Rhythmic Human-Robot Social Interaction

Marek Piotr Michalowski

CMU-RI-TR-10-19

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Robotics

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

May 2010

Thesis Committee:
Reid Simmons (chair)
Jodi Forlizzi
Hideki Kozima
Illah Nourbakhsh

Contents

List of Figures

Abstract

Social scientists have identified and begun to describe rhythmic and synchronous properties of human social interaction. However, social interactions with robots are often stilted due to temporal mismatch between the behaviors, both verbal and nonverbal, of the interacting partners. This thesis brings the theory of interactional synchrony to bear on the design of social robots with a proposed architecture for *rhythmic intelligence.* We have developed technology that allows the robot Keepon to perceive social rhythms and to behave rhythmically. We have facilitated constrained social interactions, and designed experimental protocols, in which a robot variably synchronizes to human and/or environmental rhythms—first in a dance-oriented task, and second in a cooperative video game. We have analyzed these interactions to understand the effects of Keepon's rhythmic attention on human performance. This thesis demonstrates that variations in a robot's rhythmic behavior have measurable effects on human rhythmic behavior and on performance in rhythmic tasks. Furthermore, human participants were able to assume and transition between the roles of leader or follower in these tasks.

Acknowledgements

The author's work was made possible in part by financial support from an NSF Graduate Research Fellowship, an NSF East Asia and Pacific Summer Institutes/JSPS Summer Program, NSF IGERT grant #0333420, JSPS grant #18330144, Japan's National Institute of Information and Communications Technology, Miyagi University, and the Carnegie Mellon University Robotics Institute.

The author would like to thank the hundreds of children and students around the world who have enthusiastically interacted with Keepon in our demonstrations and experiments.

The author is grateful to several students, staff, and faculty who have provided support for experiments at various field sites. At Kyushu University, Kazuhide Hashiya, Hiromi Kobayashi, Saori Tanaka, and Kentaro Yamamoto generously gave time, assistance, and advice in early experiments. At the Children's School at Carnegie Mellon University, Linda Hancock, Donna Perovich, Sharon Carver, and Annie Smith helped to shape our dance-oriented sessions and graciously opened their school to our research. At Carnegie Mellon University, Ann Shoplik and Mike Shoplik (with the C-MITES program) and Margaret Reid-Miller (with the Andrew's Leap program) allowed us to study game-oriented interactions with their summer students.

Several professors have been instrumental in helping us develop our concepts, including Sara Kiesler, Chris Atkeson, Robyn Cruz, and Bard Ermentrout. In particular, Selma Sabanovic was pivotal in the author's early interest in social rhythmicity.

The author would like to thank Illah Nourbakhsh and Jodi Forlizzi for their encouragement and advice on his thesis committee. The author is grateful to his advisor, Reid Simmons, for his faith, deep patience, and unwavering support. And the author is eternally indebted to Hideki Kozima, Keepon's illustrious designer, without whose collaboration, mentorship, and friendship this thesis would have been impossible.

Finally, the author extends his thanks and love to his friends and especially to his family, who have encouraged, supported, and inspired him.

Chapter 1

Introduction

> *Poetry in general seems to have sprung from two causes, each of them lying deep in our nature. First, the instinct of imitation is implanted in man from childhood... Next, there is the instinct for harmony and rhythm.* [Aristotle, circa 330 BC] (*Poetics* section 1, part 4)

The human experience is defined primarily through our sociality, whether we express it by forging supportive bonds, cooperating to achieve goals, or struggling to resolve conflicts. Our social interactions are dynamic processes of coordinated activity between constantly adapting participants. Most people can effortlessly negotiate the give-and-take of conversations, the fluid exchange of non-verbal communication, and the intricacies of coordinated behaviors—yet some people (and most robots) are challenged by this temporal coordination.

Rhythm—which we usually think of in the context of music and dance—is simply the repetition of temporal patterns in sound or movement. Rhythms are characteristic of almost everything our bodies do, from our heartbeats to our walking gaits to our sleep cycles. Our social behaviors, too, are particularly infused with such rhythms: accents in our speech have a tempo with varying speed, and the gestures we make with our heads and arms are often repetitive. It follows that we are also particularly attuned to rhythms in our natural and social environments. Since the 1960s, beginning with the work of William S. Condon [Condon and Ogston, 1966], social scientists have discovered not only that there are rhythmic patterns in human social behavior, but also that there is usually a synchrony between the sounds and movements of interactors. The rhythms tend to converge on each other during smooth interactions through a process known as *entrainment.* [Scheflen, 1964] points out that any dancer, musician, athlete, or lover understands that sharing a common rhythm is essential to achievement of joint tasks. This *rhythmic interactional synchrony*, along with the unconscious mirroring of behaviors, is the means by which we establish *interpersonal coordination*. It is often compared to a "dance" [Hall, 1983] that serves to mediate or regulate our social interactions with each other, serving as an important foundation, or scaffold, for establishing rapport, engagement, common ground, and emotional contagion between children and caregivers, between conversational partners, and between teammates performing joint tasks. We shall review literature from the social sciences on interpersonal coordination in section 2.1.

Despite the importance of social intelligence in our lives, its basic principles are difficult to implement in artificial systems. Since rhythm is a fundamental property of a wide range of natural and biological processes, it will certainly play an important role in the artificial life we are trying to breathe into our robots. Yet while rhythm is already central to many areas of robotic research such as legged locomotion or snake robots, it has been surprisingly lacking in human-robot interaction. Interactive robots—designed to be capable of using natural human-like social behaviors—are being built for purposes of science (performing experiments to learn about human behavior or cognition), service (assisting in labor, communication, or access to information), and influence (education, therapy, entertainment, or aesthetic experience). Yet a common deficiency in contemporary social robots is the stilted and rigid nature of their interactions. Inadequate performance with respect to responsiveness, timing, and speed even leads us to use the term "robotic" to describe our interactions with people who are socially stilted. Regardless of motivation or application, the success of proximal human-robot social interaction depends strongly on robots being able to perceive the spatial and temporal properties of the interaction and to behave in a coordinated manner. We review related work from robotics in section 2.3.2.

Perhaps one obstacle to the development of integrated, robust, dynamic, socially intelligent systems is the prevalence of a mindset that communication is a sequential activity in which information is transmitted, processed, and responded to by each interactor in turn. This model has difficulty accounting for the subtle nonverbal behaviors and cues that underlie (indeed make possible) more symbolic forms of communication. If we look not at individual agents but at the entire system of an interacting dyad, we see co-action rather than inter-action: a dance of simultaneous, periodic, coordinated behaviors that enable the negotiation of roles and activities. Some easily observable aspects of this coordination, such as turn-taking, gestures, emotional expression, and natural language, are all important for effective communication and are popular areas of research in human-robot interaction. However, an increasing amount of attention is being directed to the idea that the comfort and "naturalness" of an interaction may depend on a foundation of far more subtle and fine-grained coordination between the sounds and movements of the two interactors. If we neglect to endow our robots with the ability to perceive the subtle rhythms in the behavior of their human partners or in the environment, and if the robots are incapable of moving dynamically in temporal patterns that correspond to these rhythms, then we will never move beyond the stilted, rigid, "robotic" interaction that people have come to expect from machines. It may be useful to think of interaction less as a sequential transaction of messages and more as a coupling between social "oscillators." For example, a humanoid assistant of the future should provide feedback through backchannelling behaviors (such as nodding) when listening to human speech, and the frequency of this nodding should derive from (and contribute to) the rhythmic flow of the conversation. Since rhythmicity is a fundamental property of such interactions, it should serve as the basis upon which we build higher-level symbolic, emotional, or linguistic communication, rather than as an afterthought.

Our long-term goal is to develop techniques and technologies by which artificial systems, and specifically robots, can participate in coordinated, engaging, natural social interactions that are as rich as those we experience with one another. We believe that interpersonal coordination—and specifically rhythmic interactional synchrony, the correspondence between repetitive movements of interacting partners—is necessary for the regulation of natural, comfortable, effective human-robot interaction. The development of robust social intelligence requires new methods of perceiving, modeling, and behaving according to environmental and interaction rhythms. In Chapter 3, we describe our general approach to these components and present a general architecture for designing robots that can participate in rhythmically coordinated interactions. The modeling, perception, and generation of rhythmic behaviors forms the basis for *rhythmic intelligence*—the ability to establish and maintain interpersonal coordination. Rhythmic intelligence includes the ability to achieve specific goals in an interaction through a precise temporal selection of repetitive behaviors and to understand the dynamically changing nature of an interaction such that it is possible to select appropriate roles for different situations. We believe that the concept of rhythmic intelligence can form the basis for a robust form of social intelligence—enabling robotic devices to interact naturally and fluidly with people—and can possibly lead to a greater understanding of social intelligence in humans. Our framework consists of a set of increasingly sophisticated processes. First, the robot must have a mechanism for *rhythmic attention.* From the flood of stimuli being perceived by all of its sensors, it must determine which rhythms are meaningful or important to respond to: natural, social, multi-modal, or particularly strong in a single modality. Second, the robot must have a method for *entrainment*, or comparing and matching the rhythms it perceives to the rhythms it is currently exhibiting. It may need to evaluate how successfully it is following a person's rhythm, or it may need to recognize the extent to which the person is following its lead. Third, the robot should intelligently use its rhythmic

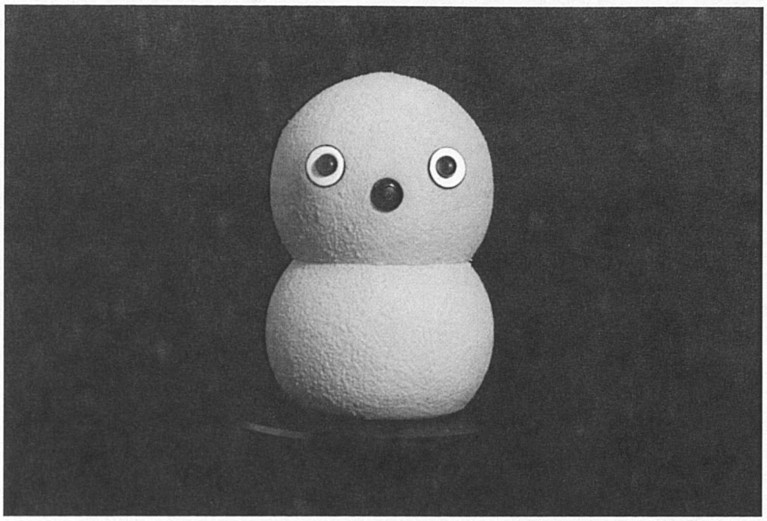


Figure 1.1: The robot Keepon.

awareness to recognize (or cause) changes in the *state* of the interaction. From dancing, to the turn-taking in verbal communication, to the coordinated manipulation of physical objects by human-robot teams, interactions have certain rhythmic trajectories that regulate the behavior of participants. Robots will either need to learn or be given such models. This undertaking can be informed by a wide range of disciplines, including artificial intelligence, neuroscience, cognitive science, signal processing, music, and dance.

Our specific goals in this thesis are to design and evaluate a robotic system whose perceptual and behavioral capabilities allow it to participate in simplified forms of rhythmically synchronous interaction. We have explored rhythm detection in a number of perceptual modalities, including vision, sound, accelerometers, and pressure sensors. We have explored a number of rhythmic modeling or entrainment techniques, including frequency domain analysis, dynamical systems, and musical metronomes. And we have designed a rhythmic behavioral mechanism on the robot Keepon,[1] a small 4-DOF robot with deformable skin and a simple appearance (fig. 1.1). The result is a system that can perceive rhythms in multiple modalities and synchronize periodic dance-like movement to these rhythms (even while under high-level attentional control of a human teleoperator). Our technology is described in Chapter 4.

We have chosen to develop and evaluate our system in the context of *play*. Play is a particularly appropriate domain for this type of work, as the principles under study (rhythm and synchrony) are magnified or emphasized in playful interactions—physical, repetitive, and exaggerated as they often are. We also work with a population of children,

---

[1] Keepon is a product and trademark of BeatBots LLC (http://beatbots.net).

both due to the design of our robotic platform, and because of the playful nature of our interaction designs. Specifically, we have chosen dance and physical video gaming as our two modes of play.

