User Perceptions of Communicative and Task-competent Agents in a Virtual Basketball Game

Divesh Lala, Christian Nitschke and Toyoaki Nishida
Graduate School of Informatics, Kyoto University, Kyoto, Japan

Keywords: Joint Activity Theory, Embodied Agents, Body Movement.

Abstract: In this paper, we describe a virtual basketball game where a human and an embodied agent can play together as a team. Our goal is to investigate whether the human prefers an agent who is highly competent at basketball or one which is not as competent but tries to actively communicate through body movements. The virtual basketball game was implemented using a Kinect to sense body movements and a pressure sensor for hands-free navigation. In order to create an agent who could react to a user's body movements, we designed an agent model based on joint activity theory. We performed an experiment where participants would play virtual basketball with each agent and evaluated them through questionnaires. It was found that participants preferred the agent which tried to communicate more with the user, even though they could distinguish that the other agent was better at playing basketball. We propose that communication capability for these types of agents is crucial, even at the expense of some task ability.

1 INTRODUCTION

The field of embodied conversational agents (ECAs) contains sophisticated agents which are designed to have a face-to-face conversation with real people, driving forward the concept of computers as social actors (Nass et al., 1994). ECAs such as Greta (Bevacqua et al., 2010) use multiple modalities such as speech, facial expression and gesture to smoothly communicate in this manner. On the other hand there are situations in the real world which are not specifically conversational in nature. For example, communication in sports often requires physical co-operation to achieve a common goal. This type of situated communication (Rickheit and Wachsmuth, 2006) has been somewhat addressed by robots and ECAs such as Max (Kopp et al., 2003; Kruijff et al., 2012), but generally the face-to-face paradigm appears to be more dominant in ECA research. Furthermore, even the environment in which Max is situated does not require extensive navigation, which is an activity commonly executed in sports. If the aim of researchers is to create an ideal virtual human, it is necessary for these agents to have this capability so they can interact within a wider variety of virtual environments.

One domain where agents can navigate freely with humans is in the context of video games. Thousands of scenarios exist where agents and humans interact with the environment itself such as picking up and using virtual objects or traveling to complex locations. Furthermore, activities in these scenarios are often done in order to achieve a common goal such as defeating an enemy. The limitation of such environments is their interactivity with the user, which is generally only done through peripherals such as a keyboard or mouse. Our proposal is that agents should exist between the face-to-face agent and video game agent paradigms by taking crucial elements of both. From ECAs we take the ability to interact using human modalities. From video games we take the ability to navigate in the world to achieve a common goal. These agents also require different methods of interaction. Figure 1 clarifies this concept.

The properties of the situated environment which such an agent inhabits are that navigation is necessary to achieve a common task and that the task should be completed using human modalities. A human in a virtualization of this environment should navigate in the world with the agent without needing a keyboard or mouse so they can effectively use body movement for communication. The agent must be physically dynamic and able to react to human input. We can assume that an ideal ECA will have the ability to hold a conversation with the user so that the experience is similar to talking with a real human. But what defines a "good" agent in the situated environmental context?

Lala D., Nitschke C. and Nishida T..
User Perceptions of Communicative and Task-competent Agents in a Virtual Basketball Game.
DOI: 10.5220/0005201200320043
In *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART-2015), pages 32-43
ISBN: 978-989-758-073-4

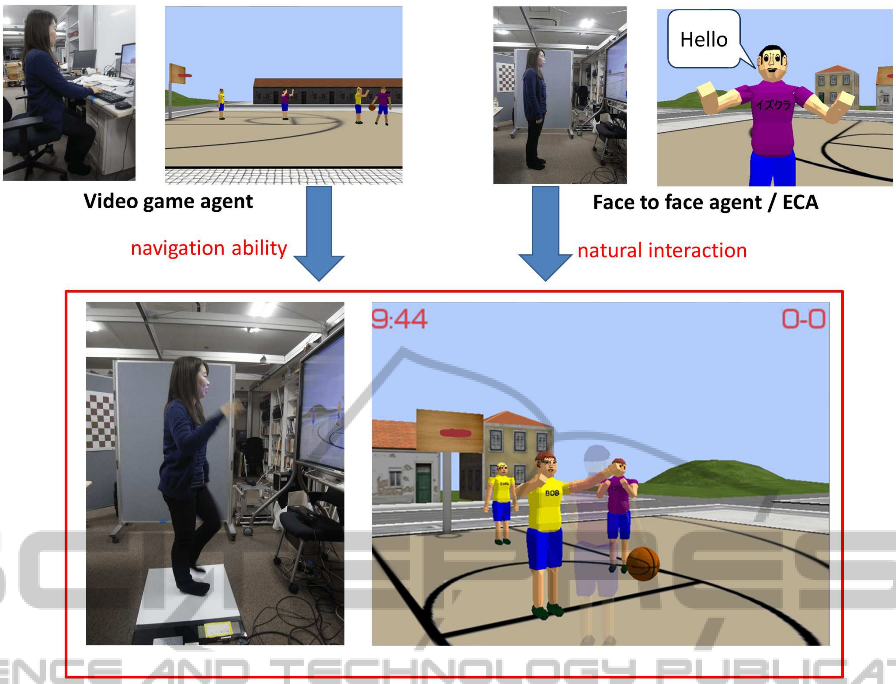

Figure 1: Navigation ability of video game agents and the human modality interface of ECAs are combined to produce an agent which travels around with an environment with the user to execute shared tasks.

