Title
An Operational Model of Joint Attention - Timing of Gaze Patterns in Interactions between Humans and a Virtual Human

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Authors
Pfeifer-Lessmann, Nadine
Pfeifer, Thies
Wachsmuth, Ipke

Publication Date
2012

Peer reviewed

An Operational Model of Joint Attention - Timing of Gaze Patterns in Interactions between Humans and a Virtual Human

Nadine Pfeiffer-Lessmann (nlessman@techfak.uni-bielefeld.de)
Thies Pfeiffer (tpfeiffe@techfak.uni-bielefeld.de)
Ipke Wachsmuth (ipke@techfak.uni-bielefeld.de)
Artificial Intelligence Group, Faculty of Technology, Bielefeld University, Bielefeld, Germany

Abstract

Joint attention has been identified as a foundational skill in human-human interaction. If virtual humans are to engage in joint attention, they have to meet the expectations of their human interaction partner and provide interactional signals in a natural way. This requires operational models of joint attention with precise information on natural gaze timing. We substantiate our model of the joint attention process by studying human-agent interactions in immersive virtual reality and present results on the timing of referential gaze during the initiation of joint attention.

Keywords: joint attention; virtual humans; social interaction

Introduction

Attention has been characterized as an increased awareness (Brinck, 2003) and intentionally directed perception (Tomasello, Carpenter, Call, Behne, & Moll, 2005) and is judged to be crucial for goal-directed behavior. Joint attention builds on attentional processes and has been identified to be a foundational skill in communication and interaction. The term joint attention is often used confusably with shared attention. We follow Kaplan and Hafner (2006) and Tomasello et al. (2005) in using the term joint attention for the phenomenon which presupposes a higher level of interactivity requiring intentional behavior and an awareness of the interaction partner. Joint attention can be defined as simultaneously allocating attention to a target as a consequence of attending to each other's attentional states (Deak, Fasel, & Movellan, 2001). In contrast, we see shared attention (as well as shared gaze) as the state in which interactants are just perceiving the same object simultaneously without further constraints concerning their mental states or their interaction history.

Mundy and Newell (2007) differentiate joint attention behaviors into two categories: *responses* to the bids of others and spontaneous *initiations*. Responding to joint attention refers to the ability to follow the direction of gaze and gestures of others in order to share a reference. On the other hand, to initiate joint attention humans use gestures and eye contact to direct the attention of others to objects, events, and to themselves.

For joint attention, interlocutors have to deliberatively focus on the same target while being mutually aware of sharing their focus of attention (Tomasello et al., 2005; Hobson, 2005). To this end, respond and feedback behaviors are necessary. Tasker and Schmidt (2008) argue that to establish joint attention a sequence of behaviors is required which has to meet certain time constraints.

We constructed an operational model of joint attention (Pfeiffer-Lessmann & Wachsmuth, 2009) for our virtual human Max (Lessmann, Kopp, & Wachsmuth, 2006) to create a more natural and effective interaction partner. The model covers four phases: the initiate-act (1), the respond-act (2), the feedback phase (3), and the focus-state (4). However, for Max to apppear believable and to use the same behavior patterns in the phases as humans do, investigations on time-frames, human expectations and insights on how humans actually perceive his behavior are indispensable. The topic of concrete reaction and duration times of feedback behaviors during the joint attention process has to our knowledge not been discussed in the area of human-computer interaction yet. The time-frames and expectations of humans for natural interactions are central subject of this paper.

In the section to follow, we provide an overview on related work covering research on joint attention in human-human interaction and in the area of technical systems. In the subsequent "Model" section, a brief summary of our own definition of joint attention is provided. Next, we present a study in immersive virtual reality concerning the exact timing of the first phase, the initiate-act, of our joint attention model. Thereafter, results are discussed and the paper ends with our conclusions and future work.

Related Work

Staudte and Crocker (2011) raise the question whether joint-attention-like behavior is unique to human-human interaction or whether such behaviors can play a similar role in human-robot interaction. They conclude that their own findings suggest that humans treat artificial interaction partners similar to humans and that it is therefore valid to investigate joint attention in settings with artificial agents.

