Video-mediated and co-present gameplay: Effects of mutual gaze on game experience, expressiveness and perceived social presence

Suleman Shahid *, Emiel Krahmer, Marc Swerts

Tilburg Center for Cognition and Communication (TiCC), Department of Communication and Information Sciences, School of Humanities, Tilburg University, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands

A R T I C L E   I N F O

Article history:
Available online 18 May 2012

Keywords:
Video-mediated communication
Mutual gaze
Social presence
Games
Fun
Children

A B S T R A C T

We study how pairs of children interact socially and express their emotions while playing games in different communicative settings. In particular, we study how such interactions can vary for environments that differ regarding the level of mediation and the associated feelings of social presence. Overall, the study compared three conditions (one face-to-face gameplay condition, and two video-mediated gameplay conditions; one allowing for mutual gaze, the other not) and focused on the social presence and non-verbal behavior of children in three conditions. The results show that the presence of mutual eye-gaze enriches the feelings of social presence, fun and game experience; conversely, the absence of mutual eye-gaze dramatically affects the quality of interaction in the video-mediated environment. The results of this study stress the importance of mutual gaze, and we therefore argue that it should become an integral component of future VMC systems, particularly in those designed for playful settings and children.

1. Introduction

These days communicating via digital media is easier than ever before; technically, it is possible to communicate with friends or colleagues anywhere in the world, and be able to see them all the while on the screen of our communication device. Researchers have long wondered to what extent this kind of video-mediated communication allows for a similar feeling of social presence as face-to-face communication does, and it has turned out that this is not an easy question to answer. It has been argued that the kind of task, and hence the kind of communication that is required, will have a strong impact on this (Whittaker, 2002), but nevertheless the majority of previous research has looked at settings in which adult users work on professional tasks. In addition, the methodology used in these studies primarily consists of usability questionnaires and task performance, so that we currently, at best, have an incomplete view of the role that social *x* plays in mediated communication.

What, for instance, if the task is a playful one, such as playing a game? Whittaker (2002) points out that face-to-face communication is different from computer-mediated communication when affect and emotion play a role, which is generally not the case in professional communication. And what if, instead of adults, children are studied? It is known from the literature that gaze behavior (an important pre-condition for perceived social presence) varies as a function of age, and it is conceivable that this impacts the comparison of computer-mediated with face-to-face communication.

In this paper, we try to answer these questions, looking at social presence in video-mediated communication of children during playful interactions, thereby extending our understanding of the comparison between video-mediated and face-to-face communication. Our methodology is as follows: we let pairs of children play games in one of three conditions (one face-to-face condition, and two video-mediated ones; one allowing for mutual gaze, the other not). We study how the children behave in these various conditions, with special attention to their non-verbal behavior. We compare the expressiveness and non-verbal cues in the three conditions, to get a better understanding of how children experience these conditions, and to what extent they experience social presence in video-mediated gameplay in the same way as they do in a face-to-face setting.

1.1. Video-mediated communication

Video-mediated communication (VMC) systems allow remote users to communicate synchronously using verbal, nonverbal and interpersonal cues (Finn et al., 1997; Giannakos et al., 2011). One of the most important features of the VMC medium is its ability to replicate a communication experience which is close to face-to-face communication, by supporting the feeling of social presence amongst its users (Kirk et al., 2010). Social presence, defined
as “the degree of awareness of another person in the interaction and the subsequent appreciation of an interpersonal relationship” (Short et al., 1976), during mediated communication is known to positively effect the feeling of psychological accessibility and enhances affective connectedness (Biocca et al., 2001; Tu and McIsaac, 2002). In VMC systems, many factors contribute to the creation of social presence (Biocca et al., 2003; Kim et al., 2011). These factors include system characteristics such as audio–video quality and delay and image size (McCulloch and Oboine, 1999; Tang, 1992), but also the extent in which VMC allows for the transmission of visual cues such as gestures, facial expressions and gaze (Annetta et al., 2009; Bruce, 1996; O’Malley et al., 1996).

Often research on VMC is couched in terms of the media richness theory, originally proposed by Daft and Lengel (1984) (and see e.g., (Whittaker, 2002), for critical discussion). According to this theory, a “medium’s capacity for immediate feedback, the number of cues transmitted, channels utilized, personalization, and language variety” (Daft and Lengel, 1986) all influence its degree of information richness. According to media richness theory, face-to-face communication has the highest richness and numeric communication (e.g., numeric spread sheets) the lowest. The question is, of course, where VMC systems fall on this continuum. Clearly, this will depend on which visual cues are enabled by a VMC system.

In this study, we zoom in on one particular visual cue, namely ‘mutual gaze’, considered to be the most sophisticated form of gaze-awareness (Cook, 1977; Fagel and Blanche, 2010; Kleineke, 1986), and we investigate its role in creating social presence and enriching social interaction in VMC systems. It has been argued that, as in face-to-face communication, VMC systems should support mutual gaze between two parties (Grayson and Monk, 2003), because lack of gaze not only effects the system’s overall usability and the naturalness of the medium (Ohno, 2005) but also diminishes the sense of intimacy between remote parties (Karahalios, 2004; Mukawa et al., 2005).

However, even though prior research explored the effect of different levels of gaze-awareness on the quality of VMC systems, the findings of these studies have not always been straightforward and a number of important aspects have not received sufficient attention to date. In many studies, the absence of mutual gaze in VMC systems was taken as a limitation of the medium (Mukawa et al., 2005; Vertegaal and Ding, 2002), and most of the time its effect was only measured on the system’s usability and users’ task performance (Grayson and Monk, 2003). It has also been reported that the success of VMC systems largely depends on the situation in which it is used in and the tasks it is designed for (O’Malley et al., 1996), but the contexts explored in VMC research have often been limited to settings where remote participants, generally adults, were involved in professional tasks related to business, work or education (Knoblauch, 1999; Masoodiana et al., 2000).