We propose that agents can have the properties of being good at a task and good at communication.

Creating an agent which is "good" at a task can simply be a matter of adjusting its parameters or optimizing its decision making. Consider an extreme case of a virtual football game, where a virtual athlete is 10 times faster and more skilful than any other character. We can say this agent is good at football, but we cannot say it is realistic or sociable. It is more like a machine or tool designed to get a task done efficiently. The opposite can also be imagined. For example a guidance agent may be extremely sociable and recognize many modalities but can't provide the requested information to the user. In this case a simple paper map might be more useful. Our long-term goal is creating an agent with both task abilities and realistic social behavior which may or may not be task-related.

Current embodied agents *do* require both task and communicative abilities, but these may be weighted in terms of importance. For example, in crowd simulations (Guy et al., 2010) or exploration agents (Little and Sommer, 2011), task ability is the highest priority. On the other hand in domains such as virtual guides (Bickmore et al., 2013) and teachers (Bergmann and Macedonia, 2013), the focus is improving communication because this is largely embedded in the task (providing information to the user in a pleasant manner). Greta and ECAs from the SEMAINE project (Schroder et al., 2012) are not even designed towards a particular task, their only function is to have a natural conversation with the user. Recently there has been a shift towards using agents for more serious applications. ECAs have been proposed to provide better information to hospital patients (Bickmore et al., 2009), taking the role of a patient to train doctors (Kenny et al., 2008) and assisting people with job interviews (Baur et al., 2013). These are arguably situations where both abilities are the most intertwined and balanced because conversation is necessary to achieve the given task. However even in these tasks the domain still remains face-to-face.

Playing basketball is a domain which is separate from these environments. The agent must be able to both execute actions and communicate with the human player to achieve the goal. We aim to discover the relative importance of both task and communication abilities. Although agents are often analyzed in respect to these abilities (Kim and Baylor, 2006; Edwards et al., 2014) we do not know of prior research on their appropriate balance. Collaborative task environments such as basketball require us to consider both task and communication contributions.

This motivates our work in which we aim to assess which properties human value in a virtual agent for a physically collaborative task, in this case basketball. For virtual agents, we define physical collaboration as a shared cooperative activity requiring both navigation and usage of virtual objects. Given the choice between an agent who is clearly proficient at basketball and one which is not so proficient but can express and understand communication signals from humans, which would a human prefer? We term the former agent *task-competent* and the latter *communication-competent*. Note that these terms are relative to each other and not absolute indicators. Our research question is "*How do people perceive embodied agents with differing proficiencies and capacities of communication in a physically collaborative task?*"

We propose that within the setting of basketball, users prefer communication-competent agents to task-competent agents. The question then turns to how to design these types of agents to help test this conjecture, in particular the communication model which is used. ECAs require modalities which are different from basketball. While ECAs make use of information on speech and facial expression, basketball agents tend to use the whole body as a communication mechanism. We cannot be certain that face-to-face communication models can account for this type of agent, so we make use of a more general framework. The theory we use to guide their design is Herbert Clark's joint activity theory, or JAT (Clark, 1996). From this theory, we use two concepts - signal identification and evidence. Signals in this context are any body movements executed in order to send a message to another party. Evidence is required to prove that a body signal is actually a signal and not a movement carrying no meaning. In Section 2.2 we describe agent communication based on an evidence model.

JAT is useful because it simplifies communication as being about signals, rather than attempting to simulate an abstract internal model. In the context of basketball, signaling is offered a greater importance, because we can already reliably assume the internal goals of the human, which is to win the game. Another advantage is that JAT can model dynamic attention because it requires the agent to perceive (i.e. see) and focus on a signal. Dynamic attention is the property of an agent which switches focus between individual and collaborative tasks during an interaction. In JAT a specific interaction is commonly termed a *joint project*. In sports, these joint projects can be identified when players can focus on individual tasks such as running, kicking or throwing, or collaborative tasks such as passing a ball or celebrating with their team mates. JAT has been applied to artificial agents in other research, but this focused on robots (Bradshaw et al., 2009). We use JAT as a basis for virtual agents in this paper and aim to test whether this theory is suitable for agent design, particularly the design of agents in basketball and other sports-like environments.

The basketball game uses body interaction as a means of communication. Literature on the use of the body as a modality can be roughly categorized into gesture and full body movement. Gesture is important for face-to-face agents with many embodied agents such as Greta being used to study gesture modulation and multimodal integration with facial expressions (Pelachaud, 2009; Martin et al., 2006). On the other hand, we focus on full body movement as a driver for interaction. This has been seen as an important function for communication in agents and robots (de Gelder, 2009; Sanghvi et al., 2011; Damian et al., 2013; Kistler et al., 2013). Extensive research on body movement and engagement has also been undertaken, particularly with regards to body movement as an engaging and affective mechanism (Bianchi-Berthouze, 2013; Kleinsmith and Bianchi-Berthouze, 2013).