While the rhythms of everyday interaction are quite complex, dance is a form of interaction that provides a context of constraints and opportunities useful in the development of rhythmically intelligent robotic technologies. Dance does not require verbal communication; it features regular structured rhythms externally reinforced by sound; it involves visual, auditory, and tactile modalities; and it is familiar and enjoyable to people of all ages. The disposition to engage in music and dance can be found throughout human cultures. Although the evolutionary processes and cognitive mechanisms behind such behavior are under debate, it is significant that music and dance are largely social activities, and it is likely that our participation in and enjoyment of these art forms (whether for the purpose of aesthetic pleasure, entertainment, communication, or ceremony) is a mode of non-verbal communication closely linked to our evolved sociality [Blacking, 1983]. In Chapter 5, we describe a series of studies involving dance-oriented play between children and Keepon. Over the course of conducting these iterative studies, we developed our technology, used various perceptual mechanisms, and refined our experimental protocol. Our methodologies consisted of both behavioral analysis of unconstrained interactions as well as controlled experiments in which the robot's configuration or other variables differed between groups of participants. We found that the robot's rhythmic behavior had a synchronizing influence on children's behavior, even when the robot deviated from an environmental rhythm. We also found that children were able to assume roles of either leader or follower in the dance interaction, but that they more closely followed environmental rhythms when acting as followers than as leaders.

In Chapter 6, we describe a study that uses video gaming as an additional playful context for rhythmic interaction. The game, played by human and robot together, requires cooperative synchronization in "jumping" on a virtual trampoline. We used the study to analyze the effects of synchrony from a perspective of task performance rather than pure rhythmic behavior. In this study, we found that the robot, as in the dance interactions, could guide the human's rhythmic performance in order to improve task performance, and that the task was easier for humans when following the robot than when leading it. We also structured the task such that the roles, as well as available information, changed over the course of the game. We found that the transitions between roles had interesting temporal properties with respect to the time required to achieve improvements in performance.

To summarize the concrete contributions of this thesis, we will present:

- a general approach for giving robots the ability to participate in rhythmically synchronized interactions;

- the development of robotic technology for rhythmic perception, behavior, and entrainment in playful, constrained social interactions;
- experimental protocols for testing the system, measuring human behavior, and evaluating the interactions; and
- an examination of the rhythmic and performance properties of leadership/initiative role transitions in different rhythmic interactions.

Our hypothesis is that a robot's rhythmic behaviors, when synchronized to environmental or human rhythms, will result in higher *engagement* (the degree of involvement or interest in the interaction) than when the robot's behaviors are unsynchronized. Secondly, we hypothesize that people can reliably take on the roles of rhythmic *leader* (demonstrating a rhythm to the robot, which imitates the rhythm) or *follower* (observing and imitating the robot's rhythm) in an interaction with a robot. We expected that people would respond more positively to the robot's responsiveness to their own rhythmic behavior, but we found instead that our tasks proceeded more smoothly or successfully when participants followed the robot's rhythms. We believe that these results point to the need for further work both on the technology and on designing richer interactive scenarios that are still amenable to experimental analysis.

Taking a broader view, this thesis aims to have a greater impact on the field of human-robot interaction by:

- providing an overview and foundation of technologies and protocols for research in interpersonal coordination;
- laying down a framework for developing more generally applicable rhythmic social intelligence; and
- suggesting rhythmic properties of role transitions and suggesting further research in other types of interaction.

In this thesis, we first review literature relevant to rhythmic synchrony, social and otherwise. We then present our general approach to the problem of creating rhythmic intelligence in robots. Next, we describe the technology we have used or developed in creating our rhythmic robotic system. We then present a series of dance-oriented studies, followed by a video-game-oriented study. Finally, we discuss our findings and future work.

Chapter 2

Related Work

We shall now review literature that applies to the problem of artificial rhythmic intelligence. This literature necessarily comes from a broad range of fields, from the social scientists who have described interpersonal coordination, to mathematicians who study oscillators, to roboticists who have worked to implement these principles in artificial systems.

The social sciences give us a body of literature that describes, and to a lesser extent experimentally demonstrates, the rhythmic properties of social behavior. These properties are a manifestation of the oscillatory properties of living beings, and more generally, of many natural phenomena. However, only recently have robots been built that begin to overcome the problems of implementing these principles.

2.1 Interpersonal coordination

The past four decades have witnessed the emergence and development of a line of research spanning disciplines such as anthropology, psychology, and cognitive science in understanding the organizing principles for interaction between intelligent embodied agents. We now understand interaction not as a sequential process of give-and-take, but as a dynamic process of coordinated activity between constantly adapting participants. [Bernieri and Rosenthal, 1991] group this body of research under the term *interpersonal coordination*, which is divided into the general principles of interactional synchrony and behavior matching. Following them, [Burgoon *et al.*, 1995] integrate these "biologically based models" in their formulation of a theory of *interpersonal adaptation.* [Hatfield *et al.*, 1994] propose that "emotional contagion" is driven by an automatic and continuous tendency for people to "mimic and synchronize their movements with the facial expressions, voices, postures, movements, and instrumental behaviors of others." Additionally, "common ground theory" describes the process by which interactors coordinate their understanding of linguistic communication [Clark and Brennan, 1991]. Rhythm is one of the channels through which this grounding takes place, along with other nonverbal elements such as attention, proxemics, and affect. These are all inter-related: for example, the speed and rhythmicity of bodily and gestural movement is a cue to affect, and attentive head movements and gestures are used together in a directional manner.

Many researchers have investigated the importance of rhythmicity in human-human interactions. They have postulated both postural and temporal coordination between interactors. In the following subsections, we review work on the mechanisms and uses of interpersonal coordination.

2.1.1 Behavior matching

Behavior matching, or posture mirroring, describes a (usually unconscious) correspondence between the positions or gestures assumed by two interactors. [Scheflen, 1964] first proposed that people in a group mirror each other's posture and that those who share a posture usually also share a viewpoint. [LaFrance, 1979] observed a correlation between posture sharing and verbal rapport. In addition to its role in establishing a bond or affiliation between interactors, [Bandura, 1977] proposes that mirroring plays an important role in human development, helping children to learn or "model" observed behaviors.

While behavior matching is an important element of human-human interaction, we are interested in developing robust intelligent systems that are not necessarily isomorphic with humans. While an inexact bodily correspondence does not preclude mirroring (indeed, we have observed it with the robot Keepon, as in section 5.1), we are specifically interested in the temporal and rhythmic properties of interaction; we therefore focus mainly on interactional synchrony.

2.1.2 Interactional synchrony

Interactional synchrony is the temporal coordination of communicative behaviors between interactors (often without awareness or volition) in order to achieve a sort of "goodness of fit" between them [Burgoon *et al.*, 1995]. Along with intrapersonal, or self-synchrony (between a speaker's own behaviors and vocalizations), interactional synchrony was conceptualized by [Condon and Ogston, 1966; Condon and Ogston, 1967; Condon and Ogston, 1971] as congruence between changes in a speaker's vocal stream with those of a listener's movements. This synchrony was discovered through video analysis consisting of frame-by-frame labeling of change points in the direction of bodily movements and segmented speech. [Bernieri and Rosenthal, 1991] have organized and expanded the definition of interactional synchrony into *interaction rhythm*, *simultaneous movement*, and *smooth meshing*.

2.1.2.1 Interaction rhythm

The rhythmic organization of social interaction is an expression of the oscillatory neurobiological language of the central nervous system through learned cultural patterns of interaction [Chapple, 1982]. Two or more people coordinate their rhythms, achieving synchrony, through a process known as *entrainment* [Condon and Ogston, 1966; Kendon, 1990], whereby multiple different rhythms converge on, or capture, each other. The social "metronomes" of multiple interactors are considered to be in sync with each other when they exhibit similar temporal behavioral patterns; the tempo or style of vocalizations or bodily movements of an interacting individual may become the *zeitgeber*, or "time giver," for a partner's vocalizations or bodily movements [Bernieri and Rosenthal, 1991]. These rhythms that govern our interactive behavior are a fundamental precursor to our communicative development; [Condon and Sander, 1974b; Condon and Sander, 1974a] find that human neonates move in precise and sustained synchronous organizations of movement with the articulated structure of adult speech as early as the first day of life. It continues to be important in our later verbal communication: synchrony is involved in establishing mutual engagement, conversational regulation, and turn taking [Auer *et al.*, 1999], all of which are essential to the presentation-acceptance process in common ground theory [Clark and Brennan, 1991].

2.1.2.2 Simultaneous movement

In addition to a correspondence between the frequency, speed, and tempo of interaction rhythms, interactional synchrony is characterized by the co-occurence of movements, gestures, vocalizations, or body positions between two or more interactors [Burgoon *et al.*, 1995]. This is often observed as simultaneous change in direction of movement in different body parts or segmentation points in a vocal stream [Condon and Ogston, 1967]. In other words, while two rhythms may be synchronized in frequency, simultaneous movement furthermore implies that interaction rhythms are synchronized in phase. A purely reactive system cannot, theoretically, be synchronous in this way, due to delays in sensing, processing, and acting. It may be argued that one of the reasons for the rhythmic organization of social behavior is that it enables synchrony and simultaneous movement through *anticipation* on this short time scale.

2.1.2.3 Smooth meshing

Meshing refers to the coordination of two or more interactors' behavioral patterns into a single, unified, meaningful "whole" [Bernieri and Rosenthal, 1991]. The adaptation of an interactor's behaviors and patterns is seen to relate to the other's actions; for example, a listener nods and uses other backchanneling cues at appropriate times during a speaker's utterance [Burgoon *et al.*, 1995]. [Kendon, 1970] suggests that the way in which synchrony varies in an interaction is related to the individuals' roles in the interaction. One's responses to the tempo of another individual's movement and speech may indicate engagement, disengagement, acknowledgment, or the assumption of the role of speaker. For example, the transition between speakers in a dialogue is marked by the most conspicuous synchrony between speech and movement, in order to signal engagement and anticipation of the pending transition. Kendon argues that these joint changes in the nature of interactional synchrony have the important function of achieving "the delicate coordination of expectancies among participants, so essential to the smooth running of an encounter." On a longer time scale (e.g., of conversational turn-taking), this principle enables anticipation of changes in the interaction and the ability to guide these changes.

2.1.3 Use of rhythm in educational and therapeutic contexts

Given the important role of rhythmicity and synchrony in human interaction, it is clear that difficulties in establishing interactional synchrony can make face-to-face interpersonal communication and interaction difficult; in fact, marked asynchronies within an individual's own behaviors [Condon and Sander, 1974b], along with abnormal entrainment in interpersonal interactions [Condon, 1986], are often characteristic of (causal to or symptomatic of) pathologies such as autism, cerebral palsy, and schizophrenia. Such disorders are frequently treated by therapeutic methods that specifically address these rhythmic difficulties. Music, dance, and rhythmic play, in addition to being widely used in standard educational practice from infancy to adulthood, are also well-established methods in treatment and therapy for a wide range of ailments, including abuse and trauma, autism and developmental disorders, learning disabilities, neuroses, attention deficit, mood and hyperactive disorders, medical illnesses, physical disabilities, substance abuse, geriatrics, and psychiatric, behavioral and emotional disorders [Fledderjohn and Sewickley, 1993; Cruz and Berrol, 2004]. [Trevarthen and Malloch, 2000] speak of music as a communication of motives and experience, pointing out an inherent musicality in infant-parent vocal interaction. They discuss the use and impact of musical therapy in social development, arguing that music is therapeutic because it "attunes to the essential efforts that the mind makes to regulate the body, both in its inner neurochemical, hormonal and metabolic processes, and in its purposeful engagements... with other people." For our purposes, we can

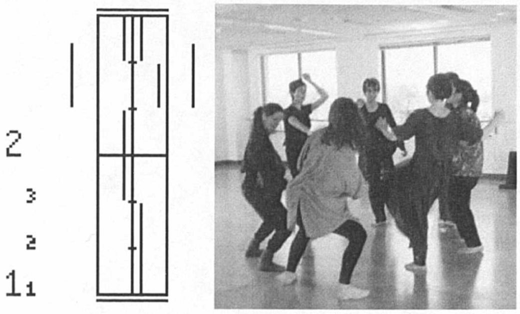


Figure 2.1: Left, a simple example of Labanotation, illustrating weight transfer between legs and a jump over two measures of time [Griesbeck, 1996]. Right, dance therapy [Lesley University, 2006]

consider interactive dance-oriented play with a robot as an activity that can be used by clinicians to study a child's behavioral patterns and to identify rhythmic abnormalities that might be useful in diagnosing and perhaps even treating certain pathologies.