In this paper, we focus on the role of mutual gaze in a non-professional, game playing context and study how the presence and absence of mutual gaze during video-mediated gameplay effects the perceived social presence and game experience of physically-apart children. Even though the focus of this study is on the role of mutual gaze during collaborative gameplay, we also compare the mediated gameplay conditions with a face-to-face condition to see to what extent the mediated experience resembles that of a natural co-present environment in this more playful context.

1.2. Video-mediated communication in playful settings

Although many VMC systems have been designed for a variety of purposes and in a number of fields, research in the area of VMC has been primarily directed to work related and business settings (Finn et al., 1997). Examples include systems for distant collaborative meetings and computer-supported cooperative work (Panteli and Dawson, 2010), healthcare related activities (Demiris et al., 2009), and for various distance education activities (Fels and Weiss, 2001). Furthermore, a number of high production and resolution systems were designed and deployed by big corporations, such as HP’s Halo and Cisco’s Tele-Presence systems (Szegedi et al., 2009). These systems were all primarily developed for adults, and associated studies for evaluating these systems also tended to focus on adult user groups (Karahalios and Donath, 2004; Karahalios, 2004; Mueller et al., 2003).

As video technology is becoming more affordable and accessible, VMC systems are started to be used for more frequently in home settings, for casual, non-work and social interactions (Wheatley and Basapur, 2009). The use of VMC technology for playful interactions in non-work settings, such as game playing, is still understudied. Furthermore, in these playful situations, children are a natural group to target, but not much research has been conducted with regard to VMC and children (Bobick et al., 1999; Giannakos et al., 2011).

There are only a handful of studies where researchers compared children’s gameplay in VMC and face-to-face settings (Batcheller et al., 2007; Yarosh et al., 2009, 2010). Yarosh et al. (2010) for instance, explored various prototypes for free play over videoconferencing. Their (mostly qualitative) results suggest that managing visibility and attention made this kind of computer-mediated play more difficult than face-to-face play. Although, studies such as these yielded interesting findings about playful interactions during VMC, they are arguably still preliminary and indecisive, as the authors note themselves. For example, it remains unclear which factors play a role in making a VMC experience as playful and fun as a face-to-face experience, data-collection was not systematic and non-verbal behavior of children was not analysed.

1.3. Mutual gaze

Do we feel our game partner is more socially present when he/she sends us a message in an online game? Does the level of social presence increase when we are engaged with him/her in an audio only or an audiovisual call? The concepts of being “socially present” and “socially absent” have become increasingly relevant after the arrival of modern communication systems (Biocca et al., 2003). These communication technologies, despite their diversity, can be said to share a common goal, which is to improve social presence during interactions (Biocca and Harms, 2002). Video-mediated communication systems are no exception in this regard.

Video-mediated communication systems capture and transmit a wide range of visual cues during dyadic interactions, and it has been argued that these visual cues contribute considerably to the feeling of social presence among remote participants during mediated interactions (Biocca et al., 2001). It has also been suggested that different visual cues in video-mediated communication are particularly helpful for children because these cues offer better means for conversation and support playfulness more strongly during remote communication (Ballasas et al., 2009). One important visual cue, which received considerable attention in the body of VMC research, is ‘mutual gaze’. Mutual gaze and its impact on perceived social presence is also the main focus of our study.

Previous research has shown the crucial role of eye gaze in establishing a social bond between partners, communicating emotional and interpersonal information, providing feedback, avoiding distraction and regulating conversations (Argyle and Dean, 1965). A review of the HCI literature also shows a strong connection between presence and absence of gaze and perceived social presence particularly in virtual environments (Bailenson et al., 2001), during human–robot interaction (Fong et al., 2003), and during avatar mediated communication (Garau et al., 2001; Steptoe et al.,
Gaze-awareness seems to have an effect on task performance and perceived social presence during VMC (Bondareva et al., 2006).

Different studies of the development of eye contact behavior in children have also found that eye contact is an important conversation resource for children (Argyle, 1972; Mirenda and Eisele, 1983) and from their early ages they start using it in a systematic manner for regulating conversations (Krantz et al., 1983). The amount of eye contact is lowest during ages 4–5 (pre-school) but rapidly increases and reaches the highest level around age 7–8. It decreases during early adolescence and after the ages 15–16 it again gradually increases to a more standard adult level (Asheer and Snortum, 1971; Levine and Sutton-Smith, 1973). Since the use and importance of gaze thus is known to vary with different age groups, it seems important to study the role of gaze during VMC with children, something which has not been done before.

A number of approaches have been used for evaluating the effect of mutual gaze on different aspects of VMC systems. Most of the time, the effect of mutual gaze has been investigated either on the system’s usability (where traditional usability measures are used for evaluating the system) or on the user’s performance (where different efficiency related measures are used). In this paper, contrary to much previous work, we focus on investigating the effect of presence or absence of mutual gaze on perceived social presence of children, in particular their non-verbal expressiveness, in playful situations. In studies on non-verbal behavior, it is usually assumed that non-verbal cues are important component of human communication (Patterson, 1990) and may offer important information about what speakers think or feel (Knapp and Hall, 2007). For example, Barkhuysen et al. (2005) found that non-verbal expressions of user’s can be used as a reliable cue to see whether problems in human–machine interaction occurred. There is also evidence that non-verbal behavior is an important factor for creating social presence and enhancing mutual understanding during mediated communication (Ruggenberg et al., 2005; Walther, 1996). For example, Bruce (1996) concluded that people use non-verbal cues during video communication for understanding each other. In this study, we investigate to what extent non-verbal cues can be indicative of the amount of fun and engagements child user’s experience in the respective conditions.