In the type of environment we are creating body movement is not only the primary modality, but is used to execute tasks and to navigate. Humans will use their bodies to send communication signals to the agents, who then react to it. Furthermore, these signals may be explicit or implicit. We make a distinction between explicit and implicit signals by stating that explicit signals are those which are *identifiable* as being a communicative signal even *without extra context*. For example, if we see somebody waving their arm above their head with no other context available, we can still reasonably assume that it represents a signal intended for someone or something. We don't know who the signal is intended for, nor do we have any idea about its meaning, but we still identify it as a signal. On the other hand, implicit signals, such as the rotation of the body, cannot be identified as such without context. A person who rotates their body could simply be turning to walk to another place with no communicative intent involved. These concepts are important because they are related to the JAT concept of identification. In order for any signal to be acted upon, it must be identified by the receiver as actually being a signal. Human players should be able to use their bodies much as they would in a real basketball game so agents must be able to identify both explicit and implicit signals from the human player.

Two main questions will be addressed in this work:

Q1 Can JAT be used as a theory for creating agents which communicate through body modality?

Q2 Are there task-based situations where agents competent at communication are preferred to agents competent at the task itself?

To answer the above research questions, we create a basketball game where a human plays with an agent team mate against two other agents. The participant plays two games, each differing only by the type of agent team mate (task-competent or communication-competent). We ask participants to play a game with each agent, evaluate them, and then determine if differences exist in regards to user perception.

This paper contains two contributions. The first is a validation of the need for communicative agents over those that are simply good at a task. The second is an agent model based on JAT which can be extended for multimodal interaction. Section 2 discusses the virtual basketball game and the design of each type of agent. Section 3 outlines our methodology for the experiment, while Section 4 presents our results and analysis. Section 5 discusses the results of the experiment before the conclusion of the paper.

2 VIRTUAL BASKETBALL

In this section, we describe the virtual basketball game to be used in the experiment. We first discuss the environment itself and then the design of the two agent types. We emphasize that the goal of this research is *not* to create a perfect basketball simulation or virtual basketball player. The basketball game exists only as a means to analyze communicative behavior. In this sense, the lessons from this work should be generalizable to environments other than basketball. Basketball is also used as a testbed because it is a co-operative game in which players must communicate to achieve a common goal. The rules are well known and do not need to be taught. We propose that basketball is viewed as a communication game, where the team with the best understanding of each other's intentions and behavior will be victorious. Another viewpoint is that playing basketball is similar to having a conversation, only with most signals being nonverbal in nature.

2.1 Environment

There are three main components of our virtual basketball environment, which is based on the VISIE system by Lala (2012). The first is the Kinect sensor placed in front of the user, for allowing body movement of the player's character and to recognize interactions. It also allows us to recognize passing, shooting and dribbling gestures to manipulate the ball during the game. We use algorithms based on Gaussian-mixture regression described by Calinon et al. (2007) and applied to basketball gestures in previous work by Lala et al. (2013). Although there are many ways to pass a ball, we limit this to a simple extension of both arms towards the target of the pass. One limitation of Kinect is that body recognition is unreliable if the user rotates their body away from the sensor.

The second main component is the pressure pad which the participants walk on in order to navigate their character around the court. We use a variant of an algorithm described in previous research by Lala (2012). This algorithm discriminates between a user's walking and non-walking states. Although the algorithm can also calculate the walking direction of the user, it was not required in this experiment because it is necessary for the user to face the Kinect sensor. Therefore any walking must be done in a forward direction and it must be possible to rotate the viewpoint of the user. This is achieved by the user stepping on the extreme left and right edges of the pressure pad. Additionally, the user can walk backwards during the game by stepping on the back edge of the pad. Through this method, the user can walk around the entire court by only using natural walking movements and steps for rotation purposes.

Finally, we describe the immersive display hardware component. Eight displays surround the user and project the virtual environment around them, allowing them to quickly visually perceive the whole of the court. In a dynamic game such as basketball, this is critical. With a flat screen display, the users would have to constantly rotate to check the position of other players and the court. This is undesirable in a dynamic movement game such as basketball. Images of the system in use are shown in Figure 2.

Gameplay in virtual basketball operates similarly to real basketball. The players dribble the ball around the court, can pass the ball to each other and can score by shooting the ball in the hoop. Players may steal the ball from their opponent by touching it with their hand while in the opponent is in possession. However, there are no fouls. One major difference is the use of a restart location. When possession swaps (such as the ball being out, stolen, or a goal scored), the other team takes the ball to the restart location to begin play again. The game operates much like street basketball, where both teams take turns to score in one hoop. Another difference is that players cannot jump during the game. The reason for this is to prevent physical damage to the pressure pad.

2.2 Basketball Agent Design

In our experiment we design two types of agent - one which is task-competent and one which is communication-competent. We make use of an existing basic agent and then modify it to produce differing behaviors. The basic agent has several capabilities. It

Figure 2: Virtual basketball (left) being played through the use of a Kinect sensor and a foot pressure pad. These sensors are combined and used in the immersive display setup (center). The right image shows a screenshot of the game.

can move around the court to a free position and avoid collisions, and make decisions on when to shoot. On defense, the agent can block opposing players. These decisions are made entirely without any direct input from the participant except for their relative position on the court. The opposing team in the experiment consists of two of these basic agents.

