Connection Science
Vol. 00, No. 00, Month 2009, 1–32

Taylor & Francis
Taylor & Francis Group

Drum-mate: interaction dynamics and gestures in human–humanoid drumming experiments

Hatice Kose-Bagci*, Kerstin Dautenhahn, Dag S. Syrdal and Chrystopher L. Nehaniv

Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, UK

(*Received 25 July 2008; final version received 29 September 2009*)

This article investigates the role of interaction kinesics in human–robot interaction (HRI). We adopted a bottom-up, synthetic approach towards interactive competencies in robots using simple, minimal computational models underlying the robot's interaction dynamics. We present two empirical, exploratory studies investigating a drumming experience with a humanoid robot (KASPAR) and a human. In the first experiment, the turn-taking behaviour of the humanoid is deterministic and the non-verbal gestures of the robot accompany its drumming to assess the impact of non-verbal gestures on the interaction. The second experiment studies a computational framework that facilitates emergent turn-taking dynamics, whereby the particular dynamics of turn-taking emerge from the social interaction between the human and the humanoid. The results from the HRI experiments are presented and analysed qualitatively (in terms of the participants' subjective experiences) and quantitatively (concerning the drumming performance of the human–robot pair). The results point out a trade-off between the subjective evaluation of the drumming experience from the perspective of the participants and the objective evaluation of the drumming performance. A certain number of gestures was preferred as a motivational factor in the interaction. The participants preferred the models underlying the robot's turn-taking which enable the robot and human to interact more and provide turn-taking closer to 'natural' human–human conversations, despite differences in objective measures of drumming behaviour. The results are consistent with the temporal behaviour matching hypothesis previously proposed in the literature which concerns the effect that the participants adapt their own interaction dynamics to the robot's.

Keywords: social robots; humanoids; robot drumming; human–robot interaction; interaction kinesics; emergent turn-taking

1. Introduction

The development of socially intelligent and adaptive robots in human–robot interaction (HRI) is an emerging interdisciplinary field across the boundaries of robotics, engineering and computer science on the one hand, and psychology, ethology and social sciences on the other (Dautenhahn 2007a). The primary goal of our research is to design a 'successful' HRI, whereby the robot is engaged in certain tasks and carries out these tasks in a manner that is socially appropriate, for example, enjoyable and acceptable for its users (Dautenhahn 2007b). It remains an open research challenge to design such 'successful' HRI: success is here defined in terms of both performance

*Corresponding author. Email: h.kose-bagci@herts.ac.uk

ISSN 0954-0091 print/ISSN 1360-0494 online
© 2009 Taylor & Francis
DOI: 10.1080/09540090903383189
http://www.informaworld.com

Copyedited by: KS

Techset Composition Ltd, Salisbury CCOS438496.TeX Page#: 32 Printed: 14/10/2009

of the human–robot pair in a task-based scenario, as well as in terms of the user's subjective experience of the interaction. Intuitively one may assume that what matters in human–human interaction should also matter in human–machine interaction. And indeed, research by Nass and his colleagues (e.g. Reeves and Nass 1997; Nass and Lee 2000) has shown that people treat interactive artefacts socially. However, robots and computers are not exactly like people and it remains open when and to what extent models and theories of human–human interaction are directly applicable to HRI (Dautenhahn 2007b).

In this article, we are particularly concerned with the *dynamics* of HRI. Specifically, we address the question of whether details of the dynamics of interaction that have been shown to play a fundamental role in human–human interaction are equally important in HRI. In human–human interaction, details of timing and synchronisation of gestures, speech, turn-taking in interaction, etc. influence the nature and meaning of interaction. But is the same also true of HRI? Implementing sophisticated dialogue and interaction models between humans and machines requires significant computational and research effort. In order to decide whether this effort is justified, we need to demonstrate that details of HRI kinesics matter. To address this issue, we used in our experiments simple and (algorithmically) arbitrary, minimal computational models underlying the robot's turn-taking dynamics, rather than trying to model faithfully complex mechanisms of cognition and learning in humans. We argue that if our simple models show an effect, that is, if we find that the details of simple interaction dynamics significantly influence the 'success' of the interaction (both in terms of objective performance and subjective user evaluation), then these results suggest that future research in HRI design needs to take into account the details of robot interaction dynamics even when not strictly based on cognitively plausible models of turn-taking and interaction.

The work discussed in this article is related to our wider research agenda where we study the importance of timing, rhythms, turn-taking and entrainment, which are key factors in the development of communication (cf. Robins et al. 2005; Robins, Dautenhahn, te Boekhorst, and Nehaniv 2008). Communication is an integral part of human social interaction. Developmental psychologists distinguish between: (a) a primary, expressive system which has semantic and intentional content but does not take account of the communication partner,[1] and (b) a pragmatic, referential system which can predict, and infer intention in the communication partner (Nadel, Guerini, Peze, and Rivet 1999). These two key processes are involved in supporting a transition from primary to pragmatic communication which requires mastering interpersonal timing and the ability to communicate about a shared topic. Research has identified the importance of contingency in rhythm, timing and inter-subjectivity in early communicative interaction of infants with a caregiver. Such protoconversation plays a key role in the natural developmental progression of human infants (Trevarthen 1999). Detailed analyses of infant–caretaker interactions show that turn-taking between adult and infant in these protoconversations are closely coordinated and reach rapid mutual entrainment.

Even before the link has been made to infant development, researchers studying human–human interaction had long recognised the importance of timing, turn-taking and synchronisation dynamics (Condon and Ogston 1967; Kendon 1970; Hall 1983). Goldin-Meadow argues that the gestures the people produce in their conversation are tightly intertwined in their timing and meaning, and that non-verbal gestural components of people's communication cannot be separated from the content of conversation (Goldin-Meadow and Wagner 2005). According to Bernieri and Rosenthal, '[i]nterpersonal coordination is present in nearly all aspects of our social lives, helping us to negotiate our daily face-to-face encounters… We also coordinate our non-verbal behavior with others to communicate that we are listening to them and want to hear more' (Bernieri and Rosenthal 1991, p. 401). In this context, interpersonal coordination is loosely defined as '…the degree to which the behaviors in an interaction are nonrandom, patterned, or synchronised in both timing and form' (Bernieri and Rosenthal 1991, p. 403).

Within the wider context of interpersonal coordination, in our work we focus on *interaction kinesics*, which can be described as the study of the role and timing of non-verbal behaviour, including body movements, in communicative and interactional dynamics. While numerous studies have investigated how people adapt to other humans (e.g. Pickering and Garrod 2004), non-human stimuli (e.g. Schmidt, Richardson, Arsenault, and Galantucci 2007) or computers (e.g. Suzuki and Katagiri 2007), interaction kinesics in HRI is a relatively unexplored area of research (Robins et al. 2005, 2008). And only few studies have focussed on experimental investigations of this important topic. For example, Watanabe (2004) investigated the embodied entrainment between speech and body motions such as nodding in face-to-face communication involving robotic and virtual characters engaging with people. Yoshikawa, Shinozawa, Ishiguro, Hagita, and Miyamoto (2006) highlighted the role of responsive gaze in human–humanoid interaction. Yamamoto and Watanabe (2003) found the differences in people's preferences concerning the timing of utterances in human–robot greeting interactions. Robins et al. (2008) explored interaction kinesics in child–robot interaction in a play context involving a robotic dog (Reeves and Nass 1997) and the child-sized humanoid KASPAR.[2] Yamaoka, Kanda, Ishiguro, and Hagita (2007) showed in an experiment with the Robovie robot and student participants how the contingency of interaction impacts participants' perception of the autonomy of the robot, depending on the degree of complexity of the interaction. The role of Robovie's response time as well as strategies of how a robot can cope with delays has been investigated by Shiwa, Kanda, Imai, Ishiguro, and Hagita (2008). A recent study by Yamaoka, Kanda, Ishiguro, and Hagita (2008) with Robovie studies the effect of the robot's body position and orientation on people's proxemics behaviour in joint attention scenarios. Outside the context of interactive robots, the importance of timing and synchronisation has also been studied in human–computer interaction (Suzuki and Katagiri 2007) and has been applied to therapeutic walking devices (Miyake 2003), as well as in evolved artificial social turn-taking agents (Iizuka and Ikegami 2004). The earlier-mentioned examples indicate the growing interest of the HRI community in interaction kinesics.

The particular experimental context chosen in our work is that of human–robot drumming. We decided to choose a joint drumming task since collaborative music performance, in general, lends itself to the study of interaction between humans and robots involving a variety of social aspects including imitation, gestures, turn-taking and synchronisation, occurring in an overall playful and enjoyable context. From a robotics point of view, drumming is a very suitable means of performing music, since it is relatively straightforward to implement and test, and can be realised technically without special actuators like fingers or special skills or abilities specific to drumming. Thus, the drumming scenario provides a playful and interactive context that allows to constrain and manipulate different experimental parameters easily.

Several researchers have studied drumming in the context of human–robot music performance. In Weinberg, Driscoll, and Parry (2005), Weinberg and Driscoll (2006) and Crick, Munz, and Scassellati (2006), robotic percussionists play drums in collaboration with interaction partners. In Weinberg et al. (2005), an approach based on movement generation using dynamical systems was tested on a Hoap-2 humanoid robot using drumming as a test case. Similarly, in Kotosaka and Schaal (2001), humanoid drumming is used as a test bed for exploring synchronisation. However, none of the prior work has specifically studied the socially interactive aspects in general, or interaction kinesics in particular, in the context of human–humanoid drumming, which are the focus of this article.

In this article, we present the results from two empirical studies involving adult participants[3] interacting with the humanoid robot KASPAR in an imitation-based interaction game based on drumming. The two experiments highlight the different aspects of HRI: (a) the role of (non-verbal) gesture communication in a joint drumming task, and (b) the dynamics of emergent turn-taking games. In Section 1.1, we will motivate the first experiment based on gesture communication which used non-verbal gestures as social cues. This approach is discussed in the light of related work on several robotic percussionists, as well as other work in the wider context of social robotics. In Section 1.2, we motivate our work on emergent dynamics of turn-taking interaction, in the context of literature highlighting the importance of turn-taking in conversations and interaction games. The actual experiments will be described in Sections 2 and 3. Note that the field of social robotics and HRI is very active, with a variety of different robotic systems used in interaction studies. A complete review of the literature in this field goes beyond the scope of this experimental paper; so we will focus our discussion of related work on research specifically relevant to our research questions. For a very recent review of the field of HRI, see Goodrich and Schultz (2007).

1.1. *Gesture communication: motivation and related work*

A robot that engages with people in interaction games could benefit from behaviour that specifically motivates the user and sustains the interaction while coping with a wide range of users. One way of motivating people to interact is through the use of social cues such as gestures. In human–human interaction, gestures play an important role in communication, coordination and regulation of joint activities. Indeed, in the related field of virtual agents, researchers have shown the beneficial effects of gestures and expressions used by virtual agents, both in short-term and in long-term interactions, in maintaining user involvement with the tasks encouraged by the agent (Bickmore and Cassell 2005; Bickmore and Picard 2005).

Applied to robotics, this suggests that a robot may require social cues and gestures to motivate users to interact with it, for example, in the field of assistive robotics (Tapus and Matarić 2006). A variety of robotic systems have been using social cues and gestures to encourage HRI. A well-known example is KISMET, where facial expressions were used to regulate the interaction with people inspired by interactions of infants with their caretakers (Breazeal 2002). Other recent examples include small cartoon-like robotic 'creatures' such as KEEPON and ROILLO, designed to be used in interaction with children (Kozima, Nakagawa, Yasuda, and Kosugi 2004; Michalowski, Sabanovic, and Michel 2006). These small robots have a limited action repertoire, but can produce selected gestures to engage in interaction with children in the playground. The fixed gestures are either random or tele-operated by a hidden puppeteer via a Wizard of Oz technique, as a part of social interaction. ROILLO is a simple robot with a rubber coated foam head, body and an antenna. It has three wires connected to simple servos, which move the head and body in various directions. It is used in experiments to study the interactions between the robot and the children (Michalowski et al. 2006). KEEPON is a minimalist expressive robot that only has a rubber head and an oval body. It has a small CCD camera and a microphone on it. It can move its head, turn its body and make bobbing actions to show its 'feelings'. It has both attentive and emotive actions. It is simple but robust enough to be used in play rooms in interaction with children (Kozima et al. 2004; KEEPON 2007, http://univ.nict.go.jp/people/xkozima/infanoid/robot-eng.html#keepon).