These artificial agents can consist, on the one hand, of robots (Deak et al., 2001; Imai, Ono, & Ishiguro, 2003; Breazeal et al., 2004; Doniec, Sun, & Scassellati, 2006; Nagai, Asada, & Hosoda, 2006; Yu, Schermerhorn, & Scheutz, 2012; Huang & Thomaz, 2011; Staudte & Crocker, 2011) and, on the other hand, of virtual humans (Peters, Asteriadis, & Karpouzis, 2009; Zhang, Fricker, & Yu, 2010; Bailly, Raidt, & Elisei, 2010).

Kaplan and Hafner (2006) point out that research in robotics concentrates only on partial and isolated elements of joint attention (e.g. gaze following, simultaneous looking or simple coordinated behavior) covering solely the surface of the process but not addressing the deeper, more cognitive

aspects of the problem. The same authors stress that no system achieved true joint attention between a robot and a human or between two robots according to their definition yet. This appears to be still the case, however progress has been made with respect to investigating joint attention behaviors.

A number of researchers in cognitive science and cognitive robotics use developmental insights as a basis for modeling joint attention showing how a robot can acquire joint attention behaviors by supervised and unsupervised learning (Deak et al., 2001; Nagai et al., 2006; Doniec et al., 2006). However, the aspect of intentionality and explicit representation of the other's mental state are not accounted for in these approaches.

Another area of research investigates the impact of artificial agents' joint attention behavior on humans. Here, real interaction scenarios can be distinguished from humans rating video material. Huang and Thomaz (2011) argue that video-based experiments offer the advantage of studying humans' perception of joint attention behaviors without dealing with technical challenges of identifying the humans' behaviors. According to Staudte and Crocker (2011), it has been shown that video-based scenarios without true interaction yield similar results to live-scenarios and can therefore provide valuable insights into humans' perceptions and opinions.

Huang and Thomaz (2011) use videos to investigate humans' judgements of robots initiating and ensuring joint attention behavior. Their results suggest that humans overall preferred robots showing joint attention behavior. Staudte and Crocker (2011) also follow a video-based approach; they conclude that participants robustly follow the robot's gaze and use it to anticipate upcoming referents. Bailly et al. (2010) try to quantify the impact of deictic gaze patterns of their agent. They explicitly instructed participants not to take the agent's behavior into account, but the participants were drastically influenced by the agent's gaze patterns anyway.

In a real interaction scenario, Peters et al. (2009) study how human participants perceive the virtual agent's simple shared attention behavior of non-verbal cuing and how subtle changes of this behavior affect the gaze-following of human participants. Huang and Thomaz (2011) investigate the respond-act of their robot and the resulting impact on a human-robot collaborative task using a task-based metric. They find that the robot responding to referential foci significantly outperforms the one staying focused on the human.

The robot of Breazeal et al. (2004) keeps a representation of its current focus of attention calculated by saliency values. Additionally, it monitors the human participant's focus of attention. It is thereby able to notice when both interactants focus on the same object simultaneously. However, the robot appears to miss feedback mechanisms on a higher level of interactivity covering intentional behavior and the awareness of the interaction partners in the joint attention process.

Many researchers investigating the impact of artificial agents which show joint attention behaviors do not account for the necessary time courses. As an exception, Yu et al. (2012) try to investigate the exact time course of multi-modal interaction patterns occuring naturally as part of joint attention processes.

In our own approach, we let human participants engage with our interactive virtual agent Max in an immersive virtual environment. As with human-robotic and human-human interactions, the interactants thus share the same three-dimensional environment and reciprocal interactions are possible. However, other than today's robotic systems, the virtual agent has (more than) human-like reaction times and is controlled by a cognitive architecture which goes from basic activation processes up to concepts of epistemic modal logics to model mutual beliefs which are essential for joint attention.

For human-human interactions, Tasker and Schmidt (2008) postulate a time frame of 5 s for the addressee of an initiate-act to respond appropriately. According to them, the duration of the respond-act has to last for at least 3 s in order to establish evidence that the partner's attention has been captured. The focus-state of joint attention has to last for a minimum of 3 s, too. This duration is in accordance to the results found by Vaughan et al. (2003) for maintaining focus on the same object in an episode of joint engagement.