2.1. Material: the card game

In this study, we used a card game, developed in Shahid et al. (2008a), as a tool for investigating our research questions. In this game, players have to guess whether the next number of a sequence will be bigger or smaller than the previous (reference) number. When the game starts, players see a row of six cards on the screen where the number of the first card is visible (‘1’ in the case of the example in Fig. 1) and the other five cards are placed upside down so the numbers are hidden. All the numbers on the cards are between 1 and 10, and a number displayed once is not repeated in a particular game. Players have to guess whether the number on the next card will be higher or lower than the previous number. Once players have guessed the number, the relevant card is turned around and the correct answer is visible on the screen. Players are also informed about the correctness or incorrectness of their guess via a characteristic non-speech sound. If players make a wrong guess, the game is finished and they move to the next game. If players guess the number correctly then they are asked to guess the next number and players only win a game if they guess all numbers in the row correctly.

The card game was developed using Adobe Flash®. Appropriate colorful images were chosen for the game background and different animations were used to turn cards around in an attempt to make the game more attractive for children. During the session, children played six games, and could in theory win six coins (one for each game that the won). However, unknown to the children, each game was completely deterministic, and two different game alternatives were used. In the first alternative, a rational decision procedure would result in winning the game, and in the second alternative, being rational would result in losing the game. Fig. 1 is an example of the losing variant: the most probable outcome for the final card would be a “higher” number than 3, but guessing “higher” would make this a losing game. One winning scenario is shown in Fig. 2 where 10 is shown on the second last card. No number can be bigger than 10 and that is why children were expected to win this game. Winning and losing games were mixed in the sequence, starting and ending with a variant in which children were likely to win.

Before the actual gameplay session started, children played a practice game. During the gameplay, children were encouraged to discuss every next guess with their partner.

2.2. Participants

Eighty-eight Dutch children ($M_{age} = 7.87$ years, $SD = .53$) participated in this study of which 44 children (22 pairs) played the game in the ‘mutual gaze’ condition and the remaining 44 children played the game in the ‘no gaze’ condition. To compare the two mediated conditions (‘no gaze’ and ‘mutual gaze’) with a non-mediated, co-present condition, we used the data from our earlier study. In that study, 44 children ($M_{age} = 7.95$ years, $SD = .71$) partic-
ipated in the experiment and played the card game while sitting next to each other. All together this resulted in 132 participants spread over three conditions (mediated-mutual gaze, mediated-no gaze, non-mediated/co-present). The gap between mediated experiments and non-mediated experiment was approximately 1.5 years. All children were native speakers of Dutch and were in the group 4 of Dutch elementary school system. In all conditions we balanced for gender (equal number of boys and girls). Parents and teachers gave prior written permission for their child to participate, and signed a consent form stating that the recordings could be used for research purposes.

2.3. Setup

2.3.1. Mediated conditions

The experimental setup was the same for both mediated conditions, except for the gaze manipulation. Children played the game in pairs, and each child was directed to a different room. Communication between the rooms was made possible via a wireless Internet connection. The experimental setup for children was the same in both rooms. Two adults supervised the experiment. During the actual experiment, one was in a separate space, and could control the Card Game and observe the whole experiment, without being noticed by the children.

In both rooms, two chairs were placed in front of a desk on which a Windows XP laptop computer was placed, on which the card game ran. Behind the desk, a camcorder was placed on a tripod in such a way that it could record the children’s face and upper part of their body. Before the start of each experiment, the camera was adjusted to the height of the participating child.

The two chairs placed in front of the desk were turned to each other with an angle of about 45 degrees. One chair was reserved for the child participant, on the other chair a 24-in. computer monitor was placed vertically (comparable to children’s height). A live video stream of the second participant, sitting in the other room, was shown on this monitor. This monitor was placed in such a way that the participant sitting next to the monitor could see the live video stream properly (Fig. 3). This was all done to make the sitting environment and experimental setup as similar as possible to the co-present condition.

The 24-in. monitor, placed on the chair, was attached to a MacBook, which was placed behind the monitor. We used iChat (Mac’s embedded audio–video conferencing utility) for enabling the audiovisual communication between the two children. A different Mac computer, connected wirelessly with the MacBooks in the room, allowed us to control the connection between the two rooms, without actually entering.

A high definition (1080p), separate webcam was attached to the MacBooks in the respective rooms. In the case of the mutual gaze condition, it was placed on top of the 24 in. monitor as shown in Fig. 3. By looking at the monitor where the partner’s video steam was broadcasted, the participants were able to see each other’s faces simultaneously. In the case of the no gaze condition, the web-
2.3. Co-present condition

Data for the co-present condition was collected in essentially the same way, except that both children were in the same room. For this a quiet room in the school was used. The children were positioned on two chairs, placed in front of desk on which a computer monitor was placed. The computer monitor was used to show the card game. Behind the computer monitor a camcorder was placed on a tripod in such a way that it could record the children’s faces and the upper part of their body. The experimenter was positioned behind a screen, outside the visual field of the participants. A laptop was used by the experimenter to control the game.

2.4. Procedure

The experiment was conducted in three elementary schools located in Tilburg and Helmond, the Netherlands. Children were asked to form pairs themselves. In all conditions, children received instructions in pairs. The experimenter always welcomed the children and started a small discussion by asking a few questions to break the ice (“How old are you? Do you like to play games?” etc.). After this introductory phase, the experimenter gave instructions orally, telling the children about the game, the game rules and the coins they could win. They were told that they are required to discuss the final guess with each other and they have to come to a shared decision before announcing the number orally. They were told that once they reveal their guess (by saying either ‘higher’ or ‘lower’), the next card would appear automatically. In the case of mediated conditions, an audio–video session was already established and participants were informed about the audiovisual setup and how they could see and talk to each other using this communication system. When children seemed to understand the rules and the setup, children were asked to take their position. In the mediated conditions, each child was brought to a different room, and asked to sit down on the chair next to the chair containing the monitor showing their friend. In the non-mediated condition, children stayed in the same room and sat down in the chairs next to each other.