The agents themselves are fairly low resolution avatars and textures with simple animations. In fact they are well below the standards of visual realism in commercial games and other research applications. There are a couple of reasons for this design decision. Firstly, a user may expect a highly realistic agent to behave to a more human-like standard than a non-realistic one, in line with the work by Garau et al. (2003). Unlike video game agents, this agent has to react to human modality input and abnormal behavior would greatly affect user perception if the avatar was highly realistic. Furthermore, the focus in this work is on full body modalities. In order to reduce the extent of facial expressions as an influencing variable the faces, while containing some expression, are obviously static and not an indicator of any emotional state.

2.2.1 Task-competent Agent

To create the task-competent agent (T-C agent), we improve the physical characteristics of the basic agent by simply parameterizing certain features. In our experiment, this agent is 50% faster at walking, running, sidestepping and turning, while 75% faster at dribbling. Additionally, the T-C agent is 75% better at shooting and is more likely to successfully steal the ball from an opposing player. The reasons for the values of these parameters are fairly arbitrary. They are simply values which allow the T-C agent to be successful at basketball without the need for communication. The design of the T-C agent is done to provide a contrast to the other players of being more competent at basketball.

The behavior of the T-C agent also differs. Its preference is to attempt to shoot a goal. Rather than attempting to collaborate with its team mate, it will dribble towards an empty space near the goal. When this agent's team mate has the ball, it will simply find a free space, rotate towards the goal and wait. While defending, the T-C agent always approaches the opposing player with the ball. If the ball is free, it will always attempt to capture it regardless of where its team mate is located. The T-C agent is good enough to win basketball by itself. We consider this agent as being similar to a tool because the strategy to win is to simply give it the ball and let it do the work.

2.2.2 Communication-competent Agent

The communication-competent agent (C-C agent) has the same physical characteristics as a basic agent, but is 25% worse at shooting. In contrast to the T-C agent, it will react to body movement signals given from the human player and attempt to pass the ball to them if possible. It will signal for the ball if it is in a free space and can be seen by the human. Unlike the T-C agent, it displays its focus of attention through the rotation of its body. When defending, it will approach the nearest opponent in a man-to-man basketball marking strategy. If the ball becomes free, it will only attempt to capture it if it is closer than its team mate, or if its team mate is not moving towards the ball.

We implemented an evidence-based model for this agent inspired by JAT. The concept of this model is that each signal must be perceived (observed by the agent in the virtual world), identified (discriminated as a signal as opposed to ordinary movement behavior), recognized (the meaning understood) and accepted. To facilitate this process, the agent makes use of common ground, which is the knowledge shared between agent and human. This must be programmed into the agent beforehand. In this case, common ground knowledge is limited to the following:

- Rotation of the body indicates the focus of attention of a player.
- Wide spatial movement of the arms when the team

mate has the ball indicates a call for a pass.

- Wide spatial movement of the arms after a goal is scored indicates celebration.

We generated this knowledge of common ground from analyzing interactions between human team mates in the same environment (Lala and Nishida, 2014). In the previous experiment it was found that human team mates reacted to both implicit and explicit signaling. This justifies the discrimination of these signal types. For example, a turn towards a team mate possessor in open space was an implicit signal of a request to pass. An explicit waving of the arms signal after a goal was scored indicated celebration. The above actions were used in almost all games even with no previous interactions between team mates and little experience in basketball, which shows some generalizable behavior that can be implemented into agents.

The agent infers the meaning of these signals through the context of the game. For explicit signals, a specific gesture model is not used because there are a variety of user signals that would need to be recognized. From previous observations, we found that a wide spatial arm movement is common among these gestures, and so include it as common knowledge for the agent. Additionally, it also expresses its own intentions by calling for a pass in a similar manner (raising its hand). If it identifies a signal that a human player is looking to pass, it will hold up its arms, readying itself to catch the ball. The C-C agent also expresses a celebration action when their team scores a goal, an apology action when it makes a mistake such as throwing the ball out, and an encouragement action for when the human player makes an error.

Another feature which differentiates face-to-face and basketball agents is dynamic attention. For designing these agents according to JAT, there must be some method with which agents can engage and disengage from a collaborative act. As stated previously, in JAT terminology this is a joint project. The agent should have the ability to recognize joint projects and participate in them. To do this quantitatively, a simple evidence model was constructed for the C-C agent. It continuously looks for evidence that their partner wishes to engage in communication with them. To do this, it sums n number of f variables weighted by w over a certain period of time t, to produce evidence, ε^*. It then checks this evidence against a threshold value, thr. If this threshold is exceeded, it identifies a signal, σ:

$$\varepsilon = \sum_{i=1}^{n} w_i f_i \qquad \varepsilon^* = \sum_{j=1}^{t} \varepsilon_j \qquad \varepsilon^* > thr \rightarrow \sigma \quad (1)$$

The C-C agent uses two variables (f_1 and f_2) to recognize if the user is waiting or ready to throw a pass. f_1 is the relative rotation of the body of the user towards the player. f_2 is the relative movement of the user (towards the agent, away from the agent, or stationary). f_3 is used as a third binary variable to identify explicit signals, by identifying the raising of the arms higher than the shoulders.

3 METHODOLOGY

We used the virtual basketball environment to compare the two agent types. The only difference between the two games was the type of team mate agent used. One limitation of this work which we acknowledge is that there is no control agent (i.e. an agent with poor communication and task ability). The major reason for this is that after preliminary testing we found such an agent to be extremely useless. We make the assumption that this agent would generally be rated very low. Furthermore, having each user play three basketball games would likely increase fatigue, both physical and mental. Ideally a control agent would have been used but for these reasons we deemed this experimental design to be suitable.