Choreographers, dancers, and movement therapists have developed techniques for describing and annotating the full range of human movement. One of the most popular systems of human movement analysis used by dance and movement therapists is Laban Movement Analysis, or LMA [Laban, 1980] (fig. 2.1). LMA represents the *shape* of movement (e.g. position, distance, mass, paths) as well as the *effort* behind such movements (with respect to "space," "weight," "time," and "flow"). The system aims to understand movement comprehensively, addressing the structural and physical characteristics of the body parts that are moving, the dynamics and rhythms of movements as related to intention, the overall shape of the body, and the geometrical properties of paths through the surrounding spatial environment [Laban, 1980; Lamb and Watson, 1979; Bartenieff and Lewis, 1980]. This system is amenable to computational encoding, and [Chi *et al.*, 2000] have worked toward a character animation system that applies these ideas to generating natural synthetic gestures. [Swaminathan *et al.*, 2009] have used a Bayesian approach to identifying LMA qualities in motion capture data in real time.

[Kalish, 1968] presents an early discussion of the use of LMA to identify abnormal movement patterns, and the beneficial effects of movement therapy in autism. That research specifically uses rhythmic and mirroring movement to establish a relationship with a child, reducing rigid and abnormal movements while exhibiting more involved behaviors and purposeful vocalizations. [Rider and Eagle Jr., 1986] discuss the use of rhythmic entrainment in therapy for cerebral palsy, autism, and learning disorders. It was found that stereotyped behavior in these various disorders could be reduced by reinforcing entrainment to music matched by tempo to the speed of the behaviors; affected persons were able to effect control of their behaviors by their association with musical stimuli. We expect that a rhythmically capable social robot would afford similar opportunities for productive entrainment to both visual and auditory stimuli; furthermore, novel therapeutic methods might be afforded by a robot capable of encouraging, reinforcing, or correcting such entrainment through its own ability to entrain to rhythms.

Neuroscientists have also investigated the relationship between music and language, the coupling between the auditory and motor systems, and the mechanisms involved in the perception of rhythm by the human brain [Patel *et al.*, 2005]. The hierarchical nature of these perceptual mechanisms should inform the design of models of rhythmic behavior, shed light on the role these rhythms play in establishing communicative engagement between people, and explain the communicative and therapeutic benefits of dance and music.

2.2 Rhythmic synchrony in biological systems

Synchrony is ubiquitous among living things, but one organism has been of particular interest due to the dramatic nature of its rhythmic behavior. For centuries, travelers to Southeast Asia have reported seeing thousands of fireflies flashing in perfect unison [Strogatz, 2003]. Explanations for this remarkable coordination ranged from the weather to the existence of a "leader," until [Buck and Buck, 1968] discovered that the fireflies adjusted their rhythms in response to the flashes of others: sometimes speeding up, and sometimes slowing down. (Leadership roles have, on the other hand, been identified in other species. For example, [Naguib *et al.*, 1999] have studied how nightingales respond to perceived rivals who asymmetrically "lead" in a sequence of alternating songs with other counterparts "following.")

The large-scale social synchrony seen in firefly swarms also occurs among human crowds, most often seen in the explicitly rhythmic activities of walking or clapping. For example, on the opening of the London Millennium Footbridge in 2000, a crowd of pedestrians spontaneously synchronized their gaits to the natural resonance of the bridge itself, thereby further driving the bridge's oscillation. This caused it to wobble uncomfortably, and two days later the bridge was closed for two years of modifications [Strogatz *et al.*, 2005].

The activity of clapping illustrates interesting cultural differences in the propensity for social synchrony. Audiences in the United States generally applause in a disorderly manner until it dies out. On the other hand, Eastern European audiences often become synchronized into waves of rhythmic applause. Through recordings of natural audiences as well as controlled clapping studies, [Néda *et al.*, 2000] found that there are two distinct modes of clapping: fast (as is found immediately at the end of a performance, with a large dispersion in observed clapping frequencies) and slow (with double the period—or half the frequency—and a smaller dispersion in frequencies). The audience synchronizes its disorderly (fast) clapping by skipping every other beat to begin slow clapping; however, as this results in a lower average noise intensity (with half the number of claps per unit time), more enthusiastic members speed their clapping to increase the noise. This increases the frequency dispersion and destroys the synchronization. These 10-15 second "waves" between rhythmic and disorderly applause are the result of a tug-of-war between desires for synchrony and volume. United States audiences generally do not play the "game" of slowing down to achieve synchrony; on the other hand, during speeches by totalitarian Eastern European leaders, Barabási suggests that synchrony was maintained for even longer periods of time due to a *lack* of enthusiasm [Néda *et al.*, 2000].

The models that govern these crowd-level synchronization phenomena—whether in firefly blinking, human walking, clapping, chanting, etc.—also govern a wide range of other rhythmic biological phenomena. Female friends and co-workers who spend much time in close proximity often synchronize their menstrual cycles, perhaps through chemical signals in pheromones [Stern and McClintock, 1998]. Human circadian rhythms (sleep and body temperature cycles) entrain to the 24-hour period of the earth's rotation [Strogatz, 2003]. *Central pattern generators* are network of neurons that produce rhythms governing a wide range of biological functions. The heart's natural pacemaker, a cluster of about 10,000 cells called the sinoatrial node, generates a single, synchronized electrical rhythm from a coupling between thousands of intrinsically oscillatory cells [Michaels *et al.*, 1987]; such networks also govern vertebrate respiratory function [Marder and Bucher, 2001]. In addition to pacemaker systems, rhythms can also emerge from synaptic connections between neurons that are not themselves intrinsically rhythmic (e.g. through reciprocal inhibition), and these have been identified in many types of muscular locomotion [Marder and Bucher, 2001]. Central pattern generators inspired by biology have been used extensively in robot motion control [Ijspeert, 2008].

2.3 Rhythmic synchrony in artifacts

While biologists and sociologists were discovering and describing these forms of rhythmic synchrony, scientists in other fields—mathematicians, engineers, physicists, chemists—were also studying similar properties of nonlinear dynamical systems over the course of the twentieth century. What began to emerge is a unified science of *coupled oscillators* that applies equally to firefly rhythms as it does to the rotation of planets [Strogatz, 2003].

2.3.1 Mathematical foundations of synchrony

Christiaan Huygens, in the 17th century, discovered that clocks mounted on the same wall would become synchronized. He eventually recognized that the clocks were influencing each other's periodicity through the imperceptible movements of the connecting wall [Pikovsky *et al.*, 2002]. Even this weak coupling was sufficient to bring clocks with different intrinsic frequencies (imperfectly fabricated as they were) into synchrony with each other.

Oscillators are entities—physical, chemical, biological—that cycle automatically and actively (using their own energy) at some regular time interval. One formulation of such a self-sustained oscillator is as a dynamical system with a rotating phase point. External forces can perturb the point and shift its phase. Oscillators are coupled if they can influence each other through some physical process. Coupled oscillators impart forces on one another that either push or pull their phases closer together. Just as phase transitions and crystallization represent a spontaneous creation of *spatial* order from disorder, synchronization represents the creation of *temporal* order from groups of incoherently behaving oscillators. And just as crystallization is inevitable under the appropriate physical conditions, synchronization is inevitable under certain conditions: namely, when the intrinsic frequencies of the oscillators are within a certain range of each other, and when the coupling between them is sufficiently strong. A theoretical review of these principles is given in [Pikovsky *et al.*, 2002], while a more casual description can be found in [Strogatz, 2003].

2.3.2 Interpersonal coordination in robotics

[Restivo, 2002] emphasizes rhythmic entrainment, in both speech and body, as the foundation of any socially interactive capabilities that a robot may have. We have identified problems with rhythmicity as the cause of failures in interaction or engagement with robots [Sabanovic *et al.*, 2006; Michalowski *et al.*, 2006b]. Although temporal and postural synchrony has been demonstrated to affect engagement with virtual (animated) agents [Bailenson and Yee, 2005; Kang *et al.*, 2008; Reidsma *et al.*, 2008], only a handful of robots have been developed that address the challenge of synchronizing rhythmically to humans. Our goal is to expand significantly on this limited body of work.

2.3.2.1 Mimicry and posture mirroring

[Sidner *et al.*, 2005] have used the imitation of behaviors such as nodding and gesturing to increase the engagement between a robot and a human. (They use the term *engagement* to signify the process by which interactors start, maintain, and end their connection to each other, while we use it slightly differently to signify an interactor's degree of involvement in, attraction to, or interest in the interaction. They point out that engagement is poorly understood in a human-human context, and it has not been clearly defined in relation to measurable behaviors such as gaze or vocalization. For our purposes, we consider engagement as expressed through energy level and/or retention.) [Breazeal and Scassellati, 2002; Breazeal *et al.*, 2004] have investigated the use of imitation in social learning by humanoid robots, focusing on the necessity of an attention system to determine what to sense and a theory of mind to infer intentions. In the Aurora project, [Dautenhahn, 1999] used posture mirroring to create a sense of emotional empathy between a robot and children for applications in autism therapy. [Tanaka *et al.*, 2006] created non-interactive and interactive dance modes for a QRIO robot in a playroom with children. They found a significant difference between the two conditions with respect to the time children spent with the robot, yet this experiment concerned contingency rather than the type of rhythmic synchrony we are developing. The robot's interactive movement was an attempt at posture mirroring, which was severely limited by the difficulty in sensing posture. We have observed similar differences between contingent and non-contingent behaviors in rhythmic interaction between children and a robot-like puppet [Michalowski *et al.*, 2006a], and furthermore observed a higher incidence of rhythmic play by children when the puppeteer responded contingently to their behaviors (including posture mirroring). Because of the difficulties of automatic posture recognition, and because our robots are not isomorphic with people, we focus on rhythmic synchrony. Nevertheless, posture mirroring plays a role in the human-robot social interactions we observe.

2.3.2.2 Rhythmic synchrony

Attention to the rhythmic characteristics of nonverbal interaction has not been widely adopted in social robotics research. General ideas of turn-taking in conversation are widely implemented, but fine-grained rhythmic perception and synchrony by a robot has been difficult to develop. [Penny, 1997] developed *Petit Mal*, an art robot that reactively and nimbly evokes a rhythmic, proxemic, nonverbal interaction with people by positioning itself relative to its interacting partners. [Ogawa and Watanabe, 2001] recognize the importance of rhythmic entrainment to non-verbal cues such as nodding and gesture in vocal communcation. In their InterRobot humanoids, which are used in pairs for embodied telecommunication between remote interactors, they use nonverbal rhythmic cues such as head and body movements that are automatically synchronized to the speech input of a remote human interactor. [Andry *et al.*, 2001], with the view of interpersonal coordination as a method of learning, use synchronized imitation as a way for robots to learn sensory-motor correspondences and rhythmic motion sequences. The humanoid robots Nico [Crick *et al.*, 2006] and Haile [Weinberg and Driscoll, 2006] perform drumming synchronized to another person or a conductor as an example of the kind of rhythmic entrainment that might be applied to other social tasks. In these projects, rhythmic interaction is generally based on auditory cues and limited perception of embodied movement. However, robots should also be able to "tune in" to the bodily rhythms of their interaction partners and to generate such nonverbal behaviors (e.g., nodding, swaying, gesturing) themselves. [Robins *et al.*, 2008] hypothesize "temporal behavior matching" by children as they adapt to a robot's temporal behavior, as represented by delays in an expression imitation game and in a drumming game. A few commercial products, such as the Hasbro iDog, attempt to display musical rhythm through embodied movement. As we will discuss, the ability to detect beats in music is an important component of our concept of rhythmic intelligence, especially in the context of possible applications to musically-oriented therapy. However, singular techniques (such as signal-processing methods) are by themselves inadequate for anticipating and adapting to the rich, multi-modal rhythms of human interactive behavior.