Related work on human–robot drumming includes HAILE (Weinberg et al. 2005; Weinberg and Driscoll 2006), a robot arm designed specifically to drum in dynamic and musically sophisticated collaboration with creative human musicians. HAILE does not use fixed deterministic rules, but uses autonomous methods to create variant rhythms. It perceives a variety of complex features of the human partner's drumming, analyses the sound patterns and produces rhythms in response. Compared with HAILE, in Crick et al. (2006) a less musically sophisticated humanoid robot called NICO with an upper half body torso plays a drum together with human drummers. It has visual and audio sensing to determine an appropriate tempo adaptively using a simple threshold mechanism to parse the human partner's beats, and can distinguish its own performance with audio sensing, integrating the two sources of information to predict when to perform the next beat.

The above motivation and background led to our first experiment, where the humanoid robot KASPAR plays the drums autonomously with a human ‘partner’ (interactant), trying to imitate the rhythms produced by the human while using non-verbal gestures to motivate the human. In this experiment, KASPAR’s behaviour is deterministic in the sense of producing the same (actuator) output given the same input from its sensors.[4] KASPAR produces non-verbal (head) gestures from a limited repertoire and eye-blinking as it drums. Our approach is tested using different degrees of such non-verbal gesturing with adult participants in several drumming sessions, and the experimental results are reported and analysed below (Section 2) in terms of imitation, turn-taking and the impact of non-verbal gestures as social cues.[5]

1.2. *Emergent turn-taking dynamics: motivation and related work*

Turn-taking is an important ingredient of human–human interaction and communication, whereby the role switch (‘leader’ and ‘follower’) is not determined by external sources but emerges from the interaction. Human beings generally ‘know’ when to start and stop their turns in the social interactions, based on various factors including the context and purpose of the interaction, feedback from the social interaction partners, emotional and motivational factors, etc. They use different criteria for these decisions. In this work, our aim is to build a framework which enables emergent turn-taking, and role-switching between a human and a humanoid in an imitation game, and to understand how differences in robot turn-taking strategy can influence the emergent dynamics of HRI. We do not aim to produce psychologically plausible models of human turn-taking behaviour in this work, but employ simple, minimal generative mechanisms to create different robotic turn-taking responses/strategies.

Related work that studied turn-taking in games and conversations focussed on different aspects. An example from developmental psychology is described in Hendriks-Jansen (1996), which discusses emergent turn-taking between a mother and a baby without any explicit ‘control’ mechanism (e.g. the mother starts jiggling in response to her baby’s sucking to encourage her baby to resume sucking). This results in emergent turn-taking between the jiggling and the sucking actions. Turn-taking also has important implications in robot-assisted therapy. Indeed, one therapeutically relevant issue in teaching and education of children with autism is to teach children the concept of ‘turn-taking’. Turn-taking games have been used to engage children with autism in social interactions (Dautenhahn and Billard 2002; Robins, Dautenhahn, te Boekhorst, and Billard 2004a).

Another example of turn-taking games is given from a cognitive robotics view in R.A. Brooks (personal communication, August 28, 1997). In this work, a ball game between a humanoid robot COG and the human experimenter is described. COG and the human were reaching out and grasping a ball in alternation. Note that in this case the experimenter led the turn-taking behaviour in reaction to the robot’s visually driven actions. Ito and Tani (2004) studied joint attention and turn-taking in an imitation game played with the humanoid robot QRIO, where the human participants tried to find the action patterns, which were learned by QRIO previously, by moving synchronously with the robot.

From a linguistics point of view, some of the important features of turn-taking in human conversation identified are as follows (Sacks, Schegloff, and Jefferson 1974):

- Speaker-change recurs, or at least occurs.
- Mostly, one party talks at a time.
- Occurrences of more than one party speaking at the same time are common but brief.
- Transitions (from one turn to the next) with no gap and no overlap are common (slight gap or slight overlap is accepted).
- Turn order is not fixed, but varies.

- Turn size is not fixed, but varies.
- Length of conversation is not specified in advance.
- What parties say is not specified in advance.
- Relative distribution of turns is not specified in advance.
- Number of parties can vary.
- Talk can be continuous or discontinuous.

Built on these features, Thórisson (2002) developed a turn-taking mechanism for conversations based on his previous work on the so-called Ymir mind model for communicative creatures and humanoids. He proposed, implemented and tested a generative, multi-modal turn-taking model for a face-to-face dialogue. The model was based on literature in human–human dialogue. The above-mentioned expressive humanoid robot KISMET (Breazeal 2002, 2003) which used social cues for regulating turn-taking in non-verbal interactions with people used a sophisticated robot control architecture modelling motivation, emotions and drives to satisfy KISMET's internal 'needs'. Turn-taking between KISMET and humans emerged from the robot's internal needs and goals and its perceptions of cues from its interaction partner. Rather than trying to model any particular turn-taking behaviour as observed in human–human dialogue (as it has been done e.g. in Thórisson's (2002) work mentioned above), we pursued a synthetic, bottom-up approach by defining very simple models of turn-taking based on basic mathematical functions. Such a bottom-up approach is in line with other approaches in the research field of Embodied Artificial Intelligence (Steels and Brooks 1995; Pfeifer and Scheier 1999) and is here applied to human–humanoid interaction aiming at developing socially interactive behaviour for a humanoid robot.

Also, different from the above-mentioned work with KISMET, where the interaction was the goal in itself, we wanted to include a certain (enjoyable) task that needs to be achieved jointly by the human–robot pair, to provide the overall context.

Important in this context is the temporal behaviour matching hypothesis as proposed in Robins et al. (2008), which predicts that in HRI games, people will adapt to and match the robot's temporal behaviour, similar to the effects that can be found in the literature of human–human interaction. The hypothesis has been supported in experiments with children who were playing imitation games with KASPAR (the same robot as used in our experiments; Robins et al. 2008). While this hypothesis may at first seem trivial since people and other animals are very adaptive and adapt to the dynamics of a variety of stimuli (see, e.g. Schmidt et al. 2007), for roboticists it is very important to actually know whether people do indeed adapt and respond to the dynamics of robot behaviour – if it were false then one would not need to take robot interaction dynamics and kinesics into account – which would substantially simplify HRI design. Moreover, what types of impact robot kinesics can have on interaction and the degree and manner in which different people might be influenced differently are open issues. Thus, for HRI researchers, this is an important question to study experimentally, and, as discussed in more detail above, it has only recently attracted attention in the field of robotics and HRI (c.f. Robins et al. 2005, 2008; Crick et al. 2006; Yoshikawa et al. 2006).

Based on the above motivation and background, we designed a second experiment where KASPAR plays the drums autonomously with a human 'partner' (interactant), trying to imitate the rhythms produced by the human (as a follower) and trying to motivate (as a leader in the game) the human to respond. Using different simple, probabilistic models, KASPAR decides when to start and stop its turn. It observes the human playing and uses its observations as parameters to decide whether to listen to the human or to take the turn actively in the game. This is different from Experiment I where we tested deterministic turn-taking. This work was tested with adult participants and the results were studied in terms of imitation, interaction and turn-taking.[6]

The two experiments are described below in detail separately due to clear differences in research questions and implementation of the interaction games. However, both experiments share a common methodological approach.

We chose a within-participant design for both studies for two main reasons: (a) the study of individual differences as such is an interesting challenge in HRI research (Breazeal 2004) and (b) previous research has indeed found significant individual differences in HRI studies, for example, concerning personality traits (Walters, Syrdal, Dautenhahn, te Boekhorst, and Koay 2008), gender and personality (Syrdal, Koay, Walters, and Dautenhahn 2007), human and robot personality matching (Tapus, Tapus, and Mataric 2008), and user personality and robot personality style (Wrede, Buschkaemper, and Li). Since the literature shows individual differences of how people respond in HRI studies (e.g. based on the participants' gender, age, individual personality traits, etc.), a within-participant design approach thus seemed most suitable for understanding the range and variability, and impact of robot kinesics on interactions.

In both experiments, we evaluate the objectively measured performance of the human–robot pair as well as the subjective interaction experience as judged by the human participants.

The rest of this article is organised as follows. In Section, the first experiment on deterministic turn-taking is presented, followed by Section 3, which describes the second experiment on emergent dynamics of turn-taking. Each of these two experimental sections includes the corresponding research questions as well as descriptions of the experimental setup, experimental results and discussions of the results. Section 4 presents the overall conclusion. The final section of this article outlines the ideas for future work.

2. Experiment I: deterministic turn-taking

2.1. *Methodology*

In the first experiment, the human partner played a rhythm which KASPAR tried to replicate, in a simple form of imitation (mirroring). KASPAR has two modes: listening and playing. In the listening mode, it recorded and analysed the played rhythm, and in the playing mode, it played the rhythm back by hitting the drum positioned in its lap. Then the human partner played again. This (deterministic) turn-taking continued for the fixed duration of the game. KASPAR did not imitate the strength of the beats but only the number of beats and duration between beats. For beat frequencies beyond its skill, it used instead minimum values allowed by its capabilities.[7] It also needed a few seconds before playing any rhythm to get its joints into correct reference positions.

Figure 1 presents the basic model of KASPAR–human interaction. The model requires the gestures of both human and humanoid for social interaction, as well as drumming. Human gestures or body movements were not detected in our experimental setup and were therefore not considered in the implementation.

One of the fundamental problems in this scenario is the timing of the interaction; as discussed above, timing plays a fundamental role in the regulation of interaction. It is not always clear when the robot or human partner should start interaction in taking a turn. In this experiment, the model used some predefined fixed time duration heuristics for synchronisation. KASPAR started playing if the human partner was silent for a few seconds, and tried to motivate the participant with simple gestures.

2.2. *Research questions and expectations*

Our primary research question concerned the possible impact of robot gestures on the imitation and turn-taking game (in terms of performance), but also on the participant's subsequent evaluation of

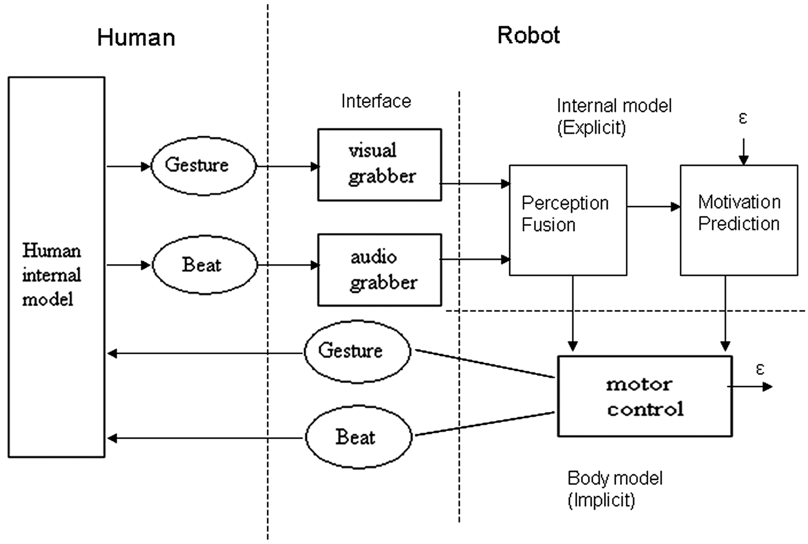


Figure 1. The model for KASPAR–human interaction.

the game. We expected that the participants would be more engaged and would evaluate the interactions more positively in experimental conditions where KASPAR used non-verbal head gestures. Moreover, we expected that too many gestures may distract people from the drumming task.

2.3. *Experimental conditions*

We studied three conditions with increasing amounts of robot gesturing:

(1) No-gesture: KASPAR did not use any gestures, it only imitated the human drumming beats it detected.
(2) Gesture: Simple head gestures (e.g. moving the head to the right or left, moving the head up or down, tilting the head slightly to different angles) and eye blinking were included in KASPAR's movements. KASPAR started drumming using one of a fixed set of gestures. If the human partners did not play their turn, then KASPAR did not respond either, and then the turn passed back to the partner. A fixed order of n gestures was used, and this order was repeated for every n turns. It was intended that the value for n should be large enough so that the participant would not realise that this was a fixed pattern but rather that the gestures seem either 'meaningful' or random (in the experiment, n was set to seven based on simulated experiments, i.e. carried out with the experimenter as the interaction partner).
(3) Gesture+: This condition is the same as *gesture*, except that KASPAR displayed on its turn in the interaction gestures even when neither the robot nor the participant played the drum. The gestures used were the same as in the *gesture* condition, and the drumming part was the same in all the three conditions.