Questionnaires were used to evaluate and compare the agents and consisted of three types of items. The first was a semantic differential scale, much of which is based on the Godspeed questionnaire described by Bartneck et al. (2009). Originally designed for robot interaction, we use this measurement to test the latent variables of perceived intelligence, animacy, and likeability. To tune the questionnaire more towards virtual characters, we decided to only use three items to measure the animacy construct: stagnant-lively, mechanical-human and artificial-lifelike. In addition, we also included our own semantic differential scales to measure task ability. These items asked the participants for the subjective interpretation of the agent's running speed (slow-fast), dribble speed (slow-fast) and defensive ability (not good at-good at).

The second type of questions were 5-point Likert scales (strongly disagree-strongly agree) which measured various aspects of the agents and the game itself. These were used to compare the agents with respect to collaboration, showing of intent and general likeability:

2A I was good at playing the basketball game.

2B When I tried to show my intention toward my team mate, it understood me.

2C When my team mate tried to show its intention toward me, I understood it.

Table 1: Analysis of semantic differential scale items.

	T-C Agent median	JAT Agent median	Wilcoxon S-R p-value
Perceived Intelligence			
Competent	4	4	0.176
Knowledgeable	3	3	0.911
Responsible	4	3	0.392
Intelligent	3	3	0.988
Sensible	3	3	0.815
Animacy			
Lively	4	4	0.372
Human-like	3	3	0.384
Life-like	2.5	3	0.374
Likeability			
Like	3	4	**0.011**
Friendly	2	4	**0.001**
Kind	3	3	**0.029**
Pleasant	3	4	**0.005**
Nice	3	4	**0.014**
Task ability			
Run	4.5	3	$<$**0.001**
Shoot	5	3	$<$**0.001**
Defend	4	3	$<$**0.001**
MyGoal	5	1.5	$<$**0.001**
OppGoal	0	2	$<$**0.001**

2D It was fun playing basketball with this team mate.

2E I want to play basketball with this team mate again.

2F I collaborated effectively with my team mate.

2G The other team collaborated effectively.

Semantic differential scales and Likert scale questions were answered after the end of each game. The final set of questions were a direct comparison of the two agents. These were answered at the end of both games. In these questions, participants were asked to directly compare the agents in terms of intelligence, likeability, showing of intention, basketball ability and future interaction:

3A Which character was more intelligent?

3B Which character did you like the best?

3C Which character actively tried to show its intention?

3D Which character was better at playing basketball?

3E If you had to play the game again, which character would you select?

A total of 32 participants were recruited for the experiment, comprised of 8 female and 24 male Japanese university students. None of the participants had previously used a Kinect before but all of them had played basketball in some capacity, although none were playing at a competitive level. Participants were given a free training session before the experiment in order to familiarize themselves with the environment, navigation, and interaction system. These training sessions took around 5-10 minutes until the participant was satisfied with using the system. Each participant played with both types of agents for 10 minutes. The order of the games was randomized to negate an ordering effect.

4 RESULTS

4.1 Semantic Differential Scale Items

To compare specific items, we use Wilcoxon signed-rank to compare differences in medians of T-C and C-C agents for each participant. These results are shown in Table 1. There was no significant difference between the two agents in any of the items related to intelligence or animacy. There was a significant difference between the two in terms of the items related to likeability. The C-C agent was shown to be rated higher in all these items. On the other hand, the T-C agent was rated as a faster runner, and a better shooter and defender. In terms of task output, the T-C agent scored more and conceded less goals.

Table 2: Analysis of summated scale items.

	T-C Agent summed mean	C-C Agent summed mean	Paired t-test p-value
Perceived intelligence	16.91	16.31	0.600
Animacy	9.09	9.31	0.769
Likeability	14.38	18.00	0.002*

Table 3: Analysis of Likert scale items.

Item	T-C Agent median	C-C Agent median	Wilcoxon S-R p-value
2A: Self-competence	2	2	0.309
2B: My intention recognized	2	3	0.017*
2C: Team mate intention recognized	2	4	<0.001*
2D: Fun	3	4	0.036*
2E: Would play again	3	4	0.080
2F: Collaboration own team	2	3	0.171
2G: Collaboration other team	3	4	0.1110

We then constructed summated scales to measure differences in means of the perceived intelligence, animacy, and likeability. To ensure reliability, Cronbach's alpha was calculated for the set of items comprising these constructs. It was found to meet internal consistency requirements for the construct of perceived intelligence ($\alpha = 0.8909$), animacy ($\alpha = 0.8066$) and likeability ($\alpha = 0.9360$). Items measuring task ability were found not to meet this requirement ($\alpha = 0.6333$) and so were not summated. Through these summated scales, we use a paired sample t-test to check for differences in mean. These results are provided in Table 2.

The summated scales verified our earlier individual item tests. There was no significant difference in the two agents in the perceived intelligence and animacy constructs. The C-C agent was ranked significantly higher than the T-C agent in the likeability construct.

4.2 Likert Scale Items

The Likert scale items were also individually tested using both Wilcoxon signed rank and Mann-Whitney U. Items were scored from 1 (strongly disagree) to 5 (strongly agree). These results are shown in Table 3.