In our work, we consider a form of rhythmic "leadership" in which a rhythm is demonstrated by one partner and imitated by the other. This notion has not been extensively or quantitatively studied in a human-human interaction context, although haptic technologies have been used to study human-robot dance interactions [Gentry *et al.*, 2003; Kosuge, 2010]. However, while these projects focus on directionality of movement, our work focuses on frequency or tempo of periodic behavior.

Social rhythmicity requires that interactors are capable of both active and passive participation in a mutually adaptive process of coordination. Our approach is modality-independent and aims to use rhythmicity intelligently to allow a robot to decide when and how to synchronize with human interacting partners. Our work builds upon existing research by: a) varying the rhythm to which the robot attends and synchronizes; b) explicitly measuring the synchrony between robot, person, and music over the course of the interaction; and c) examining the relationship between synchrony and leadership roles through qualitative and quantitative measures.

The reviewed literature suggests several design variables that must be considered when creating robots for rhythmic social interaction, for example: the modalities to which the robot is equipped to attend; the representational, reactive, or generative mechanisms for perceiving or driving periodic activity; the selection of actuators or other forms of behavioral expression (e.g. sound or light); and the degree of postural or morphological correspondence of the robot's behaviors to human behaviors. The designer's goals, together with the interactive context, determine how these decisions are made and the difficulty of implementation. In the following chapter, we outline our general approach to achieving rhythmic intelligence through a framework that takes into account both the social scientific principles and the technology requirements for implementing them in artificial systems, in such a way as to be generalizable to a variety of robotic platforms. Due to the multidisciplinary nature of this work, it is necessary both to operationalize social science findings so that they are amenable to implementation, and in the other direction, to evaluate our technology in a way that makes sense in a social science context.

Chapter 3

General Approach

3.1 Overall problem

As discussed in Chapter 2, the theory of interactional synchrony attempts to explain the mechanisms behind the temporal complexity of human social behaviors, proposing that they are often rhythmic (exhibiting periodic temporal patterns), synchronized (matched in frequency and phase between partners), and regulatory (providing feedback or direction in various social contexts). Although social scientists have observed these structures in the temporal organization of human interaction, many aspects of this rhythmicity are poorly understood. Our understanding of social rhythms is vital to our relationships with both the natural and the artificial: From a qualitative perspective, how does rhythmicity serve in the establishment of engagement and the negotiation of roles? From a quantitative perspective, what are the temporal rules underlying the synchronization of social rhythms? From an experimental perspective, how can we probe the causes and effects of these phenomena in a controlled manner? From a design perspective, how will (or should) rhythmicity play a role in our interactions with artificial systems? Finally, from an engineering perspective, how can we model the rhythmicity of social behavior for implementation in these artificial systems?

A variety of approaches might be considered to begin answering these questions. For example, we might create a range of simulated agents in order to cover a design space of possible morphologies and techniques for perception, behavior, and synchronization. However, the very physical nature of the rhythmic behaviors involved in face-to-face interactions is not likely to transfer naturally to a virtual agent. Alternatively, we might incorporate rhythmic considerations into the design of conversational and gestural perception and behaviors on interactive humanoid robots. However, these systems are costly and complex to control, and the rhythmic temporal properties of human conversational interaction are still poorly understood.

Therefore, the approach we have selected is to create a simplified (regular and exaggerated) version of interactional synchrony in a human-robot interacting pair. While we must consider that simple interactions may be qualitatively different from the natural, unconstrained interactions that we normally think of (e.g. verbal conversation), we believe that it is instructive to begin by addressing a more easily digestible version of the problem before attempting to understand and implement more general (and more difficult) social tasks. To this end, we have taken the following approach:

1. Design a framework by which a socially interactive robot can synchronize with regular environmental and social rhythms;

2. Develop an experimental methodology for facilitating, manipulating, measuring, and analyzing rhythmic human-robot interaction; and

3. Interpret the outcome of these experiments as they relate to more general questions about human social rhythmicity.

Our overall problem, then, is to create technology for rhythmic synchrony, a context within which to observe and study the effects of this synchrony, and an interpretation of the results of these studies as they relate to interaction generally. We believe that the resulting technology and methodologies will be a useful stepping-stone when later scaling up to more general social tasks.

3.2 General hypothesis

We believe that interpersonal coordination, and specifically rhythmic interactional synchrony, will be important for the regulation of natural, comfortable, effective human-robot interaction. Our hypothesis is that rhythmic synchrony by a robot, even if not consciously perceived by a human, will have a measurable effect on the human's behavior. Specifically, we expect that a robot's ability to synchronize with environmental or human rhythms will result in increased engagement in interactions—that is, humans will be more likely to engage in rhythmic behavior or will engage for longer periods of time. These hypotheses are suggested by the social science findings (discussed in section 2.1) relating synchrony to engagement and interpersonal coordination.

We also hypothesize that the robot's attention to different rhythms will contribute to regulating or reinforcing various roles in the interaction (such as leader or follower), that people will be able to assume these roles in rhythmic interactions with a robot, and that these roles will have an effect on behavior and task performance. Specifically, we expect to find a difference between leading or following the robot with respect to environmental rhythms, and that following the robot will result in higher fidelity to these rhythms.

3.3 General solution

Our interests run bi-directionally between analytic and synthetic approaches to investigating social rhythmicity. On one hand, we would like to use what we know about human social rhythmicity in order to improve the interactive capabilities of artificial systems. On the other hand, we would like to take advantage of the controllability of artificial systems in order to study and probe the properties of human social rhythmicity. These goals provide constraints on both the system and the context within which it operates. First, we must build our systems on interactive platforms that people find attractive and relatable, yet they must be controllable in an agile manner. Second, the activities should be interesting to people, yet simple enough that rhythmic behavior is relatively regular and clearly expressed. Third, people should be able to behave comfortably and naturally, yet the behaviors must be easily and clearly measurable.

3.3.1 Building rhythmic synchrony

With these considerations in mind, we elected to work with an embodied robotic system (rather than an animated agent) so that people would interact with a partner that would share the same physical space and whose physical behaviors would be visually and auditorily salient. The robot Keepon is described in Chapter 4. We selected this robot as it is vaguely anthropomorphic or zoomorphic, nonverbal, navigationally immobile, minimally expressive, and aesthetically appealing. Essentially, Keepon is capable of directing its attention, assuming various emotive poses, and expressing rhythm through full-body movement.

The development of a rhythmically intelligent social robot requires the development of a framework for integrating various existing technologies and new techniques to allow participation in rhythmic social patterns. It is necessary for the robot to perceive or select social rhythms from multiple modalities, to behave in different ways to convey rhythms and synchronize to interacting partners, and to appropriately represent and guide the relationship between its own rhythms and those of its interacting partners. We see the modeling, perception, and generation of rhythmic behaviors as forming the basis for *rhythmic intelligence*—the ability to establish and maintain interpersonal coordination. Rhythmic intelligence includes the ability to achieve specific goals in an interaction through the selection of behaviors and to understand the dynamically changing nature of an interaction such that it is possible to select appropriate roles for different situations.

Specifically, it is important for the system to anticipate future rhythms and to expect changes in the nature of an interaction's rhythmicity. It is also necessary to differentiate, and to appropriately choose, between being passive or reactive to a partner's rhythms, and being active in guiding the interaction. Rhythmic entrainment can be seen as a process of negotiation, grounding, and compromise that requires not just an automatic response to the environment, but an intelligent understanding and choice of action.

3.3.1.1 Perception

Detecting and generating social rhythms first requires methods for perceiving and operationalizing rhythmic phenomena. Perception is also necessarily closely tied to representation, and our methods of perceiving rhythm bear on the way in which we control the robot. As discussed in Chapter 2, techniques from signal processing, computer music, neuroscience, choreography, etc. can provide computational, representational, or phenomenological insight into our design of such a system. Humans exhibit social rhythms across a number of modalities, most notably in movement and sound. In addition to being able to take advantage of the richness of available sensory information, it is important to perceive in multiple modalities in order to balance the relative strengths and weaknesses of different sensory technologies. We can therefore use cameras and techniques from computer vision such as feature tracking, optical flow, and pose recognition to sense people's movement. We can furthermore use microphones and techniques from signal processing in order to hear environmental rhythms such as music, percussion, or vocal prosody. The problem of beat detection in audio signals has been approached with nonlinear oscillators [Large and Kolen, 1994] and psychoacoustically-inspired filter banks [Scheirer, 1998]. Such methods can be applied to rhythmic perception from complex or noisy sensor data in other modalities as well, such as vision or other sensors capable of providing information not available to human senses (such as accelerometers, pressure sensors, tactile force data, and motion capture). In this work we have used, at various times, audio, video, accelerometers, and pressure sensors to perceive environmental and social rhythms. For any of these modalities, there are well-established techniques for sensing changes in the environment; however, only some techniques (such as beat detection in music analysis) readily account for the perception of periodic change in the presence of noise and complexity. In order to make use of other sensor data, such as cameras and accelerometers, we needed to adapt or extend existing techniques and to analyze data over time in order to identify the temporal, dynamic qualities of a person's movement.

For our later experiments, we narrowed our use of sensors to video-game input devices with accelerometers and pressure sensors that directly and accurately recorded human movement through direct contact with the body without being invasive or cumbersome.

These input devices allowed us to capture and record human movement at high temporal resolution, and with low noise, while providing affordances for play-like interactions.

3.3.1.2 Behavior

Humans exhibit social rhythms across a range of frequencies, from fractions of a second to minutes. To exhibit analogous rhythms requires that the robot possess sufficiently dynamic behavioral capabilities. "Lifelike" speed and smoothness of motion requires appropriate mechanical and electrical design. Furthermore, the physical appearance of such a robot, particularly for nonverbal interaction with children, must be compatible with the robot's capabilities and the qualities we would like to have attributed to the robot. Keepon is so minimally designed yet physically capable of moving in a nimble and lifelike manner due to its small size, fast and agile motors, and electronics that allow fine control of position, velocity, and acceleration.

The next challenge is to control the robot such that it exhibits rhythmic behaviors. For each of the robot's degrees of freedom (DOFs), we can move its motors in a periodic manner by sending oscillatory commands to repeatedly move between two positions. Furthermore, we can use select particular DOFs for use at different times, and we can change their parameters of movement (such as range, velocity, and acceleration) in order to control the perceived meaning (e.g. emotion) and relationship to environmental rhythms. We must also allow rhythmic movements to co-occur with other (e.g. attentional, intentional) movement, which requires a mechanism for composing multiple behavioral directives (whether autonomous or human-controlled). Our system's dancing behavior is controlled independently of the direction of attention commanded by a teleoperator.

Finally, since humans exhibit rhythms in sound as well as movement, the robot should also nonverbally convey rhythms through the medium of sound. In addition to enhancing the attractiveness of the resulting artifact, sound reinforces the visual stimulus and ensures continuous communication of rhythm even when the human's gaze is interrupted. In addition to the sounds of the motors themselves, Keepon uses a simple "language" of popping sounds that accompany certain movements.

3.3.1.3 Computation and representation

Given the ability to perceive rhythms and to generate them in its own behavior, the robot must represent these rhythms in a way that it can relate them to each other, with respect to differences in tempo, phase, and influence (of one interactor over another). Furthermore, the robot should be able to alter its rhythmic behavior based on changing social roles and the perceived effects of its rhythmic behaviors on those of interacting partners.