2.4. *Experiment, results and analysis*

2.4.1. *Robot*

The experiment was carried out with the humanoid robot called KASPAR (Figure 2). KASPAR is a humanoid robot that has been designed specifically for HRI studies. It possesses a minimal set

Figure 2. The humanoid robot KASPAR and its toy drum that were used in the experiments.

of expressive robot features (cf. Blow, Dautenhahn, Appleby, Nehaniv, and Lee 2006) for more information on its design rationale. KASPAR has eight degrees of freedom (DOF) in the head and neck, and six in the arms and hands. The face is a silicon-rubber mask, which is supported on an aluminium frame. It has 2 DOF eyes fitted with video cameras, eye lids that allow blinking and a mouth capable of opening and smiling; see Blow et al. (2006) for a more detailed description.

2.4.2. *Experimental setup*

The experiment was carried out in a separate room isolated from other people and noises which could affect the drumming interaction. KASPAR was seated on a table with the drum positioned on its lap. The participants were seated in front of the robot using another drum that was fixed on the table (Figure 3). The participants used a pencil to hit the drum. Although we suggested to

Figure 3. A video snap shot from the experiments.

the participants to use one pencil and hit on the top of the drum, sometimes they used two pencils with a single hand or with both hands, and several times they used the tambourine-style bells around the drum's sides.

2.4.3. *Software features*

The implementation of robot perception and motor control used the YARP environment (Metta, Fitzpatrick, and Natale 2006). YARP is an open-source framework that supports distributed computation that emphasises robot control and efficiency. It enables the development of software for robots, without considering a specific hardware or software environment. Portaudio (2007; http://www.portaudio.com/trac/wiki/) software was used to grab the audio from the audio device, within the YARP framework. See Appendix 1 for details of the audio analysis.

2.4.4. *Participants*

Twenty-four participants (7 female and 17 male) took part in the study. Due to logistical reasons, the trials were carried out in 2 sets (a few months apart) with 12 participants each. All the participants worked in computer science or similar disciplines at the university. Only six of them had interacted with KASPAR prior to the experiment, and most of the participants were not familiar with robots in general. Note that we initially did not plan to study the influence of gender in the experiment; for this reason, the sample is not gender-balanced. However, where appropriate we mention gender differences that were observed. Four of our participants had children.

2.4.5. *Interaction game setup*

We used a 1 min demonstration of the robot without any drumming game play, where the participants were shown how to interact with KASPAR. This was followed by three games reflecting the three experimental conditions described above each lasting 3 min, without pointing out to the participants any differences between the conditions. We presented the game conditions in all the possible six different orders to analyse the effect of the order of the games. To account for possible fatigue or habituation, in the sequential order section below, we analysed the games according to their order number in the sequence experienced by the participants (independent of the particular experimental condition), as being the first game, second or third, disregarding their game types, for example, for one participant the first game (number 1) would be the no-gesture game, and for another participant, no-gesture would be the third game (number 3). After each participant finished the three games, they were asked to complete a questionnaire to assess how they subjectively evaluated the three different games.

2.4.6. *Results*

2.4.6.1. Evaluation of questionnaire data. The participants were invited to evaluate their interaction with KASPAR using a questionnaire. There were two items inviting the participant to choose which game was the most and least preferred overall. There were also three five-point Likert scales which allowed the participant to rate each drumming game in terms of (1) how much they enjoyed the game, (2) how well KASPAR drummed and (3) how sociable they perceived KASPAR to be. Open-ended questions were included to allow participants to explain their reasoning for their preferences. Most and least preferred games according to game types and sequential order were statistically analysed using a χ^2 test.

Most and least preferred games according to game type: The frequencies of participants which rated each game as most preferred and least preferred are presented in Table 1 along with residuals based on an expected count of 7.7. The differences from the expected counts were significant for both the most preferred game type ($\chi^2(2) = 6.61$, $p = 0.037$) and the least preferred game type ($\chi^2(2) = 9.74$, $p = 0.008$). The majority of the participants preferred the *gesture* game and disliked the *no-gesture* game. Their general opinion was that the game without gestures was also poor in terms of social interaction and enjoyment, which encouraged them to play more. For the *gesture* game, they said they prefer the right balance of drumming and interaction.

Most and least preferred games according to sequential order: A significant difference was found between the first and third games in terms of sequential order ($\chi^2(1) = 4.57$, $p = 0.033$). There is no significant difference overall between the three games if the second game is included (Table 2). Open-ended responses highlighted that the majority would become more familiar with the game as they played more, allowing them to interact more efficiently with KASPAR in terms of the drumming tasks. Another issue raised in the open-ended responses was that the participants would become fatigued and bored after doing the repetitive drumming task for a prolonged period of time, which may explain the lack of a significant difference between the second and third games.

Preferences: While the method of counterbalancing is an accepted means of protecting against confounders due to presentation order (Miller 1984), the clear main effect of presentation order was considered a threat to this assumption. To control for this threat, mixed model ANOVAs were run using game type to investigate possible interaction effects of presentation order and game type on both questionnaire responses and behavioural data. These were mainly non-significant, supporting the notion of independence between presentation order and game-type overall in the sample. The one exception is addressed in Section 2.4.6.2.

Sample similarities: In terms of differences between the first sample of 12 and the second sample of 12 participants, a mixed-model ANOVA found no significant differences in terms of preferences ($F(1,22) = 0.772$, $p = 0.39$). Thus, in the following we present the results from the overall sample of 24 participants.

Table 1. Most and least preferred games according to game types.

	Participants			
Game Type	Most preferred	Residual	Least preferred	Residual
No-gesture	3	−4.7	12	5.3
Gesture	12	5.3	1	−6.7
gesture+	7	−0.7	9	1.3
No preference	2	N/A	2	N/A

Table 2. Most and least preferred games according to sequential order.

	Participants			
Order	Most preferred	Residual	Least preferred	Residual
1	3	−4.3	10	−4.3
2	8	5.3	5	−0.7
3	11	−0.7	8	3.7
No preference	2	N/A	1	N/A

Preferences according to sequential order: The preferences according to order within the sample as a whole were assessed using a repeated measures ANOVA. There was an effect approaching significance in how participants rated KASPAR's drumming according to the order of the game ($F(2,46) = 3.11$, $p = 0.054$). No significant effects on game order were found in terms of the robot's sociality or enjoyment ratings. Participants tended to rate the last game more favourably across the different rating types (despite the fatigue reported by some participants during later games), see Figure 4. The results from the ANOVA, as well as the descriptives described in Figure 4, suggest that this trend was the most pronounced in the way the participants rated KASPAR's drumming.

Preferences according to game type: The repeated measures ANOVA for preferences dependent on game type found an effect approaching significance in terms of how KASPAR's drumming was rated according to game type ($F(2,46) = 2.71$, $p = 0.077$) as well as for general enjoyment of the game ($F(2,46) = 2.81$, $p = 0.07$). We found a significant effect for game type in terms of how KASPAR's sociality was rated ($F(2,46) = 5.01$, $p = 0.011$), see Figure 5.

Figure 5 suggests different trends for the different game types. The trend approaching significance for KASPAR's drumming suggests that the drumming aspect of the interaction for the *no-gesture* game was rated the most favourable, followed by the *gesture* game, with the *gesture+* game receiving the lowest rating.

In terms of the social aspect of the interaction, the opposite effect was found. The *no-gesture* game was rated the lowest, with the *gesture* and *gesture+* games rated higher. For overall enjoyment, the *gesture* games were rated the highest, followed by *gesture+*. The *no-gesture* game was rated the lowest.

2.4.6.2. Evaluation of behavioural data. The behavioural data required for the evaluation of the participant's and the robot's performance during the games were collected based on the data on the robot's own drumming behaviour and video recordings of the human's drumming behaviour which were annotated manually and then analysed quantitatively. The behavioural data include the number of turns in a specific game, the number of drumming bouts performed by the participants and the robot, and the 'drumming errors'. The errors are the differences between KASPAR's

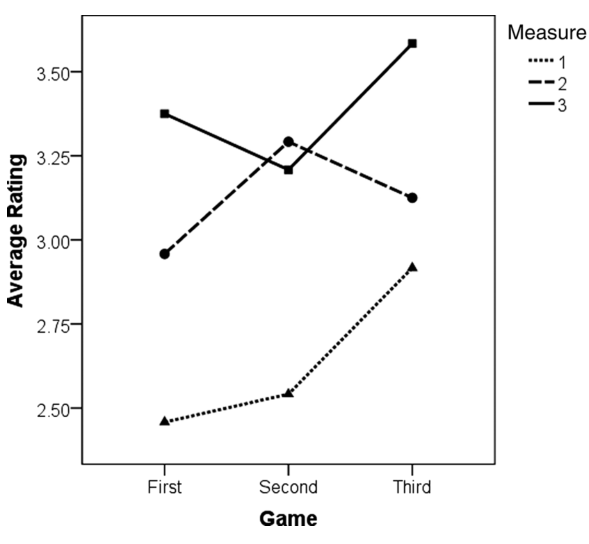


Figure 4. Ratings for games according to order in terms of (1) KASPAR's drumming, (2) KASPAR's sociality and (3) enjoyment of the game.

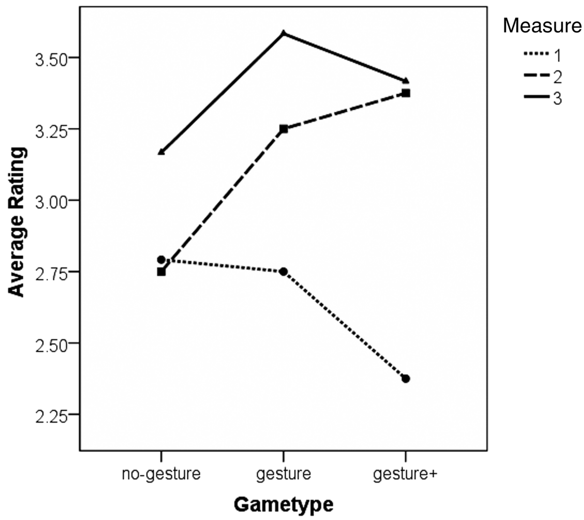


Figure 5. Ratings for games according to game type in terms of (1) KASPAR's drumming, (2) KASPAR's sociality and (3) enjoyment of the game.

actual drumming (i.e. the number of beats KASPAR plays in a particular turn) and the number of beats the participant plays. We calculated an average error per turn. Thus, 'errors' do not reflect any mistakes in the system as such, but reflect the discrepancy between human's and robot's drumming performance.

Behavioural data according to sequential order: We found a significant effect for sequential order in terms of average number of errors ($F(2,46) = 6.18$, $p = 0.004$). This effect is seen in Figure 6 and suggests that the errors were in general lower for later games.

Generally, the participants either tried very long and fast patterns or they did not beat loud enough to be detected reliably (KASPAR uses a high-level noise filter to filter out high inner noise coming from its joints, so it can only sense loud beats) when they started to play initially. Interestingly, without any external encouragement, as they got used to the game, they progressively synchronised their drumming to the robot. Details of the results are presented in Table 3. As such,

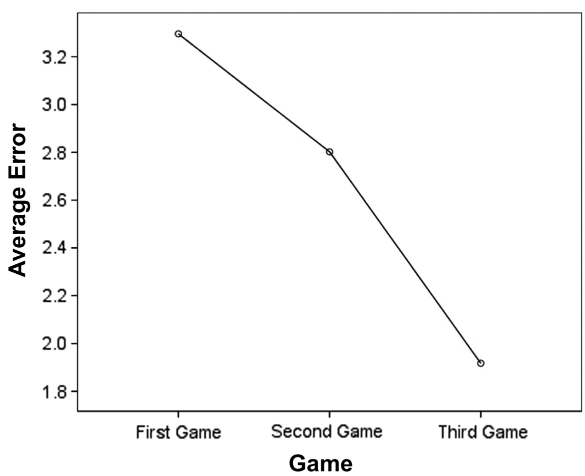


Figure 6. Average error according to sequential order.