When both children were in their chairs, the experimenter started a practice game (“So you only have to say whether the next card is higher or lower. This is just an exercise and it doesn’t really count”). After this exercise, the experimenter asked the children whether they had any questions. If not, the experimenter left the children’s field of vision and the first experimental game was started. The experiment did not have a fixed duration because different children took different time. On average, each session lasted for approximately 10–15 min.

After the sixth and final game, the experimenter asked the participants to fill in a post-experiment questionnaire. After this, the experimenter asked the children some open questions. To finish the whole session, the experimenter congratulated the children and informed them that they could trade in their coins for an

Fig. 4. No gaze condition – Webcam on the top of the front screen where card game is projected (red square shows the camera position). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Reaction of children after winning the game in the mutual gaze (top left), no gaze (top right) and co-present (bottom) conditions.
individual gift (a key ring or chocolates). Overall, the game worked well. All children made the logical choices they were expected to make in most of the games, so that each pair of children lost at least two games and won at least two games. Furthermore, not a single child reported any suspicion that the game was not real, but a deterministic simulation. Fig. 5 shows some representative stills of children winning a game in each of the conditions.

2.7. Results

Table 1 summarizes the results on the fun, engagement and endurability (interest in doing an activity or playing the game again and again) questions and shows a rather consistent picture. The children reported having most fun while playing in the mutual gaze condition \( (M = 6.47, SD = .64) \) least fun while playing in the no gaze condition \( (M = 5.57, SD = .92) \) with the co-present condition in between \( (M = 6.15, SD = .73) \), \( F(2,119) = 17.09, p < .001, \eta^2_g = .22 \). All pairwise comparisons are statistically significant.

The results of the engagement dimension showed that children were roughly equally engaged with each other in the mutual gaze \( (M = 4.60, SD = .89) \) and co-present \( (M = 4.47, SD = .60) \) conditions, while their level of engagement was relatively low in the no-gaze condition \( (M = 4.07, SD = .94) \), \( F(1,119) = 7.93, p < .01, \eta^2_g = .12 \). With the exception of the co-present vs. mutual gaze comparison, all pairwise comparisons are statistically significant.

The results for the endurability dimension revealed that children who played in the mutual gaze condition showed the strongest interest in playing again \( (M = 4.72, SD = .63) \), followed by children who played in the co-present condition \( (M = 4.45, SD = .81) \), while children in the no gaze condition were least interested in playing again \( (M = 3.80, SD = 1.03) \), \( F(1,119) = 15.72, p < .001, \eta^2_g = .21 \). Pairwise comparisons revealed that all comparisons are statistically significant.

Next we describe the results of the NMSM scale for mediated conditions (no gaze and mutual gaze) of which Table 2 summarizes the results. The results for the perceived co-presence category reveal that children were more involved and felt the presence of their partner stronger when playing in the mutual gaze condition \( (M = 4.57, SD = .51) \) than in the no gaze condition \( (M = 4.02, SD = .73) \), \( F(1,79) = 15.34, p < .001, \eta^2_g = .16 \). We did not find a significant difference between the mutual gaze and no gaze conditions on the message understanding dimension \( (F < 1) \). However, children who played the game in the mutual gaze condition indicated that they were better able to understand the emotional state of their partner \( (M = 4.47, SD = .76) \), than children who played in the no gaze condition \( (M = 3.80, SD = .89) \), \( F(1,79) = 20.66, p < .001, \eta^2_g = .21 \).

Finally the game playing condition had a significant effect on the partner’s evaluation. Children reported their game partners more and appreciated their involvement more in the mutual gaze condition \( (M = 4.63, SD = .69) \) than in the no gaze condition \( (M = 4.15, SD = .50) \), \( F(1,79) = 13.91, p < .001, \eta^2_g = .17 \).

Table 1
Average Fun scores with standard deviation (SD) on a 7-point Likert scale (1 = very negative, 7 = very positive) and average Engagement and Endurability scores (with SD) on a 5-point Likert (1 = very negative, 5 = very positive).

<table>
<thead>
<tr>
<th>Fun Toolkit Categories</th>
<th>Conditions</th>
<th>Co-presence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No gaze</td>
<td>Mutual gaze</td>
</tr>
<tr>
<td>Fun</td>
<td>5.57 (.92)</td>
<td>6.47 (.64)</td>
</tr>
<tr>
<td>Engagement</td>
<td>4.07 (.94)</td>
<td>4.60 (.89)</td>
</tr>
<tr>
<td>Endurability</td>
<td>3.97 (1.03)</td>
<td>4.72 (.63)</td>
</tr>
</tbody>
</table>

Table 2
Mean (and standard deviation) of the four categories of network minds social presence measure on a 5-point Likert (1 = very negative, 5 = very positive).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Conditions</th>
<th>Co-presence</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No gaze</td>
<td>Mutual gaze</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-presence</td>
<td>4.02 (1.73)</td>
<td>4.57 (.51)</td>
<td>15.34</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Message understanding</td>
<td>4.30 (1.72)</td>
<td>4.41 (.63)</td>
<td>&lt;1</td>
<td>1</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Affect understanding</td>
<td>3.80 (.89)</td>
<td>4.48 (.76)</td>
<td>20.66</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Partner evaluation</td>
<td>4.15 (.50)</td>
<td>4.63 (.69)</td>
<td>13.91</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
3. Perception test

In the previous section we presented children’s self-reports on how they appreciated the game playing experience in two mediated conditions, and we also compared the results with those obtained from children playing in the non-mediated (co-present) condition. We found that children in the mutual gaze condition scored significantly higher on fun and endurability than children in the co-present condition. They (children in the mutual gaze and co-present conditions), in turn, scored significantly higher than children in the no gaze condition on all dimensions of the NMPSP, with the exception of message understanding. But how do the children actually behave in the various conditions? Can we tell from their non-verbal behavior that they seem to have most fun in the mutual gaze condition? To test this, we ran a perception test in which we showed short clips (taken from all three conditions) and presented them to judges, who were asked for each clip to guess whether the child appearing in the clip had just won or lost the game. In this way, we can estimate the expressiveness of the children in the various conditions; more correct guesses indicate that children are more expressive. Running such perception tests is a standard procedure in many research disciplines (e.g., speech and emotion research) and allows us to collect objective judgments of non-verbal behavior.

