

(続紙 1)

京都大学	博士（情報学）	氏名	石井亮
論文題目	Designing Conversational Interfaces for Facilitating Conversation using User's Gaze Behaviors (人間の視線行動を利用した会話促進インタフェースのデザイン)		
(論文内容の要旨)			
This thesis addresses interface design that can exploit gaze behaviors to facilitate smooth conversation between humans and agents. Both face-to-face interaction and computer-mediated remote interactions such as voice chat, video conference and human-agent interaction, are taken into consideration. The thesis consists of eight chapters. Chapter 1 provides an introduction to the thesis. It begins with the background, motivation and research objective, followed by a list of design policies of the conversational interface aimed at the thesis. The outline of the thesis is shown in the final section. Chapter 2 reviews major previous work that is relevant to the incorporation of gaze behaviors into conversational interfaces. The role of gaze behavior is discussed in general based on the literature study in nonverbal communication. Previous work on gaze behaviors in the conversational interface is reviewed from the viewpoints of speech communication system, video communication system, and conversational agent system. Chapter 3 addresses controlling avatar's gaze behaviors based on the user's speaking states in multi-party avatar-based video conferencing. A probabilistic state transition model consisting of three states, i.e., during-utterance, right-after utterance and silence, is proposed for describing the gaze behaviors reported in the conversation analysis, such as the speaker's gaze at the next speaker in turn-taking. The results obtained from a series of evaluation experiments suggest that the proposed method allows the participants to recognize the avatar as an embodied agent of their communication partner so that utterances are motivated by the use of gaze control in avatars. Chapter 4 addresses full gaze awareness for enabling the users of a video teleconferencing system to effectively share references to virtual objects around themselves in local virtual spaces. A system called MoPaCo is proposed and implemented. MoPaCo realizes a widow metaphor connecting the local virtual places around participants by exploiting motion parallax. Experimental evaluation was conducted to investigate how much MoPaCo improves the communication with deictic nonverbal behaviors such as gaze and pointing gesture. The results suggest that precision of the indication actions performed through MoPaCo is close to that in a physical environment, regardless of the distance between the user and the display. In addition, the participants' predictive behaviors were observed, which suggests that the full gaze awareness was effectively supported by MoPaCo.			

Chapter 5 addresses estimating participants' engagement in human-agent interactions. The Wizard-of-Oz method was employed to build a multimodal conversation corpus. The collected data were annotated in terms of gaze transition pattern, mutual gaze occurrence and gaze duration. A collection of gaze transition 3-grams was constructed to analyze correlation between gaze transition and engagement. It was found that the distribution of gaze transition 3-grams, gaze pattern durations, eye move distances and pupil sizes have a strong correlation with human judgment of engagement/disengagement by observation. The engagement estimation model using tree learning achieved F-measure 0.854 and 0.704 for estimating engagement and disengagement, respectively.

Chapter 6 addresses improving the smoothness of human-agent communication using the engagement estimation method proposed in the previous chapter. An engagement-sensitive conversational agent system was implemented. It allows for detecting user's disengagement in real time. An experimental evaluation was conducted to compare the participants' responses to the agent's probing based on engagement estimation and periodic probing. It was found that the agent's probing based on engagement estimation significantly reduced the participants' disengagement compared with periodic probing, and in addition the participants' subjective evaluation of the conversation agent was significantly improved in terms of engagement awareness, appropriateness of behaviors, smoothness of conversation, favorability and perceived intelligence compared with periodic probing.

Chapter 7 discusses the implication of the proposed methods and presents future directions including introduction of sensing other communication modality to improve the engagement estimation.

Chapter 8 concludes the thesis by reinforcing the results achieved in the study.

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、注視情報を活用した会話インタフェースを実現することを目的とした、遠隔コミュニケーションにおけるユーザの視線行動の円滑な伝達と、人間とエージェントのインタラクションにおけるユーザの視線行動の解釈と会話行動への反映に関わる研究について述べている。主要な成果は次のとおりである。

1. アバタを用いた遠隔コミュニケーションにおいて音声情報のみからアバタの視線行動を制御する方法を提案し、実装と評価実験を行った。この方法は、発話終了直後のアバタのユーザへの視線行動が発話促進に寄与するという会話分析の知見に基づくものである。評価実験により、この方法により実際にユーザの発話が促進されることを確認した。

2. ユーザの視点位置に応じて2次元ディスプレイに投影された会話参加者の周囲空間表示を変化させることにより、注視への気づきを生成するという窓越しコミュニケーションのメタファーを用いた遠隔映像会話インタフェースMoPaCoの考案、実装、評価実験を行った。MoPaCoを用いることにより、会話における指示対象の同定が、単なる2次元表示を用いた場合よりも顕著に改善され、物理的な窓を用いた場合と同程度の円滑さで行われることを確認した。また、ユーザが対話相手の視線行動を察知し、移動したり、注意を向けたりするという予測行動が観察された。

3. 注視対象遷移パターン、注視時間、注視点の移動距離、瞳孔径がユーザの会話参加態度と高い相関があり、会話参加態度の推定に有用であることを実験により明らかにした。注視時間長を考慮することで会話参加態度の推定精度が向上すること、ユーザの会話参加態度が積極的であるとき注視点移動距離が小さくなり、瞳孔径が大きくなる傾向があること、などの知見を得た。これらを会話エージェントに組み込むことによりエージェントの印象が向上することを実験的に明らかにした。

以上、本論文は、会話インタフェースにおける視線利用について、音声情報に基づくアバタの視線行動の制御がユーザの発話の促進に寄与すること、遠隔映像コミュニケーションにおける窓越しコミュニケーションのメタファーの実現が注視への気づきに寄与すること、視線行動からの会話参加度の推定が可能であり、それに基づくモデルを会話エージェントに搭載することにより会話エージェントの印象が改善されることを実証したものであり、会話インタフェースの進歩に少なからぬ貢献をしている。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成25年8月23日に実施した論文内容とそれに関連した試問の結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。

要旨公開可能日： 26年 9月 23日以降。

本論文（第5章、第6章）に記載された内容が他に帰属する著作権を侵害することがないことの確認が完了次第公開する。