Table 3. Observed human drumming behaviour according to order.

Order	Average error	Maximum no. of beats	Average no. of beats	Average no. of turns
1	3.30 ± 3.15	41	6.67 ± 4.22	15.88 ± 5.23
2	2.80 ± 3.36	37	5.58 ± 3.57	17.63 ± 5.84
3	1.92 ± 1.86	20	4.70 ± 2.61	19.13 ± 4.64

the preference for the third game among the participants could be explained by the lower number of errors for this game.

Behavioural data according to game type: Figure 7 shows a trend approaching statistical significance ($F(2,46) = 2.15$, $p = 0.13$) where the *gesture+* game had the highest average error, followed by the *gesture* game. The *no-gesture* game had the smallest error rate.

The maximum number of beats decreased with the increasing amount of gestures in the game (Table 4). There was a slight increase in the average number of beats with the increasing amount of gestures in the game, but this was not significant. The average number of turns tended to decrease as the amount of gestures in the game increased. This significant effect ($F(2,46) = 4.41$, $p = 0.018$) is described in Figure 8. The only interaction effect observed in this experiment between order of presentation and game type occurred for this variable ($F(2,44) = 6.020$, $p = 0.005$). This effect is described in Figure 9 and suggests that for participants who were introduced to the *gesture+* condition in the first or second game had a higher number of turns for the *no-gesture* and *gesture* game than those who encountered this game type last, while the reverse was true for the *no-gesture* condition.

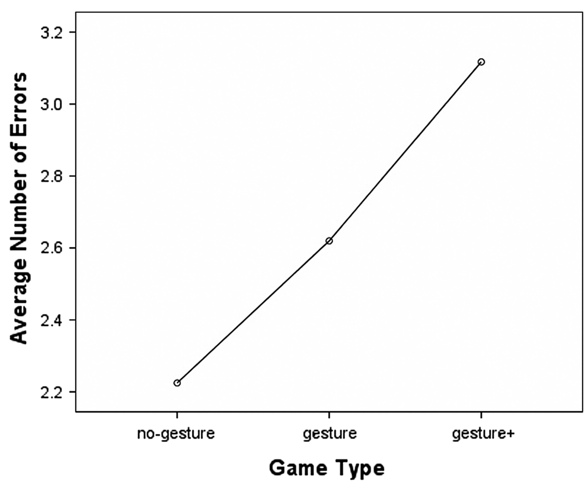


Figure 7. Average number of errors according to game type.

Table 4. Observed human drumming behaviour according to game type.

Game type	Average error	Maximum no. of beats	Average no. of beats	Average no. of turns
No-gesture	2.22 ± 2.52	41	5.24 ± 3.54	19.00 ± 5.49
Gesture	2.62 ± 3.16	37	5.60 ± 3.67	17.83 ± 4.63
Gesture+	3.12 ± 3.01	31	6.21 ± 3.89	15.58 ± 5.61

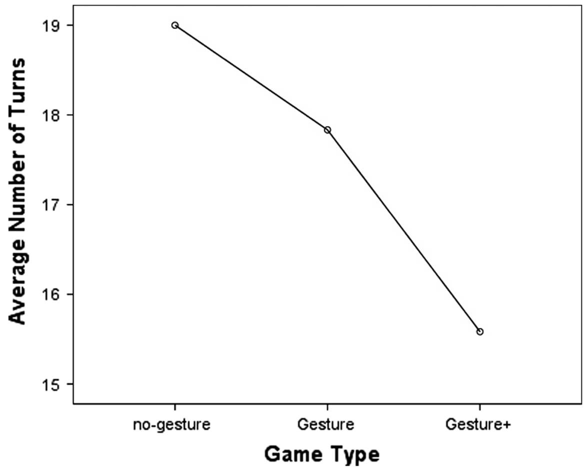


Figure 8. Average number of turns according to game type.

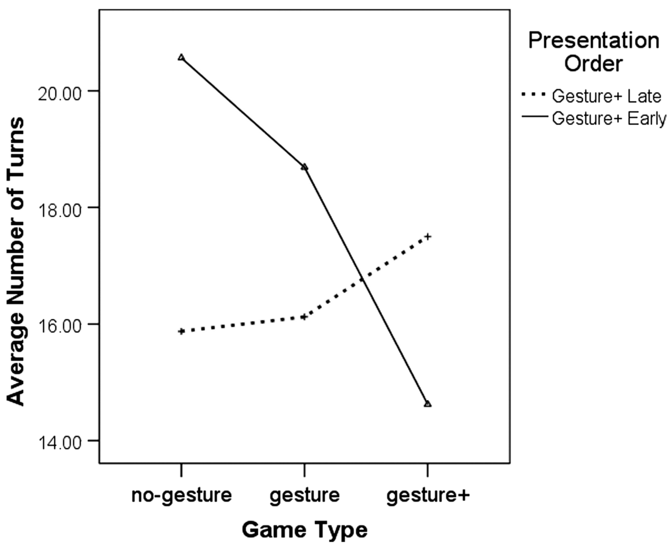


Figure 9. Interaction between game type and presentation order for number of turns.

2.4.7. *Discussion of results*

Experiment I not only investigated the possible impact of using robot gestures on drumming game (in terms of performance), but also on the participants' subsequent evaluation of the game. We expected that an intermediate level of gestures would benefit the interaction game.

Results show that the humans were indeed motivated by gestures and did, overall, enjoy the drumming experience. There did, however, seem to be a saturation level for the amount of gestures used to encourage interaction, where the amount of gestures in the gesture+ condition seemed to interfere with the participants' concentration. Drumming with no gestures, while considered efficient in terms of the drumming task, was not rated as successful in terms of social interaction. The reason for the high error rates at the start of the games is likely in part due to the participants' high expectations from the game. According to the questionnaire results, male participants appeared to view the experiment not as a game, but rather as a task to complete. Participants also may have tried to 'test' the robot's limitations during the initial stages of the trials, leading to higher error-rates, as this could involve playing rapidly in long sequences, or using different parts of the drum to create different sounds and enriching their play. They also expected that KASPAR could watch, understand and imitate them (most thought that the robot could detect them with its cameras, positioned in the eyes, and that the gestures were meaningful). As the game progressed, the understanding of the limited capabilities of the robot would increase, leading them to modify their drumming to synchronise more efficiently with the robot. This effect might have been mitigated by participant fatigue, however, as boredom was also mentioned by some participants when answering questions regarding the later games.

The data also suggest that the participants changed their style of play with the increasing level of robot gestures, playing fewer, yet longer sequences of beats.

Our sample, overall, rated the *gesture* condition as the most enjoyable, which, interestingly, had worse error rates in the evaluations of the objective performance than those without gesture. This is likely due to the *gesture* condition incorporating gestures making the interaction enjoyable to those participants who valued this aspect of the interaction, while having a lower error rate than the *gesture+* condition, and so is less adversely impacted by a task-based evaluation than this condition.

This shows that the right amount of gestures would serve to attract the attention of one portion of the participants, and make their experience enjoyable, although it did not actually help their drumming (in objective terms). This draws attention to the marked distinction between the subjective evaluations and objective performance measures.

Overall, the results from Experiment I confirmed our initial expectations (see Section 2.2), but pointed out the different effects of gesture on the dynamics of drumming performance and participants' subjective evaluation. These results helped in designing the next study (Experiment II).

3. Experiment II: emergent turn-taking

3.1. *Methodology*

As motivated earlier, one of the fundamental problems in the human–robot drumming scenario is the timing of the interaction, as timing plays a fundamental role in the regulation of human interaction. It is not always clear when the robot or human should initiate interaction in taking a turn. Therefore, in Experiment I, some predefined fixed time duration heuristics were used for synchronisation, whereby KASPAR started playing if the participant was silent for a few seconds, and would also try to motivate the participant with simple non-verbal gestures.

In Experiment II, we took a different approach and used a novel, probability-based mechanism for timing and turn-taking so that the temporal dynamics of turn-taking *emerge* from the interaction between the human and the humanoid. As explained earlier, the computational models were deliberately chosen to be simple, minimal and (algorithmically) arbitrary. Thus, these models are not meant to faithfully model turn-taking, cognition or learning in humans. Our research agenda is to study whether even such simple and arbitrary computational models will evoke different types of interaction and adaptation of people to the robot's behaviour.

We selected three different simple and minimal computational models to control the starting and stopping of the robot's regular drumming beats. This response is based on the duration time of the previous turn and on the number of beats played in the previous turn by the interaction partner. We denote the models as *Model 1, Model 2* and *Model 3*. *Model 1* uses a step function, *Model 2* a simple triangular function and *Model 3* a hyperbolic function that generates probabilities for starting or

Algorithm 1 The turn-taking algorithm

Human plays (turn # i=1)
Kaspar plays after waiting 2 seconds when human stops
FOR i=2 to *n* **DO**
ThTime$_i$= KasparPlayingTime$_{i-1}$
IF model$_j$ (HumanPlayingTime$_i$,ThTime$_i$) = 1
THEN KASPAR STARTS PLAYING
ThBeat$_i$= # of HumanBeats$_i$
IF model$_j$ (# of KasparBeats$_i$,ThBeat$_i$) = 1
THEN KASPAR STOPS PLAYING
END FOR (end of the game)

Figure 10. The turn-taking algorithm used in Experiment II.

stopping the robot's drumming based on these inputs from previous interaction (Figure 10). The output is bounded by maximum and minimum limits to ensure that KASPAR and the participant have time to play at least once in every turn. For every turn, the robot assesses the probability of start or stop, and takes action accordingly. For starting, the robot uses the time duration of its last bout of playing and for stopping it takes the number of beats of the human participant from the previous turn into account. The minimum number of beats KASPAR will play is one even if the resulting number of the beats recommended by any of the models is below one. The participant starts the game and KASPAR uses its turn-taking strategy when the human participant is silent for 2 s (only for the first turn). After the first turn, the turn-taking strategy is always determined by the robot's probabilistic models. Depending on the previous duration and number of beats in the interaction, according to their respective probability functions (1), (2) and (3), the return value of the three models triggers the starting or stopping in the turn-taking algorithm (Algorithm 1 in Figure 10). The probability functions for the three computational models are presented in Equations (1), (2) and (3), and visualised in Figure 11.

$$p(x) = \begin{cases} 0, & x < \mathrm{Th} \\ 1, & x \geq \mathrm{Th} \end{cases} \quad \text{(Step: Model 1)}, \tag{1}$$

$$p(x) = \frac{x}{\mathrm{Th}} \quad \text{(Linear: Model 2)}, \tag{2}$$

$$p(x) = 1 - \frac{1}{x} \quad \text{(Hyperbolic: Model 3)}. \tag{3}$$

Here, x is measured in units of time for the case of starting, or, respectively, as the number of beats for stopping. Similarly, Th represents the threshold parameter of time for starting and the

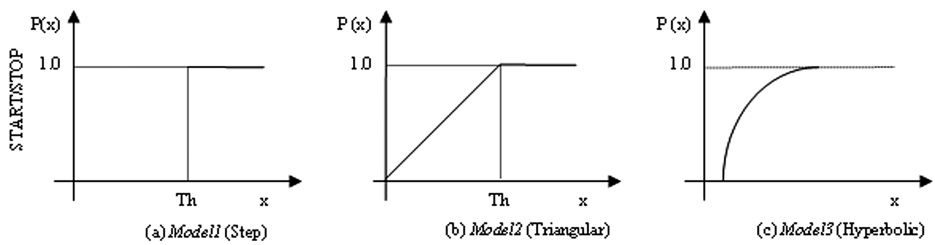


Figure 11. Computational models for START/STOP actions. For START actions, Th=ThTime, since the x-axis variable is the time (t). For STOP actions, Th=ThBeat. The x-axis variable is the number of beats (b). For START, Th is the duration of KASPAR's previous drumming bout, and for the STOP action, Th is the number of beats in the human's previous drumming bout; except that the minimum value for Th is 1.5 s (experimentally determined) for START and 1 beat for STOP actions. The only model which does not have the threshold limitations is Model 3 due to its hyperbolic nature. The y-axis gives the probability of START/STOP as a function of time/number of beats based on previous interaction.

number of beats for stopping, respectively. For each model, a decision function is called returning a 0 ('no') or 1 ('yes') is called to decide whether to change the robot's current behaviour. That is, the function *model i* (x,Th) is called to start or stop KASPAR playing using the respective $p(x)$ function for that model as in Algorithm 1. In *Model 1,* if $p(x)$ is 1 then the model triggers starting or stopping, and this depends only on Th and the current value of x. *Models 2* and *3* have probability functions that can take values other than just 0 and 1, so a random value r in [0,1] is generated and if r is not less than the function output, then the model returns 1 (otherwise 0). Thus, in effect, in all three of these simple models, a starting or stopping action, given the current values of parameters x and Th, occurs at appropriate points with probability $p(x)$ according to the respective model, so that the model then triggers the start or stop of drumming, or otherwise no change in the behaviour occurs – see the conditionals (IF-statements) of the robot control in pseudocode of Figure 10.[8] In future, other models could also easily be assessed.