The items with significant differences between the agents were items measuring the recognition of intention, the expression of intention, and fun. In all these items, the C-C agent was rated higher. Interestingly, there was no significant difference between the agents in terms of how participants rated their collaboration.

4.3 Direct Comparisons

Finally, we use simple frequency analysis to show direct comparisons of the agents. This is presented in Figure 3. Like our previous analyses, there was no significant difference in the two agents in terms of intelligence. The C-C agent was liked by more participants, more likely to actively show its intention, and was preferred for future interaction. More participants rated the T-C agent as being good at basketball. The frequency graph is shown in Figure 3.

We wished to investigate if there were any underlying differences in participants who chose different agents as being more intelligent. To investigate this issue further, we performed a crude analysis by dividing the participants into two groups according to their answer. We then ran statistical tests on each item within these two groups. Table 4 shows these results. As expected, there was a significant difference in perceived intelligence items according to each group's choice of agent. For participants who chose the T-C agent as being more intelligent, there was no significant difference in likeability items. These participants also did not find any differences in regard to the recognition of intention, the expression of intention, or fun. On the other hand, participants who thought the C-C agent was the most intelligent thought this agent was more life-like and that it collaborated better with them than the T-C agent.

5 ANALYSIS AND DISCUSSION

We now present several findings from the experiments:

- The agent designed using JAT theory is distinguishable from the task-competent agent.
- There is a significant difference in the likeability of the two agents.
- The task-competent agent is recognized by the

Table 4: Sub-group analysis of participants according to agent chosen as most intelligent

Agent chosen as most intelligent	**T-C Agent** ($n = 15$)		**C-C Agent** ($n = 17$)	
	Best agent	p-value	Best agent	p-value
Perceived Intelligence	T-C	<**0.001**	C-C	<**0.001**
Likeability	-	0.831	C-C	<**0.001**
Lifelike	-	0.246	C-C	**0.037**
My intent recognized	-	0.501	C-C	<**0.001**
Team mate intent recognized	-	0.349	C-C	<**0.001**
Collaboration own team	-	0.186	C-C	**0.0040**

participants at being more competent at basketball.

- The communication-competent agent is recognized as being more able to recognize and express user intention.
- The communication-competent agent is preferred by the majority of participants.

Firstly, it is clear that the two agents are distinguishable. We found a number of statistically significant differences between the two. It can be said that at the very least participants could recognize that they were interacting with different agents so could evaluate them separately. This finding is important because it means that the preferences of the users were more likely to be based on differences rather than guesswork.

In terms of likeability, there was a statistically significant difference between the two, supported by the individual items tests, summated scale and direct comparison. The C-C agent was found to be much more likeable. Additionally, it was found to be much more fun than the T-C agent, which could also be related to the likeability aspect.

The T-C agent was objectively better at the task, in both scoring and preventing goals. Through the item tests the T-C agent was found to be rated a faster runner, better at shooting the ball, and better at defending. Additionally, the direct comparison test showed the T-C agent to be better at basketball in general. This confirmed the expectations of the experiment.

Our Likert scale and direct comparison test provided support to show that the C-C agent had better communication capabilities in terms of body movement recognition and expression. Along with the previous finding it showed that users perceived the agents as being different based on the design. The C-C agent was more communicative but worse at basketball.

If we assume that preference can be measured through future interaction intention, then the C-C agent was much more preferred than the T-C agent, supported by the direct comparison test.

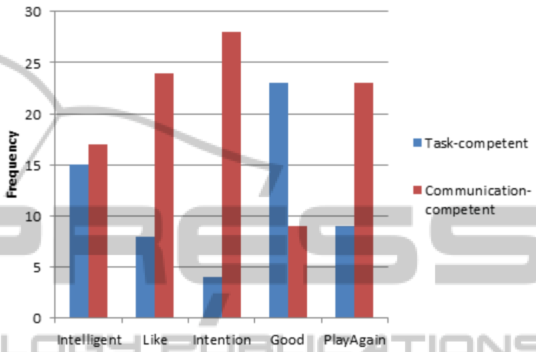

Figure 3: Direct comparison of T-C and C-C agents.

5.1 Agent Differences

Overall, we can conclude that for our basketball game, an agent which attempts to communicate with the user is much more preferred over an agent who merely does the task well, which goes some way to answering our research question **Q2**. In the virtual basketball scenario task ability is important to achieving the goal. However participants still preferred the C-C agent even if it was poorer at the task. We have also shown that the differences between the agents can be identified by participants while playing the game.

There is no obvious answer to exactly why the C-C agent was preferred, but there can be several explanations. The most obvious one is simply that people like to interact. In a real life basketball game, this phenomenon would likely occur. We can imagine that a real game would be incredibly boring if only one player did everything. It would seem that this is the same for agents. Our experiment provides evidence that participants transferred human properties to the agents, such as kindness and friendliness. They did not just consider the agent as a tool, but as a social actor. If participants were only interested in the outcome of the game, the T-C agent would be highly valued, but we have shown that even with a task with a clearly measurable outcome, social properties of an agent have a more positive effect.