Mueller-Tomfelde (2007) takes a closer look at the research literature to figure out appropriate time scales for referential actions. Since an initiate-act or respond-act could be characterized as such, his results should be highly relevant for natural time-scales of joint attention behaviors. He argues that since a pointing action includes cognitive aspects, it is more than a basic movement-primitive and thus more than a basic physical act constrained by the nature of cognitive operations at a time period of about a 1/3 of a second. Therefore, Mueller-Tomfelde (2007) expects an appropriate temporal scale of referential primitives to be greater than 300 ms while being less than the temporal scale of actions of a higher cognitive level with a temporal time window of 2-3 s.

Model of Joint Attention

The model of joint attention presented here is in agreement with the model of Tasker and Schmidt (2008), except that we do not adopt their time constraints for joint attention. Our model also meets the requirements of Kaplan and Hafner (2006), for a longer discussion see Pfeiffer-Lessmann and Wachsmuth (2009). However, our model differs in that we do not require interactants to perform a certain sequence of behaviors. Instead we define the effects of joint attention behaviors on their mental states. Thereby, different behaviors can be performed counting as joint attention behaviors. However, to realize a natural interaction partner, we are now investigating valid joint attention behaviors performed by humans to be implemented in our artificial agent.

We define four phases characterized by the mental states of the interactants (see Figure 1). In order to engage in joint attention, the interaction partners need to have a certain kind of psychological engagement with each other, which can be described as involving a species of perception as well as a species of emotional responsiveness (Hobson, 2005). This

can be defined as the precondition for joint attention. To establish joint attention, certain behaviors leading to certain mental states need to take place. The first phase can be described as the *initiation-phase*; one of the interactants performs an initiate-act, which the other interactant can recognize. The second phase can be described as the *respond-phase*. Now the addressee of the initiate-act needs to perform a respond-act. The third phase is characterized as the *feedback-phase*; the interactants affirm that they have recognized the interaction attempts of their interaction partners. The forth and last phase consists of the *focus-phase*; now both interactants focus on the object of attention and are aware of the joint attention state (see also Pfeiffer-Lessmann and Wachsmuth (2009) for a formalized definition of the required mental state for joint attention).

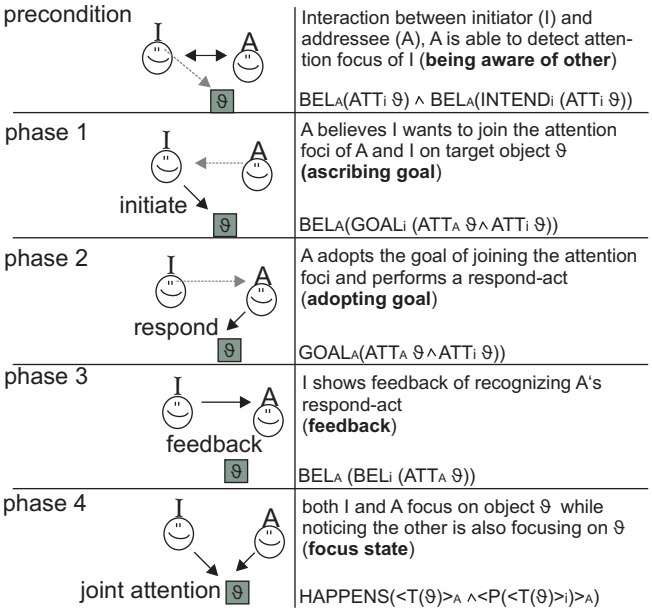

Figure 1: Phases of the joint attention process - Initiator (I) and addressee (A) wear hats according to their roles.

The model of Huang and Thomaz (2011) with five steps shares many features with our model except that we define to-be-aware of the interactant as a prerequisite and not a step in the joint attention process and that they concentrate in their step 4 on verifying the response of the addressee whereas we lay more emphasis on the required feedback mechanisms between both interactants.

Study on the Timing During the Initiate-Act

While the review of related work has brought up timing data on the phases 2 to 4 of our model, little has been found on the internal timing of events during the initiate-act. With the following study, we address the question on the timing of the initiator's referential act in which she first introduces the target of the joint attention process. Additionally, we investigate acceptable response times of the addressee for a referential act to be considered successful. As a first step, we thereby focus on referential acts via eye gaze.