Consequently, at every turn, the robot decides when to start and stop according to the performances of both the human player and itself. Thus, the game and its dynamics are not deterministic but emerge from the moment-to-moment status of both KASPAR and the participant.

Complementary to Experiment I, we decided not to introduce any robot gestures in Experiment II but to focus our analysis on the turn-taking behaviour. Therefore in Experiment II, KASPAR did not use any gestures.

3.2. *Research questions and expectations*

In order to investigate the effect of three different generative computational models on emergent turn-taking dynamics in an imitation game, our primary research questions were as follows:

- How do different robot turn-taking strategies based on different minimal computational probabilistic models impact on the drumming performance of the human–robot pair?
- How do the different robot turn-taking strategies impact on the participants' subjective evaluation of the drumming experience?

We expected to have 'successful' games in terms of turn-taking emerging from the interaction between the human and the humanoid, and that the different computational models would show different degrees of success in terms of engaging and sustaining interaction. Our 'success' criteria were as follows: (1) the number of turns with no or slight overlaps and gaps and (2) the number of human beats detected by the robot and the number of beats played by the robot itself that will give us hints about the quality of the games.

3.3. *Experimental conditions*

We studied three models with different parameters (Figure 11) in three different experimental conditions. We set up simulated experiments before the live experiments, to define the maximum and minimum limits and thresholds for the actual experiments with humanoid and human participants. Each model is used both for starting and stopping the robot's play and represents an experimental condition. For *start* the time duration of the previous turn is used, and for *stop* the number of beats of the previous turn is used as a threshold. As described in the previous section, *Model 1* was a step function, where the return value of the function is '1' if the input value of the function is not smaller than the threshold; thus, we expect this model to give more play time and a higher number of beats than the other models. Ideally, if the human beats long sequences, this model would reach very high values so we put a maximum time limitation (both interactants cannot play longer than 10 s per turn). Unlike *Model 1*, *Model 2* has a triangular shape which has the threshold as an upper boundary. Since we have a probabilistic approach we can have values smaller than the threshold. In fact, we expect this model to give the least play time and lowest resulting number of beats for the participants; so we foresee that the model would not be as popular as the other two models. The last condition is *Model 3*, a hyperbolic model, which cannot be limited by the thresholds. It reaches high values (close to one) very fast compared with *Model 2*. Therefore, we predict that it would result in more play time (i.e. enable the robot to play more beats than *Model 2*). Also, in our simulations, we noticed that it could enable 'coordinated games' (i.e. with a very low number of overlaps and conflicts between the human's and the robot's drumming) if we played short sequences, but since the model is not limited by thresholds, it 'reacts' to the human but does not exactly 'imitate' the human's drumming games, which we suspected that the participants might not find acceptable.

3.4. *Experiment, results and analysis*

3.4.1. *Robot*

The experiments were carried out with the humanoid robot KASPAR that was also used in Experiment I (see Section 2.4.1).

3.4.2. *Experimental setup*

The experimental setup was the same as in Experiment I (see Section 2.4.2).

3.4.3. *Software features*

The same software features were used as in Experiment I (see Section 2.4.3).

3.4.4. *Participants*

Twenty-four participants (8 female and 16 male) took part in the study. Due to logistical reasons, the trials were carried out in 2 sets (a few months apart) with 12 participants each. All participants worked in computer science or similar disciplines at the university. Only two of them had interacted with KASPAR prior to the experiment, and most were not familiar with robots in general. Six of our participants had children. (Regarding gender balance of the sample, see comment in Section 2.4.4).

3.4.5. *Interaction game setup*

We used a 1 min demonstration of the robot without any drumming game involved, where the participants were shown how to interact with KASPAR. This was followed by three games reflecting the three experimental conditions described above each lasting 3 min, without indicating to the participants anything about the differences between the conditions. The participants were simply instructed that they could play drumming games with KASPAR. As we did in Experiment I, we used all six possible different presentation orders of the games to analyse the effect of the order of the games on the humans. To account for possible fatigue, habituation or learning by the participants, in the *sequential order* section below, we analysed the games according to their order number in the sequence experienced by the participants (independent of the particular experimental condition): thus calling them the first game, second or third, disregarding their game types, for example, for one participant the first game (order 1) would be the *Model 1* game, and for another participant, *Model 1* would be the third game (order 3). After finishing the three games, the participants completed a questionnaire.

3.4.6. *Results*

3.4.6.1. Evaluation of questionnaire data. The participant evaluations were elicited in a questionnaire in the same manner as in Experiment I (see Section 2.5.1).

Most and least preferred games according to game type: See Table 5 for the number of participants which rated each game as most preferred and least preferred. There was a significant deviation from the expected counts for the most preferred game type ($\chi^2(2) = 7.76$, $p = 0.021$) as well as for the least preferred game type ($\chi^2(2) = 10.89$, $p = 0.004$). Table 5 shows that both the *Model 1* and *Model 3* games were preferred by a comparable amount of participants, while fewer participants preferred *Model 2* most.

Table 5 also shows that the highest number of the participants considered the *Model 2* game as the least preferred, while the M*odel 1* and *Model 2* games had a small number of participants which considered them the least preferred. The *Model 3* game was slightly more popular than the *Model 1* game.

Most and least preferred games according to sequential order: The number of participants which rated each game as most preferred and least preferred according to the sequential order can be seen in Table 6. The deviations from the expected count were approaching significance for the most preferred game ($\chi^2(2) = 5.25$, $p = 0.07$). Table 6 suggests that the most popular game type was the third game, while first and second games were less preferred. Table 6 also suggests that all ordinal positions of occurrence in the sequence of the games had a similar number of participants which considered them the least preferred.

As for Experiment I, in order to control for the threat against the assumptions of the counterbalancing method, mixed model ANOVAs were run using game type to investigate the possible interaction effects of presentation order and game type on both questionnaire responses and behavioural data. These were non-significant, supporting the notion of independence between presentation order and game-type overall in the sample.

Other preferences: The order of the games did not have a significant impact on the participants in terms of evaluation of the game. There were, however, significant differences according to the model used in terms of how participants evaluated the games. The participants did not rate KASPAR's drumming significantly differently across the models ($F(2,46) = 1.64$, $p = 0.20$). There was an effect approaching significance for how they rated KASPAR in terms of sociality

Table 5. Most and least preferred games according to type.

	Participants			
Game Type	Most preferred	Residual	Least preferred	Residual
Model1	9	1.7	6	−4.3
Model2	2	−6.3	17	−0.7
Model3	13	4.7	4	3.7

Table 6. Most and least preferred games according to sequential order.

	Participants			
Order	Most preferred	Residual	Least preferred	Residual
1	4	−4.0	9	0.035
2	7	−1.0	8	−0.7
3	13	5.0	9	0.035

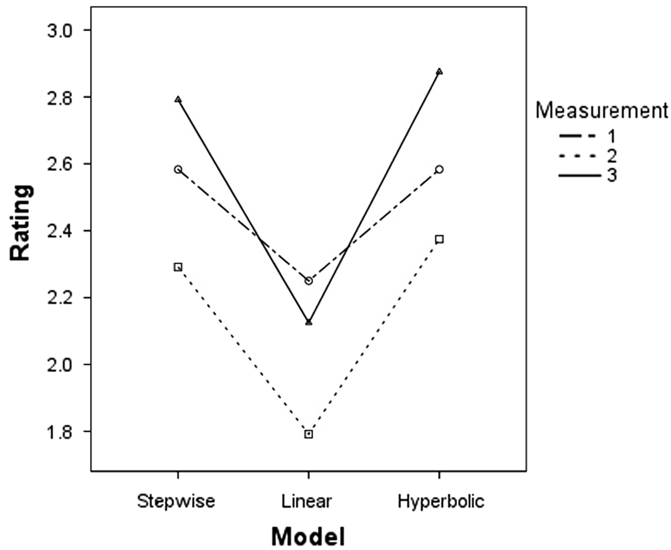


Figure 12. Ratings for the three measurements: (1) KASPAR's drumming, (2) KASPAR's perceived sociality and (3) participant enjoyment.[9]

($F(2,46) = 3.12$, $p = 0.054$), and participants significantly differentiated between the models in terms of enjoyment ($F(2,46) = 7.59$, $p = 0.001$). These effects are shown in Figure 12, which suggests that for all three, there was a tendency for the participants to rate the interactive aspects of the games lower when the linear model was used.

Sample similarities: The possibility of systematic differences between the first sample of 12 and the subsequent sample of 12 was assessed using mixed-model ANOVA. This ANOVA found no significant systematic differences between the two groups ($F(1,22) = 0.070$, $p = 0.79$). Since an identical experimental protocol was used for both groups of participants, this result supports the analysis of both samples as one larger sample. **Q5**

3.4.6.2. Evaluation of behavioural data. The behavioural data regarding the performance of the human partner during the games consisted of KASPAR's own detection of the human's drumming (denoted as 'KASPAR's view'), and video recordings of the human's drumming that were annotated and analysed manually (referred to as 'human's view'). The behavioural data includes the number of *zero turns* (where KASPAR could not register any beat performed by the human partner but played at least one beat, and passed the turn to the human), *non-zero* turns (KASPAR would register at least one drum beat of the human participant), the number of drum beats performed by human participant and KASPAR, and turn durations (referred to as 'time' in the text).

Behavioural data according to sequential order: There was no significant difference between the games according to the order (e.g. for number of turns, $F(2,22) = 0.007$, $p = 0.99$, with ANOVA). Only the human's total number of beats per game increased with the order of the games as they got used to the scenario while they played more (Table 7, KASPAR's perspective, and Table 8, human's perspective).

Behavioural data according to game type: The game types are compared in detail in Tables 9 (human's drumming) and 10 (KASPAR's drumming).

Table 7. Observed behaviour of KASPAR according to order.

Order	Average no. of beats per turn	Maximum/minimum no. of beats	Total no. of beats	Average time per turn	Maximum/minimum time per turn	Total time
1	1.7 ± 0.8	6/1	135 ± 33	1.08 ± 0.1	3/1	97.6 ± 41
2	1.72 ± 0.8	6/1	136 ± 30	1.07 ± 0.1	3/1	96 ± 40
3	1.76 ± 0.7	7/1	138 ± 26	1.07 ± 0.1	4/1	95.8 ± 41

The repeated measures ANOVA found significant differences between *Model 2* (linear model) and the other models, across a range of variables. In terms of the total number of beats there was a marked difference in the number of beats by the human registered by KASPAR ($F(2,46) = 58.95$, $p < 0.001$), as well as the total beats by KASPAR ($F(2,46) = 470.63$, $p < 0.001$), between the models used. There was no difference, however, between the models in terms of the actual number of beats *played* by the human participants ($F(2,46) = 0.037$, $p = 0.96$). Referring to Figure 13, we can see the relationship between detected human beats, beats *produced* by KASPAR and *the* actual *beats played by the participants across* the models.

The graph suggests that while the actual number of beats played by the humans remains more or less constant across the models, the registered number of beats decreases dramatically between the stepwise model and the other two models, while the number of beats by KASPAR increases. Thus, in the cases of linear and hyperbolic models KASPAR appeared less responsive to the playing of the participants. This result may account for the participants' higher evaluation scores for the stepwise model, compared with the linear model.

