Reactive Animated Pedagogical Agents: Exploring Dyadic Gaze Interaction

(反応的ペダゴジカル・エージェント：視線インタラクションの研究)

Hanju Lee

Table of contents

Chapter 1. Introduction 5

1.1 Review of pedagogical agent research 5

1.2 Brief review 6

1.2.1 The Persona Effect Theory 6

1.2.2 The Split-Attention Effect Theory 7

1.2.3 The Modality Effect Theory 8

1.2.4 The Social Agency Theory 8

1.3 Reciprocal interaction in human learning 9

1.3.1 Temporal contingency 11

1.3.2 Gaze interaction 11

1.4 The purpose of this thesis 13

Chapter 2. Designing PAGI (Pedagogical Agent with Gaze Interaction) 14

2.1 Main functions and goals 14

2.2 Learning material selection 15

2.2.1 Technical specification 17

Chapter 4. Experiment 1: Persona Effect and Split Attention Effect 19

4.1 Introduction 19

4.2 Method 20

4.2.1 Pedagogical Agent 20

4.2.2 Participants 20

4.2.3 Apparatus 20
4.2.4 Procedures 21
4.3 Result 23
4.4 Discussion 24
Chapter 5. Experiment 2: Temporal Contingency Effect 25
5.1 Introduction 25
5.2 Method 26
5.2.1 Participants 26
5.2.2 Design 27
5.2.3 The Pedagogical Agent with Gaze interaction (PAGI) 27
5.2.4 Materials 28
5.2.5 Apparatus 30
5.2.6 Procedure 30
5.3 Result 31
5.4 Discussion 36
Chapter 6. Experiment 3: Anthropomorphic image of pedagogical agent and Temporal Contingency effect 40
6.1 Introduction 40
6.2 Method 42
6.2.1 Participants 42
6.3 Result 43
6.4 Discussion 45

Chapter 7. Experiment 4: Saliency of pedagogical agent and Temporal Contingency effect 46

7.1 Introduction 46

7.2 Method 46

7.2.1 Participants 47

7.3 Result 47

7.4 Discussion 50

Chapter 8. Experiment 5: Lead-in vs Follow-in Joint attention 54

8.1 Introduction 54

8.2 Method 56

8.2.1 Participants 56

8.2.2 Design 56

8.2.3 Joint Attention PAGI (JA-PAGI) 57

8.2.4 Materials 58

8.2.5 Apparatus 58

8.2.6 Procedure 59

8.3 Result 60

8.4 Discussion 63

Chapter 9. General Discussion 64

9.1 Summary of the findings 64

9.2 Does reactivity of pedagogical agent facilitate learning effect? 65

9.3 Social cue vs non-social cue in relation with individual difference: age difference and competence. 66

9.4 The merit of using virtual agents in psychological experiments 69

Chapter 10. Conclusion 70

Reference……..71

Acknowledgments 83

Chapter 1. Introduction

1.1 Review of pedagogical agent research

Pedagogical agents are characters presented on computer screen designed to facilitate learning in computer-based learning environment. The defining feature of pedagogical agents is their visual presence on screen (Heidig & Clarebout, 2011), but otherwise can be largely diverse. Pedagogical agents can be as simple as pictures of characters which communicate by on-screen texts, to as complex as 3-D computer graphics generated characters with voiced narrations (Heidig & Clarebout, 2011; Schroeder, Adesope, & Gilbert, 2013). Pedagogical agents can be animated or static, and animated ones are often called “animated pedagogical agents”.

The research of pedagogical agent began as an attempt to innovate computer-based learning. While advances in computer technology hold great promise for learning, they are yet to be utilized to their full potential. For instance, Moreno (2001) criticized that computers are often used as high-tech textbooks, which uses great amounts of on-screen text to convey information. Pedagogical agent was announced as a potential tool to innovate learning, by incorporating social aspect of learning environment by utilizing social cues, such as gestures, voice and gaze (Atkinson, 2002; Craig, Gholson, & Driscoll, 2002; Johnson, Rickel, Stiles, & Munro, 1998; J. C. Lester, Stone, & Stelling, 1999; Moreno et al., 2001).

In spite of high claims made by initial studies, the pedagogical agent research has not yet succeeded to yield decisive results. In their review, Heidig & Clarebout (2011) reported that majority of the studies yielded no difference in learning when comparing agent and control groups. More recent review concluded that the effect of pedagogical agents was small (Schroeder et al., 2013). Considering the large cost required for developing pedagogical agents, the results seem to disapprove the need for more pedagogical agent research. We should consider, however, that research of pedagogical agents is still in its early stage and we need more research to answer the question, "*Do pedagogical agents facilitate learning?*" There are still neglected features of pedagogical agents that require urgent attention. For example, temporal aspect of social cues has not been tested, despite the fact that features such as temporal contingency and temporal congruity is known to be crucial in human communication.

In this chapter, a review of pedagogical agent literature is provided. Then reciprocal aspects of human interaction and their relation to learning are explained. Finally, the purpose of this thesis is presented.

1.2 Brief review

Pedagogical agent research has been influenced by contemporary multimedia educational theories. This sub-chapter introduces pedagogical agent literature and its major theories.

1.2.1 The Persona Effect Theory

Lester et al. (1997) presented the pedagogical agent "Herman the Bug", a cartoon 2-D character that teaches structure of plants to middle school students. In the study, five types of

Herman, which differed in communication mode (e.g. fully expressive, muted), were employed. Participants acquired better scores in the post-test than in the pre-test, but no significant differences were found between condition groups; all groups scored equally better in the post-test, even the group which worked with the muted version of Herman. As most types of learning would result in better scores in post-test than pre-test, this result indicate that interacting with Herman may had minimal effect on learning. Lester, however, claimed that the mere physical presence of the pedagogical agent leads to better learning, and termed it "*the persona effect*". Soon after, the study was criticized for not including a control group but is still frequently cited (Heidig & Clarebout, 2011). Since then, conflicting evidences surrounding the persona effect have been found. Several groups have failed to confirm the Persona Effect (Andre, Rist, & Muller, 1999; R. E. Mayer, Dow, & Mayer, 2003; Moreno et al., 2001), while some found positive effect on learning, especially on learner's motivation (Baylor & Ryu, 2003; Dunsworth & Atkinson, 2007; Moundridou & Virvou, 2002).

1.2.2 The Split-Attention Effect Theory

The split-attention effect theory stems from the cognitive load theory. The cognitive load theory explains effectiveness of multimedia learning by measuring the cognitive load placed on users by multimedia material. The theory divides cognitive load into three types; germane, intrinsic and extraneous (F. Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Fgwc Paas & Vanmerrienboer, 1994; Sweller, 2010). Germane cognitive load is the result of learning itself, hence is considered as effective cognitive load. Intrinsic cognitive load is caused by the difficulty of learning material and the prior knowledge of learners. Finally, cognitive load due to poor instructional design that hinders learning is extraneous cognitive load. The cognitive load theory has been popular among education literature, and has been used as framework in multimedia

learning research (Kirschner, Sweller, & Clark, 2006; R. E. Mayer & Moreno, 2003; Mousavi, Low, & Sweller, 1995).

According to the cognitive load theory, there lays a concern that pedagogical agents may split the limited attention of learners thus causing extraneous cognitive load (Dehn & van Mulken, 2000), thus causing the split-attention effect . Indeed, Baylor & Kim (2009) found that too rich animations of pedagogical agents could harm learning. However, Mayer & DaPra (2012) supports otherwise, as pedagogical agent displaying more lively gestures were found to be more effective. The split-attention effect theory is in obvious conflict with the persona effect theory, and has also produced mixed results.

1.2.3 The Modality Effect Theory

The modality effect theory states that by simultaneously using dual channel—visual, audio—inherent in human working memory, the learner can process more information in limited time (Ginns, 2005; R. E. Mayer & Moreno, 2003; Moreno & Mayer, 1999). The modality effect has also been found with pedagogical agents (Atkinson, 2002; Craig et al., 2002; Moreno et al., 2001).

1.2.4 The Social Agency Theory

The social agency theory claims that incorporating social agency in computer-based learning facilitates deeper learning from students (Moreno et al., 2001). Pedagogical agent, anthropomorphic image of computer system, is largely supported by the social agency theory. As result, pedagogical agent literature has focused on making more human-like agents, in attempt to elicit more social agency. Various features of pedagogical agent have been studied, such as the personalities expressed by speech style of agents (e.g. expressiveness, politeness) (Kim, Baylor,

& Shen, 2007; Moreno & Mayer, 2004; Veletsianos, 2009; Wang et al., 2008), the gender (Kim et al., 2007), the role in learning environment (e.g. colleague, student, expert, novice) (Biswas, Leelawong, Schwartz, Vye, & Teachable Agents Grp, 2005; Bodenheimer et al., 2009; Chase, Chin, Oppezzo, & Schwartz, 2009). The quality of voice is known to affect learning and is well replicated; human voice is more effective compared to machine-generated voice (Atkinson, Mayer, & Merrill, 2005; Richard E. Mayer & DaPra, 2012).

Nonverbal social cues such as gesture and gaze have also received attention from literature. Several studies reported positive effect of social cues (Baylor & Kim, 2009; Baylor & Ryu, 2003; Richard E. Mayer & DaPra, 2012). However, the quality assessed by previous studies is limited to the life-likeness of agents' animation. That is, while human nonverbal social cues are reciprocal, no research has been made on how pedagogical agents should react to learners' social cues.

In the next sub-chapter, previous studies on reciprocal aspects of human learning and the theoretical basis of this thesis is provided.

1.3 Reciprocal interaction in human learning

Facilitating social interaction has been one of the central topics in the literature focusing on pedagogical agents. However, the approach has largely concentrated on the quality of animation of agents (Baylor & Kim, 2009; Richard E. Mayer & DaPra, 2012). Recent developments in social neuroscience suggest the need for a reciprocal approach, suggesting that social cognition may be fundamentally different when individuals are interacting with others rather than merely observing (Anders, Heinzle, Weiskopf, Ethofer, & Haynes, 2011; Leonhard Schilbach et al., 2013). For example, Redcay et al. (2010) showed that live interaction with a

human experimenter, as compared to viewing video recordings of the interaction, displayed greater activation in brain regions involved in social cognition and reward. In addition, Schilbach et al. (2010) demonstrated that forming joint attention with a virtual character stimulates areas of brain associated with social cognition, while avoiding joint attention recruited areas related to control of attention and eye-movements.

Moreover, reciprocal interaction is argued to affect multimedia learning. Not surprisingly, children under three years of age learn less from screen media than from engaging in live social interaction with adults (M. Krcmar, Grela, & Lin, 2007; Kuhl, Tsao, & Liu, 2003); the phenomenon called the video deficit effect(Anderson & Pempek, 2005). While the cause of video deficit is not yet fully understood, several studies have found that providing live social interaction through screen media mitigates the effect (Nielsen, Simcock, & Jenkins, 2008; S. Roseberry, Hirsh-Pasek, & Golinkoff, 2014; Troseth, Saylor, & Archer, 2006). For example, Nielsen, Simcock, & Jenkins (Nielsen et al., 2008) demonstrated that when two year olds communicated with experimenter through closed circuit TV system, the children were as likely to success in imitating the experimenter's actions as the children who interacted directly with an experimenter.

Human social interaction is complex, and involves various features. From these features, this thesis focused on temporal contingency and joint attention. The reason for doing so is both theoretical and practical; they are well studied thus strong theoretical basis can be obtained, and they occur frequently in learning environments thus the results will have practical application.

1.3.1 Temporal contingency

Social interaction holds a distinct feature, temporal contingency. In this thesis, the term temporal contingency points to the responsiveness of agents involved in human interaction. Human interaction is inherently responsive and deteriorates when either participant fails to respond in temporally acceptable window. Indeed, human social cues involve high level of temporal regulation. When humans communicate, gestures (Shockley, Santana, & Fowler, 2003) and eye movements (Richardson, Dale, & Kirkham, 2007) become temporally coupled.

From early stage, humans are sensitive to temporal contingency. Deligianni, Senju, Gergely, & Csibra (2011) found that non-human objects that display contingent movement to gaze elicit gaze following behaviour from 8-month-olds. Also, recent evidence from development psychology suggests that temporal contingency influences learning. Longitudinal studies have provided evidences that maternal temporal contingency affects child development in language (Tamis-LeMonda, Bornstein, & Baumwell, 2001), and general skills such as social skills and problem-solving (Landry, Smith, & Swank, 2006). Short-term studies also presented evidence that temporal contingency affects language development. For example, Goldstein, King & West (Goldstein, King, & West, 2003) demonstrated that the immediate reactions of mothers to infant vocalizations or gestures facilitate speech production from infants.

1.3.2 Gaze interaction

Social interaction involves many domains, such as gesture, facial expression, and voice tone. Among these domains, this thesis focuses on gaze interaction as gaze is one of the most well described social cues (Emery, 2000) , can be easily measured, and be added to virtual characters with relative ease. Gaze is also known to have strong effect on learning. For example,

previous research from the field of developmental psychology indicates that gaze interaction is critical for early language learning (Brooks & Meltzoff, 2005; Morales et al., 2000; Striano, Chen, Cleveland, & Bradshaw, 2006).

Gaze interaction consists of complex features and each is used to convey rich information, such as attention, emotion and threat (Emery, 2000). Of those features, three basic features that are most prominent in learning environment were selected; mutual gaze, joint attention, and gaze following. These features are determined by dyadic gaze direction (Figure 1-1).

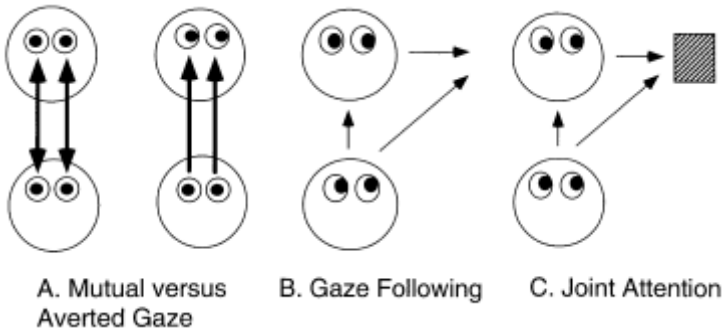

Figure 1-1. Three types of dyadic social cues determined by gaze direction, excerpted from (Emery, 2000). A. Mutual gaze is where the attention of individuals A and B is directed to one another. Averted gaze is where individual A is looking at B, but the focus of their attention is elsewhere. B. Gaze following is where individual A detects that B's gaze is not directed towards them, and follows the line of sight of B onto a point in space. C. Joint Attention is the same as Gaze Following except that there is a focus of attention (such as an object), so individuals A and B are looking at the same object.