3.1. Stimuli

We randomly selected 20 pairs of children from each of the mediated conditions (mutual gaze and no gaze). From these 40 pairs, we selected the response of their first winning game (in which they made a correct prediction for the last card) and the response of their first losing games (in which the final guess turned out to be incorrect). From each pair, we quasi-randomly selected one child, making sure that half of the children came from one room, and half from the other, and that half of the children were boys and half were girls. In all clips, we zoomed in on the face of the child, making sure that the computer screen, placed next to the participants, was never visible in the final clip (see Fig. 6).

Fig. 6. Representative stills of children for winning (right) and losing games (left) during no gaze (top), mutual gaze (middle) and co-present (bottom) conditions.
For comparisons, we also used clips of 20 pairs of children from the ‘co-present’ condition, where winning and losing responses were selected in the same way as above. Furthermore, from the clips of the co-present condition, we randomly selected one child from each pair, and zoomed in on his or her face. Half of the children sitting on the right chair and half of the children sitting on the left chair were selected. In all cases, the video snippets were cut from the moment the final card of the game was turned until the main response of the child was finished. The average length of the stimuli was similar across the three conditions. This resulted in three sets (one for each condition), each consisting of 40 clips (2 (win/lose) × 20 children).

3.2. Participants

Ninety Dutch adults (judges) participated (M<sub>age</sub> = 21.39, SD = 2.37) in the perception (judgement) tests, with a roughly equal number of men and women.

3.3. Procedure

Three group perception tests were conducted and in each test 30 Dutch viewers viewed one set of stimuli. For every test, groups of participants were invited into a quiet classroom where a computer screen was projected on the classroom wall using a beamer. Stimuli were projected without sound, to avoid that participants could pick up auditory cues for winning or losing. Participants were told that they would see a set of 40 stimuli in which children were showing their emotions after winning or losing a game and that their task was to determine whether the children had just won or lost their game. Clips were presented to participants in a fixed random order. Each stimulus was preceded by a number displayed on the screen indicating the upcoming stimulus, and followed by a 6 s pause during which participants could fill in their score on the answer form. The actual perception test was preceded by a short training session in which three clips were shown (different from the ones shown in the actual set of stimuli) to make participants familiar with the stimuli and the experimental task. If everything was clear, the actual test started. During the test there was no interaction between participants and experimenter.

3.4. Statistical analysis

To test for significance we performed a repeated measurement analysis of variance (ANOVA) with two within-subjects factors, namely game status (levels: win, lost), and gender (levels: boy, girl) and with one between-subjects factor: presence (levels: no gaze, mutual gaze, co-presence) and with percentage of correct classification as the dependent variable. All pairwise comparisons were performed using Tukey's HSD.

3.5. Results

First of all, we found that the average number of correct classifications is highest for children playing in the mutual gaze condition (M = .84), lowest for children playing in the no gaze condition (M = .68), with the scores for children playing in the co-present condition in between (M = .74), F(2,87) = 66.28, p < .001, η<sup>2</sup><sub>p</sub> = .60 (all pairwise comparisons significant).

In addition, judges found it easier to see whether children lost their game (M = .80) than whether they won (M = .70), F(1,87) = 28.21, p < .001, η<sup>2</sup><sub>p</sub> = .245, and judges made more correct classifications for boys (M = .77) than for girls (M = .72), F(1,87) = 12.40, p < .01, η<sup>2</sup><sub>p</sub> = .125. Besides these main effects, there was one significant two-way interaction, between game status and presence (F(1,70) = 24.177, p < .001, η<sup>2</sup><sub>p</sub> = .257). This interaction can be explained by inspection of Fig. 7: there it can be seen that for both no gaze and co-present conditions, losing clips are more often classified correctly than winning clips, but in the case of the mutual gaze condition both winning and losing are classified correctly about equally often.

4. Observational analysis

In this section, we describe a detailed observational analysis of the children's actual non-verbal behavior, and correlate these with the scores from the perception test, to see which cues our judges relied on. We also code cues of their social interactions in the three conditions, to see how these conditions influenced children's' social behavior.

4.1. Development of observation scheme

To quantify behavior in terms of (1) their affective responses to winning or losing a game and (2) the children’s social interactions during gameplay, we developed two new observation schemes. The first focused on observing and quantifying positive (joy after winning) and negative (disappointment after losing) emotions. This part of the coding scheme was developed on the basis of an existing coding scheme (Shahid, 2012), extended based on informal observations of video clips taken from the various conditions.

This resulted in two main categories (see Appendix A for details): (1) positive affect (subcategories: smile, laughter, clapping, jumping/moving, winning arm gesture), and (2) negative affect (subcategories: sadness (frown), anger, closing eyes, covering face, head down). In addition to these visual features, we also added one auditory feature, namely the verbal response to winning or losing. For this part of the coding scheme, a number of facial features are roughly comparable with Action Units (AU's) of the Facial Action Coding Systems (FACS) proposed by Ekman and Friesen (1978), e.g. AU 6 and 12 for smile/happiness.