Significant differences were found between the models in terms of the ratio of turns in which KASPAR registered the beats from the human participant to the total number of turns ($F(2,46) = 77.18$, $p < 0.001$), see Figure 14.

Figure 15 suggests that KASPAR registered more human activity in terms of turns with both the stepwise and the hyperbolic models than with the linear model. According to Table 9, this is also clear in terms of the actual number of non-zero turns, despite the much higher number of total turns with the linear model. The difference in the actual number of turns was highly significant as well ($F(2,46) = 28.78$, $p < 0.001$). The above results suggest that in terms of turn-taking, KASPAR was more 'aware' (in terms of detection of beats) of the participants' behaviour in the stepwise and hyperbolic conditions than in the linear condition. The time spent drumming by the participant as registered by KASPAR may also serve to differentiate between the linear models and the two other models. There were significant differences between the three models ($F(2,46) = 1897.71$, $p < 0.001$), see Figure 15.

Figure 15 suggests that the amount of time in which KASPAR registered the human participant as drumming differs dramatically across the three models. The stepwise model is the most effective in this sense, followed by the hyperbolic model with the linear model being the least efficient.

These measures do suggest that some of the participants' preferences for the stepwise and hyperbolic model can be explained by objective measures of KASPAR's responsiveness to the actual drumming of the human participants. They do not, however, explain why the participants equated the stepwise and hyperbolic models in terms of enjoyment.

3.5. *Discussion of results*

Overall, the results confirm our initial expectations, namely that different computational models will lead to different human–humanoid drumming interactions (as evaluated subjectively and objectively).

Table 8. Observed drumming behaviour of human according to order.

Order	No. of turns	No. of non-zero turns	Maximum no. of beats	Total no. of beats (KASPAR's view)	Total no. of beats (human's view)	Average time per turn	Maximum/minimum time per turn	Total time
1	93 ± 44	28 ± 13	5	43 ± 23	110.2 ± 35	0.99 ± 0.6	3.11/0.01	70 ± 27.3
2	92 ± 42	29 ± 16	4	45 ± 29	113.7 ± 39	0.99 ± 0.6	2.06/0.01	69.7 ± 27.2
3	92 ± 44	31 ± 17	5	52 ± 34	115.7 ± 38	0.99 ± 0.6	3.11/0.01	68 ± 25.4

Table 9. Observed behaviour of human's drumming according to game type.

Game type	No. of turns	No. of non-zero turns	Maximum no. of beats per turn	Sum of beats (KASPAR's view)	Total no. of beats (human's view)	Average time per turn	Maximum/minimum time per turn	Total time
Model 1	66.2 ± 4	36.4 ± 15.7	5	66.7 ± 30.3	113.8 ± 33.1	1.52 ± 0.02	3.1/1.5	101 ± 5.3
Model 2	152.1 ± 3.1	21.1 ± 12.3	3	25.5 ± 14.8	114.2 ± 44	0.25 ± 0.01	0.61/0.01	37.3 ± 1.3
Model 3	58.7 ± 1.4	$30.5 + 13.8$	5	48.29 ± 23.9	111.5 ± 33.3	1.2 ± 0.01	1.8/1	70 ± 1.7

Table 10. Observed behaviour of KASPAR's drumming according to game type.

Game type	Average no. of beats per turn	Maximum/minimum no. of beats	Total no. of beats	Average time per turn	Maximum/minimum time per turn	Total time
Model 1	1.47 ± 0.28	5/1	96.5 ± 12.4	1.02 ± 0.03	3/1	67.5 ± 3.2
Model 2	1.01 ± 0.01	3/1	154 ± 3.42	1 ± 0.003	3/1	152 ± 2.8
Model 3	2.69 ± 0.1	7/2	157.9 ± 3.6	1.19 ± 0.04	4/1	69.5 ± 1.8

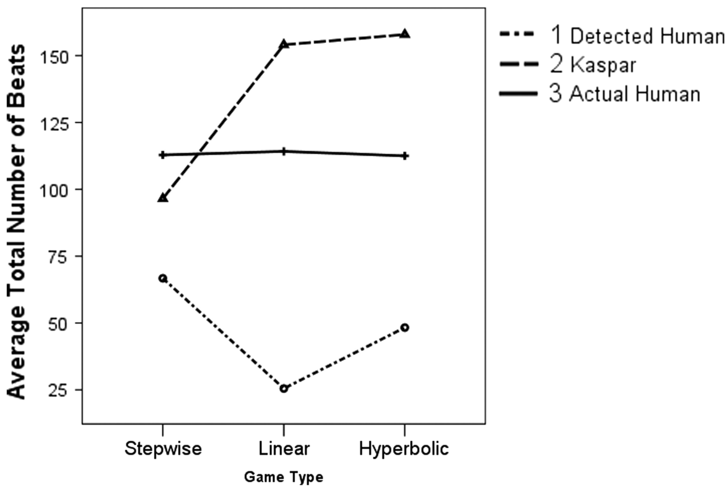


Figure 13. Total number of beats for (1) detected human beats (2) KASPAR's beats and (3) actual human beats.

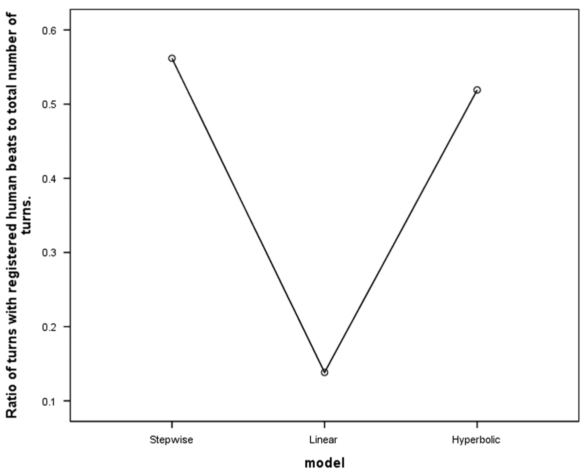


Figure 14. Ratio of turns with registered human beats to total number of turns according to the model.

As stated in the previous section, *Model 2* gave the least play time to the human and KASPAR. The impression that the participants may have got is one where KASPAR did not seem to imitate the human participants' game at all, but rather 'played on its own' (KASPAR would play at least one beat even when it did not detect a response from the human participant; Figure 16). As a consequence, KASPAR acted as a leader in the game most of the time. There were also many overlaps between KASPAR's play turns and the human participants' play turns in *Model 2*. This could be because either KASPAR or the human participants interrupted each other. More importantly, this would also cause the loss of detection of humans' beats (as described above, KASPAR would not 'listen' when it was playing). Replies to the open-ended questions in the post-game questionnaires related to this game described KASPAR's behaviour using the terms like 'annoying' or 'rude'. Thus, both the behavioural data as well as the questionnaire results describe an interaction in which the interaction's rules for turn-taking was not apparent to the

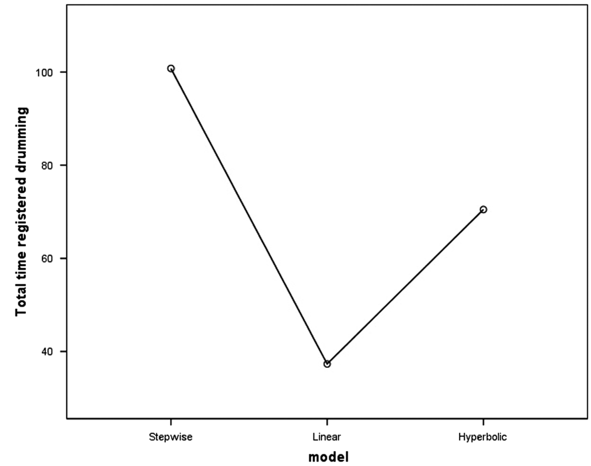


Figure 15. Total time registered for drumming according to the models.

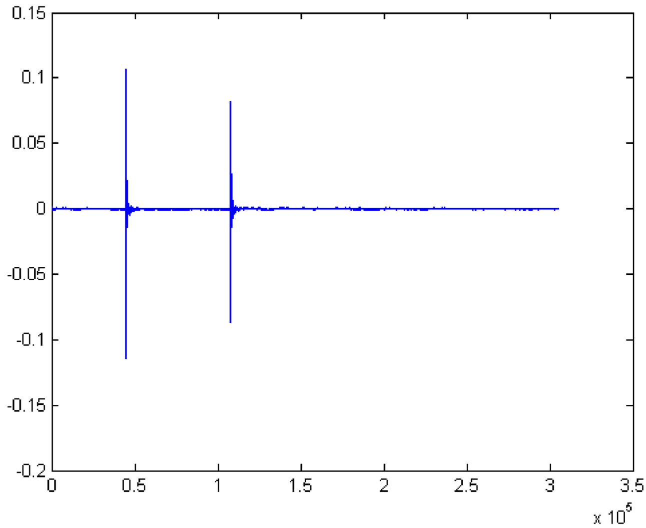
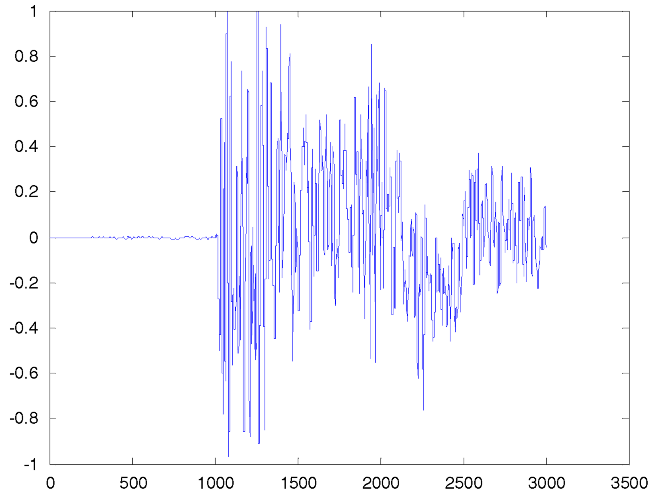


Figure 16. Representation of two beats in an example sound wave. Q6

human participant leading to repeated breakdowns in the social interaction, which in a human–human interaction would be described as impolite and a source of stress. Together, these measures provide an explanation as to why the participants disliked the *Model 2* game.

As stated in the previous sections, since *Model 1* uses the previous play time as a threshold, it ensures that the current play time is at least as long as the previous play time for the human player. This longer play time (compared with other games) led to both players playing longer turns which may have created the impression that the tempo of the game was slower than in the other games. This could explain the preferences for *Model 3* since the tempo of this game would be experienced as faster, having more exchanges and being perceived as more interactive. While the observed play time for the human participants was shorter than for *Model 1*, it was still long enough to allow for a coordinated game. This, coupled with the emergent nature of KASPAR's drumming in *Model 3*, led it to being viewed as more 'natural' by participants. In this game, both the human and the KASPAR played 3–4 beats in every turn (the model's probability distribution favours



Q6 Figure 17. A single drumming bout is presented in detail. It is not a single peak but consists of many local minima/maxima.

high values), with fewer gaps in-between the participants' drumming compared with *Model 1*, and far fewer overlaps compared with *Model 2* between two turns. But *Model 3* was not bound by thresholds by nature, so it seemingly exhibited a degree of independence in regards to the human participants' performance, which some of the participants reported as being annoying. Some participants, however, did express a liking for this, though, for example, one participant described this phenomenon like 'teaching her son to play a drum'. Similarly, another participant asked if she should consider KASPAR as a professional drummer or a child while she commented on the games, since it 'looks like a child drumming rather than a professional' (Figure 17). Statements like this support the notion, suggested by the quantitative data, that the emergent turn-taking of *Model 3* was perceived to have more in common with a human–human interaction than that of the other models.

In *Model 1* the human participant was given more play time than KASPAR, but KASPAR played more beats than the human participants. However in *Model 3*, KASPAR and the participant were given almost equal durations and opportunities to play. So in the case of *Model 3*, KASPAR could act equally as a follower as well as a leader and thus had more impact on the play and played longer rhythms.

One should also note that there is a considerable amount of zero turns in all the three models. However, only in the case of *Model 2* was this amount high enough to affect the overall game. When these turns were distributed among normal turns as in *Model 1* and *Model 3*, they did not dominate the behaviour but were compensated for by the non-zero turns. But for *Model 2*, zero turns seemed to dominate the whole game and were described by the participants as a source of dislike for the model/game type.

