2003, 493-497.

Fujita, K. (2004). Wearable Locomotion Interface using Walk-in-Place in Real Space (WARP) for Distributed Multi-user Walk-through Application, in Proc. IEEE VR2004 Workshop, 29-30.

Bowers, J., Pycock, J. & O'Brien, J. (1996). Talk and Embodiment in Collaborative Virtual Environments, in Proc. CHI’96, 58-65.

Cassell, J. & Vilhjalmsson, H., (1999). Fully embodied conversational avatars: making communicative behaviors autonomous, Autonomous Agents and Multi-Agent Systems, 2(1), 45-64.

Wray, M. & Belrose,V. (1999). Avatars in Living Space, in Proc VRML’99, 13-19.

Argyle, M. & Cook, M. (1976). Gaze and Mutual Gaze, London: Cambridge University Press.

Okada, K., Maeda, F., Ichikawa, Y. & Matsushita, Y. (1994). Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design, in Proc. CSCW’94, 385-393.

Sellen, A. J. (1995). Remote conversations: the effects of mediating talk with technology, Human Computer Interaction, 10(4), 401-444.

Vertegaal, R. (1999). The GAZE GroupWare System : Mediating Joint Attention in Multiparty Communication and Collaboration, in Proc. CHI'99, 294-301.

Lee, P.S., Badler, B. J. & Badler, I.N. (2002). Eyes Alive, ACM Trans. Graphics, 21(3), 637-644.

Bilvi, M. & Pelachaud,C. (2003). Communicative and Statistical Eye Gaze Predictions, in Proc. AAMAS, 2003.