Abstract

This thesis attempts to design an interface that facilitates conversation using human gaze behavior in order to facilitate smooth remote computer-based human communication and human-agent communication as well as human face-to-face communication.

In human face-to-face communication, nonverbal behaviors have an important function. It is known that gaze behavior, in particular, plays an important role in transferring the speaker's turn, encouraging speech, and conveying attitudes and interests. At the same time, progress in computers and information technology has led to the widespread use of remote communication wherein conversation takes place through transmitted voice and video between parties at distant locations. Progress in artificial intelligence has realized conversational agents with the body of a human in which conversation functionality has been incorporated. The smooth, natural communication found in face-to-face communication is not possible in these communication systems. The main reason for this disadvantage is that nonverbal communication, particularly the gaze behavior, cannot be sufficiently transmitted. Specifically, gaze cannot be correctly transmitted through a video image in remote communication, and the receiver cannot read the attitude or interests of the sender, which is usually conveyed by a person's gaze behavior, through a conversational agent. As this shows, remote communication and conversational agent systems do not have a functionality to actively transmit and recognize gaze behavior; thereby, natural communication such as that found in face-to-face communication has not been achieved.

In this thesis, we work with the structure of the advanced conversational interfaces that allows the human gaze behavior to be transmitted and the attitudes expressed by that activity to be understood in order to facilitate conversation in voice chat, video communication and conversational agent system.

First, focusing on the conversation promotion and speech-right transfer functions of gaze in multiple-user voice chat system, we propose a method for automatically controlling the gaze behavior of an avatar based on voice sound to encourage the user to speak. As human gaze behavior varies depending on the speaking state of the user, we model the human gaze behavior in three states: speaking, right-after-speaking, and silence. We structure a conversational interface that allows the avatar gaze behavior to be automatically controlled based on user speech information. We perform evaluation experiments on participants to determine whether the avatar gaze behavior facilitated speech. The results suggest that avatar gaze behavior encourages users to speak.

Second, we propose an advanced video display method that would correctly transmit

gaze and pointing gesture in a video communication system. An issue in video communication is that although user gaze can be displayed using video images, camera placement and 2D video display make it impossible to correctly convey the gaze direction. We propose MoPaCo, a window interface that displays video images as if the display were a window connecting two spaces, in an attempt to realize smooth communication through the correct transmission of the gaze direction. We conduct experiments to determine whether the gaze was transmitted correctly in MoPaCo. Results indicate that the system allowed gaze to be transmitted similar to that found in face-to-face communication. We perform evaluation experiments involving remote collaboration work in order to determine whether correct gaze transmission has effects on the smoothness of communication. Results suggest that users could refer to target objects smoothly in MoPaCo, as if they were speaking in a face-to-face manner through an actual window, and that it was possible for conversation partners to predict the next target. Subjective assessments indicate that MoPaCo encouraged smooth conversation and communication, and strengthened memories of the conversation, demonstrating that the system contributes to conversation quality.

Third, in face-to-face conversations, the speakers are continuously checking whether the listener is engaged in the conversation and change their conversational strategy if the listener is not fully engaged in the conversation. With the goal of building a conversational agent that can adaptively control conversations, we analyze listener gaze behaviors and develop a method for estimating whether he or she is engaged in the conversation based on those behaviors. We collect measured data on the user gaze behavior, user introspection and observation of others related to decreased interest in a conversation, and speech information using the Wizard of Oz method. We analyze user gaze data to determine whether the user shows active conversational engagement and to identify gaze transition patterns associated with active conversational engagement. In addition, we propose a model for performing the estimation of conversational engagement using gaze data obtained in real-time during a conversation with an agent. Evaluation experiment results indicated an estimation precision of F-measure above 0.70.

We implement a system to calculate conversational engagement in real time by using the estimation model to create a conversational agent adjustable to varying attitudes. Evaluation experiments demonstrate that the conversational agent capable of varying its behavior depending on the user's conversational engagement had an effect on interactions.

We design a new conversation promotion interface using the gaze behavior in three

conversational scenes in order to verify the conversation promotion effect. We believe that the results of this study provide a guideline for actively taking advantage of the gaze behavior in the design of future conversation systems. We also believe that our findings, in which the gaze function implied in the field of psychology so far have been practically verified, will contribute to resolution of the problems in the field of psychology.