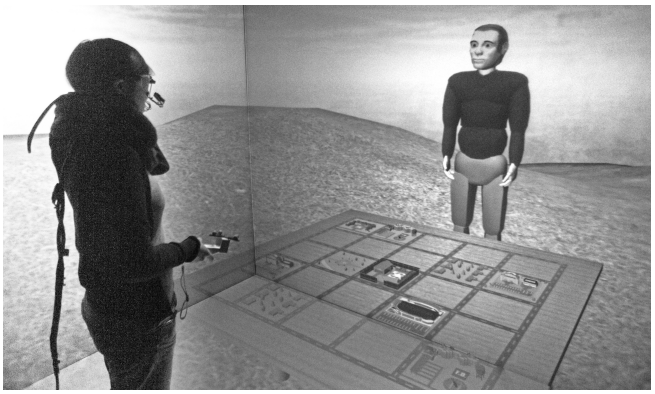

Figure 2: In the study, the human participant faces the virtual agent Max in a fully immersive virtual reality environment. Between the two interlocutors is a table with ten objects from a city planning scenario, which serve as reference objects. The eye gaze and the head movements of the human participant are tracked and the line of gaze onto the objects in the virtual environment is computed in real-time.

Scenario We investigate joint attention in a cooperative interaction scenario with the virtual human Max, where the human interlocutor meets the agent face-to-face in 3D virtual reality (see Figure 2). The human's body movements and gaze are picked up by infrared cameras and an eye tracker (Pfeiffer, 2011). This enables Max to follow the human's head movements and gaze in real-time, the two aspects of human joint attention behavior considered in this study.

Participants

Altogether data from 20 participants (10 women, 10 men) has been collected. All participants were students or employees of Bielefeld University. The age of the participants was between 21 and 45 years, with a mean of 28 years and a SD of 5.17.

Method

The participants were invited to our lab and given a brief introduction to the study. At this time, they filled out a short questionnaire and read the written instructions for the tasks. After that, they were equipped with the tracked stereo glasses required for the immersive virtual reality setup. For controlling the experiment, they were given a Wii Remote to step through the trials. After the participants had entered the virtual environment, they had time to get accustomed to the scenario. Finally, the eye-tracking system was calibrated and the participants repeated verbally the procedure of the study. After all questions had been answered, the trials started.

The two possible roles of an interlocutor (initiator or addressee), are reflected by the study design: two blocks I and A are repeated, where the human participant is the initiator

in block I and the addressee in block A. The blocks were repeated three times, for the first ten participants in the order IAIAIA, for the second ten in the order AIAIAI. The tasks within each block are described below. The ten items in block I and block A had a pre-randomized sequence, which was static between participants but different for the first, second and third presentation of the block.

After all blocks were completed, the participants were debriefed. Before departing, all participants received a recompense for taking part in the experiment.

I: Dwell Time of Referential Gaze Produced by Initiator
The aim of block I is gathering data about the typical dwell time of the referential gaze act of an interlocutor when attempting an initiate-act. During the initiate-act, the interlocutor focuses on the target object for a certain amount of time α_r (r for reference) until she checks back by focusing on the face of the interaction partner for time β_r. The total duration of the initiate-act is $\alpha_r + \beta_r + 2\varepsilon_r$, with ε_r being the very short time needed to shift the gaze focus.

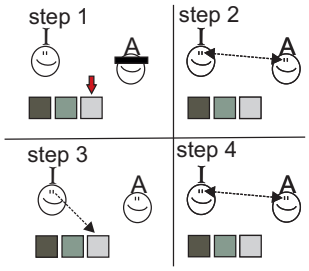

Figure 3: The sequence of steps for one item in block I. I=initiator (human) and A=addressee (agent)