These three components of eye interaction are known to play crucial roles in learning. An extensive body of literature from developmental psychology have been dedicated to finding how they affect child development (Brooks & Meltzoff, 2005; Morales et al., 2000; Okumura, Kanakogi, Kanda, Ishiguro, & Itakura, 2013; Senju & Csibra, 2008; Striano et al., 2006; Tomasello & Farrar, 1986). Recently, Csibra & Gergely (2009) presented an interesting hypothesis, *natural pedagogy*. The natural pedagogy hypothesis states that human are born with

ability to respond to the social cues and interpret them as meta-information essential for selecting valid and generalizable information from surrounding environment. Several studies supporting the hypothesis shows that three gaze cues (i.e. mutual gaze, gaze following and joint attention) may be inherent in natural human learning; Human neonates can detect mutual gaze from birth (Farroni, Csibra, Simion, & Johnson, 2002), mutual gaze elicits gaze following from 6-month-olds (Senju & Csibra, 2008), 9-month-olds shows understanding of object-directed gaze (Senju, Csibra, & Johnson, 2008).

1.4 The purpose of this thesis

The main goal of the present study is to implement features of social interactions with animated pedagogical agents and verify their learning effects. The study focused on reciprocal aspect of social interaction and its important features, temporal contingency and joint attention.

To achieve this goal, we implemented an experimental-purpose pedagogical agent that can be easily modifiable, easily extendable, while maintaining strict control over the behaviors of the agent itself (detailed in Chapter 2). Using the agent, we conducted five experiments detailed in the following chapters.

In experiment presented in Chapter 3, persona effect is tested. In Chapter 4, temporal contingency is implemented into gaze interaction of the pedagogical agent, and its learning effect is tested. We expand the finding in Chapter 5 and 6, by testing if temporal contingency affects non-social cues. In Chapter 7, we assess how temporal order of contribution in joint attention affects learning.

Chapter 2. Designing PAGI (Pedagogical Agent with Gaze Interaction)

2.1 Main functions and goals

PAGI was designed as an experimental purpose pedagogical agent. It had to be easily modifiable, easily extendable, while maintaining strict control over the behaviors of the agent itself. The foremost goal set at the beginning of this research was to develop a pedagogical agent system that could be reused in successive experiments. At the time of this research, temporal aspect of pedagogical agent had been hardly studied, thus trial-and-error method had to be employed to search various factors.

In spite of the advance in computer technology, developing a pedagogical agent system still requires steep technical skills. Accordingly, several studies resorted to using commercially available agent systems, such as MS agent (Atkinson, 2002; Atkinson et al., 2005). While the systems are valid, the studies were limited by preexistent functions of the system. Several other groups developed their own system, but the available level of modification was rather small; they were restricted to easily implementable features (e.g. control over pace, feedback type, gesture

and the visual image) (Craig et al., 2002; Richard E. Mayer & DaPra, 2012; R. E. Mayer et al., 2003; Moreno & Mayer, 2004, 2005).

To achieve greater degree of freedom, two design principles were established. First, agent system should be independent from learning material. This enables to use the same agent system in different experimental context. Second, the sequence of agent's animations should be modifiable with ease; not to mention the orders and rules of the agent behavior, qualitative parameters such as reaction timing should be easily modified. This means pre-recorded sequences cannot be used, unlike previous studies.

2.2 Learning material selection

While the first principle of PAGI was to separate agent system from learning materials, learning material still occupies large proportion of experimental design. That is, learning material determines how pedagogical agents are incorporated in the context.

As stated in previous section, this research was planned to take trial-and-error approach, thus we wanted to start with a material that adds less complication to experiment design. Adopting features for more advanced materials would propose another research topic, as we have to consider how to adopt each feature to fit the materials.

For above reasons, we used foreign language vocabulary as learning materials. While it is not a suitable teaching material for pedagogical agents to establish superiority over traditional teaching formats (e.g. books), it was simple enough that we could focus on PAGI's behaviours.

In sum, to focus on the features of PAGI (i.e. temporal contingency, joint attention) and examine its educational effect per se as much as possible, we started with a simple learning material, foreign language words. Korean nouns with less than four syllables were used. The

words were selected by a Korean-Japanese bilingual based on two criteria, low resemblance between the pair and familiarity to general population.

System Architecture

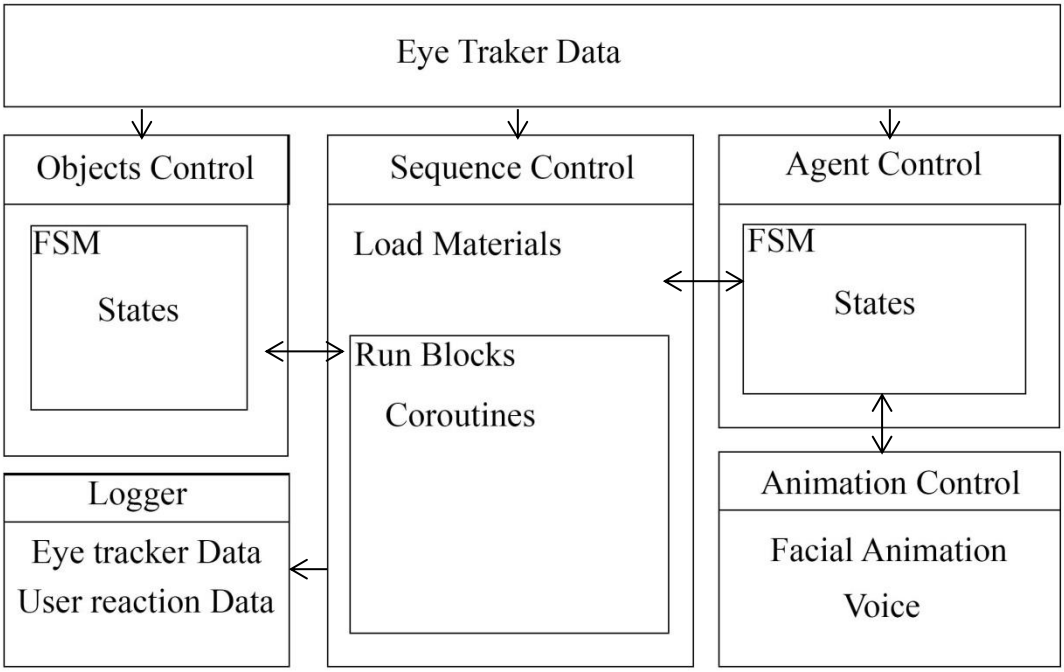

Figure 2-1 System architecture of PAGI

The system architecture of PAGI is provided as **Figure 2-1**. Following our design principle 1 (i.e. agent system should be independent from learning material), experimental sequence, agent and other objects are controlled by separate components. The agent and other objects consist of finite state machines. Sequence Control component communicates with other components by triggering state machine to the next state, and other components send messages to Sequence Control component when their state has been changed. To meet the design principle 2 (i.e. the sequence of agent's animations should be modifiable with ease), behavior of the agent

was divided into small parts, and each part was modifiable with parameters such as speed, onset and offset. Eye-tracker data was passed to each component, and fixation identification was done locally at each component.

2.2.1 Technical specification

The system was developed using the commercial game engine, Unity3D (http://unity3d.com/). Unity3D was selected as it is simple to use and fits the needs of one-man-development team than any other candidates.

For 3-D model which constructs the visual image of the agent, various tools were used. For Experiment 1 (Chapter 3), Softimage (Autodesk) was used. For Experiments 2, 3, and 4 (Chapters 4-6), DAZ Studio (DAZ 3D), 3ds MAX (Autodesk) was used.

The Lip-sync process was done using facial motion-tracking. This was to ensure the lips matched the voice of the agent. Using pre-recorded lip-sync animation is in conflict with the design principle 2 (i.e. the sequence of agent's animations should be modifiable with ease) as the animation cannot be modified during runtime, but natural human voice had to be used, as previous studies suggests that machine voice have disadvantage over human voice (Atkinson et al., 2005; R. E. Mayer et al., 2003). While machine voice has seen some technological advance since previous studies, implementation of machine voice technology exceeds the scope of this research. For facial motion-tracking, Oqus (Qualisys) and FaceRobot (Autodesk) was used for Experiment 1, and Kinect (Microsoft) and FaceShift Studio (FaceShift) was used for Experiments 2, 3, and 4. While the method used for Experiment 1, using state-of-the-art motion capture system, provided more accurate and smooth result, the process took too much effort for a one-man-development team. Hence, later experiments switched to secondary method that uses pre-defined blend shapes, and were able to greatly reduce the facial motion-capture process.

Eye-tracking was done using Tobii T60 and Tobii Rex (Tobii, Sweden). Experiment 1 used Tobii T60 and Experiments 2, 3, and 4 used Tobii Rex, which became commercially available in 2013.

Chapter 3. Experiment 1: Persona Effect and Split Attention Effect

3.1 Introduction

As described in Chapter 1, the persona effect and the split attention effect are two of the major theories in pedagogical agent literature. The persona affect theory suggests that the visual presence of pedagogical agent itself elicits positive learning effect, while the split attention argues that pedagogical agent splits limited attention of learners thus inflicts harm to learning.

Research surrounding two theories has produced mixed results. This may have been affected by the great variety in functions and properties of agents used in previous studies. As the first step towards developing PAGI, we felt the need to confirm that basic elements of PAGI do not hinder user's learning. To answer this question, we conducted an experiment with PAGI at the initial stage of development. The agent has similar scheme as PAGI but is not equipped with reactive interaction functions.

The type of learning materials was identical with PAGI. The participants learned Korean words. Participants were presented with words in their native language (Japanese) and

foreign language (Korean) in audio, and were asked to remember the latter. A photo was presented for each word.

The experiment used a repeated measures design including one within-subject variables: pedagogical agent's physical presence. During learning phase, two conditions were provided: with physical image of pedagogical agent (agent condition), and without physical image of pedagogical agent (no-agent condition).

3.2 Method

3.2.1 Pedagogical Agent

The agent was developed as three-dimensional character using FaceRobot (Autodesk), and photorealistic textures were used to enhance graphical quality (see **Figure 3-1**). Although the agent is capable of behaviours such as blinking, eye movement, and complex facial expression, only lip movement was applied for this experiment. The agent's lip movement was implemented using facial motion-tracking, with Oqus (Qualisys) and FaceRobot (Autodesk).

3.2.2 Participants

Nine participants (5 females) were recruited. Their mean age was 20.33 (SD = 1.58) years. Participants were all native Japanese without Korean language experience. The condition order was counterbalanced among participants. They all served as unpaid volunteers.

3.2.3 Apparatus

The learning phase was presented on 17-inch LCD monitor. For the test phase, a 17-inch CRT monitor was used. Sound stimuli were presented through two speakers (BOSE Media Mate II).

3.2.4 Procedures

After arriving at the laboratory, participants were guided to seat in front of a computer screen then instructed to memorize Korean words. They were told there would be a test afterwards. No instructions were given on how to interact with the agent.

The experiment was divided into two phases, learning phase and test phase. In the learning phase, fifteen words were presented for each condition, total of thirty words. A photo was displayed, and a word describing the photo was presented by voice, first Japanese then Korean (**Figure 3-2**). Each word was presented twice, successively. Every word was a noun, consisting of less than five syllables (for both Japanese and Korean). Words were presented in random order.

After working with the agent, participants were tested after 1-min break. The test phase was carried out immediately after the learning phase. Four pictures were presented on the screen and a Korean word was verbally given (**Figure 3-3**). Participants were instructed to pick the picture that corresponded to the word. Participants used number keys 1–4 on a keyboard during the test to choose the picture. Participants were informed that there was no time limit during the test phase. The experiment was conducted in soundproof environment. Whole process took around ten minutes.

Figure 3-1 A close-up view of the pedagogical agent

Figure 3-2 Sample frame from the learning phase; agent condition (left) no-agent condition

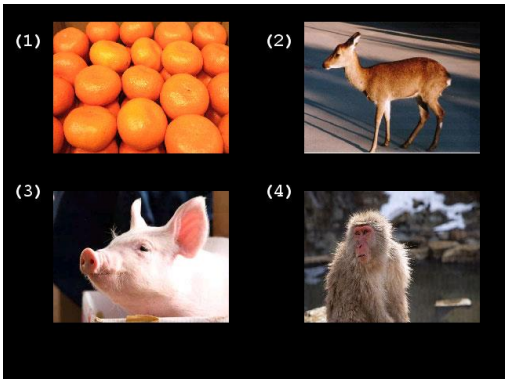

Figure 3-3 Sample frame from test phase

3.3 Result

Repeated measures analyses of variance were conducted on test scores. The main effect of condition did not reach significance (F (1,7) = .01, p > .5). The ANOVA revealed order effect did not affect the result (F (1,7) = .47, p > .5). All participants scored higher than chance level; 7.5 words as the test phase consisted of thirty four-choice questions. (agent condition: t (8) = 8.004, p < 0.001; no-agent condition: t (8) = 7.803, p < 0.001, one-sample t test). The result is shown in **Figure 3-4**.