Although there were gaps between the humans' and the robot's turns in *Model 1*, while in *Model 3* KASPAR did not seem to imitate the human participants in every turn, both models were successful in terms of emergent turn-taking. As a consequence, according to the participants' questionnaire feedback, they preferred *Model 1* and *Model 3* to *Model 2*.

As seen in the previous study, the participants actively explored the limits of KASPAR's drumming as well as the rules of the game, and adapted themselves to the games over time, which resulted in better games in terms of turn-taking and synchronisation in the later games. Thus, we observed longer sequences of playing without any overlaps or gaps between the turns. This suggests that the human participants were not passive participants in this game, but actively adapted themselves to the capabilities of the robot on their own initiative. This finding is consistent with the notion of recipient design, a concept from ethnomethodology, where we find that natural speech is always designed for its recipient (i.e. the interaction partner) and interpreted as having been so designed. Here, the speaker creates his or her turn 'with recipients in mind, and listeners are motivated to "hear" a turn that is for them and all participants closely and constantly track the trajectory of the talk to hear "their" turn' (Boden 1994, p. 71). According to conversation analysis, this turn-taking is integral to the formation of any interpersonal exchange (Boden 1994, p. 66). While in our study the robot's behaviour was controlled and based on simple computational models, we found that the participants used their recipient design skills in the interaction. The issue of recipient design will be explored further in our future research.

4. Conclusions

This article presented basic research into the regulation of interaction dynamics during social/playful HRI. We introduced an experimental setup based on human–humanoid drumming games as a suitable scenario for HRI research on non-verbal cues, synchronisation, timing and turn-taking using drumming games. Generally, the results showed that believable and enjoyable human–humanoid interaction dynamics can be created with minimal models underlying the robot's turn-taking behaviour.

Specifically, the results from this experiment suggest that there was active adaptation on the part of the participants, throughout the games. However, the efficiency of such adaptation may be countered by the participant fatigue/boredom reported in the later games, which highlights the essential role of research into how to maintain a user's interest in the interaction with a robot. One should note, however, that the results also indicate a trade-off between the subjective evaluation of the drumming experience from the perspective of the participants and the objective evaluation of the drumming performance, as well as individual differences in how the participants approached the game. The participants as a whole preferred a certain amount of robot gestures as a motivating factor in the drumming games that provided an experience of social interaction. However, the sample was divided in terms of what degrees of gestures were appropriate. The results highlight the need to ascertain to what degree the strategies used by a robot to encourage and maintain interest in such interactions, interfere with the task the interaction is centred around, as well as consider the role of individual differences in the appropriateness of these strategies. Experiment II showed that the different minimal, probabilistic models that controlled the robot's interaction dynamics led to different subjective evaluations by the participants and different dynamics in the performances of the games. The results from the questionnaires and behavioural data analysis suggest that the participants preferred the models which enable the robot and human to interact more and provide turn-taking closer to 'natural' human–human conversations, despite the differences in objective measures of drumming behaviour.

Overall, the results from our studies are consistent with the *temporal behaviour matching hypothesis* (Robins et al. 2008) which concerns the effect that the participants adapt their own interaction dynamics to the robot's. Note that our child-sized robot KASPAR, despite some human-like features such as a face, arms and few facial expressions, is still mechanical in nature (e.g. the movements are not following the biological models of movement generation, the facial expressions are minimal and not based on the models of human facial expressions, and in terms of its appearance the robot has a slightly cartoon-like appearance where we deliberately did not cover up metals and wires, e.g. protruding from the neck and wrists). But participants still adapted to the dynamics of this robot which highlights the importance of considering interaction kinesics in HRI design in general, not only in research attempting to exactly copy human-like appearance and behaviour.[10] A systematic study of the impact of robot appearance on participants' behaviour in human–humanoid drumming experiments is an interesting area of research but goes beyond the scope of the current article.

There are several noteworthy limitations of this work including methodological as well as technological limitations. Ideally, in order to generalise the results towards a wider user group the study could be repeated with participants of different age ranges, personality traits, cultural background, gender, etc. Such studies would help to explain group differences (e.g. concerning why the subjective evaluation of the participants in our study differed). Different subjective rating scales could be used. Qualitative analysis of the human–robot behavioural data (e.g. by using conversation analysis)[11] could flesh out further details of the interaction. The timing algorithms used in Experiment II could be refined in future work alongside a systematic variation of different types of robot gestures in order to find out which of these gestures have the most impact on the interaction. It may also be interesting to replicate the experiment with a different robot that had a broader spectrum of possible drumming behaviours, as this may not only enrich the interaction but also provide additional data for the performance evaluation. Last but not the least, an electronic drum could be used in order to ease the detection of the beats.

5. Future work

The HRI experiments presented in this article were based on a drumming scenario and we found that this is a very suitable task for the study of HRI and adaptive behaviour. However, our long-term research aims to go beyond a simple drumming synchronisation task and to develop richer social interaction between the robot and the human partner, which would not simply focus on synchronisation to produce the same tempo, but could provide a successful (in terms of the task) as well as enjoyable social experience to people, while allowing us to gain insight into the role of non-verbal interaction kinesics in sustaining and regulating HRI.

Based on these results, future work will investigate further issues related to interaction kinesics in general, and recipient design in particular. As mentioned above, several factors regarding robot non-verbal gestures as well as computational models underlying the robot's turn-taking behaviour seem to influence the objective performance and subjective evaluation of the interaction experience. Future work needs to investigate these further, including other factors such as the consideration of individual participants' preferences, personality profiles, as well as long-term effects.

In light of our promising results from using gestures, we foresee a system wherein KASPAR's behaviour may be motivated and rewarded by the human partner, through the partner's gestures and other expressive actions, and respond to these by playing novel acoustic rhythms and using its own repertoire of expressions and gestures to provide feedback to the human interaction partner, and, importantly, to become a 'partner' in the interaction that is not only responding but also taking the initiative proactively. If our results can be extrapolated, then such a system will be capable of motivating and sustaining interaction.

One interesting direction for future work concerns eye gaze, which plays an important role in regulating human–human interaction and communication (e.g. Kendon 1967; Farroni, Johnson, and Csibra 2004), and possibly also HRI kinesics (Mutlu, Shiwa, Kanda, Ishiguro, and Hagita 2009). While the study of gaze cues goes beyond the scope of the article, in our future work we aim to study the role of eye gaze (mutual gaze, eye gaze direction, etc.) in HRI games.

Research on interaction kinesics, as exemplified in this work, can potentially contribute to a wide range of application areas of social robots, in particular those that require long-term and repeated interaction (e.g. robots as assistive companions in the home, or robots as therapeutic or educational playmates for children). In such situations, the social acceptance of the robot, including the users' enjoyment of the interaction as well as the performance of the system in collaborative tasks, is crucial to the success of a particular application.

Acknowledgements

This work was conducted within the EU Integrated Project RobotCub ('Robotic Open-architecture Technology for Cognition, Understanding, and Behaviours'), funded by the EC through the E5 Unit (Cognition) of FP6-IST under Contract FP6-004370. We thank our colleagues in the Adaptive Systems Research Group for their help in preparing the experiments and the paper: Michael L. Walters, Joe Saunders, N. Assif Mirza, Qiming Shen, Aris Alissandrakis, Nuno Otero and Ester Ferrari.

Notes

1. In this article, we use the terms 'interactant' and 'interaction partner' (or 'partner' in short) synonymously. Thus, the term 'partner' does not imply a long-term relationship or affective bonding between human and robot.
2. KASPAR stands for kinesics and synchronisation in personal assistant robotics. The robot has been developed by our research group.
3. Note: KASPAR has previously been used successfully in studies involving children (Robins et al. 2008), including children with special needs (e.g. Robins, Dautenhahn, and Dickerson 2009) as well as adults (Kose-Bagci, Dautenhahn, Syrdal, and Nehaniv 2007; Kose-Bagci, Dautenhahn, and Nehaniv 2008). The work presented in this article is focussed on adult participants.
4. In this article, the terms 'deterministic' and 'probabilistic' turn-taking refer to the robot's control algorithm, that is, whether the robot behaves according to a deterministic or probabilistic algorithm that determines how it responds in a given moment given its sensory input. This point deserves clarification since any interaction involving humans has non-deterministic interaction dynamics as far as the *overall* human–humanoid interaction dynamics is concerned, since one cannot predict exactly how humans will behave in the interaction.
5. Preliminary results from the first 12 participants were summarised in Kose-Bagci et al. (2007).
6. Preliminary results with only an initial analysis based on 12 of the 24 participants are presented in Kose-Bagci et al. (2008).
7. KASPAR needed at least 0.3 s between beats to get its joints ready, so that, even if the human played faster, KASPAR's imitations still required minimum pauses of at least 0.3 s between the beats.
8. Note: We had also tried to *start* using beats and *stop* using time with simulated data, but the current combination resulted in more drumming time and a higher number of beats for both human and KASPAR, so this combination was preferred in the current implementation.
9. See footnote to Figure 6 above
10. For example, see android research (MacDorman and Ishiguro 2006) or other studies into the importance of robot appearance in HRI experiments (e.g. Walters et al. 2008).
11. See Robins, Dickerson, Stribling, and Dautenhahn (2004b) for an example of using conversation analysis in HRI research.
12. Similar to Kotosaka and Schaal (2000). Synchronised robot drumming by neural oscillator. *International Symposium on Adaptive Motion of Animals and Machines.*

References

Bernieri, F.J., and Rosenthal, R. (1991), 'Interpersonal Coordination: Behaviour Matching and Interactional Synchrony', in *Fundamentals of Nonverbal Behavior*, eds. R.S. Feldman and B. Rimé, Cambridge: Cambridge University Press, pp. 401–432.

Bickmore, T.W., and Cassell, J. (2005), 'Social Dialogue with Embodied Conversational Agents', in *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems*, eds. J. van Kuppevelt, L. Dybkjaer and N. Bernsen, New York: Kluwer Academic, pp. 23–54.

Bickmore, T.W., and Picard, R.W. (2005), 'Establishing and Maintaining Long-Term Human Computer Relationships', *ACM Transactions on Computer-Human Interaction*, 12(2), 293–327.

Blow, M.P., Dautenhahn, K., Appleby, A., Nehaniv, C., and Lee, D. (2006), 'Perception of Robot Smiles and Dimensions for Human-Robot Interaction Design', in *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, pp. 469–474.

Boden, D. (1994), *The Business of Talk: Organizations in Action*, Polity Press. **Q7**

Breazeal, C. (2002), *Designing Sociable Robots*, MIT Press.

Breazeal, C. (2003), 'Towards Sociable Robots', *Robotics and Autonomous Systems*, 42(3–4), 167–175, T. Fong (guest ed.).

Breazeal, C. (2004), 'Social Interaction in HRI: The Robot View', *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(2), 181–186.

Condon, W.S., and Ogston, W.D. (1967), ‘A Segmentation of Behavior’, *Journal of Psychiatric Research*, 5, 221–235.

Crick, C., Munz, M., and Scassellati, B. (2006), ‘Synchronization in Social Tasks: Robotic Drumming’, *Proceedings of IEEE RO-MAN 2006*, pp. 97–102.

Dautenhahn, K. (2007a), ‘Methodology and Themes of Human-Robot Interaction: A Growing Research Field’, *International Journal of Advanced Robotic Systems*, 4(1), 103–108.

Dautenhahn, K. (2007b), ‘Socially Intelligent Robots: Dimensions of Human–Robot Interaction’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 679–704.

Dautenhahn, K., and Billard, A. (2002), ‘Games Children With Aautism Can Play With Robota, A Humanoid Robotic Doll’, in *Universal Access and Assistive Technology*, eds. S. Keates, P.M. Langdon, P.J. Clarkson and P. Robinson, London: Springer-Verlag, pp. 179–190.

Degallier, S., Santos, C.P., Righetti, L., and Ijspeert, A. (2006), ‘Movement Generation Using Dynamical Systems: A Humanoid Robot Performing a Drumming Task’, in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS06)*.

Farroni, T., Johnson, M.H., and Csibra, G. (2004), ‘Mechanisms of Eye Gaze Perception During Infancy’, *Journal of Cognitive Neuroscience*, 16(8), 1320–1326.