The second part of the coding scheme focused on observing social interaction during gameplay. This scheme was designed on the basis of the informal consultation of video recordings from all conditions. All together this resulted in two categories: (1) connectedness (subcategories: thumbs up, waving, virtual (touching)), and (2) responsiveness (subcategories: chair moved (for sitting closer to the partner), eye fixation (looking at each other for 1.5 or more seconds), turn to partner, and turn taking). We coded the presence and frequency of a feature and the number of times a particular behavior (e.g. smiling, touching, hugging, etc.) occurred in the selected fragments. In a few cases, we coded the duration of the

![Fig. 7. Average percentage of correct classifications as a function of presence and game status.](image-url)
behavior for a fixed interval (e.g. eye-fixation). During the coding, the unit of analysis was always one (player).

4.2. Stimuli

The stimuli used for the non-verbal analysis of positive and negative affect were the same as those presented previously in the perception test. For coding social interaction, essentially the same set of stimuli was used, but in a extended duration (including the discussion leading up to the final higher or lower decision) and with sounds not muted. Furthermore, for these analyses we did not zoom in on one participant, but looked at both participants and their interaction in the way it was recorded.

4.3. Procedure

The first author and one independent researcher (trained by him) performed the labeling of the observational data. Both coders worked independently, after which the inter-rater reliability for each feature was measured and was found to be substantially high for all categories (ranging from 61% to 92%). The coding was performed in two rounds: first focusing on non-verbal cues related to affect, and then focusing on social cues.

4.4. Data analysis

To test for significant differences we conducted separate ANOVAs with presence (levels: no gaze, mutual gaze, co-present) as the between subjects factor and mean number of visual features as the dependent variable.

4.5. Results

Table 3 summarizes the number and type of visual features related to affect displayed by children in three game-playing conditions. In general, it can be seen that both for positive and negative effect the number of features is the highest for the mutual gaze category, slightly lower for the co-present category and the lowest for the no gaze category.

In general, it can be seen that both for positive and negative effect the number of features is the highest for the mutual gaze category, slightly lower for the co-present category and the lowest for the no gaze category. The one-way ANOVA for the affective dimension revealed that children showed most (non)verbal cues while playing in the mutual gaze condition, the least while playing in the no gaze condition, with the co-present condition in between, $F(2119) = 14.15, p < .001$, $\eta^2_p = .19$. All pairwise comparisons are statistically significant.

We explored to what extent these visual (positive and negative) cues correlate with the percentage of correct classifications. To do so, we correlated the number of different cues observed in a stimulus with the percentage of correct classifications in the perception test and we indeed found a strong and statistically significant positive correlation between the two (Pearson $r = .81$, $N = 120$, $p < .001$). Fig. 8 clearly shows that the increase in the amount of visual features present in stimuli gives rise to better classification.

We also explored to what extent the number of visual cues correlate with the fun scores. We correlated the average number of visual (positive and negative) cues shown by a child with reported fun score and we found a strong and statistically significant positive correlation for three conditions: Mutual gaze ($r = .83$, $p < .001$), no gaze ($r = .66$, $p < .01$), co-presence ($r = .78$, $p < .001$).

The one-way ANOVA for the social dimension revealed that the number of social cues and behaviors shown by the children in the mutual gaze condition are much higher compared to the no gaze and co-present conditions, $F(2119) = 35.04, p < .001$, $\eta^2_p = .37$. With the exception of the co-present vs. no gaze comparison, all pairwise comparisons are statistically significant. Table 4 summarizes the number and type of social cues displayed by children in three game-playing conditions.

![Fig. 8. Correlation between percentage of correct classification and number of visual features shown by children.](image)

### Table 3

Number of positive and negative features shown by children during three game playing conditions.

<table>
<thead>
<tr>
<th>Affect category</th>
<th>Features</th>
<th>Game playing conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No gaze</td>
</tr>
<tr>
<td>Positive</td>
<td>Smile (happy)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Laugh</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Clapping</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Jumping/moving</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Winning gesture</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Verbal response</td>
<td>3</td>
</tr>
<tr>
<td>Total positive features</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Negative</td>
<td>Sad (frown)</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Closing eyes</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Covering face</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Head down</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Verbal response</td>
<td>5</td>
</tr>
<tr>
<td>Total negative features</td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>Total visual features</td>
<td></td>
<td>78</td>
</tr>
</tbody>
</table>

### Table 4

Number of social cues shown by children during three game playing conditions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
<th>Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No gaze</td>
</tr>
<tr>
<td>Connectedness</td>
<td>(Virtual) Touching</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Thumbs up</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Waving</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total features</td>
<td>5</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Stool moved</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Turn to partner</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Eye fixation</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Turn taking*</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Total features</td>
<td>80</td>
</tr>
<tr>
<td>Total social features</td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>

* Children were instructed to guess the next number turn by turn (first guess by player A, second guess by player B, so on). We observed whether or not turn taking process was respected during a particular game.
5. Discussion

In this paper, we studied how children experienced playing a card game with a friend, comparing three conditions: two video-mediated conditions, one which allows for natural, mutual gaze, and one which does not, and one condition in which game-playing children were sitting next to each other without a mediating interface. The general results are remarkably consistent: children seem to enjoy playing the card game least in the no gaze condition, while their appreciations for the mutual gaze and co-present condition are comparable (often with a preference for the former over the latter). Below we give a more elaborate discussion of several of our findings.

5.1. Effect on game experience and social presence

The post game questionnaires revealed that children had more fun, and were more engaged in the mutual gaze than in the no gaze condition, and children in the mutual gaze condition were also more interested in playing again. These results are in line with the media richness theory, and strengthen the results of previous studies where researchers reported similar effects with different user groups and in different domains, i.e., the presence of mutual gaze increases users' satisfaction and pleasantness of tasks (Bailenson et al., 2001). In addition, we found that children in the co-present condition always scored higher for these scales than those in the no gaze condition but not quiet as high as those in the mutual gaze condition. However, keeping into account the differences of subjective ratings between co-present and mutual gaze conditions, it is clear that both conditions are closely related to each other with overall high positive scores.