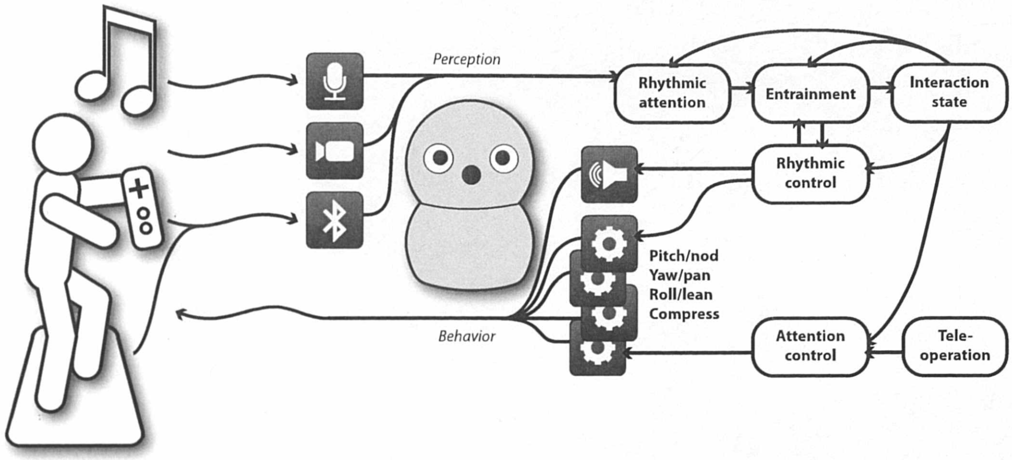


Figure 3.1: A framework for a robot that synchronizes to social rhythms.

Existing paradigms for programming artificial interactive systems and selecting behaviors, cues, or vocalizations are inadequate for our purposes since they are typically based on a discrete, sequential model of communication. In contrast, we need to represent, reason about, and manipulate continuous and co-occuring interaction rhythms. Implementing interactional rhythmicity and synchrony in computational systems is a relatively unexplored area, though various fields inform our selection of appropriate methods. For example, signal processing provides us with techniques for isolating the dominant or significant components of a signal in the frequency domain; however, in our experience, these techniques perform poorly on low-frequency signals in the short time windows that we are working with. Neuroscience can inform us about the mechanisms by which oscillating neural structures, or central pattern generators, maintain and change their periodic signals in response to sensory input. Computer music gives us methods for encoding, manipulating, and synchronizing rhythmic information in real time by representing tempos and using metronomes to drive periodic activity.

With insight from these various areas, we explored a framework for generating behavior that is synchronized with perceived rhythms in a directed manner (fig. 3.1). We propose that such a framework might consist of three components, corresponding to the *rhythmic attention*, *entrainment*, and *interaction state* components in figure 3.1. These stages build on each other, and might be developed and evaluated sequentially. If the robot can perceive rhythms in multiple modalities, it must either integrate information from multiple modalities or select the appropriate modality at different times (*rhythmic attention*). The robot must then relate or compare the rhythms it is perceiving to the rhythms it is exhibiting (under "rhythmic control," e.g. by an oscillator or metronome), so as to identify synchrony, asynchrony, and the changes in a human partner's rhythm in response to changes in its own—this is necessary for *entrainment*, or the establishment of rhythmic synchrony. Finally, the robot must use information about this relationship, as well as knowledge of the *state* of the interaction, in order to direct its attention (spatially and rhythmically), to intentionally modify its behavioral rhythms, and to anticipate changes in the interaction.

In the context of this thesis, we have made several simplifications in addressing the layers of this framework. We experimentally controlled the robot's rhythmic attention either to an environmental or a human rhythm. Entrainment entailed having the robot follow, as closely as possible, the perceived rhythm with its rhythmic behavior. We then compared the multiple rhythms in our analysis of the recorded data, looking for periods of synchrony or asynchrony. Finally, we explicitly controlled the state of the interaction—in the form of switching leader and follower roles—through explicit instructions or clear behavioral changes. These simplifications allowed us to create technology that afforded meaningful rhythmic social interactions with the robot. They also suggested constraints for experimental contexts that allowed us to answer questions about specific quantifiable elements of these interactions. Since they are simplifications, we must be careful in applying our findings to social interaction more generally.

We expect such a general framework to facilitate the creation of autonomous robots capable of participating in service-related, knowledge-producing, or experiential interactions with humans. However, the required higher-level abilities such as natural language and skill acquisition remain technologically difficult. The capacity for rhythmic synchrony alone is limited in its ability to create and sustain a compelling and engaging interaction. While entrainment to environmental rhythms might give an impression of awareness, joint activity, and even intelligence, we require a more sophisticated human-level contextual understanding of the situation in order to maintain interest and engagement. While investigating the rhythmic principles of interaction, we often used a "partial teleoperation" system (described in greater detail in Chapter 4) to create a context for what we would like to observe (see fig. 3.1). The underlying rhythmicity of the robot is autonomous, but a human operator has control over the selection of behaviors and the overall attentional focus of the robot. Allowing a human to provide high-level control of the robot allows us to create more interesting interactions, gives us insight into changes and additional techniques we might apply, and relieves us of developing more difficult aspects of autonomy while allowing us to isolate the aspects of the interaction that involve its rhythmicity.

3.3.2 Facilitating rhythmic synchrony

The development of socially interactive technology calls for an iterative process of interaction and design. Identification of relevant issues, development of approaches, and evaluation of the technological artifacts we design all benefit from repeated feedback from interaction with people in real settings. Therefore, bringing the robot to children and observing real interactions has been an important component of our approach. This required that we design interactive contexts that are natural and intuitive to people while providing for more regularity than open-ended conversational interaction. Our challenge was to facilitate nonverbal rhythmic social interactions while constraining them such that they would be consistent and measurable.

With this in mind, we set out to create playful interactions that are nonverbal (involving physical movement and sound), engaging, compelling, and fun for children. Several factors contributed to the selection of play as a good domain in which to perform this research. Child-robot interaction is already a popular area of research for the purposes of pedagogy, cognitive development, and therapy [Michaud and Theberge-Turmel, 2002; Dautenhahn, 1999]. Play-oriented interactions are a particularly appropriate context since the principles under study (rhythm and synchrony) are magnified or emphasized in playful interactions—physical, repetitive, and exaggerated as they often are. For example, dance-oriented play to the accompaniment of music consists of more structured and regular behavioral rhythms than those observed in social interactions, so it may serve as a more tractable domain for social robots to perceive and exhibit rhythmic behaviors, and as a stepping-stone to more open-ended interactive processes of coordination and synchrony—particularly when these processes have strong periodic elements, such as nodding or turn-taking. The appeal of Keepon's interaction with music is validated by the popularity of several videos on the internet. Video games similarly encourage playful behavior, including regular rhythmic movement, and recent developments in physical gesture-based gaming (exemplified by the Nintendo Wii) provide us with associated input devices, such as the Wii Remote and Wii Balance Board that we have used.

Working with young children allows for more plausible nonverbal interaction, allowing us to avoid the difficulties of autonomous natural language processing. It may also be argued that children are a more natural audience for the artifacts we will create, as they are more interested in these types of toys and perhaps less susceptible to biasing pressures of performing under observation. Finally, our goals of engagement and comfort are preeminently important in play, since it is essential to maintain a child's interest and attention. Therefore, we believe that focusing on play is a good method of isolating the behaviors and phenomena in which we are interested. We selected music-led dancing and a goal-oriented game (based on the pretense of jumping on a trampoline) as two experimental domains that provided structure for joint physical rhythmic behavior by human and robot.

The development of appropriate experimental protocols has been a challenging, but important, part of this work. We began with hand-puppeteered interactions that were recorded and analyzed using traditional social science methods. With a real robot, we then piloted open-ended interactions in order to observe a range of rhythmic interactive styles. Next, we attempted a strict experimental protocol to examine the leader/follower roles that we had observed, which was unfortunately insufficient to encourage natural rhythmic play. We modified our protocol to allow ample facilitation and comfort, while still constraining the context to allow for comparisons between trials. Finally, we developed a rhythmic game that allowed us to measure success in a non-musical task. We believe that this progression may be informative in designing experimental protocols for examining various aspects of non-verbal interaction.

We began this work with the design of a robot named Roillo [Michalowski *et al.*, 2006a]. We followed a sequential design process involving a variety of exploratory activities: rendering, animation, surveys, physical prototyping, and observation and analysis of puppeteered interactions with children. We began with the goal of designing a character with a morphology that would, like Keepon, be minimally representative of its interactive capabilities, both to minimize the cost and complexity of the robot and to reduce the inappropriate attributions that might be made to a more visually elaborate form. In the interest of using a shape that would provide a cue for directionality—missing in Keepon when it is turned away—we selected the Reuleaux tetrahedron for the body and head. Roillo was meant to have an articulated "antenna" on top of its head as a simple means of performing deictic gesturing and expressing affect. Since Roillo was to be simple, inexpensive, and robust to the sort of rough handling that can be expected from children, we avoided designing a complex armature that would be directly driven by motors. Instead, our approach was to use inexpensive servos to pull wires attached to points in a foam body, and to use the springiness of the foam to bring those points back to their starting positions.

In the early stages of our design of Roillo, we performed an exploratory study aimed at gaining a general sense of children's initial impressions of the robot—how children would perceive and react to Roillo's minimal design and to the range of movements the robot would be able to make. Since our mechanical prototype enabled only a limited range of the desired motions, we separately constructed a 1ft tall rubber Roillo puppet (fig. 3.2, right). The puppet sat atop a 2x2ft box that housed the puppeteer. A one-way mirror allowed the puppeteer to see interactors. The week-long study was conducted with 22 children, ages 4-6, from a local kindergarten. The interactions were videotaped and reviewed by the researchers to find themes in children's behaviors and statements about Roillo.

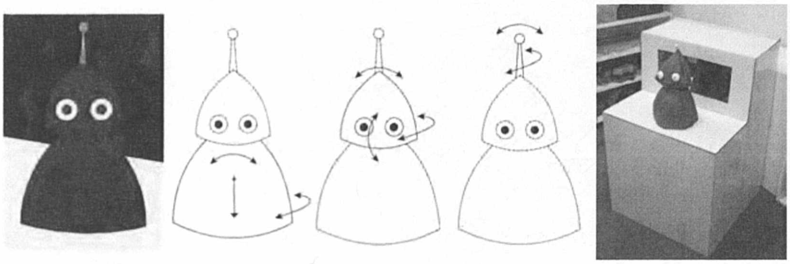


Figure 3.2: Desired degrees of freedom for Roillo's body, head, and antenna (left). Our puppet and puppeteer's box (right).

Along with children's general reactions, we were interested in how the rhythm of Roillo's movements and their contingency to the actions of the children affected the interaction. Our aim was to see when and how children engaged in rhythmic entrainment, turn-taking, and imitation. For one group of children, the puppeteer watched children through a one-way mirror and moved the puppet in a rhythmic manner contingent to the children's location and manner of movement. For the other group of children, the puppeteer was blindfolded, listening to music, and moving the puppet in a dance-like manner without knowing the children's location or behavior. We found that the incidence of children's rhythmic behavior was higher, and their perceptions of the puppet more positive, when it was behaving contingently to the children [Michalowski *et al.*, 2006a]. This suggested that rhythmic awareness and synchrony would be an important component of human-robot interaction and encouraged our further experimentation.

As the next step in our iterative experimental design process, we facilitated open-ended interactions in order to explore styles of play that would be attractive to children. We also wanted to gather data that could, to some extent, be quantified such that it would help us to formulate more concrete experimental hypotheses and methodologies by suggesting what we might find. During our early software development on the robot Keepon, we conducted a pilot study of the robot's interaction with children during an open house (discussed in section 5.1). The robot was set up to synchronize its rhythmic movement to children's movement as visually perceived using optical flow. Under these circumstances, due to imperfect perception, the robot's rhythmic behavior was synchronized to children only occasionally. However, we did not consider the system's poor perception and unpredictable behavior to be a liability, as it resulted in a variety of situations that could be separated and analyzed after the fact. By situating this system in an unconstrained and busy setting (with many people engaged in different activities, so that children would be comfortable exploring Keepon as an exhibit) over many hours of interaction, we gathered a rich collection of video data of different types of interactions. In post-analysis, we observed the effects of the robot's synchrony to music on children's interactive behavior. We noticed some children that attempted to imitate the robot, while other children were more independent in their behavior.