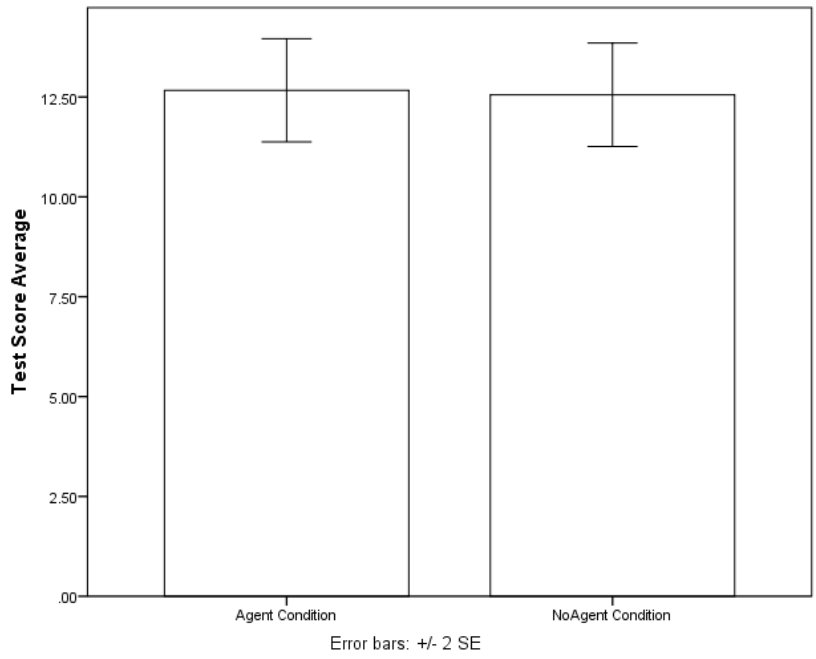

Figure 3-4 Test score average for each condition

3.4 Discussion

The goal of this experiment was to assess if the persona effect or the split attention effect is elicited by basic elements of PAGI. The result showed that the physical image of pedagogical agent did neither damage nor enhance learning outcomes. This concurs with previous studies that suggested physical presence of pedagogical agent does not affect learning (R. E. Mayer et al., 2003).

The remaining question is why some previous groups succeeded in replicating the two effects. One possible explanation is that other features of the pedagogical agents affected the result. For example, quality the agents' animation—how the agents move and act in learning environment—may have been a critical uncontrolled variable.

In this chapter, we demonstrated that mere presence of pedagogical agent is not enough to make a difference. In the following chapters, we describe series of experiments focusing on more advanced features of pedagogical agents.

Chapter 4. Experiment 2: Temporal Contingency Effect

4.1 Introduction

As discussed in Chapter 1, one of the main purposes of pedagogical agent is to prompt social stance from learner, which is argued to facilitate deeper learning (see Chapter 1, section a-i. Social agency theory). The supporters of this particular theory attempted to elicit social conversation schema in learner, by adding social cues to anthropomorphized computer programs; pedagogical agents (Atkinson, 2002; Craig et al., 2002; Richard E. Mayer & DaPra, 2012; Moreno et al., 2001).

While previous studies have found that social cues improve the way agents are perceived (Louwerse, Graesser, Lu, & Mitchell, 2005), and even trigger involuntary mimicking as in human-to-human interaction (Bailenson & Yee, 2005), there is yet no definite evidence that the display of social cues enhance the quality of learning (Heidig & Clarebout, 2011; Schroeder et al., 2013).

There remains huge gap in the quality of social cues between those used by human and those implemented in pedagogical agents. Perhaps the most important feature that has been

neglected is temporal contingency. This is surprising as reactivity is the key aspect of social interaction, and temporal contingency is a compulsory feature in implementing reactive social interaction. Social cues in human interaction are highly reactive, and it is doubtful that simply adding animation to agents would prime social stance. That is, essentially there is no difference between pedagogical agents that play pre-recorded gesture animation and animated characters from traditional television shows.

In this chapter, we present an experiment which tested the hypothesis that the temporally contingent gaze interaction of animated pedagogical agents would enhance word learning. We developed an animated pedagogical agent capable of temporal contingent gaze interaction, called PAGI (Pedagogical Agent with Gaze Interaction). PAGI simulates mutual gaze, gaze following, and joint attention with students while teaching foreign language words.

4.2 Method

4.2.1 Participants

Thirty participants (7 women) were recruited from a subject pool at the University of Tokyo. Their mean age was 20.19 (SD = 1.47) years. Participants were all native Japanese without Korean language experience. Additional five participants were not included in the sample due to the failure of the eye-tracking system during the experiment, which caused the agent to malfunction. Participants were randomly assigned to either live (n = 15; women = 4; mean age = 20.0) or recorded (n = 15; women = 3; mean age = 20.4) group.

All participants provided written informed consent. The experiment was approved by the University of Tokyo Ethics Review Board.

4.2.2 Design

A between-subjects yorked-condition design was used, in which participants learned Korean words with the temporally contingent agent (Live group) or the recorded agent (Recorded group). The live group was paired with the live-interacting pedagogical agent, and the recorded group with the agent replaying behaviour sequence recorded during live group sessions. Thus, recorded group was provided with the same agent exhibiting the same behaviours in the exact sequence as the live group, except without temporal contingency.

4.2.3 The Pedagogical Agent with Gaze interaction (PAGI)

As explained in Chapter 2, PAGI is an experimental animated pedagogical agent designed to teach Korean words to Japanese students. Changes have been made from Experiment 1 (Chapter 3). Firstly, the visual image was replaced by a less realistic 3D male cartoon character. This was due to participants reporting the uncanny valley effect. The uncanny valley effect, originally termed "*Bukimi no Tani*" in Japanese (Mori, 1970), is a theory that argues people have an unpleasant impression of a humanoid robots or CG agent, when it has an almost, but not perfectly, realistic human appearance. It was voiced by a male Korean-Japanese bilingual speaker, lip-synced to the voice using predefined visemes.

PAGI started with an opening narration, explaining that he will be teaching Korean words, while gazing at participant's eyes, initiating eye contact (Figure 4-1). After the narration, PAGI initiated word learning phase. First, two pictures were presented (Stage 1). PAGI waited for an eye contact and then shifted his gaze to the target picture. He then waited for the participant to follow his gaze and fixate on the target picture, and form joint attention. After joint attention was formed, PAGI returned his gaze to the participant and spoke a frame sentence (the

first portion of the sentence leading to a target word, (e.g., “this is” or “next is” in Japanese)(Stage 2). Finally, PAGI spoke the target Korean word twice (Stage 3). PAGI repeated Stage 1-3 for each word.

In Stage 2, if the participant did not form the joint attention within 3 second time limit, PAGI looked at the participant and delivered attention-redirecting dialogue (“Please follow my lead”) in Japanese. This was to mimic the behaviour of human tutor delivering attention-redirecting dialogue and to prevent the live group participants from taking advantage of the system and taking too much time to memorize each word. However, a pilot test revealed that for the replay group, the dialog impaired the perceived reliability of the system, which is a confounding factor that could critically damage the experiment. Thus recordings containing attention-redirection dialogues were not used for recorded group. As the result, seven recordings were distributed to fifteen recorded group participants.

4.2.4 Materials

All dialogues except target words were presented in Japanese. Korean nouns with less than four syllables were used for the lesson. Each word was presented with a corresponding picture and written Japanese word (Figure 4-1). The word list was identical for all participants, and was presented in the same sequence. The words were selected by a Korean-Japanese bilingual based on two criteria, low resemblance between the pair and familiarity to general population.

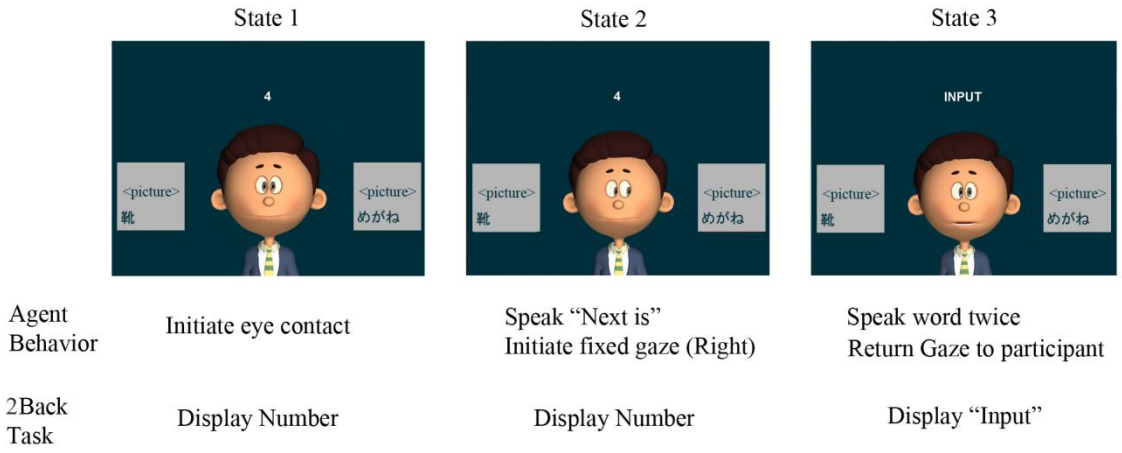

Figure 4-1 Gaze interaction scheme of PAGI.

A distracter picture was presented with each target pictures to force gaze following. If only the target picture was presented, gaze following would be unnecessary. To avoid this, a random picture from the word list was simultaneously presented as a non-target word. As the result, to obtain the correct meaning of the word, participants needed to watch and follow PAGI's gaze. The target and distracter picture pairs were presented randomly to the left or right of PAGI (Figure 4-2b). Participants learned 60 words, which were divided into two blocks of 30 words, with a 1-min rest between the blocks.

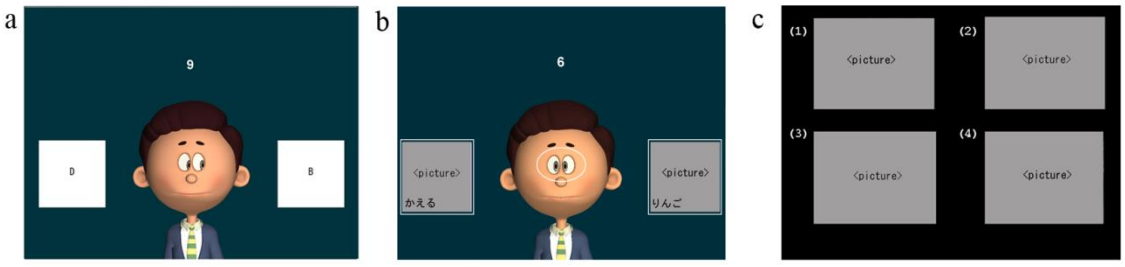

Figure 4-2 Selected frames from the stimuli: (a) the practice session, (b) the learning phase, PAGI State 1 (the white lines represent areas of interest and were not visible during the experiment), and (c) test phase. (The pictures are grayed-out due to copyright restrictions.)

As the task was simple and repetitive, a pre-test revealed a ceiling effect, and participants reported that the difference between two conditions was too minor to be noticed. To solve these two problems, we employed dual-task design using digit span two-back task. A single-digit number appeared above PAGI and then was replaced by “input”, thus forcing participants to take their eye away from PAGI from time to time. This enabled participants to notice whether PAGI was responding to them or not. Participants were instructed to input the digit using keyboard numpad while the two-back sign was showing “input”. When the answer was inputted during the input time window, the sign changed to blank to inform participants that their input was handled, regardless of its correctness. Key inputs made when the sign showed otherwise (a number or blank) were ignored. The two-back task was temporally synced to word learning task; started and ended at the same time as each word (Figure 4-1).

4.2.5 Apparatus

The eye-tracker (Tobii, Sweden) was integrated with a 17-inch LCD monitor, on which stimuli were displayed. A nine-point calibration was administered at the start of every block. A webcam was placed under the eye-tracker focused on the participant’s eyes to monitor the gaze interaction. For the test phase, a 17-inch CRT monitor was used. Sound stimuli were presented through two speakers (BOSE Media Mate II).

4.2.6 Procedure

When the participants arrived at the laboratory, each was led separately to a room and seated in front of a monitor. The experimenter told each participant that the purpose of the

experiment was to assess how humans learn foreign language and instructed them to engage in gaze interaction with the agent, by forming mutual gaze and following the agent's gaze.

The experiment was divided into two phases: learning and test. In the beginning of the learning phase, each participant was given a practice session, to get accustomed to PAGI and rehearse digit span two-back task. The practice session was identical to the learning phase, except that instead of the Korean words, ten English alphabet letters—from 'a' to 'j'—were presented (Figure 4-2a). Thus, each participant was given 10 trials (each alphabet counted as one trial) in the practice session. For two-back task, participants were instructed to use any fingers of their preference, and to return all fingers to starting position after each input; all fingers placed in line below numpad.

The test phase was carried out immediately after the learning phase. Four pictures were presented on the screen and a Korean word was verbally given (Figure 4-2c). Participants were instructed to pick the picture that corresponded to the word. Participants used number keys 1–4 on a keyboard during the test to choose the picture. Participants were informed that there was no time limit during the test phase. The entire experiment lasted 25 - 30 min.

4.3 Result

The test score was composed of the number of correct answers. Participants from both groups scored higher than the chance level; 15 words as the test phase consisted of sixty four-choice questions. (live group: t (14) = 10.193, $p < 0.001$; recorded group: t (14) = 7.572, $p < 0.001$, one-sample t test). The mean test scores differed significantly between the two groups (t (28) = 3.372, $p = 0.002$, d = 1.24) (Figure 4-3).

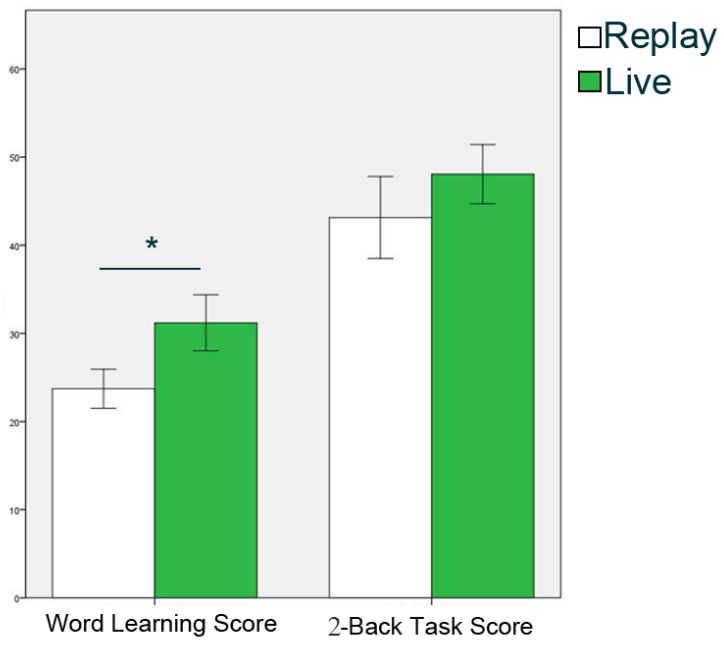

Figure 4-3 The group average of test results. Left: word learning. Right: two-back task. The asterisk indicates statistical significance. Error bars represent standard errors.