The results of the networked minds social presence measure (NMSPM) scale showed that the absence or presence of mutual gaze during video-mediated game play has a number of effects on the way the games were experienced by the child players. The mutual gaze had a direct effect on the perceived intimacy and closeness of children. Children reported that they felt strongly connected with each other and felt the presence of their partner in the same physical space in the mutual gaze condition more, as compared to the no gaze condition. This is an interesting new finding in the context of children's remote game play, and is in line with the findings of Bondareva et al. (2006), who showed the direct influence of presence or absence of mutual gaze on perceived social absence of adults.

However, regarding the ‘message understanding’ dimension, we observe that children in the two mediated conditions were able to understand each other's messages without problem. Mutual gaze or lower level of social presence did not have an effect on the perceived message understanding. Although mutual gaze was not present in one condition, other visual cues, system characteristics, and children's own capacity could have helped them in understanding each other (Wise et al., 2004). Furthermore, the task in hand did not require extensive explanation or argumentation, which might have positively affected the ‘perceived understanding’ scores.

Results also show that mutual gaze enabled remote players to understand and synchronize each other's emotional responses in an adequate manner, while lack of mutual gaze negatively influenced 'affective understanding'. Children in the no gaze condition scored lower on the 'affective understanding' dimension than the children in the mutual gaze condition. These findings strengthen the already established but less emphasized argument that gaze not only has a communicative function but it also assists in regulating joint arousal, a function which is also closely connected to intimacy and tele-presence (Vertegaal, 1999; Vertegaal and Ding, 2002). Our results show that the findings of previous studies are not restricted to formal tasks performed by adults but can also be seen in more playful situations with children.

Overall the results of three categories of the NMSPM and their relationship (e.g. in the no gaze condition, children's rating for the ‘affective understanding’ dimension was the lowest among all dimensions) can be interpreted in terms of what Kim et al. (2011) have previously reported, as they showed that social presence is developed on the basis of shared attention, mutual support, open communication and affective connectedness. According to their model, mutual attention and understanding are relatively easy and affective connectedness is quite difficult to achieve in a mediated environment. This is also in line with the argumentation of Biocca et al. (2003), creators of the NMSPM construct (Harms and Biocca, 2002), who claim that social presence follows a pattern, that starts with spatial presence and develops into psychological involvement and affective connectedness.

The different game conditions revealed some other social consequences as well. Children trusted their partners more in the mutual gaze condition than in the no gaze condition, and found their partners friendlier and attentive, compared to the no gaze condition. In the mutual gaze condition, children also respected the logical guesses of partners more often and did not stick to their own guesses even when it was their turn to decide. Apparently, higher level of trust in each other's opinion and friendly attentiveness also led to more number of games won during the mutual gaze condition. In view of VMC and children's gameplay, it would be interesting to explore the relationship between trust and game status more in future work.

5.2. Effect on emotional expressions

In the perception test, we found that overall the number of correct classifications of the child data in terms of whether they had won or lost was rather high, showing that facial expressions and other nonverbal features have clear cue value in that respect. Interestingly, the percentage of correct classification was the highest for the mutual gaze condition and the lowest for the no gaze condition, which implies that children were very expressive while playing games in the presence of mutual gaze and absence of gaze had a negative effect on expressiveness. In the co-present condition, children were also rated more expressive than the no gaze condition but not quite as high as in the mutual gaze condition. The perception test scores are very much in line with the fun scale and fun toolkit self-reports in the sense that the children's appreciation of the game is reflected in the degree to which they become expressive. Moreover, the results of the perception test support the social presence self-reports, particularly in as far as the ‘affective understanding’ dimension of the NMSPM scale is concerned. In sum: mutual gaze not only had an effect on children's subjective feelings but it also influenced their expressive behavior.

The perception test also showed that in the case of no gaze and co-present conditions, it was easier for judges to correctly judge losing games (higher percentage correct) than winning games which is in accordance with the results of a previous study using the game play setting (Shahid et al., 2011) and with data of adult participants (Shahid et al., 2008b). However in the mutual gaze condition, the percentage of correct classification was equally high for both winning and losing games, which is not in accordance with our previous findings. One possible explanation for this result could be that this is due to a ceiling effect, namely that in this condition children were so excited about the game that it made them very expressive overall, irrespective of the game status and this higher expressiveness positively influenced the judgment of participants.
The results of the observational analysis also agree with the results of self-reports and perception test scores. Observational codings for the affective dimension revealed that children showed the highest number of visual cues, both negative and positive, in the mutual gaze condition, the lowest number of visual cues in the no gaze condition and the number of visual cues shown in the co-present condition was in between two mediated conditions. There was also a strong positive correlation between percentage of correct classifications and the number of visual features shown by children after winning or losing a game. The results of a correlational analysis not only strengthen the results of perception test but also endorse it as a practical method to get an objective response about the affective states of participants.

The results of the social dimension revealed that children showed more social cues and felt more connected with each other in the mutual gaze condition than in the no gaze and the co-present conditions, whereas the no gaze and the co-present conditions did not differ. Interestingly, only for the mediated gaze condition, children produced typical gestures, like thumbs up and waving, to give feedback or get attention from the other player, which could have strengthened the feeling of social connectedness. Similarly, it turned out that children turned to their partner more often in the mediated conditions. Furthermore, in the mutual gaze condition, players always waited for the response of their partner for having a true eye contact (eye-fixation). In the co-present condition, players may have felt less of a need to constantly turn to their partner and establish eye contact as players were sitting side by side in the same physical space and were sure about each other’s response. In the case of mediated conditions, players may have felt a stronger desire to give each other feedback and synchronize their actions. This appears to be in line with existing studies in the context of VMC (Bondareva et al., 2006; Bordia, 1997), where it has been shown that people in mediated conditions gesture more (hyper-gesturing) than the co-present condition for conveying the message.