With these observations in mind, as well as the notions of role and leadership in social interactions [Kendon, 1970], we worked to develop a more rigorous experimental method. Through tighter control of the protocol, our goal was to carefully minimize all experiential differences between subjects other than the manipulation. Specifically, we were interested in the different ways that a robot might take an "active" or "passive" role in leading or following the rhythms of an interaction. In order to improve the perception of human movement, we developed an accelerometer-instrumented toy to which the robot could synchronize its movements. We also prepared the robot to be able to dance synchronously to a number of songs, and enabled the teleoperator to switch between having the robot synchronize to the toy or to the music. We attempted to create a controlled scenario in which children experienced exactly the same context for interaction, in a sparse lab setting, with consistent behaviors by the researcher (described in section 5.2). We found, instead, that the robot's novelty, the children's ages, and individual variance in attitudes made it difficult to evoke the desired rhythmic behaviors. We believe that the holistic nature of social interaction requires that we allow each child's interaction to unfold naturally, with appropriate scaffolding from caregivers, if we want to control for a particular variable.

Since the previous protocol inhibited the emergence of comfortable play, we modified our protocol to allow for individual facilitation while still generating quantifiable per-session data that could be directly compared to each other. Our modifications were designed to alleviate the stifling and intimidating environment of our controlled lab study. We also introduced the use of the Wii Balance Board as the sensor for perception of bodily movement. In a series of introductory sessions with the robot in the children's classroom, the children were able to examine, play, and dance with the robot in an exploratory manner. The children were then taken out of the classroom individually to participate in two experiments (described in section 5.3) in which the robot danced in synchrony either to the music or to their movements; in the second experiment, the tempo of the music changed and there was a period in which Keepon deviated from the music. The protocol successfully elicited continuous rhythmic movement from nearly all the children. We found suggestive results that children's activity level was higher when the robot was dancing in synchrony with them, that their synchrony to music was higher when they were "following" keepon, and they tended to follow Keepon when it deviated from the rhythm of the music.

Since music-driven dancing is one limited form of interaction, we wanted to develop an additional interactive context that would incorporate regular rhythmic behavior and would furthermore allow us to quantifiably measure performance in a different way. We believed that a game would offer affordances for play, would allow effective use of our perceptual tools, and would enable us to measure success in a particular task. After considering a number of game scenarios, we chose a virtual trampoline as a good setting for rhythmic bouncing and a cooperative task. This experiment (described in Chapter 6) required synchrony in order to achieve the goal of hitting a target, and there were switching modes of play in which robot and human each took the role of leader or follower according to changes in availability of game state information. We found that performance was improved when people followed the robot, and we identified interesting differences between performance improvements (or "learning curves") after particular role transitions.

3.3.3 Analyzing rhythmic synchrony

3.3.3.1 Observational behavioral analysis

As discussed in Chapter 2, social scientists have discovered and described social rhythmic phenomena, and we have drawn from their ideas in building our robotic system. Similarly, we have used social science techniques of naturalistic observation and fine-grained behavioral analysis (of recorded video) [Michalowski *et al.*, 2006a; Michalowski *et al.*, 2007b] at points in the design of our technologies and interactive scenarios. These techniques provide both qualitative insights into human activities that are otherwise difficult to isolate and quantitative data in which we can identify significant differences between conditions. We initially found failures in rhythmicity in other robots (GRACE and the Roboceptionist) through the use of frame-by-frame coding of position, pose, and interactive behavior of people interacting in naturalistic environments [Sabanovic *et al.*, 2006]. Noldus Observer software [Noldus, 2006] allowed fine-grained, continuous labeling of the robots and participants behaviors as events and states (with durations of time) for temporal analysis. We used this method again, and additionally used a longer-time-scale labeling of social engagement and disengagement, for hundreds of children freely interacting with an early rhythmic system on Keepon [Michalowski *et al.*, 2007b]. We found differences in the elicitation of rhythmic play in children when the robot was in sync with music in the environment (later described in section 5.1). While these methods yield quantifiable data from recordings of natural behavior, they are extremely labor-intensive to perform, and require multiple coders for reliability. Ideally, we would always be able to perform such thorough analyses on our systems, both to evaluate what we have developed and to identify interaction patterns that would guide further development. However, if the time or personnel are not available, we must try to develop perceptual technologies that generate rich quantified measurements and recordings of the interaction, to reduce the burden of manual coding and to change the analysis into a problem of automated data mining.

3.3.3.2 Quantitative analysis

Although we would like to develop automated methods for analyzing these interactions, social scientists have not extensively operationalized the principles of social rhythmicity, so we lack a synthetic prescription for implementing them. We also lack an analytical toolbox for measuring periods of synchrony, degree of synchrony, and other quantifiable manifestations of rhythmic social interaction. In general, our challenge is to evaluate the "goodness" of an interaction according to some measures. There are a number of ways that the quality of an interaction may be quantified. The time that children maintain interest in the interaction, their expressed or vocalized affect or emotions, the duration and quantity of certain behaviors such as smiles, and the overall amount of activity or movement can be quantified. Recordings of the robot's sensor readings and motor commands are available as data that can be related to the manually coded aspects of rhythmic synchrony; these can also be used to compare the robot's perception of the rhythmic events in its environment with the evaluations of a trained human observer. As discussed above, a robot must take into account multiple streams of rhythmic data from various sensory modalities, as well as the rhythmic behavior of the robot itself, and compare these rhythms to evaluate the past and to shape the future. The same sorts of techniques could be used on recorded data in order to analyze a particular interaction. In fact, such tools would be useful even in analyzing human-human interaction. We have worked to build various analytical tools for determining frequency and amplitude of movement, periods of synchrony while taking harmonics into account, and the effects of role transitions on changes in rhythmic movement. By running these analyses on recorded sensor data from experimental sessions, we generate statistics on overall performance (that can be compared between subjects) or changes in performance in response to changing conditions (that can be compared within subjects). While we have used (and in many cases developed) these analyses after the fact, they can be adapted for use in the course of real-time interactions, in order for the robot to understand the relationship between rhythms (in the "rhythmic comparison" component of fig. 3.1).

3.4 Summary

We have outlined a framework for rhythmic intelligence, consisting of perceptual, cognitive, and behavioral technologies organized by a set of components that would enable increasingly sophisticated participation in rhythmically coordinated interactions. The "rhythmic attention," "entrainment," and "interaction state" components illustrated in figure 3.1 are roughly analogous to the elements of interpersonal coordination outlined in section 2.1: interaction rhythm, simultaneous movement, and smooth meshing. "Rhythmic attention" accounts for the presumption that perceived social cues will have a rhythmic quality, and chooses between various modalities; "entrainment" compares and matches rhythms so as to achieve simultaneous, coordinated movements; and "interaction state" recognizes or uses the rhythmic properties of roles or transitions to track and guide the trajectory of an interaction. Although we envision this framework to be generally applicable to creating rhythmic intelligence in unconstrained social interactions, in this thesis we studied simplified aspects of these components while adapting or developing technologies for the perception, representation, and behavior of social rhythms in a robot. Our experimental contexts, as well, are simplified in the form of dance- or game-oriented play in order to constrain the setting and resulting data such that our problem is tractable.

While our simplifying assumptions and constraints restrain us from generalizing our conclusions to unconstrained interaction, we believe that our approach is useful and worthwhile. These simplified interactions allow us to use inexpensive and widely available perceptual technologies, such as the accelerometers and pressure sensors we have used, and to develop techniques for processing this sensor data that might be expanded or refined to handle more unconstrained input. We can also find limitations (or advantages) in the influence that a small robot, with limited degrees of freedom, may have on human behavior in nonverbal interactions.

In the following chapter, we describe the perceptual, cognitive, and behavioral technology we have developed for use in our experiments.

Chapter 4

Technology Development

We have developed a general-purpose architecture for enabling and studying rhythmic human-robot social interaction, including behavior, perception, entrainment, analysis, and a context for facilitating measurable interactions. The technology we have developed addresses each of these components in the context of simplified and constrained rhythmic interactions. Our technology allows the robot Keepon to perceive rhythms in a variety of modalities and to synchronize periodic dance-like movements to these rhythms. Our experimental protocols allow for observation of these rhythmic interactions, and we can analyze the resulting recorded data to understand the properties of the synchronicity of the interaction.

We first describe our hardware platform and the software for driving rhythmic behavior. Next, we describe the perceptual hardware and software we have used to detect rhythms in human behavior. Finally, we describe the methods we have used (and attempted to use) to establish entrainment and to detect synchrony.

4.1 Rhythmic behavior

Our interactive robotic system is a combination of hardware and software that creates a coherent and clear impression of coordinated movement and sound.

4.1.1 Hardware

We selected the creature-like robot Keepon (fig. 4.1), designed in 2003 by Dr. Hideki Kozima [Kozima and Nakagawa, 2006], as our hardware platform. Keepon was designed to perform emotional and attentional communication with human interactors, especially children. Keepon has a yellow snowman-like body composed of two vertically arranged

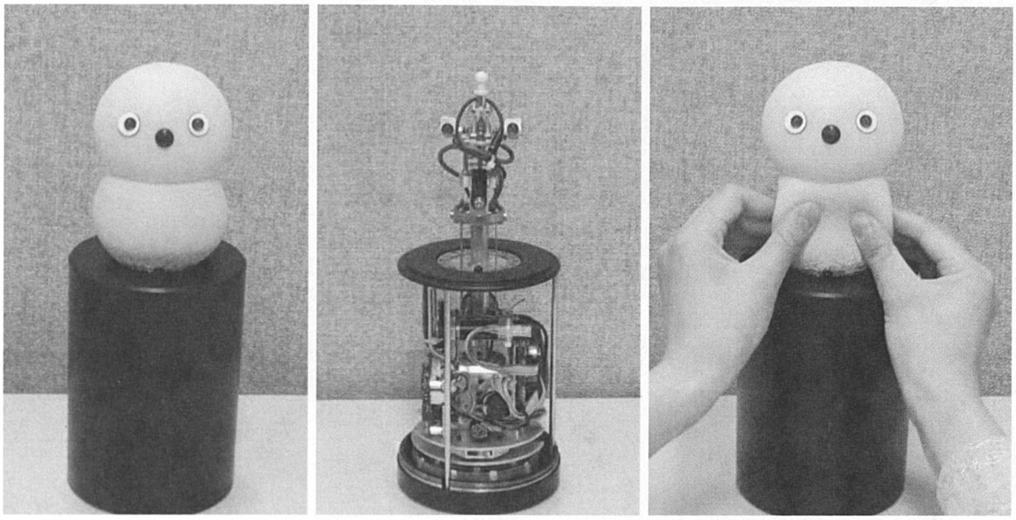


Figure 4.1: Keepon's external and internal structure and its deformable body.

spheres with a height of 120mm. The upper part (the "head") has two eyes (each of which is a 120° wide-angle color CCD camera) and a nose (which is a microphone). The lower part (the "belly") contains a small gimbal and four wires by which the body is manipulated like a marionette. Keepon sits atop a black cylinder that contains four motors and two circuit boards (a PID controller and a motor driver). Since Keepon's body is made of silicone rubber and is relatively hollow, the head and belly deform whenever Keepon changes posture or when someone touches it. Keepon's skin is flecked with a "fuzzy" texture that is appealing to the touch; indeed, the robot is designed to be handled by children and to withstand rough contact. The skin protects Keepon's mechanism from human contact just as it protects human skin from the mechanism.

The body has four degrees of freedom (DOFs, fig. 4.2): pitch (nodding, leaning, or tilting forward and back) ±40°, yaw (panning, "shaking" the head, or turning left and right) ±180°, roll (rocking or leaning side-to-side) ±25°, and compression (bobbing or shrinking vertically) with a 15mm stroke. For each DOF, the PID controller can be given a command position as well as optional parameters for maximum velocity and acceleration. Given a position command, the controller generates a trapezoidal velocity profile that smoothly accelerates the motor at the desired rate until it reaches maximum desired velocity, and decelerates the motor at the same rate prior to stopping at the desired position.