Live group also performed better on two-back task, with marginal significance (t (28) = 3.046, p = 0.095). This was expected as the two –back task was temporally synced to the word learning task, thus was also under temporal contingency effect, albeit less explicit than the main task.

As the experiment used dual-task design, the difference of attention allocation on two tasks may have affected the result (i.e. the live group assigned larger proportion of cognitive resource on the word learning task). The difference could not be compared between the two groups, however, as both tasks were affected by temporal contingency. Nonetheless, analysis on each group showed no significant correlations between the test scores and two-back scores. Also, group difference of test scores cannot be explained by attention allocation as live group scored better on both tasks.

Generally, memory task performance increases when participants are given more time, and the time spent on the learning phase was not controlled in our experiment. To assess the

influence of the time spent on learning, we examined the relationship between participants' learning time and test scores. The duration of the learning phase did not differ significantly between the two groups (live group: M = 368.7 sec, recorded group: M = 352.8 sec; t (28) = 1.391, $p > 0.1$). There was no significant correlation between learning phase duration and test scores. These analyses revealed that time spent on learning did not significantly affect the test scores.

Because the attention-redirecting dialogue was provided only for the live group, it may have influenced the result. To assess this, we conducted an analysis using the live group data. The overall number of received attention-redirecting dialogues was small (M = 1.27); seven participants did not receive any, four received one, two received two, and two received four. As the distribution of attention-redirecting dialogue counts were outside of the limits of normality (standardized skewness coefficient > 2), a nonparametric procedure, Spearman's rho was used for the analysis. The correlation between the number of received dialogues and test scores was not significant (r = 0.439, $p > 0.1$, Spearman's rho). In addition, we observed no significant difference in test scores between participants who received the attention-redirecting dialogue and those who did not (t (13) = 1.298, $p > 0.2$). This provides some evidence that the attention-redirecting dialogue did not significantly affect the result.

The eye tracking data were gathered using commercial software (Tobii Studio, Sweden). As eye tracking data is inherently noisy, we conducted strict pre-selection of data. The samples containing less than 50 valid fixations (data points classified as fixation inside the area of interest, that does not contain missing gaze points and was longer than 100ms—which is argued to be the minimum fixation duration; Tobii Fixation Filter(Olsson, 2007) was used for fixation classification) were excluded from the analysis for statistical validity. As the result, three

participants from the live group and five from replay group were excluded for eye tracking data analysis.

The eye tracking data analysis revealed no immediate difference between the two groups. Reaction time for mutual gaze and joint attention was calculated by measuring the time elapsed from PAGI's gaze shift and subsequent fixation on the target object (mutual gaze: PAGI, joint attention: target picture). There was no significant difference in the reaction time between groups either in mutual gaze (live group: $M = 852.33$ ms, recorded group: $M = 828.46$ ms, $p > 0.8$) or joint attention (live group: $M = 896.99$ ms, recorded group: $M = 780.43$ ms, $p > 0.1$). To analyse participants' commitment to each gaze interaction, proportion of each gaze behaviour was assessed. The proportion of direct gaze was calculated as the total duration that the participant fixated on PAGI while PAGI was looking at participant / total duration of PAGI looking at participant; joint attention as total duration that the participant fixated on the target picture while PAGI was looking at the picture / total duration of PAGI looking at the picture. There was no significant difference between the two groups, for mutual gaze (live group: $M = 0.291$, recorded group: $M = 0.180$, $p > 0.1$), or joint attention (live group: $M = 0.437$, recorded group: $M = 0.290$, $p > 0.1$). None of these factors were correlated with the test scores.

To further evaluate the differences in visual search patterns, the average fixation duration was measured. The fixations on PAGI and pictures were subjected to the analysis. There was no significant difference between the groups (live group: $M = 556.54$ ms, recorded group: $M = 475.52$ ms, $p > 0.2$), and there was no significant correlation with the test scores. However, additional analysis revealed that a certain time-window of fixation duration was related to higher test scores (Figure 4-4). The average fixation duration of the higher scoring group (split by the test score median) was inside 350~750ms, with only two samples outside 400~700ms window.

Indeed, participants with average fixation duration within 400ms~700ms scored significantly better (N = 11, M = 31.55) than those with shorter or longer gaze duration (N= 11, M = 24.64, t (20) =3.086, p = 0.006). The distribution of samples regarding the duration window differed significantly between the two groups; nine of twelve members of the live group were inside the time window compared to two out of ten from replay group (Fisher's exact test, two-tailed, p = 0.03).

Interestingly, the correlation between the test scores and the fixation duration was in the opposite direction (live group: r = -0.626, p = 0.03, recorded group: r = 0.644, p = 0.045). The replay group was more likely to have fixation average shorter than 350 ms (Fisher's exact test, two-tailed, p = 0.029), which appears to reflect the need for more frequent visual search. As for the live group, excessive fixation duration was related to lower test scores, which we initially believed to be consequence of attention gaps. As attention gaps should have increased the likelihood of receiving attention redirecting dialogues, we assessed if there was a correlation between the fixation duration average and the attention redirecting dialogue count, but it did not reach significance (r = -0.041, p > 0.1, Spearman's rho).

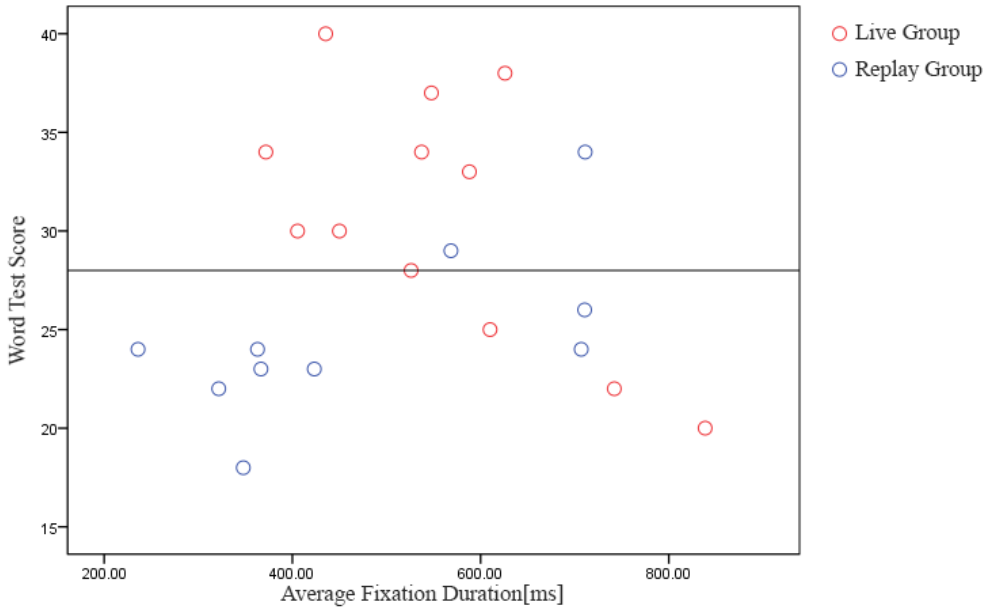

Figure 4-4 Scatterplot illustrating the relation between word test scores and average fixation duration. The horizontal line represents the median test score. Eight participants were not included in the eye tracker data analysis (but included in test scores) due to lack of valid data, as explained in Result.

4.4 Discussion

The present experiment demonstrated that temporally contingent gaze interaction improves the word learning with animated pedagogical agents. The result supports the importance of live social interaction on human learning and extends this to computer-based education.

Because the attention-redirecting dialogue was presented only for the live group, we initially suspected that the dialogue might have influenced the result. However, our auxiliary analysis suggests that the influence was negligible. This seems at glance to contradict with previous finding of D'Mello et al (2012), who reported that the use of attention redirecting dialogue improved learning of high school biology. However they also reported that the effect was confined to deep reasoning, and the effect was not found in overall learning, especially in

directly transferred knowledge. As attention redirecting dialogue is strongly related to boredom or drift of attention, its effect is understandably influenced by complexity of learning material and the length of lesson. The facts that the learning material of our experiment was relatively simple and short should have minimized the effect of the attention redirecting dialogues, as shown in our result.

If the attention-redirecting dialogue did not improve learning, it remains unclear why the temporal contingency of gaze interaction had such a significant effect. One possible explanation is that lack of temporal contingency induced larger cognitive load by inflicting the need for more frequent visual search; the replay group participants had to constantly check for the agent's cues whereas the live group participants could progress in their own terms. The analysis of the fixation duration average showed that the live group was more likely than the replay group to have fixation average inside the time window 400~700ms which was linked to higher test scores, and replay group was more likely to have shorter fixation average. While fixation duration cannot always be directly linked to cognitive processes, in regards to our experimental task, shorter fixation durations can be linked to greater devotion towards visual search. Therefore, it can be assumed that temporal contingency reduced extraneous cognitive load, contributing to less burden for visual search.

The analysis also showed that for the live group, excessive fixation duration was related to lower test scores. This is probably attributable to a lack of motivation. The participants may have thoughtlessly followed the instruction by exhibiting gaze behaviours, but did not invest full attention to word learning. If this was the case, fixation duration may be used to counter low attention in real world settings, especially for students with low motivation. To this end, how

fixation duration is affected by motivation is an interesting research topic and requires future research.

Another major factor influencing the result may be social presence. In social agency theory, social cues of pedagogical agents prime a feeling of social partnership in the learner, which leads to deeper cognitive processing during learning, and results in a more meaningful learning outcome (Richard E. Mayer & DaPra, 2012; Moreno et al., 2001). Previous studies suggest that the social agency of an pedagogical agents have impact on learner motivation (Baylor, 2011; Heidig & Clarebout, 2011), and the design of these agents could alter the motivational outcomes (Baylor, 2011; Baylor & Kim, 2009). In addition, recent evidence from social neuroscience implies that social cues from a virtual agent— direct gaze and socially relevant facial expression—recruit brain regions related to emotional processing (L. Schilbach et al., 2006). The temporal contingency of gaze interaction may have influenced perceived social agency, affecting motivation of participants. While D'Mello, Olney, Williams, & Hays (D'Mello et al., 2012) reported gaze reactivity of a pedagogical agent did not produce motivational outcomes, the gaze interaction in the study was limited to attention redirecting dialogue which the agent vocalized when student's gaze was away for a specific time. PAGI's interacting behaviours were more reciprocal and real-time, thus may have yielded greater social presence, resulting in more motivational outcomes.

While the result of this experiment was straight-forward, care is required when extending discussion to real-world learning. In this experiment, we employed dual-task design, which tasked participants to perform two-back task simultaneously with word learning task. Participants had to distribute their attention among these two tasks, and while our analysis showed that attention allocation was not the cause of temporal contingency effect, there remains

possibility that temporal contingency helped participants deal with attention allocation process, rather than learning itself. Future work is needed to verify how temporal contingency affects learning when less attention allocation process is required.

In conclusion, our experiment has led us to conclude that temporal contingency of gaze interaction is a key feature in the improvement of the effectiveness of animated pedagogical agents. Our data suggest that temporal contingency should be considered when designing animated pedagogical agents. The mechanism behind temporal contingency effect is not made clear by this experiment, but we propose two hypotheses, 1) temporal contingency reduced extraneous cognitive load related to visual search, 2) temporal contingency primed social stance in learners which may have enhanced learning.

Chapter 5. Experiment 3: Anthropomorphic image of pedagogical agent and Temporal Contingency effect

5.1 Introduction

In the previous chapter, we revealed that temporal contingency facilitates learning, and proposed two hypotheses that may explain the result; 1) temporal contingency reduces extraneous cognitive load related to visual search, 2) temporal contingency prime social stance in learners which enhance learning. The first hypothesis is based on the cognitive load theory and the second hypothesis is based on the social agency theory.

To assess more deeply into this matter, we evaluated whether temporal contingency effect is exclusive in social cues or also found with non-social cues. As explained in Chapter 1, the natural pedagogy theory states that humans are born with ability to respond to the social cues and interpret them as meta-information, which is essential for selecting valid and generalizable

information from surrounding environment (Csibra & Gergely, 2009). Also, Wu & Kirkham (2010) reported that infants show signs of deeper learning when exposed to social cues, compared to non-social cues.

Also, how humans perceive social cues and non-social cues (non-biological cues; e.g. arrow, direction words) have been the focus of large body of literature in field of neuroscience (Nummenmaa & Calder, 2009). Gaze and non-social cues are known to elicit similar behaviours (i.e. reflexive attention shifts) (Deaner & Platt, 2003), but recent studies using EEG and fMRI suggest that gaze and non-social cues may operate on different attention systems (Hietanen, Leppanen, Nummenmaa, & Astikainen, 2008; Hietanen, Nummenmaa, Nyman, Parkkola, & Hamalainen, 2006; Kingstone, Tipper, Ristic, & Ngan, 2004).

Following these theories from the field of developmental science and neuroscience, which suggest that social cues such as gaze are unique and facilitate natural learning, we hypothesized that temporal contingency would elicit stronger benefits when incorporated in social cues than non-social cues. From variety of non-social cues, we focused on arrow cue. Arrow cue is functionally close to the gaze cue used by PAGI, and is most extensively studied non-social cue (Nummenmaa & Calder, 2009).

In pedagogical agent literature, one study compared the learning effect of human agent and arrow agent. Choi & Clark (Choi & Clark, 2006) reported no overall difference between agent types; only students with low prior knowledge benefited from human agent. This may be due to the fact that arrow is an extremely overlearned directional cue which elicits almost identical behaviours as gaze. Although gaze and arrow may use different neurocognitive system as explained above, they may not immediately affect learning, especially for adults who are well adapted to directional meaning of arrow.

5.2 Method

Other than the visual image of the agent, overall design and procedure was identical to Experiment 2 (explained in Chapter 4). PAGI was replaced by a 3-D arrow. (Figure 5-1).