Goldin-Meadow, S., and Wagner, M.S. (2005), ‘How our hands help us learn’, *Trends in Cognitive Science*, 9(5).

Goodrich, M.A., and Schultz, A.C. (2007), ‘Human-Robot Interaction: A Survey’, *Foundations and Trends in Human-Computer Interaction*, 1(3), 203–275.

Hall, E.T. (1983), *The Dance of Life: The Other Dimension of Time*, Doubleday: Anchor Press.

Hendriks-Jansen, H. (1996), *Catching Ourselves in the Act Situated Activity, Interactive Emergence, Evolution, and Human Thought*, Cambridge, MA: MIT Press.

Iizuka, H., and Ikegami, T. (2004), ‘Adaptability and Diversity in Simulated Turn-Taking Behavior’, *Artificial Life*, 10(4), 361–378.

Ito, M., and Tani, J. (2004), ‘Joint Attention Between a Humanoid Robot and Users in Imitation Game’, in *Proceedings of the 3rd International Conference on Development and Learning (ICDL’04)*, La Jolla, USA.

Kendon, A. (1967), ‘Some Functions of Gaze Direction in Social Interaction’, *Acta Psychologica*, 26, 22–63.

Kendon, A. (1970), ‘Movement Coordination in Social Interaction: Some Examples Described’, *Acta Psychologica*, 32, 100–125.

Kose, H., and Akin, H.L. (2001), ‘Object Recognition in Robot Football Using a One-Dimensional Image’, in *The Tenth Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN 2001)*, pp. 291–300.

Kose-Bagci, H., Dautenhahn, K., Syrdal, D.S., and Nehaniv, C.L. (2007), ‘Drum-Mate: A Human-Humanoid Drumming Experience’, in *IEEE-RAS Humanoids2007*.

Kose-Bagci, H., Dautenhahn, K., and Nehaniv, C.L. (2008), ‘Emergent Turn-Taking Dynamics in Drumming Games with a Humanoid Robot’, in *IEEE RO-MAN 2008*.

Kotosaka, S., and Schaal, S. (2001), ‘Synchronized Robot Drumming by Neural Oscillator’, *Journal of the Robotics Society of Japan*, 19(1), 116–123.

Kozima, H., Nakagawa, C., Yasuda, Y., and Kosugi, D. (2004), ‘A Toy-Like Robot in the Playroom for Children with Developmental Disorder’, in *Proceedings of the International IEEE Conference on Development and Learning (ICDL-04)*, San Diego, USA.

MacDorman, K.F., and Ishiguro, H. (2006), ‘The Uncanny Advantage of Using Androids in Social and Cognitive Science Research’, *Interaction Studies*, 7(3), 297–337.

Metta, G., Fitzpatrick, P., and Natale, L. (2006), ‘Yarp: Yet Another Robot Platform’ (special issue on Software Development and Integration in Robotics), *International Journal on Advanced Robotics Systems*, 3(1), 43–48.

Michalowski, M.P., Sabanovic, S., and Michel, P. (2006), ‘Roillo: Creating a Social Robot for Playrooms’, in *Proceedings of IEEE RO-MAN 2006*, pp. 587–592.

Miller, S.H. (1984), *Experimental Design and Statistics*, London: Routledge.

Miyake, Y. (2003), ‘Co-creation in Man-Machine Interaction’, in *Proceedings of the 12th IEEE International Workshop on Robot & Human Interactive Communication (RO-MAN03)*, IEEE Press, pp. 321–324.

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., and Hagita, N. (2009), ‘Footing in Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues’, in *Proceedings of the 4th International Conference on Human-Robot Interaction (HRI09)*, March 2009, San Diego, CA.

Nadel, J., Guerini, C., Peze, A., and Rivet, C. (1999), ‘The Evolving Nature of Imitation as a Format of Communication’, in *Imitation in Infancy*, eds. J. Nadel and G. Butterworth, Cambridge University Press, pp. 209–234.

Nass, C., and Lee, K.M. (2000), ‘Does Computer-Generated Speech Manifest Personality?’ in *Proceedings CHI2000*, pp. 329–336.

Pfeifer, R., and Scheier, C. (1999), *Understanding Intelligence*, MIT Press.

Pickering, M.J., and Garrod, S. (2004), ‘Toward a Mechanistic Psychology of Dialogue’, *Behavioral and Brain Sciences*, 27, 169–225.

Reeves, B., and Nass, C. (1997), *The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places*, New York: Cambridge University Press.

Robins, B., Dautenhahn, K., te Boekhorst, R., and Billard, A. (2004a), ‘Effects of Repeated Exposure of a Humanoid Robot on Children with Autism’, in *Designing a More Inclusive World*, eds. S. Keates, J. Clarkson, P. Langdon and P. Robinson, London: Springer Verlag, pp. 225–236.

Robins, B., Dickerson, P., Stribling, P., and Dautenhahn, K. (2004b), ‘Robot-Mediated Joint Attention in Children With Autism: A Case Study in Robot-Human Interaction’, *Interaction Studies*, 5(2), 161–198.

Robins, B., Dautenhahn, K., Nehaniv, C.L., Mirza, N.A., Francois, D., and Olsson, L. (2005), 'Sustaining Interaction Dynamics and Engagement in Dyadic Child-Robot Interaction Kinesics: Lessons Learnt from an Exploratory Study', in *Proceedings of IEEE RO-MAN'05*, IEEE Press, pp. 716–722.

Robins, B., Dautenhahn, K., te Boekhorst, R., and Nehaniv, C.L. (2008), 'Behaviour Delay and Robot Expressiveness in Child-Robot Interactions: A User Study on Interaction Kinesics', in *Proceedings of the ACM/IEEE 3rd International Conference on Human-Robot Interaction* (*HRI '08*), ACM, New York, pp. 17–24.

Robins, B., Dautenhahn, K., and Dickerson, P. (2009), 'From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot', in *Proceedings of the Second International Conferences on Advances in Computer-Human Interactions, ACHI 09*, 1–7 February 2009, Cancun, Mexico, IEEE Computer Society Press, pp. 205–211.

Sacks, H., Schegloff, E.A., and Jefferson, G. (1974), 'A Simplest Systematics for the Organization of Turn-Taking for Conversation', *Language*, 50, 696–735.

Schmidt, R.C., Richardson, M.J., Arsenault, C., and Galantucci, B. (2007), 'Visual Tracking and Entrainment to an Environmental Rhythm', *Journal of Experimental Psychology*, 33(4), 860–870.

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. (2008), 'How Quickly Should Communication Robots Respond?', in *ACM/IEEE 3rd Annual Conference on Human-Robot Interaction (HRI2008)*, pp.153–160.

Suzuki, N., and Katagiri, Y. (2007), 'Prosodic Alignment in Human-Computer Interaction', *Connection Science*, 19(2), 131–141.

Steels, L., and Brooks, R.A. (eds.) (1995), *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Syrdal, D.S., Koay, K.L, Walters, M.L., and Dautenhahn, K. (2007), 'A Personalised Robot Companion? The Role of Individual Differences on Spatial Preferences in HRI Scenarios', in *IEEE International Symposium on Robot and Human Interactive Communication (Ro-man)*, Jeju Island, Korea, pp. 1143–1148.

Tapus, A., and Matarić, M.J. (2006), 'Towards Socially Assistive Robotics', *International Journal of the Robotics Society of Japan*, 24(5), 576–578.

Tapus, A., Tapus, C., and Mataric, M.J. (2008), 'User-Robot Personality Matching and Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy' (special issue on Multidisciplinary Collaboration for Socially Assistive Robotics), *Intelligent Service Robotics Journal*, 1(2), 169–183.

Thórisson, K.R. (2002), 'Natural Turn-Taking Needs No Manual: Computational Theory and Model, from Perception to Action', in *Multimodality in Language and Speech Systems*, eds. B. Granström, D. House and I. Karlsson, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 173–207.

Trevarthen, C. (1999), 'Musicality and the Intrinsic Motive Pulse: Evidence from Human Psychobiology and Infant Communication' (special issue), *Musicae Scientae*, 155–215. **Q9**

Walters, M.L., Syrdal, D.S., Dautenhahn, K., te Boekhorst, R., and Koay, K.L. (2008), 'Avoiding the Uncanny Valley: Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion', *Autonomous Robots*, 24(2), 159–178.

Watanabe, T. (2004), 'E-cosmic: Embodied Communication System for Mind Connection', in *Proceedings of IEEE RO-MAN'04*, Kurashiki, Japan.

Weinberg, G., and Driscoll, S. (2006), 'Robot-Human Interaction With an Anthropomorphic Percussionist', in *Proceedings of International ACM Computer Human Interaction Conference (CHI 2006)*, Montreal, Canada, pp. 1229–1232.

Weinberg, G., Driscoll, S., and Parry, M. (2005), 'Musical Interactions With a Perceptual Robotic Percussionist', in *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, Nashville, TN, pp. 456–461.

Wrede, B., Buschkaemper, S., and Li, S. 'Do You Like This Robot? The Role of Robot Behavior, Robot Personality and User Personality', in *ACM/IEEE Human-Robot Interaction Conference – HRI2006*, Salt Lake City, Utah, USA. **Q10**

Yamamoto, M., and Watanabe, T. (2003), 'Time Lag Effects of Utterance to Communicative Actions on Robot-Human Greeting Interaction', in *Proceedings of IEEE RO-MAN'03*, pp. 217–222.

Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2007), 'How Contingent Should a Lifelike Robot Be? The Relationship Between Contingency and Complexity', *Connection Science,* 19(2), 143–162.

Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2008), 'Developing a Model of Robot Behavior to Identify and Appropriately Respond to Implicit Attention-Shifting', in *Proceedings of ACM/IEEE HRI 2009 Conference*, pp. 133–140.

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto, T. (2006), 'Responsive Robot Gaze to Interaction Partner', in *Proceedings of Robotics: Science and Systems*.

Appendix 1. Audio analysis

The acoustic sound waves recorded by the sound grabber module are converted to digital music samples, which allows the use of mathematical computations and sample-based techniques. To detect the patterns of a sound wave, a filter-based method is used, based on the work of Kose and Akin (2001) originally used to detect visual patterns. This method which is called Audio Analyser was used in the drumming experiments with KASPAR as well as a different humanoid robot (iCub) in real time. Also, in work not reported in this article, it was integrated to Webots software (Cyberbotics) to be used in a simulated drummer modelled after the iCub robot. The real power of the method comes from its being computationally efficient, simple, fast and real time. The drumming experiments are real time, and to have games which appear 'natural' with short durations between turns, we need to identify the bouts of drumming as soon as they are produced. Therefore, it is not possible to record them first and perform off-line analysis, or use efficient but complex methods in terms of computational resources and time. Also, although the human participants are expected to use either the end of a pencil or one hand to hit the toy drum, many different strategies were observed (and people were not discouraged to use them): they were observed to use the tambourine-style bells around the drum, use both hands or sometimes use a pen or a stick to hit the drum. Therefore, it is not trivial to train the system with 'normal' drumming bouts. Also, the high inner noise of the humanoid, besides the high noise around the drumming area (due to people present in the room), makes the environment very challenging and require us to set up high noise filters. The noise filters should be high enough to filter out the inner and outer noise, but low enough to pick up as many drumming bouts as possible. Since we use participants from both genders and all age groups, we could observe very frequent or very light bouts of drumming which are even harder to analyse. In the current implementation, we only use audio feedback to detect the drumming bouts, but in future work, we plan to use visual feedback also. However, as we mentioned earlier, the participants were allowed to use various different ways to produce sound during their drumming games; so even the addition of visual feedback would not bring optimal success in bout detection.

To detect the patterns inside a sound wave, a filter-based method is used.[12] In this method, a four-item mask is applied to every sample in the sound wave, and a filter is constructed. The peaks in this filter show the edges in the sound wave. A mask of $[-1\ -1\ 1\ 1]$ is used to detect rising edges, and another mask of $[1\ 1\ -1\ -1]$ is used to detect falling edges. Any part of the sound wave between a rising and a falling edge is a region which represents the beat in the sound wave. This is because a beat is represented by a set of points and not a single point. Once the regions are detected, a threshold is applied on the average value of the points in the region, to detect the 'real beats' and discard noise. This method is computationally simple but fast and efficient.