These four degrees of freedom were designed to produce two qualitatively different types of actions:

- **Attentive action:** Keepon orients toward a certain target in the environment

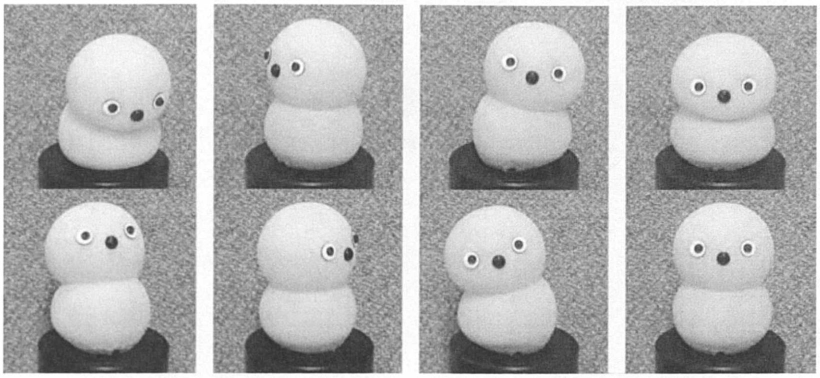


Figure 4.2: Keepon's four degrees of freedom (pitch, yaw, roll, and compression).

by directing its head (i.e. gaze) up/down and left/right. This action enables eye contact and joint attention.

- **Emotive action:** Keepon rocks from left to right and/or bobs its body up and down while keeping its attention fixed on a target. This gives the impression of expressing Keepon's internal rather than perceptual state. These behaviors suggest emotions such as pleasure and excitement about the target of Keepon's attention. These behaviors are accentuated by short "popping" sounds appropriate to the movement from a speaker below the robot's body.

With these two actions, Keepon can express *what* it perceives and *how* it evaluates the target. These communicative functions of Keepon's actions can easily be understood by human interactants, even babies and toddlers. Research with Keepon [Kozima *et al.*, 2009] can be summarized to have resulted in the following findings:

- Age, experience, and developmental level have effects on children's conception of Keepon as moving object, lifelike creature, or social agent.
- Group effects and nurturing motivations help to shape individual children's interactions with the robot.
- Keepon can serve as a social "pivot" for facilitating interactions between children and peers or caregivers.
- Keepon's simple appearance and behavior promote comfort and understanding in children, particularly those with developmental disorders such as autism.
- The perceptual and behavioral technology embodied in Keepon makes it suitable for mediating and recording interactions between a "puppeteering" therapist and the children in whose playroom Keepon is situated.

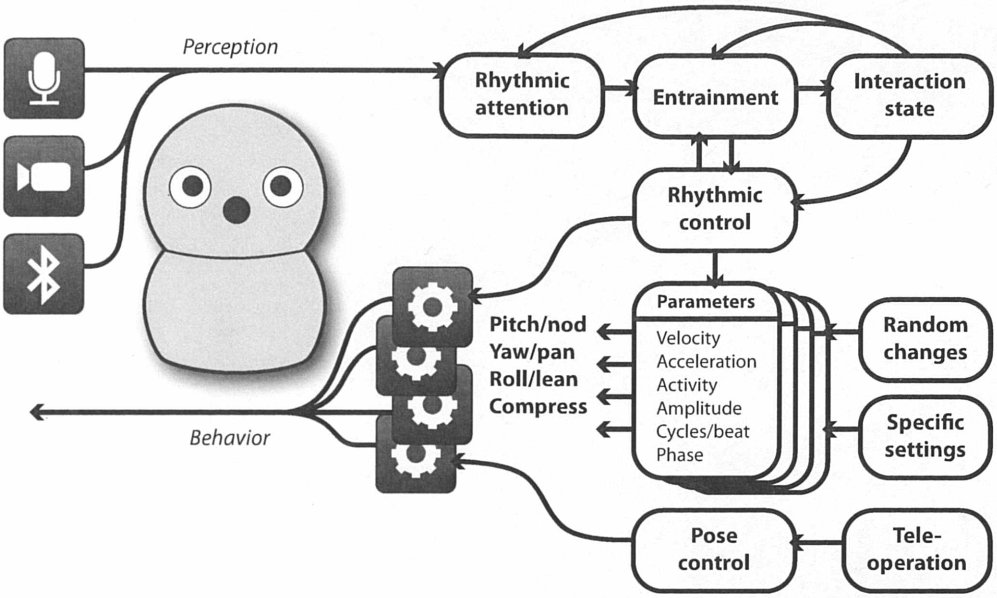
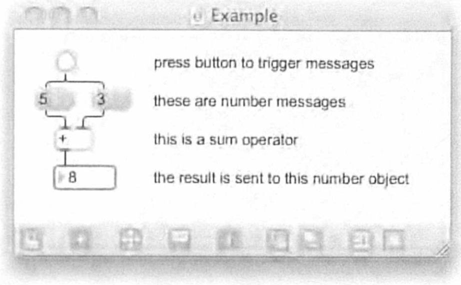


Figure 4.3: The control architecture for rhythmic movement and partial teleoperation.

This thesis research introduces a third type of action (to the attentive and emotive actions just described):

- **Rhythmic action:** Keepon's body moves in an oscillatory manner. As we describe below, each DOF is independently controllable and can be parameterized in a variety of ways. When the DOFs perform repetitive, back-and-forth motions in a coordinated way—in other words, when they are internally synchronized to a common rhythm—the robot's skin integrates these separate movements into full-body movements that might be perceived as a variety of different "styles," none of which were programmed beforehand.

Keepon is connected to a PC (in our case, a MacBook Pro) via a USB-serial adapter. On board the robot, the RS-232 signal is converted to RS-485 for communication with firmware on the robot's 32-bit SH-2 microcontroller. Keepon's NTSC camera feed is digitized from S-Video to a IEEE 1394 signal using a converter box. The microphone has a standard 3.5mm TRS connector but was not normally used in our research, as music was played from the controlling PC and was therefore directly accessible.



press button to trigger messages

these are number messages

this is a sum operator

the result is sent to this number object

Figure 4.4: A sample program ("patcher") in Max/MSP that adds two numbers.

4.1.2 Software

Our architecture for rhythmic perception and control is illustrated in figure 4.3. It consists of various perceptual inputs from which rhythms may be detected and extracted, a system for selecting a rhythm to which to entrain Keepon's movement, a controller for each DOF, a set of parameters for each DOF, and optional input from a teleoperator.

Our high-level control software is written in Max/MSP, a visual programming language originally designed for music composition, digital instrument control, and multimedia performance [Cycling '74, 2006]. It is highly modular, consisting of "objects" (which act like self-contained programs, but which are in fact dynamically linked libraries) with inputs and outputs graphically arranged and connected on a "patcher" canvas. Messages or data are processed as they "flow" through the objects in the patcher. For an example, see figure 4.4. Max/MSP is suited to our purposes due to its native support and tools for musical constructs, timing (e.g. clocks and metronomes), signal processing, and graphical user interfaces. It is also flexible enough that we can build our own objects (e.g. in C/C++) for arbitrary message processing, interfacing with hardware, and linking to external libraries (e.g. OpenCV).

Each of Keepon's four DOFs has its own control module, made up of a cluster of Max/MSP objects. Upon receiving direction either from autonomous processes or a human teleoperator, these control modules generate commands to set position, velocity, and/or acceleration for the appropriate motor. The commands may be independent, isolated events (e.g. saccade to gaze in a particular direction) or repetitive and oscillatory (e.g. dancing control). The control modules pass these commands on to a serial control interface, also written in Max/MSP, which formats them into the appropriate 6-byte packets and sends them to the robot's serial port (and ultimately to the robot's on-board PID controller).

Each module remembers, as "state," the current position of its DOF (for example, if the robot is oriented in a particular direction). In addition to maximum velocity and acceleration, a number of dance-specific parameters may be commanded or randomly selected for each DOF: whether it is currently active, the "amplitude" of movement, cycles per beat (half a cycle or a full cycle), and phase (e.g. turning right then left, or left then right)[1]. The amplitude parameter is used to set the positional range of motion to be covered by that DOFs dancing behavior. That is, the repetitive movements are specified to cover an angle (for the pitch, yaw, and roll DOFs) or a distance (for the compression DOF). The control modules correspondingly set their DOFs' speed and velocity to ensure that the range of motion can be covered in the available time. That is, when the range of motion is small, Keepon moves slowly, and when the range is larger, Keepon moves more quickly. The mapping between range and velocity/acceleration was determined empirically.

For autonomous dancing behavior, Keepon is controlled by a metronome that drives the robot's periodic movement with a stream of "beats." This metronome is built with Max/MSP's `metro` object, which generates a train of pulses separated by a specified time interval. The tempo of the metronome can be set in various ways (described below). Each beat is a message that is passed to the four control modules. Upon receiving a beat, a control module uses its DOF's current position and parameter settings to generate a command to move in a particular direction towards a limit of the desired range of motion. On the next beat, another command causes the DOF to move in the opposite direction.[2] In this way, the robot performs oscillatory movements while maintaining an average pose around the DOF's original position.

The wide range of possible parameter settings makes possible a nearly limitless variety of dance styles, even on such a simple robot with only four DOFs. In some of our interactions, we have randomly selected and modified these parameters at random time intervals in order to keep the dancing interesting. In other interactions, we have constrained the parameters to particular settings to fit the context or to match the stimuli (e.g. musical, or physical human movement) in an appropriate way. In practice, we have identified two interesting physical properties of the hardware that impact the dancing behavior:

- Dorsoventral (front-back) and anteroposterior (up-down) movements are more attractive at high frequencies, and with a full cycle per beat, than are lateral (side-to-side) movements. Therefore, we typically constrain the pitch and compression

---

[1] This makes a difference when multiple DOFs are composed together: for instance, the movement *TurnLeft/RollLeft* → *TurnRight/RollRight* looks different than the movement *TurnLeft/RollRight* → *TurnRight/RollLeft*.

[2] In the case that the DOF should move a full cycle per beat of the metronome, the module must send the return command in between beats. For this purpose, the DOFs internally maintain timers to track the period between the previous two beats. By using half of this period, the modules can generate oscillatory commands at effectively twice the frequency of the metronome.

DOFs to a full cycle per beat, and the yaw and roll DOFs to a half cycle per beat. Incidentally, we believe that this may also hold for human dancing; it seems more natural to nod or bounce with a full cycle per beat, while side-to-side movements are more comfortable with one movement per beat.

- The compression DOF is most salient and responds nearly instantaneously to commands. As we shall discuss, this supports the emphasis of the compression DOF when Keepon's dancing is reactive to the environment.

For the purposes of natural interaction, we have developed a system of *partial teleoperation* whereby a remote human teleoperator (i.e., the experimenter, or "puppeteer") controls the overall direction of attention of the robot. An interface built in Max/MSP allows the teleoperator to see the view from Keepon's camera eyes and to use a virtual joystick-like interface to control the robot's pose (pan and tilt). Alternatively, the teleoperator can use buttons on an interface device such as the Nintendo Wii Remote to control the robot's attention. In order that the teleoperator does not exert undue influence on the social interaction, it is important to have a set of rules, or a script—namely, in our case, that the robot should always attend to the child's face, or to a toy, or to some other relevant location. In this way, the teleoperator can be regarded as an intelligent sensor rather than a source of social cues. The robot's periodic rhythmic movements, controlled autonomously according to the particular interactive or experimental context, are then centered around the robot's direction of attention (fig. 4.3). While it would be possible to run the robot completely autonomously using, for example, face recognition, having a human provide more reliable sensory information can create more compelling social interactions while allowing us to focus on the relationship between the rhythmic behaviors of robot and interactor. Initially, Keepon was able to perform periodic movements in all four DOFs, with the yaw and pitch movements centered around the robot's direction of attention. We observed some confusion regarding the robot's attention, so we usually divided the roles of the DOFs (using yaw and pitch for attention, and compression and roll for rhythmic movement).

4.2 Rhythmic perception

In order to perceive rhythms in multiple modalities, we have developed methods of extracting rhythmic information from both sound and movement.