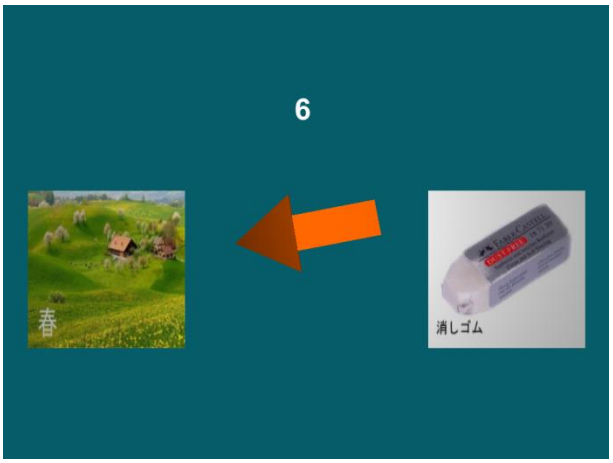

Figure 5-1 Selected frame from experiment 3

5.2.1 Participants

Thirty-two participants (15 women) were recruited from a subject pool at the University of Tokyo. Their mean age was 20.21 (SD = 1.48) years. Additional six participants were not included in the sample due to the failure of the eye-tracking system during the experiment, which caused the agent to malfunction. Participants were all native Japanese without Korean language experience. Participants were randomly assigned to either live (n = 15; women = 7; mean age = 20.2) or recorded (n = 17; women = 8; mean age = 20.1) group.

All participants provided written informed consent. The experiment was approved by the University of Tokyo Ethics Review Board.

5.3 Result

The test score was composed of the number of correct answers. No significant differences between condition groups were found in the mean test scores (t (30) = 0.626, $p > 0.5$; experiment 3-2: t (13) = -0.468, $p > 0.5$) (Figure 5-2). There was also no difference in two-back task scores ($p > 0.5$).

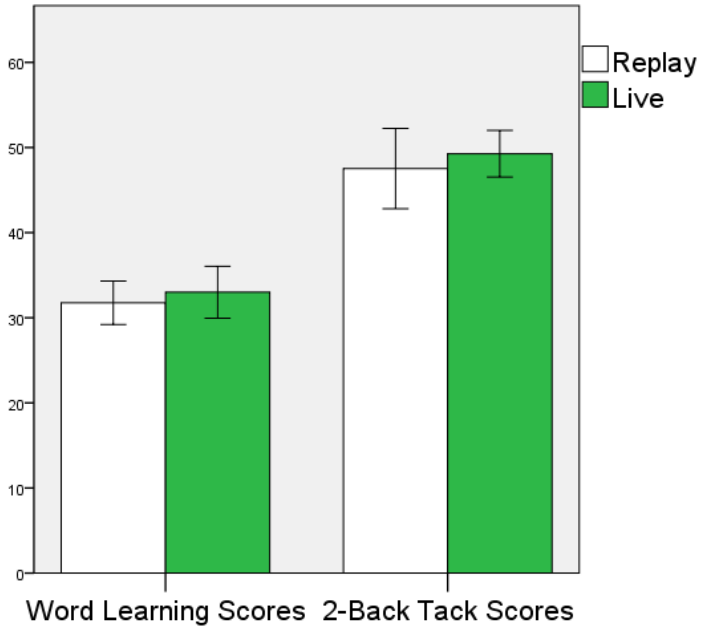

Figure 5-2 The group average of test results. Error bars represent standard errors.

To assess the effect of agent type, post-hoc analysis was conducted on experiment 2 and 3. As significant interaction was found between agent type and condition ($p < 0.05$), simple main effects analysis was employed. The analysis of word learning scores showed that the replay group participants scored significantly lower when they learned from PAGI compared to the arrow agent ($p < 0.005$). The main effect of agent type of not significant among the live group participants ($p > 0.1$).

As in Experiment 2, average fixation duration was measured (Figure 5-3). The live group participants were significantly longer in average fixation duration than the replay group participants (live group: M = 469.28 ms, recorded group: M = 378.08 ms, t (30) = 2.874, p < 0.01).

Two-way ANOVA was conducted including data from Experiment 2 to compare the effect of agent type and temporal contingency on average fixation duration. The analysis revealed significant main effect of agent type ($F(1,51) = 6.328$, $p < 0.02$); arrow agent group were significantly shorter in average fixation duration than PAGI group, and condition ($F(1,51) = 6.350$, $p < 0.02$); replay group were significantly shorter in average fixation duration than live group. The interaction of these two factors was not significant ($F(1,51) = 0.084$, $p > 0.5$).

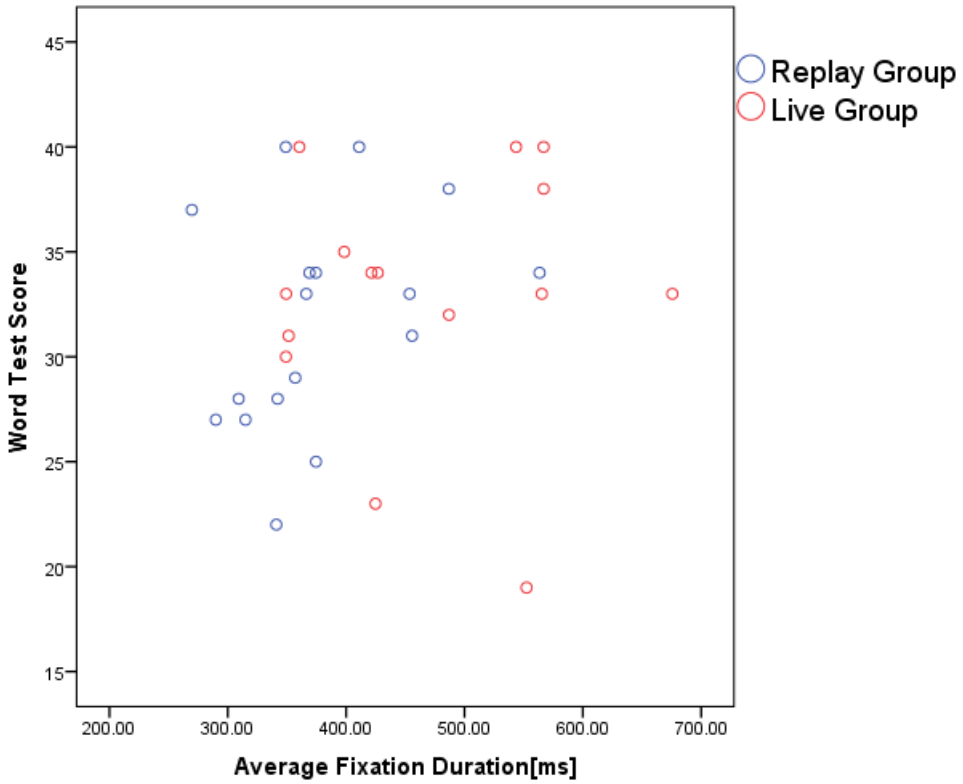

Figure 5-3 Scatterplot illustrating the relation between word test scores and average fixation duration.

Fixations average and test scores were not correlated for both groups ($p > 0.1$).

5.4 Discussion

The experiment showed that the arrow agent did not yield temporal contingency effect. This infers that the visual image of the agents affect temporal contingency effect. Two major factors may have contributed to the result; a) saliency, and b) social-ness of the agents' physical image (i.e. social cue vs non-social cue).

To investigate further, we conducted Experiment 5, which is presented on the next chapter. The question regarding the mechanism of temporal contingency effect, which was raised in Introduction will also be answered in the next chapter.

Chapter 6. Experiment 4: Saliency of pedagogical agent and Temporal Contingency effect

6.1 Introduction

Experiment 4 was conducted to corroborate Experiment 3 (Chapter 5), by testing if the saliency of the arrow agent affected learning effect of the agents. The visual image of the arrow agent was altered to match saliency of PAGI's gaze, in terms of size and color.

6.2 Method

The agent of experiment 4 was designed to match saliency to gaze of PAGI as closely as possible; the horizontal length was matched to length of eye movement of PAGI, and color was matched to that of PAGI's eyes.

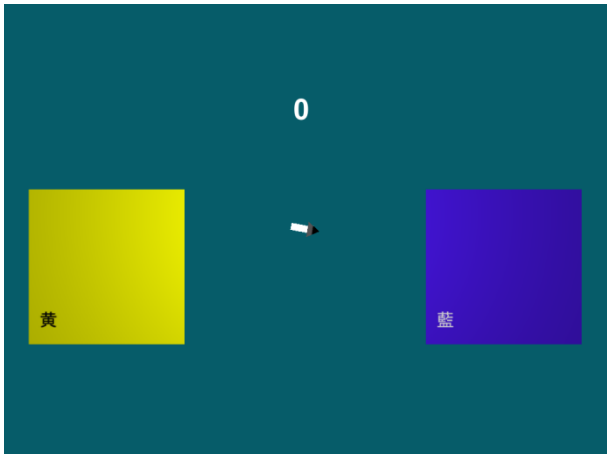

Figure 6-1 Selected frame from experiment 4

6.2.1 Participants

Fifteen participants (7 women) were recruited from a subject pool at the University of Tokyo. Their mean age was 20.2 (SD = 1.35) years. Participants were all native Japanese without Korean language experience. Participants were randomly assigned to either live (n = 7; women = 3; mean age = 20.2) or recorded (n = 8; women = 4; mean age = 20.1) group.

All participants provided written informed consent. The experiment was approved by the University of Tokyo Ethics Review Board.

6.3 Result

The test score was composed of the number of correct answers. No significant differences between condition groups were found in the mean test scores (t (13) = -0.468, p > 0.5) (Figure 6-2).

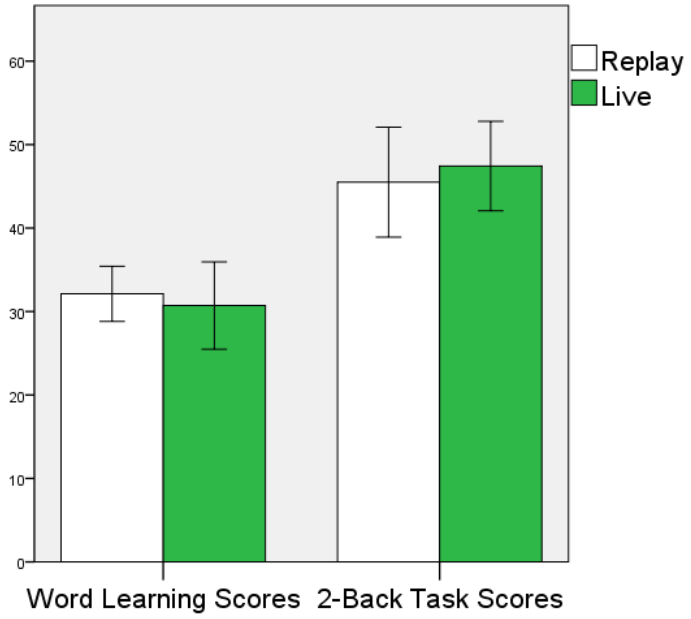

Figure 6-2 The group average of test results. Error bars represent standard errors.

To assess the effect of agent type, post-hoc analysis was conducted on experiment 2, 3 and 4. As significant interaction was found between agent type and condition ($p < 0.05$), simple main effects analysis was employed. The simple main effects analysis of word learning scores showed that the replay group participants scored significantly lower when they learned from PAGI compared to the two arrow agents ($p < 0.005$). There were no differences among the live group participants ($p > 0.1$). There were no differences in test scores between two arrow agent groups ($p > 0.1$).

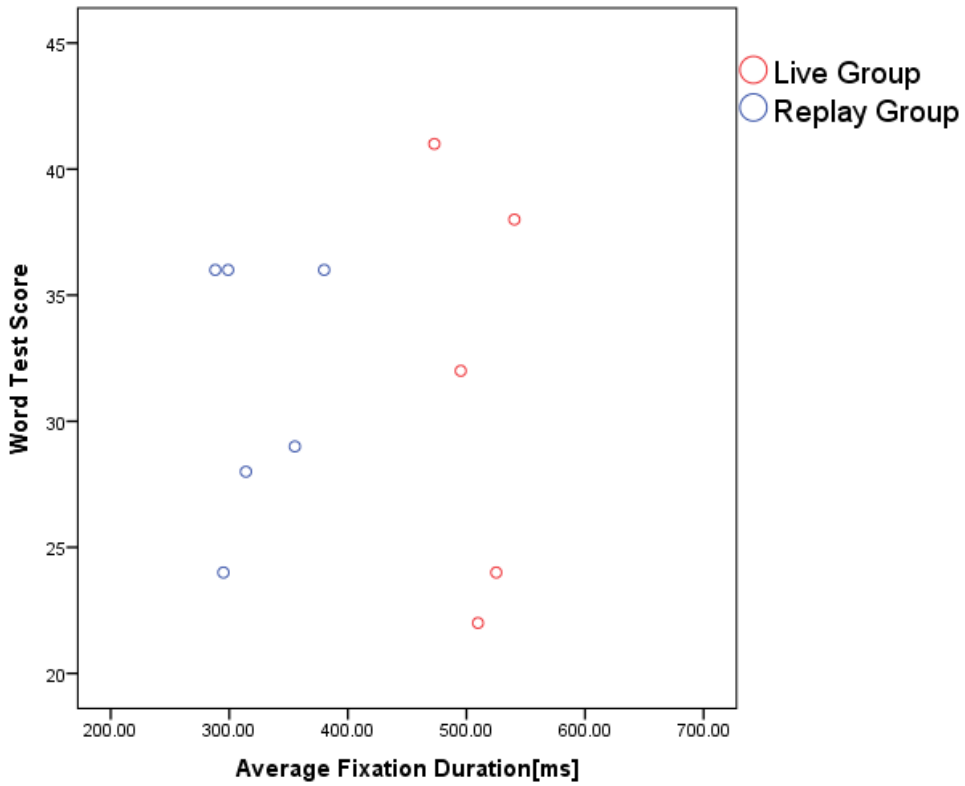

Figure 6-3 Scatterplot illustrating the relation between word test scores and average fixation duration. Four participants were not included in the eye tracker data analysis (but included in test scores) due to lack of valid data.