Another reason for the preference of the C-C agent could be related to the difficulty of the game. An ideal game must be challenging enough so the user does not get bored, but not so difficult that they become frustrated. In our experiment, this balance may have been addressed more by the C-C agent. The T-C agent may have been too good at basketball - the participant found no challenge in the game. However, we do not know how "bad" the C-C agent has to become before the user perceives it negatively. The parameterized physical abilities and communicative capabilities of the agents would need to be adjusted to assess the optimal balance of the game in terms of agent task and communication.

We also believe that the task domain is important influencing this balance. Basketball, while presented here as a communication analysis tool, is perceived by the participants as a game for entertainment. For this reason we cannot claim that the same results would be seen in other co-operative tasks. For example, a search-and-rescue agent requires both task and communication capabilities but because the task is more serious in nature users may be willing to trade smooth communication for the agent being able to find people more efficiently.

Designers should maximize both task and communication abilities because in many cases they do not overlap, but it would be useful to know what features of agents deserve the most effort. In this case it would appear to be the latter. We can say that being competent at playing basketball (which is an important goal for AI in video games) is not as useful as ensuring good communication skill.

5.2 Perceived intelligence

Perceived intelligence did not appear to be effected by the ability to communicate, nor did it effect the preference of the participants. We initially thought that the agent with the ability to read and express body language would be seen as more intelligent. This is not borne out in the results. It would also appear that more complex behaviors would need to be displayed in order for the users to believe the C-C agent was truly intelligent. This increased complexity includes both intelligent communicative behavior (e.g. recognition of more complex signals) and task-based behavior (e.g. developing an better game strategy).

One other explanation could be that humans consider both task ability and communication ability as forms of intelligence. Therefore, each agent can be seen as intelligent, but exhibit different *types* of intelligence. A similar argument can be made for humans in the real world, for example Gardner's theory of multiple intelligence (Gardner, 1983). The detailed analysis could be indicative of this. The participants who chose the C-C agent as being more intelligent also rated it as being more likeable and able to show intention. However there was no similar pattern for participants who chose the T-C agent. This could indicate that task and communication ability are metrics for different intelligences and that individuals perceive intelligence according to one of these categories. A future research avenue then becomes how we can more accurately implement and measure these different intelligence types.

5.3 The JAT Model Agent

One outcome of this research is the implementation our evidence-based JAT agent model, designed to address **Q1**. The C-C agent was found to be able to express and recognize body movements indicating intention signals from participants. This result shows that our model is at least better than having no communication functionality. We cannot definitively conclude that JAT is the best model for our purpose as many confounding factors exist. However JAT concepts, particularly identification through implicit signals, provide a perspective on human-agent collaboration which differs from simply recognizing explicit gestures. One way to improve the agent would be to learn signals and strategies from the human during the game through common ground inferences.

How would JAT fare against other agent designs? It is difficult to compare across domains but we propose that as a *conceptual framework*, JAT is satisfactory for agent implementation. In this work the concepts were signals, joint projects and an evidence model. We justified the use of JAT because it could accommodate a dynamic navigable environment and the implementation of the JAT agent appeared satisfactory. Users of the JAT agent acknowledged that intentions could be known (through the use of signals) and intentions could be recognized (through the evidence model). Identification of implicit and explicit signals is also crucial and the JAT framework's ability to explicitly model this process make it a useful consideration for designers.

The next step is to compare the JAT model with others and expand it to more modalities. A useful property of our model is that it can be used for multiple modalities by increasing the number of variables to include any desirable features. For example, speech pitch and skin temperature can be included in the same model. We are currently considering how to implement sound in our basketball system to further test different modalities and users' reactions to them.

5.4 Limitations

There are a few limitations that we should consider. First, there was no reward for winning the game. If a motivation were involved, participants may have been more inclined to play with the T-C agent. The fact that the perception of the experiment was close to a video game influenced the participant's assessments of the agents. It is likely they were looking to be entertained rather than trying to win, which meant the C-C agent was more likely to be preferred.

Secondly, we do not know the effects of these differences over a longer time period. As a participant becomes more familiar with the environment, they may prefer to play the game individually and have no need for either agent. Alternatively, they could become used to the agent and be able to predict their actions. In this case, more co-operative behavior may arise which changes their perceptions, particularly in regards to intelligence. The games used in the experiment were only 10 minutes long so there was probably not enough time for the participants to get familiar with the agent, especially given that they had to also familiarize themselves with the interface.

Finally, the environment itself has a limited communication channel, that of body movement. This experiment deliberately left out these modalities to focus on the body but in order for basketball to be useful, multimodal interaction must be implemented. It is impossible to claim that this agent is close to a virtual human. However there are many potential research directions associated with speech and gesture combination in this environment because they require different interaction protocols than ECAs.

6 CONCLUSION

In this paper we designed a virtual basketball game in which the users could control an avatar, perform basketball gestures and navigate the court without the need for hand-held peripherals. Our goal was to assess people's perceptions of an agent team mate with higher basketball ability against one with higher communication ability. We also evaluated our joint activity theory-based agent model for the communication-competent agent. We found that people were able to distinguish between the two agents, and preferred the one with higher communication ability but there existed no difference in the perceived intelligence of the agents. This would suggest that users prefer communication ability to task ability in this environment although this could largely be due to the nature of the game itself.

Another important outcome is the use of joint activity theory as a basis for agent design. The use of an evidence model to both recognize and express intention through body movements was acknowledged by the participants through questionnaires. We proposes that joint activity theory can serve as a basis for agent design in this type of environment. Our future plan is to improve the model of the so it can be used with different modalities in agents which are not ECAs.