5.3. Mutual gaze in co-present and video-mediated communication

And finally, it is noteworthy that our analyses revealed that children tend to appreciate the mutual gaze condition more than the co-present one. Both subjective se children were more expressive after losing children were more expressive after losing if-reports and objective assessments were higher for the mutual gaze than for the co-present condition. Prima facie, these findings seem to contradict a general prediction of the media-richness theory, that face-to-face communication is the ideal form of communication and that people generally prefer it over a mediated communication experience. It should be stressed, however, that with a few notable exceptions (Yarosh et al., 2009, 2010), most studies that found evidence for this assumption, were conducted in professional settings with adult participants. We think it is precisely the fact that our experiment involved young participants engaging in playful interaction that accounts for our findings.

There is also a general consensus among researchers in the computer-mediated communication (CMC) research area, that the success of a CMC system largely depends on the type of task it is designed for (Bordia, 1997; Hollingshead et al., 1993). Under right settings and over time the CMC experience cannot only match the face-to-face experience (Thurlow et al., 2004), it can even be hyperpersonal (Walther, 1992, 1996). A couple of studies conducted in the gameplay context, one with adults (Bos et al., 2002) and another one with children (Batcheller et al., 2007), also showed that mediated experience can nicely match the face-to-face experience, however with some limitations.

We explain our results along similar lines and argue that the simplicity of the game-playing task, the naturalness of the mutual gaze condition’s setup, and its close resemblance with the co-present condition played a key role in shaping the results the way they are. We also conjecture that in the mediated mutual gaze condition, the children have to “work somewhat harder” to communicate than in the face-to-face condition. Obviously, the same applies to the no gaze mediated condition, but since communication is not facilitated there, the children fail in their attempts to properly communicate. That children work a bit harder can also be seen from the frequency of social cues displayed by children in this condition. Plausibly, the extra effort pays off in fun, engagement and endurance as well and children appreciated the video-mediated mutual gaze condition over the co-presence one.

Interestingly, we found these effects even though the mutual gaze was not perfect, in the sense that players could see each other’s gaze without true eye-contact. In other words, even when mutual gaze may not be ideal, it can still improve the mediated communication. See also Grayson and Monk (2003), who argue that depending on the task and the context, mutual gaze awareness in typical video conferring systems may support similar communicative functions as true mutual gaze. This is particularly relevant with regard to young children, where previous research shows that children do not always go for direct eye contact. Rather they sometimes seem to prefer “eye-to-face” contact (or focus on a boader area around eyes) because it is unusual for them to fixate eyes on a small area (Argyle and Cook, 1976).

6. Conclusion and future work

This study aimed to explore the effect of variations in level of gaze on the perceived social presence, emotional expressions, social interaction and game playing behavior of children during video-mediated and co-present communication situations. This study consisted of two parts where the first part (Section 2) was primarily used to collect data. We used a deterministic card guessing game and asked children to play games in three different conditions. After the game playing session, we collected children’s subjective response to fun, engagement, and perceived social presence. In the second part (Sections 3 and 4), we used a perception test, as an objective evaluation method, for measuring the emotional response of children after winning and losing a game. In the last part of this study, we did an observational analysis and coded the video data along affective and social interaction dimensions.

The results show that compared to the no-gaze condition where pairs did not feel a social bond, the feeling of social presence and connectedness was high in the natural-gaze condition and children experienced as if they were physically present in the same space with their game partner. Furthermore, children had more fun and showed rich social and emotional behavior in the presence of mutual gaze than in the absence of it. Overall the co-present condition was also much more appreciated over the no gaze condition with high self-reports and expressiveness scores but not quite as high as those in the mutual gaze condition. The study yielded new and interesting insights into the video mediated gameplay of young children. The children we focused on are particularly relevant because during this age (7–8 years) they exploit mutual gaze for supporting and regulating communication. This study suggests that mutual gaze is a crucial ingredient for establishing a joyful video mediated gameplay and without it the quality of interaction is dramatically effected.
An important thing to note is that, although our mutual gaze condition enabled a reasonable level of natural, mutual gaze, it was still not an ideal form of eye-gaze, where participants can look directly into each other’s eyes, as is possible in face-to-face communication. Therefore, in another follow-up study, we would like to replicate our study using so-called Eye Catchers. An Eye Catcher is a videophone with a high-resolution camera installed behind the screen that is invisible to the participants sitting in front of it because of its half-silvered mirror. The camera is located right in the center of the screen, which offers true eye-gaze during video conferencing. Such a setup would be interesting for our purposes, because if children who communicate via an Eye Catcher are more expressive than those communicating via the original “mutual gaze” condition, then we can conclude that true eye-gaze is most important for mediated communication and the lack of it during our “mutual gaze” condition was perhaps partly compensated by the novelty effect of the communication setup. However, if children’s expressiveness while playing via Eye Catchers would be comparable to the “mutual gaze” condition, then we can conclude that a less ideal form of mutual gaze appears to be enough for playful tasks and contexts, and having a true, direct mutual gaze facilities does not add much to the overall game playing experience. Finally, our results have potential for the development of educational software, for example where remote tutoring is concerned. However, the details of this left for future research.

Appendix A. Observational scheme
Appendix B. NMSPM questionnaire

It was easy to understand my partner
My partner found it easy to understand me
Understanding my partner was difficult
I had difficulty in understanding my partner
My thoughts were clear to my partner
My partner’s thought were clear to me
I could tell how my partner felt
My partner could tell how I felt
It was possible for me to describe my partner’s feelings accurately
It was possible for my partner to describe my feelings accurately
My partner’s emotions were not clear to me
My emotions were not clear to my partner
My partner was friendly
I was friendly to my partner
I trusted my partner
I did not take a personal interest in my partner
My partner trusted me
My partner did not take personal interest in me
I was aware of my partner’s presence
My partner was aware of my presence.
My partner caught my attention
I caught my partner’s attention
I often didn’t notice my partner (felt alone)
My partner often didn’t notice me (felt alone)

References