As in Experiment 2, average fixation duration was measured (Figure 6-3). The live group participants were significantly longer in average fixation duration than the replay group participants (live group: $M = 508.66$ ms, recorded group: $M = 321.92$ ms, $t(9) = 9.405$, $p < 0.001$). Simple main effect analysis was conducted to compare average fixation duration between different agent groups. The analysis revealed that for replay groups, PAGI group were significantly longer in average fixation duration than two arrow agent groups ($p < 0.05$). For live group, PAGI group were significantly longer in average fixation duration than the bigger arrow agent group with marginal significance ($p = 0.095$). Any other combinations did not yield significant result (Figure 6-4).

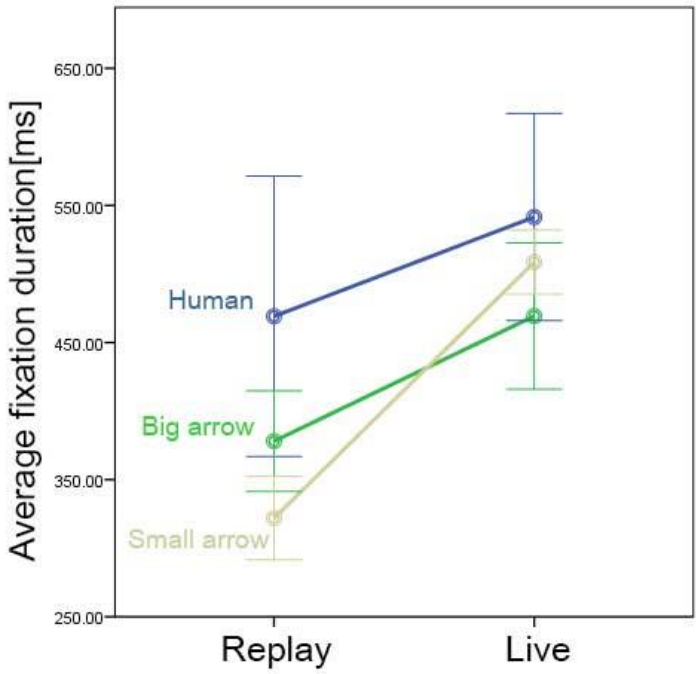

Figure 6-4 Simple main effect of agent type and temporal contingency on average fixation duration.Bars represent standard error.

6.4 Discussion

The present experiment yielded same result as Experiment 3, as temporal contingency effect was not found with the small arrow agent. Despite the small arrow agent was less salient than the big arrow agent used in Experiment 3, there was no difference in word learning scores. This implies that difference in social-ness of the agents affected temporal contingency effect, rather than saliency of the agents. Gaze—a social cue—triggered temporal contingency effect whereas arrow—a non-social cue—did not.

The post-hoc analysis showed that agent type had no effect when the agents were temporally contingent, however, when they were not temporally contingent, the anthropomorphic agent (PAGI) yielded less learning effect compared to non-human agents.

Our hypotheses on the reason of temporal contingency effect, which were raised in

Chapter 4 was: 1) temporal contingency reduces extraneous cognitive load related to visual search, 2) temporal contingency prime social stance in learners which enhance learning. The basis for hypothesis 1 was that average fixation duration was shorter in replay group compared to live group. This relation held true for all three agent types, but there were no differences in word learning scores. Also, when analysis conducted within replay groups showed that average fixation duration of PAGI-replay group was the longest, but word learning score was the lowest. This suggests that cognitive load caused by visual search was unlikely to be the cause of temporal contingency effect.

Thus, the hypothesis 2) temporal contingency prime social stance in learners which enhance learning, is a more likely explanation. Temporal contingency affects the quality of PAGI's movements as social cue thus priming greater social stance in learners.

However, the anthropomorphic agent, which provided social cues, was inferior to the arrow agents in replay group. In live group, there was no difference in learning effect. One may argue that the result implies social stance of learners inflicts no or harmful effect on learning, which in turn contradicts the social agency theory. However, we argue that the result does not imply that social cues are harmful to learning, but provides interesting depth to the social agency theory. Firstly, we suggest the results of replay group and live group were each lead by a different reason. Secondly, we suggest that social cue and non-social cues in learning is a nature versus nurture problem. That is, human infants are born equipped with understanding of social cues, but are trained to learn without social cues. Thus, while social cues may be hard-wired into human brain, adults are more comfortable using general (non-social) attention system for learning.

In replay group, the anthropomorphic agent yielded lesser learning effect compared to the arrow agents. This may have been caused by incomplete activation of gaze-related social attention system. Recent studies from neuroscience suggest that social cues trigger different brain areas when they are reciprocal (Redcay et al., 2010; Leonhard Schilbach et al., 2013; Wilms et al., 2010). Human social cues are highly reciprocal, and require participants to be sensitive to each other's cues. However, when social cues are not directed at one self, it is no longer reciprocal and it is natural for the person to neglect the social cues. Thus, when the anthropomorphic agent displayed non-reciprocal (non-temporally contingent) gaze, the natural neurocognitive function of participants was to neglect the gaze. On the other hand, non-social cues are usually not reciprocal and adults are trained to use non-social cues for learning. In sum, participants had to provide extra effort when the anthropomorphic agent displayed non-reciprocal (non-temporally contingent) gaze, as it was against their natural neurocognitive system for social cues, whereas non-social cues did not suffer as they used different system.

In live group, the agent type did not affect learning. We suggest this was due to overtraining of non-social cues. That is, as participants of the experiments were all university students, they were extremely adept learners who were accustomed to using non-social cues while learning. Thus, while social cues and non-social cues may be operated by separate cognitive systems, participants were able to learn with same efficiency using either system.

In conclusion, our hypothesis is that social cues that are not temporally contingent cause negative effect as it violates natural attention system. However, future works are needed to validate the hypothesis, using neuroimaging technics. Regarding the social agency theory, we could not find evidence that social cues facilitate learning more than non-social cues. They may

still prove to be beneficial for complex learning materials, when cognitive load caused by learning is higher.

Chapter 7. Experiment 5: Lead-in vs Follow-in Joint attention

7.1 Introduction

Joint attention refers to a type of dyadic social interaction, where individual A detects that B's gaze is not directed towards them, and follows the line of sight of B onto a focus of attention (such as an object), so individuals A and B are looking at the same object (Emery, 2000).

How joint attention influences learning has long been the focus of research, particularly in the field of developmental science. Tomasello & Farrar (1986) investigated naturalistic interaction between mother and child, and reported that interactions within joint attention facilitated child's language learning. More importantly, whether the referenced object was already the focus of attention of the child had significant effect on child language. They followed this result with an experimental study, showing that children learned words better when the referenced object was already the focus of attention, compared to when their attention had to be

redirected to the object. This was labeled maternal ‘follow-in’ vs ‘lead-in’; follow-in refers to following the child’s attention, whereas lead-in refers to the mother leading the child’s attention. The result has been replicated by several studies. For example, longitudinal studies reported that parent’s responsiveness to the infant’s focus attention predicts early language development (Carpenter, Nagell, & Tomasello, 1998; Tamis-LeMonda et al., 2001).

Although the temporal order of contribution of joint attention (follow-in vs lead-in effects) in adult learning has not received attention from previous studies, recent development from the field of neuroscience indicates the possibility that follow-in and lead-in joint attention elicit different neural processes. Schilbach et al (L. Schilbach et al., 2010) found that when participant lead gaze interaction (follow-in), the brain area related to reward system was activated, which implies having learners lead gaze interaction could enhance motivational outcomes.

However, adults are much more capable of redirecting their attention than infants. Follow-in and lead-in joint attention may elicit different cognitive system, but if the effect is confined to motivational outcomes, it is expected that the learning effect would be difficult to observe, especially within short-term experiments. Moreover, such delicate effect would easily perish during post-input process; for example, when participants rehearse information after initial input. Humans, especially adults, use phonological loop in working memory to rehearse newly learnt words. To prohibit the rehearse process and preserve initial effect caused by interaction with the pedagogical agent as much as possible, the present experiment used articulatory suppression. Articulatory suppression is a well-established technic, in which participant is required to utter repeatedly a redundant speech sound. Articulatory suppression is used to

suppress subvocal rehearsal and disrupt phonological short-term memory (Papagno, Valentine, & Baddeley, 1991).

In this chapter, we explain an experiment that focused on follow-in vs lead-in joint attention. The experiment was designed to test if redirecting joint attention could potentially affect learning of adults, using a modified version of PAGI, JA-PAGI (Joint Attention-PAGI). As adults are more capable of shifting attention, we expected the gap to be small, thus we aimed to capture micro-differences in behaviours during interaction as well as the differences in learning effect.

7.2 Method

7.2.1 Participants

Eight participants (3 women) were recruited from a subject pool at the University of Tokyo. Their mean age was 21.38 (SD = 2.5) years. Participants were all native Japanese without Korean language experience. Additional one participant was not included in the sample due to falling asleep during the experiment.

7.2.2 Design

A repeated measures design was used, in which participants learned Korean words with JA-PAGI. JA-PAGI operated in two modes, follow-in mode and lead-in mode. In follow-in mode, JA-PAGI taught the name of the object which was the focus of attention of participant. In lead-in mode, JA-PAGI taught the name of the object which was not the focus of attention of participant. The only variable of the experiment was the percentage of trials (one trial for each word) in each mode; in the follow (Follow-In) condition, JA-PAGI was in follow-in mode in

80% of the trials, and in lead (Lead-In) condition, JA-PAGI was in lead-in mode in 80% of the trials. The order of condition was counter-balanced.

7.2.3 Joint Attention PAGI (JA-PAGI)

The overall visual of JA-PAGI is similar to PAGI, with small modification to disposition of objects.

JA-PAGI started with an opening narration, explaining that he will be teaching Korean words, while gazing at participant's eyes, initiating eye contact (Figure 7-1). After the narration, JA-PAGI initiated word learning phase. First, two pictures were presented (Stage 1). JA-PAGI waited for an eye contact. After eye contact, two pictures began to jerk up-and-down (Stage 2), and continued to do so until participant fixated on one of the pictures. Then, JA-PAGI shifted his gaze to the referenced picture (follow-in mode) or the other picture (lead-in mode). After 300 ms (time including gaze shift animation), JA-PAGI returned his gaze to participant and spoke the target Korean word once (Stage 3). PAGI repeated Stage 1-3 for each word.

During Stage 2, fixation threshold was set to 500 ms. This was to prevent JA-PAGI from responding to quick saccades when participant perform quick search on two pictures. A pilot test revealed that when the fixation threshold was set below 500 ms, wrong picture was sometimes set as the target, as participants had already shifted gaze to the other picture when JA-PAGI started to move its gaze.

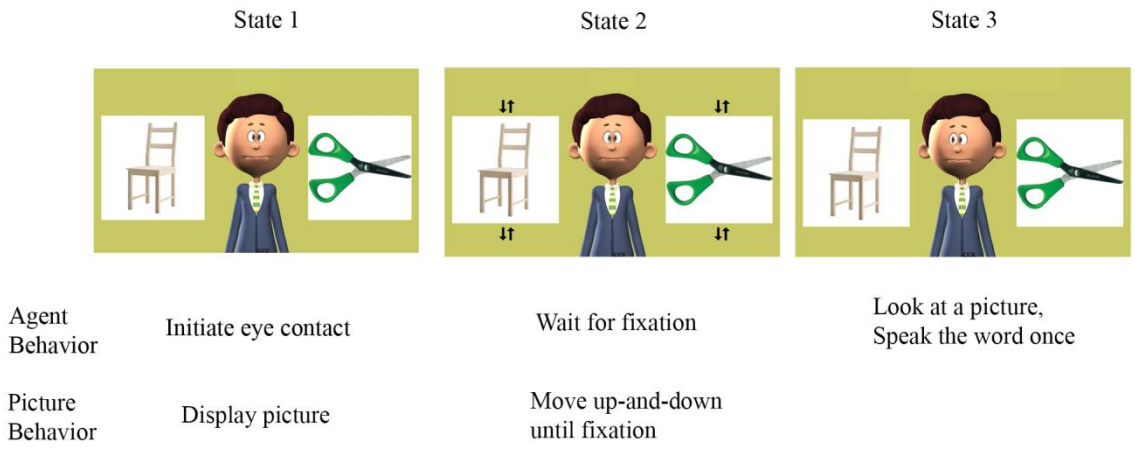

Figure 7-1Gaze interaction scheme of JA-PAGI. Black arrows from State 2 frame represents direction of movement and was not visible during experiment.

7.2.4 Materials

All dialogues were presented in Korean. Korean nouns with less than four syllables were used for the lesson. Each word was presented with a corresponding picture (Figure 7-1). The word list was identical for all participants, and presented order was randomized. The words were selected by a Korean-Japanese bilingual based on two criteria, low resemblance between the pair and familiarity to general population. Participants learned 40 words, which were divided into two blocks of 20 words, without any rest between the blocks.

7.2.5 Apparatus

The eye-tracker (Tobii EyeX, Sweden) was installed on a 23-inch LCD monitor, on which stimuli were displayed. A nine-point calibration was administered at the start of every

block. A webcam was placed on top of the monitor focused on the participant's face to monitor the gaze interaction. The same equipment was used for the test phase. Sound stimuli were presented through two speakers (BOSE Media Mate II).

7.2.6 Procedure

When the participants arrived at the laboratory, each was led separately to a room and seated in front of a monitor. The experimenter told each participant that the purpose of the experiment was to assess how humans learn foreign language and instructed them to engage in gaze interaction with the agent, by forming mutual gaze and following the agent's gaze. Also, they were instructed to fixate on one of the two pictures which they thought to be more interesting, when pictures were moving. Finally, participants were required to mouth repeatedly Japanese alphabets 'ア、イ、ウ、エ、オ' (pronounced /a/, /e/, /u/, /e/, /o/). This was a modified version of articulatory reduction. The modification was necessary as main learning material was presented in audio, the sound uttered by participant could interrupt learning. Experimenter signalled participant to start articulatory reduction 2s prior of each block.

The experiment was divided into two phases: learning and test. In the beginning of the learning phase, each participant was given a practice session, to get accustomed to JA-PAGI and rehearse articulatory reduction. The practice session was identical to the learning phase, except that instead of the Korean words, ten English alphabet letters—from 'a' to 'j'—were presented. Thus, each participant was given 10 trials (each alphabet counted as one trial) in the practice session.