The interaction scenario described above, with a human interlocutor addressing the virtual agent Max, provides the frame for this task. Max plays the role of an interlocutor, while the participant is instructed to perform an initiate-act for one of the objects located on a virtual table between the participant and Max (see Figure 2). In an orientation phase prior to each initiate-act, Max gets blindfolded and the next target object is highlighted with a red arrow (see Figure 3, step 1). This design has been chosen to make explicitly clear that Max has no prior knowledge of the target object. Once the participant has located the new target object, she has to return her gaze to Max and press a button to start the interaction. This removes the arrow as well as the blindfold of Max. Max then gives a short verbal phrase to provide the context of the joint attention act and the human participant can start her initiate-act. The participant is instructed to use gaze only to try to direct the attention of Max towards the given target object (Figure 3, step 2). She should start while focusing on Max' face, then attempt an initiate-act by focusing at the target object as long as she feels is needed (Figure 3, step 3, while we collect data on α_r). She should then interrupt focusing on the target object and check back at Max' face (Figure 3, step 4). Finally, she should press a button as soon as she feels that Max should have reacted by then (while we collect data on β_r, the expected maximum response time). Because at this point in time we do not want Max to influence the participant's timing behavior, Max does not show any reaction in response to the participant's attempts. The whole procedure is repeated for the remaining objects, until all ten objects have been covered.

A: Dwell Time of Referential Gaze Accepted by Addressee
In the second part of the study we reverse the roles of the human interlocutor and the virtual agent Max.

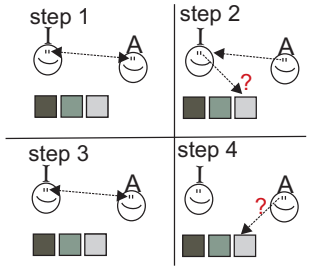

Figure 4: The sequence of steps for one item in block A. I=initiator (agent) and A=addressee (human)

Now, it is Max who is performing initiate-acts to achieve joint attention and the human interlocutor observes and evaluates these attempts. During the initiate-act, Max will stop focusing on the interlocutor and move his gaze focus to the target object for an amount of time from a predefined set ranging from 600 ms to 3000 ms in steps of 600 ms (Figure 4, step 2). These values have been selected to comprise typical non-communicative gaze durations and the findings on mean durations from the literature. Max will then focus back at the participant's face (Figure 4, step 3). The participant is asked to watch Max' gaze. Once Max focuses back at the participant, she has to decide whether Max had intended her to follow his gaze. If she decides so, she has to press a button and gaze at the target object (Figure 4, step 4). If not, she has to do nothing. After five seconds, Max will automatically continue with the next item.

During the interaction, one measurement is made. By pressing the button the human participant ascribes Max to have performed a valid initiate-act. We then count the dwell time used by Max from the given set as an acceptable dwell time for an initiate-act, α_a (a for acceptance).

Results

In a post-study questionnaire (seven-point Likert scales (1-7), median score is given here), the participants reported that they felt present in the virtual environment (score 5) and experienced the agent as being even more present (score 6). The naturalness of the communication with the agent, however, was rated 3 (SD 2). The participants also were able to fully concentrate on the task (score 6) and were not hindered by the devices. Overall, they enjoyed the experience in the virtual

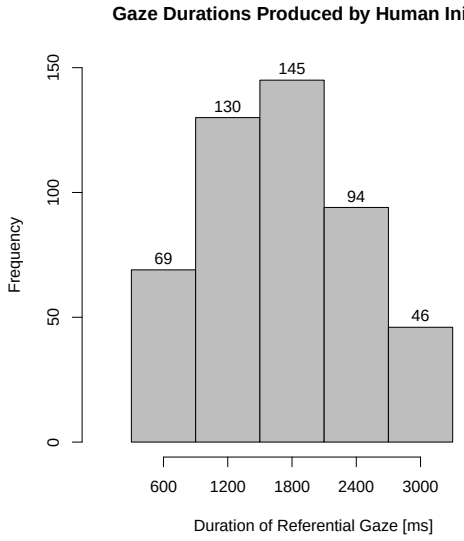

Figure 5: Dwell times of referential gaze during initiate-acts produced by the human participants in study part I.