ACKNOWLEDGEMENTS

This research is supported by the Center of Innovation Program from Japan Science and Technology Agency, JST, AFOSR/AOARD Grant No. FA2386-14-1-0005, JSPS Grant-in-Aid for Scientific Research (A) Number 24240023, and Kyoto University Design School.

REFERENCES

Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81.

Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K., and André, E. (2013). A job interview simulation: Social cue-based interaction with a virtual character. In *2013 International Conference on Social Computing (SocialCom)*, pages 220–227. IEEE.

Bergmann, K. and Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learner's memory performance. In Aylett, R., Krenn, B., Pelachaud, C., and Shimodaira, H., editors, *Intelligent Virtual Agents*, volume 8108 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.

Bevacqua, E., Prepin, K., Niewiadomski, R., de Sevin, E., and Pelachaud, C. (2010). Greta: Towards an interactive conversational virtual companion. *Artificial Companions in Society: perspectives on the Present and Future*, pages 1–17.

Bianchi-Berthouze, N. (2013). Understanding the role of body movement in player engagement. *Human-Computer Interaction*, 28(1):40–75.

Bickmore, T. W., Pfeifer, L. M., and Jack, B. W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1265–1274. ACM.

Bickmore, T. W., Vardoulakis, L. M. P., and Schulman, D. (2013). Tinker: a relational agent museum guide. *Autonomous Agents and Multi-Agent Systems*, 27(2).

Bradshaw, J. M., Feltovich, P. J., Johnson, M., Breedy, M. R., Bunch, L., Eskridge, T. C., Jung, H., Lott, J., Uszok, A., and van Diggelen, J. (2009). From tools to teammates: Joint activity in human-agent-robot teams. In Kurosu, M., editor, *HCI ('10)*, volume 5619 of *Lecture Notes in Computer Science*, pages 935–944.

Calinon, S., Guenter, F., and Billard, A. (2007). On Learning, Representing, and Generalizing a Task in a Humanoid Robot. *Systems, Man and Cybernetics, Part B, IEEE Transactions on*, 37(2):286–298.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Damian, I., Buhling, R., Kistler, F., Billinghurst, M., Obaid, M., and Andre, E. (2013). Motion capturing empowered interaction with a virtual agent in an augmented reality environment. In *2013 IEEE International Symposium on Mixed and Augmented Reality*, pages 1–6.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3475–3484.

Edwards, C., Edwards, A., Spence, P. R., and Shelton, A. K. (2014). Is that a bot running the social media feed? testing the differences in perceptions of communication quality for a human agent and a bot agent on twitter. *Computers in Human Behavior*, 33:372–376.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 529–536. ACM.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books, New York.

Guy, S. J., Chhugani, J., Curtis, S., Dubey, P., Lin, M., and Manocha, D. (2010). PLEdestrians: A least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, pages 119–128.

Kenny, P., Parsons, T. D., Gratch, J., and Rizzo, A. A. (2008). Evaluation of justina: a virtual patient with ptsd. In *Intelligent virtual agents*, pages 394–408. Springer.

Kim, Y. and Baylor, A. L. (2006). Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educational Technology Research and Development*, 54(3):223–243.

Kistler, F., André, E., Mascarenhas, S., Silva, A., Paiva, A., Degens, N., Hofstede, G. J., Krumhuber, E., Kappas, A., and Aylett, R. (2013). Traveller: An interactive cultural training system controlled by user-defined body gestures. In *Human-Computer Interaction–INTERACT 2013*, pages 697–704. Springer.

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, pages 1–20.

Kopp, S., Jung, B., Leßmann, N., and Wachsmuth, I. (2003). Max - a multimodal assistant in virtual reality construction. *KI - Künstliche Intelligenz*, 4(03):11–17.

Kruijff, G.-J. M., Janíček, M., and Zender, H. (2012). Situated communication for joint activity in human-robot teams. *IEEE Intelligent Systems*, 27(2):27–35.

Lala, D. (2012). VISIE: A spatially immersive environment for capturing and analyzing body expression in virtual worlds. Masters thesis, Kyoto University.

Lala, D., Mohammad, Y., and Nishida, T. (2013). Unsupervised gesture recognition system for learning manipulative actions in virtual basketball. In *Proceedings of the 1st International Conference on Human-Agent Interaction*.

Lala, D. and Nishida, T. (2014). A joint activity theory analysis of body interactions in multiplayer virtual basketball. In *Proceedings of the 28th British HCI Conference*.

Little, D. Y. and Sommer, F. T. (2011). Learning in embodied action-perception loops through exploration. *CoRR*.

Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S., and Pelachaud, C. (2006). Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, 3(3):269–291.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM.

Pelachaud, C. (2009). Studies on gesture expressivity for a virtual agent. *Speech Communication*, 51(7):630–639.

Rickheit, G. and Wachsmuth, I. (2006). *Situated Communication*, volume 166. Walter de Gruyter.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *6th Intlernational Conference on Human-Robot Interaction*, pages 305–311.

Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., Ter Maat, M., McKeown, G., Pammi, S., Pantic, M., et al. (2012). Building autonomous sensitive artificial listeners. *Affective Computing, IEEE Transactions on*, 3(2):165–183.