The test phase was carried out immediately after the learning phase. Four pictures were presented on the screen and a Korean word was verbally given. Participants were instructed to pick the picture that corresponded to the word. Participants used number keys 1–4 on a keyboard

during the test to choose the picture. Participants were informed that there was 6s time limit during the test phase. The entire experiment lasted 25 - 30 min.

7.3 Result

The test score was composed of the number of correct answers. Participants scored higher than the chance level; 10 words as the test phase consisted of forty four-choice questions. (t (7) = 6.907, $p < 0.001$, one-sample t test). A paired sample t-test revealed that the mean test scores did not differ significantly between the two conditions (t (7) = -0.672, $p > 0.5$). (Figure 7-2)

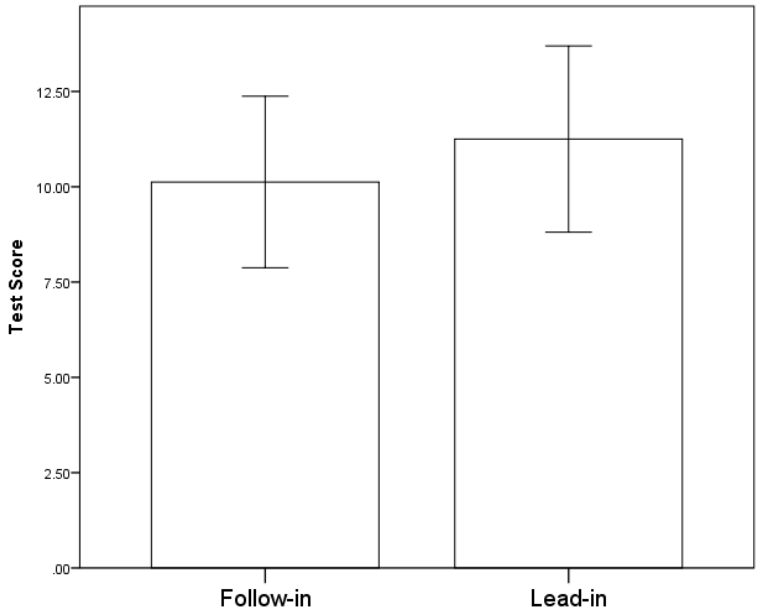

Figure 7-2 The average of test result for each condition. Error bars represent standard errors.

The eye tracking data were used to investigate gaze patterns. Each trial was grouped by 2x2 factors, block condition (B-follow/B-lead) and trial condition (T-follow/T-lead). Firstly, the average reaction time to lead-in condition trials were assessed. The average RT was 0.99 s (SD = 0.4). The lead-in condition RT was correlated to scores of B-lead (r = 0.705, p = 0.051) and

T-lead ($r = 0.685$, $p = 0.061$) trials with marginal significance, and T-follow ($r = 0.719$, $p < 0.05$) trials with statistical significance.

To investigate further, the time-course of eye movements in relation to JA-PAGI's behavior was assessed. The eye tracking data were divided into two groups by test score median, and onset plots of eye movements were measured (Figure 7-4, Figure 7-5). The comparison between two groups revealed that high-score group spent more time in Stage 2. Indeed, the duration of Stage 2 was related to the test scores ($r = 0.850$, $p < 0.01$, Figure 7-3). The duration was Stage 2 was also correlated with the average reaction time to lead-in condition trials ($r = 0.745$, $p < 0.05$). The duration of other stages were not related with the test scores or the average RT.

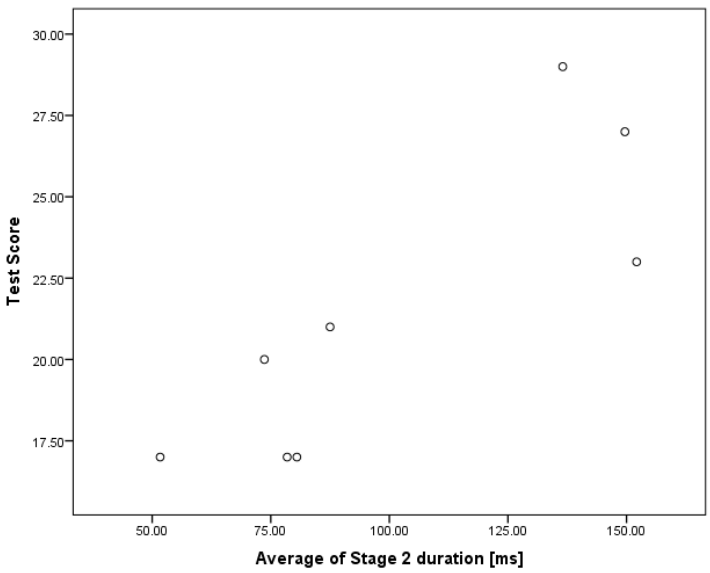

Figure 7-3 Scatterplot illustrating the relation between average Stage 2 duration and the test scores.

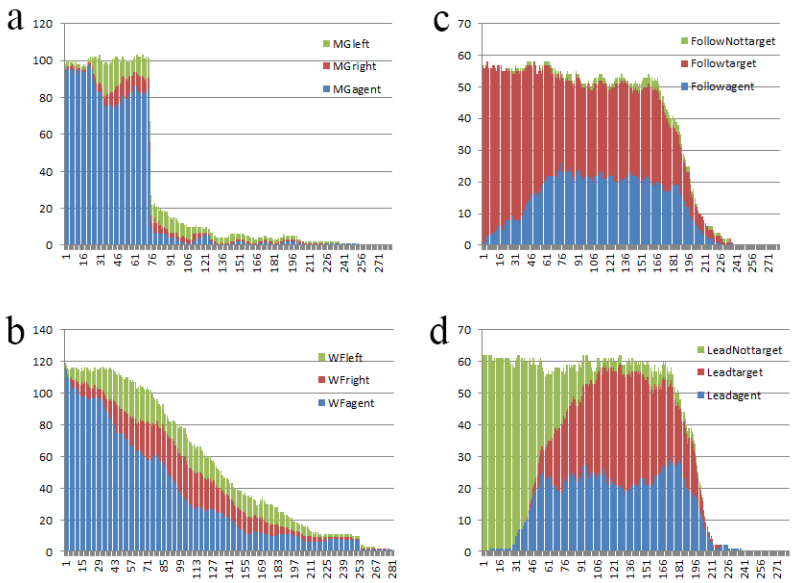

Figure 7-4 Onset plots of eye movements of high-score group. Vertical axis represents number of data samples, and horizontal axis represents time in frames. a) Stage 1 of JA-PAGI, b) Stage 2, c) Stage 3 follow trials, d) Stage 3 lead trials. For a) and b), blue-agent, red-right picture, green-left picture; for c) and d), blue-agent, red-target picture, green-non target object.

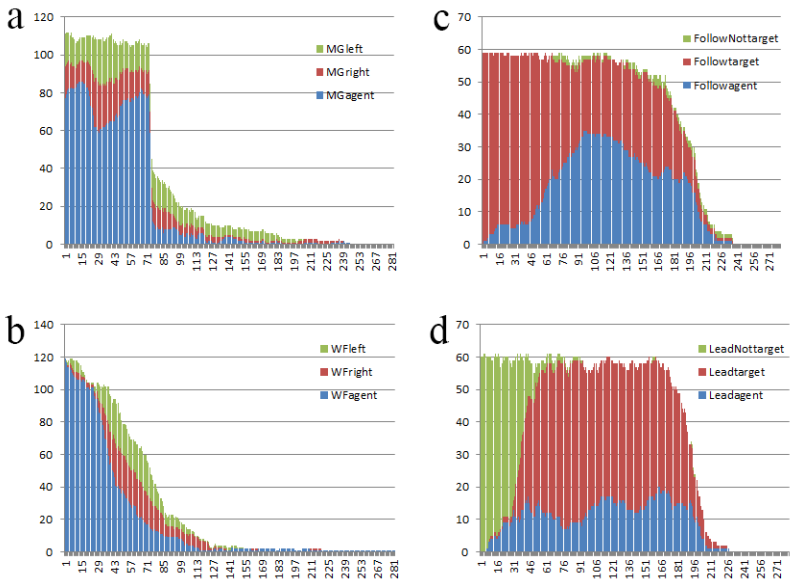

Figure 7-5 Onset plots of eye movements of low-score group. Vertical axis represents number of data samples, and horizontal axis represents time in frames. a) Stage 1 of JA-PAGI, b) Stage 2, c) Stage 3 follow trials, d) Stage 3 lead trials. For a) and b), blue-agent, red-right picture, green-left picture; for c) and d), blue-agent, red-target picture, green-non target object.

7.4 Discussion

The experiment did not reveal whether follow-in and lead-in joint attention affect learning of adults differently. By investigating eye behavior patterns, we found a major flaw in the experimental design.

As shown in the result, time spent on the stage prior to JA-PAGI's gaze movement was strongly related to higher scores. This infers that high-score participants were able to adopt a strategy using lengthy fixation threshold during Stage 2 of JA-PAGI; deliberately postponing interaction and taking time to identify both pictures before JA-PAGI turned its gaze to the target picture. The result implies that the distribution of attention during interaction was not the same among participants and it affected the test scores.

Disappointingly, the present experiment confirmed our initial concern that the effect of follow-in vs lead-in joint attention in adults would be small and the experiment can easily be affected by other cognitive processes. Future research calls for more rigorous experimental design; for example, increasing the number of pictures.

Also, future research may benefit by focusing on individual differences. Follow-in vs lead-in is known to be affected by receiver's individual difference. Differences in gaze following of infants predict language development (Brooks & Meltzoff, 2005). Scott et al. (2013) showed that infants' tendency to respond to lead-in joint attention were strongly related to vocabulary growth. While adults are generally capable of displaying joint attention, close observation of individual differences may help us predict various cognitive functions, including learning efficiency.

Chapter 8. General Discussion

8.1 Summary of the findings

The study implemented reciprocal aspect of social interaction into pedagogical agents, and tested its learning effects. Two important features were tested, temporal contingency and temporal order of contribution in joint attention. Five experiments reported in this thesis provided important finding but not without limitations. The main findings of this study are summarized as follows.

1) The mere physical image of pedagogical agent did not facilitate learning.
2) For social cues, temporal contingency facilitated learning.
3) For non-social cues, temporal contingency did not facilitate learning.
4) Social cues were not necessarily superior compared to non-social cues.
5) The short-term learning effect of temporal order of contribution in joint attention may be small.

This study provides first empirical evidence that reciprocal aspect of social interaction has practical implications for computer-based education. However, not all of the results

presented in this study are straight forward. We will discuss implications and limitations of these results in the next subchapters.

8.2 Does reactivity of pedagogical agent facilitate learning effect?

In Chapter 4, we presented Experiment 2, which showed temporally contingent gaze interaction elicits more learning effect than non-contingent gaze interaction. This finding adds to the social agency theory, and implies that reactivity is indeed a key aspect for implementing social interaction into computer-based education. The result also concurs with recent developments in social neuroscience, which suggest that social cognition may be fundamentally different when individuals are interacting with others rather than merely observing (Anders et al., 2011; Leonhard Schilbach et al., 2013).

Results from Experiment 3 and 4 (Chapter 5 and 6) suggest that the temporal contingency effect is specific to social-cues. When paired with non-social cues (i.e. arrow), temporal contingency did not facilitate learning. The possible explanation for the result is, as non-social cues are generally not reciprocal, adults are trained to react to non-social cues that are not-contingent. By comparison, non-contingent social interaction may trigger incomplete activation of social attention system, thus demanding more effort from recipients.

The social communication of human, both verbal and non-verbal, is rich and unique. Reactivity is one of its major features, which in turn requires temporal contingency. Our result suggests the possibility that absence of such major feature has greater affect than previously assumed. Not only it affects functional role as a communicational tool, it may damage

accompanying cognitive activity, such as learning. The notion proposes certain challenge for pedagogical agent literature, as it implies many more features should be implemented for an agent to function well; if an agent lacks important aspect of social communication, the agent may inflict harm on learning.

While we proposed several hypotheses, mechanism between reactivity and learning is not made clear in this study. Future work would benefit by taking brain measurement during gaze contingent interactions. For example, pre-stimulus medial temporal theta (4-8Hz) activity has been shown to be related to successful encoding of words (Guderian, Schott, Richardson-Klavehn, & Duezel, 2009), suggesting that the theta activity reflects preparatory state when anticipating information. It would be interesting to see relation between medial temporal theta activity and reactivity, and how it affects learning outcomes.

8.3 Social cue vs non-social cue in relation with individual difference: age difference and competence.

The social agency theory states that learners benefit from social stance primed by pedagogical agents. However, results from Experiment 3 and 4 (Chapter 5 and 6) seem to contradict the notion, as non-human agents produced same learning effects compared to human agent.

The biggest difference between human agent and arrow agents used in this study was that human agent represented social cues, whereas arrow agents represented non-social cues.

While two types of cues are known to elicit similar response, attention systems behind each cue are argued to be different. Moreover, the natural pedagogy theory (Csibra & Gergely, 2009) argues that human infants are born with ability to utilize social cues to gather

meta-information crucial for learning, as seen from their specific response to social cues. Indeed, humans may have dedicated or unique attention system for social cues, which is strongly connected to learning from early stages.

However, adults are capable of learning without social cues. This may be a result of training, or adaptation. This is supported by the fact that young children are affected more by the lack of social cues. Indeed, children under three years of age learn less from screen media than from engaging in live social interaction with adults (M. Krcmar et al., 2007; Kuhl et al., 2003). Moreover, the video deficit effect is mitigated when live social interaction is provided through screen media (Nielsen et al., 2008; S. Roseberry et al., 2014; Troseth et al., 2006).

If this is true, pedagogical agent, the aim of which is to provide social cues in computer-based learning environment, would be more beneficial for younger age group. Indeed, recent review on pedagogical agent research showed that pedagogical agents produced larger effect size for K-12 student than older student group (Schroeder et al., 2013). However, there has been only small number of studies that targeted younger group, and these studies targeted rather older age group within K-12, grade 4-6 (Holmes, 2007; Kizilkaya & Askar, 2008; Moreno et al., 2001; Plant, Baylor, Doerr, & Rosenberg-Kima, 2009). The literature would benefit with more studies dedicated to younger group (i.e. K-3). Especially, it would be interesting to see how agents interact with young children who are yet unable to learn intentionally.