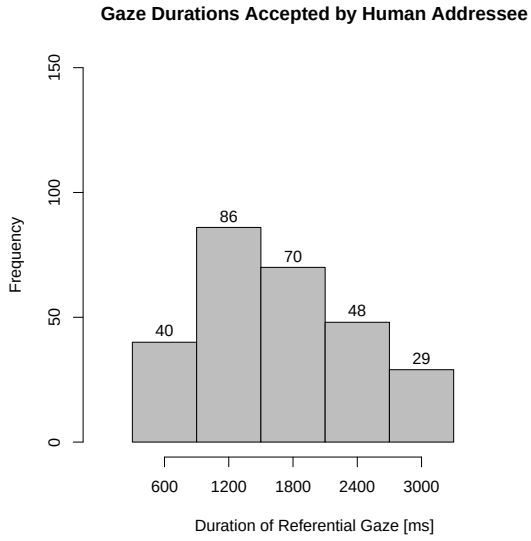

Figure 6: Dwell times of referential gaze during initiate-acts produced by Max in study part A, which have been accepted by the human participant as being intentional.

reality (score 5) and had no difficulties with the task (score 3 rating the difficulty).

I: Dwell Time of Produced Referential Gaze

During block I, 560 initiate-acts were recorded. Overall, the mean dwell time of referential eye gaze (α_r) was 1896.82 ms (SD 963.46 ms) and the median was 1796 ms. A histogram of the durations of the referential gaze is depicted in Figure 5. During the orientation phase when the participants had to identify and remember the target object, the mean dwell time of eye gaze was 1559.58 ms (SD 1029.24 ms) and the median was 1390.5 ms. The dwell time during search was significantly shorter than the dwell time of referential gaze (t-Test results in t=5.91 with p=0.001 by 545 DoF, confidence interval 215.75 ms to 430.42 ms).

Overall, the mean duration until the human participant expected a feedback after the production of a referential eye gaze towards the target object (β_r) was 2556.07 ms (SD 1721.06 ms) and the median was 2247.5 ms.

A: Dwell Time of Accepted Referential Gaze

In block A, Max produced altogether 600 initiate-acts with referential gaze of different durations (600 ms to 3000 ms in 600 ms steps). The task of the human participant was to decide, whether she accepts the gazing behavior as being intentional in that Max wanted to guide her attention to the target object. The dwell time of accepted referential gaze of the five discrete levels is α_a with a median of 1800 ms. The histogram of the accepted dwell times is depicted in Figure 6.

A chi-squared test comparing the accepted dwell times α_a in block A and the dwell times α_r for referential gaze used by the participants in block I (discretized to the discrete values used in block A) shows no significant differences (p=0.22, see also Figure 5 and Figure 6).

Discussion

For the presented study we created an immersive virtual environment and let the participants engage in joint attention with a virtual agent to have a realistic but highly controlled experimental setup to run our studies on cognitive models of joint attention. The feedback from the participants regarding their own experience of presence and the presence of the virtual agent renders this approach a success.

With this advanced setup, we aimed at substantiating our knowledge about the timing of referential gaze within the initiate-act. In block I, we found a mean dwell time of referential gaze towards the target object α_r of about 1897 ms. We take the significant differences of the α_r from the dwell time on the same target objects during the orientation phase (when the target objects are shown to the participants) as a confirmation of the different nature of gaze use in search and in referential gaze. This also shows that the design of the study is plausible to the participants regarding the different interaction states (orientation phase vs. dialog). Using these timing patterns, Max will learn to arbitrate between gaze search and referential gaze in the future.

If roles are switched and the initiate-act is performed by the virtual agent Max, we found that participants accepted the same kind of gaze patterns as natural as they themselves performed when they had the initiative. This substantiates our findings and at the same time emphasizes the high acceptance of the virtual agent Max as an interaction partner.

A respond-act of the addressee to an initiate-act of the human participant was expected before 2556 ms. This is well

below the 5 s time frame postulated by Tasker and Schmidt (2008) based on human-human interactions. However, as our study focused on the dwell times during referential gaze and Max by design only showed a response when triggered, it was difficult for the participants to decide this threshold. A more thorough investigation of this threshold should use a more complex scenario, were, e.g., Max produces respond-acts with different delays, similar to the design in block A.

Conclusion

The high acceptance of Max as an interaction partner with human-like capabilities and the comparability of our findings in human-machine interaction with those found in human-human interaction motivate us to follow this line of research further. The 1.9 s dwell time of the referential gaze act is compatible with related findings in human-human interaction. In next steps, we would substantiate our model of joint attention by incrementally increasing the complexity of the interaction scenario until the full process of joint attention can be simulated in real-time in a more natural scenario. This would also allow us to directly compare joint attention behaviors between human-human and human-agent interactions.