The PAGI fits the need nicely, as it can be easily modified for testing different age groups. The task used by PAGI is word learning, which is one of the most well-studied task paradigms with younger children. PAGI can even work with pre-verbal children, as it can utilize IPLP (Intermodal Preferential Looking Paradigm), which is an experimental paradigm used for testing young infants (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013).

Individual difference in adults is also an interesting topic. While adults are capable of learning with non-social cues (Chapter 5 and 6), social cues may still produce lower cognitive load. Thus, social cues may be beneficial when the cognitive load caused by learning is large enough to presser limited attention resource. Choi & Clark (2006) compared learning effect of human agent and arrow agent, and found that human agent produced better learning effect from students with low prior knowledge, but not from students with intermediate and high prior knowledge. This may have been caused by differences in required cognitive load between participants.

Also, competence in social interaction varies among adults. Scott et al. (2013) showed that infants' tendency to respond to lead-in joint attention were strongly related to vocabulary growth. The individual difference may also affect learning of adults. While the result of Experiment 5 (Chapter 7) was most likely caused by difference in strategy adopted by participants, it showed that distribution of attention during interaction is strongly related to learning outcomes. It would be interesting to test adult with different competence in social interaction and learning, to see how two factors correlate.

In sum, future work is needed to assess situations where social cues can be beneficial, in terms of learning material, and learner differences.

8.4 The merit of using virtual agents in psychological experiments

The present study also proposes the merit of using artificial agents in the field of psychology as a tool for assessing human-to-human interactions. The merits of utilizing virtual environments in psychological experiments has been discussed (Blascovich et al., 2002) , and has been successfully implemented in social neuropsychology(L. Schilbach et al., 2010; L. Schilbach et al., 2006). In this study, we showed that strict control over interaction features (e.g. temporal contingency, temporal order of contribution in joint attention) is possible, by using artificial agent. This would have been difficult for human experimenter. Therefore, we believe artificial agents have potential as experimental tools. For example, one construct that can benefit forthwith from using artificial agents is the video deficit effect. Currently, the most common practice when examining the video deficit effect is to use human experimenters as a counterpart to video stimuli (Marina Krcmar, 2010; M. Krcmar et al., 2007; Sarah Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009). There are two limitations to this experimental design. First, strict control over variables cannot be ensured. Second, the granularity of the experiment is limited by the precision of human perception. For example, the latency of mutual gaze cannot be controlled in milliseconds, as can be done with artificial agents. We believe artificial agents open new possibilities for assessing subtle human behaviours, precisely and cost-efficiently.

Chapter 9. Conclusion

We designed an experimental purpose pedagogical agent, capable of reciprocal gaze interaction, and tested its learning effects. We focused on temporal contingency and temporal order of contribution in joint attention. Five experiments were conducted. The main findings of this study have shown that temporally contingent social cues facilitate learning.

Although the present study succeeded in adding reactive, reciprocal aspect to the scope of social interaction displayed by pedagogical agents, several important questions both theoretical and practical were raised. Firstly, the present study indicated the possibility that social cues may produce negative effect when crucial features, such as temporal contingency is missing. As social attention system is designed for reciprocal interaction, non-contingent social interaction may trigger incomplete activation, thus demanding more effort from recipients. By comparison, non-social cues are generally not reciprocal, thus general attention system is trained to react to non-social cues that are not-contingent. This hypothesis is based on the theory from neurocognitive science which suggests social cues are handled by unique attention system. However, evidence is lacking to support the hypothesis, and future research is required, from multidisciplinary fields. Secondly, the present study suggested that adults are capable of learning with non-social cues as well as with social cues, unlike young children. The comparison between age group is needed, and future research will focus on the topic.

Reference

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *Neuroimage, 54*(1), 439-446. doi: 10.1016/j.neuroimage.2010.07.004

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist, 48*(5), 505-522. doi: 10.1177/0002764204271506

Andre, E., Rist, T., & Muller, J. (1999). Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence, 13*(4-5). doi: 10.1080/088395199117333

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2). doi: 10.1037//0022-0663.94.2.416

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology, 30*(1), 117-139. doi: 10.1016/j.cedpsych.2004.07.001

Bailenson, J. N., & Yee, N. (2005). Digital chameleons - Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16*(10), 814-819. doi: 10.1111/j.1467-9280.2005.01619.x

Baylor, A. L. (2011). The design of motivational agents and avatars. *Etr&D-Educational Technology Research and Development, 59*(2), 291-300. doi: 10.1007/s11423-011-9196-3

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25*(2), 450-457. doi: 10.1016/j.chb.2008.10.008

Baylor, A. L., & Ryu, J. (2003). The effects of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research, 28*(4), 21.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & Teachable Agents Grp, V. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392. doi: 10.1080/08839510590910200

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103-124. doi: 10.1207/s15327965pli1302_01

Bodenheimer, B., Williams, B., Kramer, M. R., Viswanath, K., Balachandran, R., Belynne, K., & Biswas, G. (2009). Construction and Evaluation of Animated Teachable Agents. *Educational Technology & Society, 12*(3), 191-205.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535-543. doi: 10.1111/j.1467-7687.2005.00445.x

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), V-143.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable Agents and the Protege Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology, 18*(4), 334-352. doi: 10.1007/s10956-009-9180-4

Choi, S., & Clark, R. E. (2006). Cognitive and Affective Benefits of an Animated Pedagogical Agent for Learning English as a Second Language. *Journal of Educational Computing Research, 34*(4), 25. doi: 10.2190/A064-U776-4208-N145

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2). doi: 10.1037//0022-0663.94.2.428

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153. doi: 10.1016/j.tics.2009.01.005

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*(5), 377-398. doi: 10.1016/j.ijhcs.2012.01.004

Deaner, R. O., & Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Current Biology, 13*(18), 1609-1613. doi: 10.1016/j.cub.2003.08.025

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies, 52*(1). doi: 10.1006/ijhc.1999.0325

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated Gaze-Contingent Objects Elicit Orientation Following in 8-Month-Old Infants. *Developmental Psychology, 47*(6), 1499-1503. doi: 10.1037/a0025659

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education, 49*(3), 677-690. doi: 10.1016/j.compedu.2005.11.010

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581-604. doi: 10.1016/s0149-7634(00)00025-7

Farroni, T., Csibra, G., Simion, G., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9602-9605. doi: 10.1073/pnas.152159999

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313-331. doi: 10.1016/j.learninstruc.2005.07.001

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America, 100*(13), 8030-8035. doi: 10.1073/pnas.1332441100

Golinkoff, R. M., Ma, W. Y., Song, L. L., & Hirsh-Pasek, K. (2013). Twenty-Five Years Using the Intermodal Preferential Looking Paradigm to Study Language Acquisition: What Have We Learned? *Perspectives on Psychological Science, 8*(3), 316-339. doi: 10.1177/1745691613484936

Guderian, S., Schott, B. H., Richardson-Klavehn, A., & Duezel, E. (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences of the United States of America, 106*(13), 5365-5370. doi: 10.1073/pnas.0900289106

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27-54. doi: 10.1016/j.edurev.2010.07.004

Hietanen, J. K., Leppanen, J. M., Nummenmaa, L., & Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain Research, 1215*, 123-136. doi: 10.1016/j.brainres.2008.03.091

Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hamalainen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMR1 study. *Neuroimage, 33*(1), 406-413. doi: 10.1016/j.neuroimage.2006.06.048

Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education, 48*(4), 523-547. doi: 10.1016/j.compedu.2005.02.007

Johnson, W. L., Rickel, J., Stiles, R., & Munro, A. (1998). Integrating pedagogical agents into virtual environments. *Presence-Teleoperators and Virtual Environments, 7*(6). doi: 10.1162/105474698565929

Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning, 23*(3). doi: 10.1111/j.1365-2729.2006.00210.x

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition, 55*(2), 269-271. doi: 10.1016/j.bandc.2004.02.037

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Kizilkaya, G., & Askar, P. (2008). The effect of an embedded pedagogical agent on the students' science achievement. *Interactive Technology and Smart Education, 5*(4), 208-216. doi: doi:10.1108/17415650810930893

Krcmar, M. (2010). Can Social Meaningfulness and Repeat Exposure Help Infants and Toddlers Overcome the Video Deficit? *Media Psychology, 13*(1). doi: 10.1080/15213260903562917

Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology, 10*(1), 41-63.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America, 100*(15), 9096-9101. doi: 10.1073/pnas.1532872100

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology, 42*(4), 627-642. doi: 10.1037/0012-1649.42.4.627

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). *The persona effect: affective impact of animated pedagogical agents*. Paper presented at the Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, Atlanta, Georgia, USA.

Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9*(1-2), 1-44.

Louwerse, M. M., Graesser, A. C., Lu, S. L., & Mitchell, H. H. (2005). Social cues in animated conversational agents. *Applied Cognitive Psychology, 19*(6), 693-704. doi: 10.1002/acp.1117

Mayer, R. E., & DaPra, C. S. (2012). An Embodiment Effect in Computer-Based Learning With Animated Pedagogical Agents. *Journal of Experimental Psychology-Applied*, *18*(3). doi: 10.1037/a0028616

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology*, *95*(4). doi: 10.1037/0022-0663.95.4.806

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/s15326985ep3801_6

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*(3), 283-298. doi: 10.1016/s0193-3973(99)00040-4

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358-368. doi: 10.1037//0022-0663.91.2.358

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, *96*(1), 165-173. doi: 10.1016/0022-0663.96.1.165

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*(1), 117-128. doi: 10.1037/0022-0663.97.1.117

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-213. doi: 10.1207/s1532690xci1902_02

Mori, M. (1970). The uncanny valley (Originally in Japanese: Bukimi no tani.). *Energy (Oxford), 7*(4), 33-35.

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*(3), 253-261. doi: 10.1046/j.0266-4909.2001.00237.x

Mousavi, S. Y., Low, R., & Sweller, J. (1995). REDUCING COGNITIVE LOAD BY MIXING AUDITORY AND VISUAL PRESENTATION MODES. *Journal of Educational Psychology, 87*(2), 319-334. doi: 10.1037/0022-0663.87.2.319

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science, 11*(5), 722-731. doi: 10.1111/j.1467-7687.2008.00722.x

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences, 13*(3), 135-143. doi: 10.1016/j.tics.2008.12.006

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). The power of human gaze on infant learning. *Cognition, 128*(2), 127-133. doi: 10.1016/j.cognition.2013.03.011

Olsson. (2007). *Real-time and offline filters for eye tracking.*, KTH Royal Institute of Technology.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71. doi: 10.1207/s15326985ep3801_8

Paas, F., & Vanmerrienboer, J. J. G. (1994). VARIABILITY OF WORKED EXAMPLES AND TRANSFER OF GEOMETRICAL PROBLEM-SOLVING SKILLS - A COGNITIVE-LOAD APPROACH. *Journal of Educational Psychology, 86*(1), 122-133. doi: 10.1037/0022-0663.86.1.122

Papagno, C., Valentine, T., & Baddeley, A. (1991). PHONOLOGICAL SHORT-TERM-MEMORY AND FOREIGN-LANGUAGE VOCABULARY LEARNING. *Journal of Memory and Language, 30*(3), 331-347. doi: 10.1016/0749-596x(91)90040-q

Plant, E. A., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education, 53*(2), 209-215. doi: 10.1016/j.compedu.2009.01.013

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage, 50*(4), 1639-1647. doi: 10.1016/j.neuroimage.2010.01.052

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination - Common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407-413. doi: 10.1111/j.1467-9280.2007.01914.x

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype Me! Socially Contingent Interactions Help Toddlers Learn Language. *Child Development, 85*(3), 956-970. doi: 10.1111/cdev.12166

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live Action: Can Young Children Learn Verbs From Video? *Child Development, 80*(5), 1360-1375.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393-414. doi: 10.1017/s0140525x12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds Made for Sharing: Initiating Joint Attention Recruits Reward-related Neurocircuitry. *Journal of Cognitive Neuroscience, 22*(12), 2702-2715. doi: 10.1162/jocn.2009.21401

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44*(5), 718-730. doi: 10.1016/j.neuropsychologia.2005.07.017

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). HOW EFFECTIVE ARE PEDAGOGICAL AGENTS FOR LEARNING? A META-ANALYTIC REVIEW. *Journal of Educational Computing Research, 49*(1), 1-39. doi: 10.2190/EC.49.1.a

Scott, K., Sakkalou, E., Ellis-Davies, K., Hilbrink, E., Hahn, U., & Gattis, M. (2013). *Infant contributions to joint attention predict vocabulary development.* Paper presented at the Proceedings of the 35th Annual Conference of the Cognitive Science Society.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9). doi: 10.1016/j.cub.2008.03.059

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition, 108*(2), 303-319. doi: 10.1016/j.cognition.2008.02.009

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology-Human Perception and Performance, 29*(2), 326-332. doi: 10.1037/0096-1523.29.2.326

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology, 3*(3), 289-299. doi: 10.1080/17405620600879779

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review, 22*(2), 123-138. doi: 10.1007/s10648-010-9128-5

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767. doi: 10.1111/1467-8624.00313

Tomasello, M., & Farrar, M. J. (1986). JOINT ATTENTION AND EARLY LANGUAGE. *Child Development, 57*(6), 1454-1463. doi: 10.1111/j.1467-8624.1986.tb00470.x

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development, 77*(3), 786-799. doi: 10.1111/j.1467-8624.2006.00903.x

Veletsianos, G. (2009). The impact and implications of virtual character expressiveness on learning and agent-learner interactions. *Journal of Computer Assisted Learning, 25*(4), 345-357. doi: 10.1111/j.1365-2729.2009.00317.x

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies, 66*(2). doi: 10.1016/j.ijhcs.2007.09.003

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes-using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience, 5*(1), 98-107. doi: 10.1093/scan/nsq024

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology, 107*(2), 118-136. doi: 10.1016/j.jecp.2010.04.014

Acknowledgments

I would like to express my special appreciation and thanks to my advisor Professor Dr. Hiraki Kazuo, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. I would especially like to thank Dr. Kanakogi for aids in experimental design and writing of my Ph.D. thesis. You have been there to support me throughout this research.