Although autonomous behaviors of Max were reduced to a minimum in our controlled setup his naturalness of communication was already rated 3. We believe this rating will increase significantly when he shows his full range of communicative and joint attention behaviors.

Acknowledgments

This research is supported by the Deutsche Forschungsgemeinschaft in the SFB 673 Alignment in Communication.

References

Bailly, G., Raidt, S., & Elisei, F. (2010). Gaze, conversational agents and face-to-face communication. *Speech Communication*, *52*(6), 598–612.

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., et al. (2004). Humanoid robots as cooperative partners for people. *Int. J. of Humanoid Robots*, 1–34.

Brinck, I. (2003). The objects of attention. In *Proc. of ESPP2003, Torino* (pp. 1–4).

Deak, G. O., Fasel, I., & Movellan, J. (2001). The emergence of shared attention: Using robots to test developmental theories. In *Proc. of the First Intl. Workshop on Epigenetic Robotics, Lund University Cognitive Studies, 85* (p. 95-104).

Doniec, M. W., Sun, G., & Scassellati, B. (2006). Active learning of joint attention. In *Proc. of 2006 IEEE-RAS int. conf. on humanoid robots (Humanoids 2006).*

Hobson, R. P. (2005). What Puts the Jointness into Joint Attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: communication and other minds* (p. 185-204). Oxford University Press.

Huang, C.-M., & Thomaz, A. (2011). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *RO-MAN 2011 IEEE* (pp. 65 –71).

Imai, M., Ono, T., & Ishiguro, H. (2003). Physical Relation and Expression: Joint Attention for HumanRobot Interaction. *IEEE Transactions on Industrial Electronics*, *50*(4), 636-643.

Kaplan, F., & Hafner, V. (2006). The challenges of joint attention. *Interaction Studies*, *7(2)*, 135-169.

Lessmann, N., Kopp, S., & Wachsmuth, I. (2006). Situated interaction with a virtual human - perception, action, and cognition. In G. Rickheit & I. Wachsmuth (Eds.), *Situated Communication* (p. 287-323). Berlin: Mouton de Gruyter.

Mueller-Tomfelde, C. (2007). Dwell-based pointing in applications of human computer interaction. In *Proc. of the 11th Int. Conf. on Human-Computer Interaction (INTERACT 2007)* (pp. 560–573). Springer Verlag.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, *16*, 269–274.

Nagai, Y., Asada, M., & Hosoda, K. (2006). Learning for joint attention helped by functional development. *Advanced Robotics*, *20*(10), 1165–1181.

Peters, C., Asteriadis, S., & Karpouzis, K. (2009). Investigating shared attention with a virtual agent using a gaze-based interface. *Journal on Multimodal User Interfaces, Kluwer Academic Publishers*, *3*(1-2), 119–130.

Pfeiffer, T. (2011). *Understanding multimodal deixis with gaze and gesture in conversational interfaces*. Aachen, Germany: Shaker Verlag.

Pfeiffer-Lessmann, N., & Wachsmuth, I. (2009). Formalizing joint attention in cooperative interaction with a virtual human. In B. Mertsching, M. Hund, & Z. Aziz (Eds.), *KI 2009: Advances in Artificial Intelligence* (pp. 540–547). Springer Verlag.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*(2), 268 – 291.

Tasker, S. L., & Schmidt, L. A. (2008). The dual usage problem in the explanations of joint attention and children's socioemotional development: A reconceptualization. *Developmental Review*, *28*(3), 263–288.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675-691.

Vaughan, A., Mundy, P., Block, J., Burnette, C., Delgado, C., & Gomez, Y. (2003). Child, caregiver, and temperament contributions to infant joint attention. *Infancy*, *6*(6), 603–616.

Yu, C., Schermerhorn, P., & Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Trans. Interact. Intell. Syst.*, *1*(2), 13:1–13:25.

Zhang, H., Fricker, D., & Yu, C. (2010). A multimodal real-time platform for studying human-avatar interactions. In *Proc. of the 10th Int. Conf. on Intelligent Virtual Agents* (pp. 49–56). Springer-Verlag.