4.2.1 Audio

For simple audio, we use amplitude peak detection (using Max/MSP's `peakamp~` object) in the audio signal to identify clear beats, such as clapping or drumbeats. For isolating rhythms in more complex music, we have integrated a more sophisticated implementation of beat detection [Jehan, 2004] based on an algorithm of [Scheirer, 1998]. The signal is divided by a number of bandpass filters, and each channel is then processed by a bank of tuned comb filter resonators. One of these resonators phase-locks to the periodicity of the band's envelope, and the peaks across the bands are used to determine the dominant tempo. However, algorithms like this one can be confused by complex rhythms and lack the high-level knowledge to label beats in the same way as human listeners. Therefore, in order to improve reliability in experimental scenarios where consistency is important, we have usually hand-labeled the beats in songs used as environmental rhythmic stimuli. The experimenter listens to the song (usually regular and simple children's music) and presses a button for each beat. This manually created beat track (recorded using Max/MSP's `mtr` object, which records and time-codes arbitrary streams of messages) is aligned to the song for later playback during experiments, when it is used as the source of beats.

4.2.2 Vision

For visual movement, we use simple methods of detecting periodic movement through intensity variation and optical flow. We integrated the OpenCV library [Intel, 2006] into the Max/MSP environment by creating an object that accepted a video stream, ran specified OpenCV functions, and output the relevant results. This enables us to apply operators to an incoming video stream and identify frames of interest. For example, for each frame, we can compute average pixel intensity in the grayscale image. By maintaining a running average of this value, we can detect dramatic changes in the scene, that is, when the intensity moves above (or below) a threshold, or when the derivative of the intensity changes sign. Alternatively, we can detect changes in the average direction of movement (with hysteresis) in the image (e.g. by using the Lucas-Kanade optical flow method or the KLT feature tracker). By treating these events as "beats" and identifying consistent spacing (e.g. by looking for a sequence of three recent beats in which the two intervals are similar), we can identify the frequency of periodic movement in the environment. We have informally experimented with such methods in order to provide rhythmic direction for Keepon's dancing. When the background is stationary, and the perceived human movement is clear, regular, and exaggerated, it is easy to detect movement. Unfortunately, we must use a stationary external camera in these situations, as the robot's self-movement is difficult to isolate from the overall movement in the video stream. Furthermore, these familiar constraints in computer vision for perception of human appearance and behavior

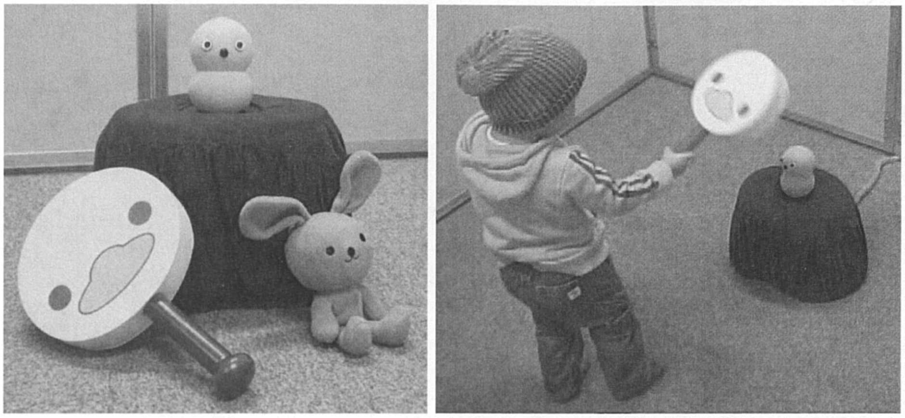


Figure 4.5: Keepon and the two toys used with the accelerometer: a stuffed rabbit (169g, 290mm long) and a soft paddle (137g, 320mm long). Keepon dancing with a child holding the paddle.

are not conducive to the dynamic and child-directed interactions that we are studying; our attempt at using vision-based rhythm detection in section 5.1 was not successful at generating synchronization to children's dancing.

4.2.3 Accelerometers

To overcome the difficulties of sensing human movement through vision, we developed a system for detecting rhythmic movement through the use of battery-powered three-axis accelerometers, with wireless Bluetooth data transfer, implanted in toys (fig. 4.5). We expected the use of a toy to carry a number of benefits:

- It might encourage more exaggerated, regular, repetitive movement, as seen in children's play with dolls (e.g. walking, talking);
- It might allow children to use their imaginations to create narratives involving the toy and the robot as characters; and
- The toy, being of similar size and form to the robot, might be seen as the robot's "peer" through which the children might develop a different type of relationship with the robot; for example, one between equals, rather than between a baby and a (larger) caregiver.

The accelerometer provides force data for three axes of movement. The magnitude of the overall acceleration is the Euclidian norm of the vector defined by these three values. We detect rhythmic movements of the toy by finding peaks in this magnitude. These peaks occur when the sensor is undergoing maximal acceleration during a change in direction.

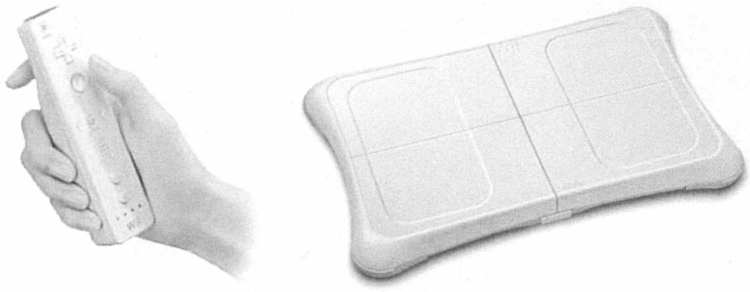


Figure 4.6: The Nintendo Wii Remote (left), used for attention/emotion control, and Wii Balance Board (right), used for rhythmic perception.

Since the sensor data is rather noisy, the peaks are found from zero-crossings in lowpass-filtered data. We can then treat these direction change points as "beats" in the same way as musical beats or visual movement direction changes were described above.

We have used the Nintendo Wii Remote, or "WiiMote" (fig. 4.6, left), as an interface for teleoperating the robot as it is portable, battery powered, communicates with a PC using Bluetooth, and has buttons that can be mapped to various behaviors. The WiiMote includes a similar three-axis accelerometer to that which we have used in our toys, and it can be used interchangeably with them. We have used it in informal dance interactions with children during demonstrations, and future work may specifically incorporate the gameplay affordances and functionality provided by the WiiMote.

Unfortunately, the accelerometer in these devices can only detect translational movement. We have observed children rotating the toys in a rhythmic manner (section 5.2), but we could not perceive this movement correctly. This can be addressed in the future with the addition of an angular rate sensor (such as that included in newer versions of Wii hardware).

Our method also has difficulty detecting beats in certain oscillatory movements (e.g. slow, smooth movements or circular trajectories made parallel to the floor) where the force vector remains roughly constant. In practice, however, movements are often jerky, and gravity provides a bias (e.g. in circular movements perpendicular to the floor) that allows us to detect beats.

4.2.4 Pressure sensors

Our final mechanism for perceiving rhythmic bodily movement, and the one used in most of our later experiments, involves pressure sensors on the floor. The Nintendo Wii Balance Board, or "WiiFit" (fig. 4.6 right), contains four pressure sensors in its corners and communicates the sensor readings over a Bluetooth connection. By interpolating

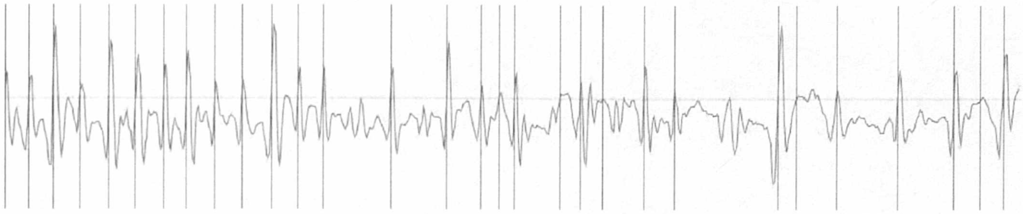


Figure 4.7: An illustration of typical data from the WiiFit board from the child's data in fig. 5.12, with detected bounces labeled with vertical lines. This segment is 20 seconds long and shows the transition from the first fast section of the song to the first slow section of the song.

between the four values, we obtain the 2D position of a person's center of gravity and their overall weight on the board. We normalize the data by zeroing it around the mean weight over the previous one or two seconds, and smooth it by filtering the value logarithmically between changes. "Beats" in the children's movements as they bounce up and down are detected as increasing zero-crossings in this normalized, smoothed weight data (as depicted in fig. 4.7), or as direction changes in the 2D position of the child's center of gravity (to detect swaying or foot-to-foot motion). When the current weight exceeds 1.05 times the mean weight, it is recognized as a bounce, and a subsequent bounce can only be detected after the weight decreases through 0.95 times the mean weight. The bounce is thus recognized as the person begins to apply downward force and before they reach their lowest (or heaviest) point.

The advantage of this perception system is that it is familiar to most children (and adults), it is non-intrusive, and it does not require the interactant to hold or manipulate any objects. It also captures gross (vertical) physical movement, well-suited to dancing behavior. Finally, it is natural to interpret the 2D position of a person's center of gravity as lateral or dorsoventral leaning, which is easily mapped to Keepon's own ability to lean in these same two directions and affords the possibility for simple posture mirroring or dance style imitation.

4.3 Rhythmic synchrony

Given a robot's ability to behave rhythmically, and to perceive rhythms in environmental stimuli or human behavior, we must create techniques by which it can establish synchrony (entrainment) and by which we can evaluate whether synchrony has occurred (analysis).

4.3.1 Entrainment

Entrainment is the process by which one rhythm, through coupling or interaction, is brought into synchrony with another rhythm. Our challenge, after perceiving beats from one of the sensory modalities we have discussed, is to direct Keepon's rhythmic behavior to gradually or immediately match its rhythmic behavior to the perceived rhythm. We have developed several methods for processing perceived beats and generating Keepon's command beats accordingly. A simple method is for Keepon's beats to be purely reactive; that is, a command beat to the robot is caused by a detected beat in sensory stimuli. From a communications standpoint, we can assume that a command is sent by the PC and processed by the robot's microcontroller nearly instantaneously (the delay is, in fact, approximately 2.5ms). However, when using the slower DOFs (yaw, pitch, roll), the delay between the motor's initial movement and the point at which a commanded limit of movement is reached makes a purely reactive system insufficient for synchrony. For example, using "natural" acceleration and velocity parameters, the time to completion of motion for these DOFs may be 300-1000ms. On the other hand, the compression DOF is completed extremely quickly (on the order of 100ms). We were able to reduce any perceptible delay even further since, in many of our experiments, we are able to report beats slightly before they actually occur—for instance, we slightly shift the music's labeled beat-track or, in the case of WiiMote/WiiFit perception, the software reports a beat as soon as the acceleration or weight exceeds a threshold (rather than waiting until it hits its peak). In these situations, by constraining the robot's dance style to accentuate the fast compression DOF, we can achieve extremely low-latency (<100ms) reactive entrainment.[3] The remaining DOFs are used to create interesting dance behaviors, but since the endpoints of their trajectories are smooth, they are dominated by the more abrupt and salient direction-changes in the compression DOF. We believe that such a reactive system can be sufficient, in practice, for achieving the appearance of synchrony if the relevant DOFs are highly responsive.

Another method evaluates a recent history of perceived beats in order to determine a "tempo." In this method (using, for example, the Max/MSP `sync~` object), the most recent three beats are evaluated for consistent spacing. If the first time interval (between the first and second beat) is within range of the second time interval (between the second and third beat), the average of the two intervals gives us the new tempo. If the second interval is different, the third beat is ignored unless the time to the following beat is consistent. The tempo, whose frequency is given in beats per minute (BPM) and is trivially converted to the time interval between beats, is used to set the metronome (discussed in

---

[3] For comparison, human reaction times are much slower: individual neuronal latencies can be at least 100ms, so reaction time can be several times this duration (depending on the sensory-motor processing pathways involved) [DiCarlo and Maunsell, 2005].

section 4.1.2) that generates a stream of behavioral beats separated by the desired time interval. This method allows Keepon to maintain a rhythm when perceptual inaccuracy misses beats, or to ignore irregularly spaced spurious beats. By increasing the window of beats considered for tempo estimation, we can also smooth out temporal inaccuracy in beat perception. Alternatively, if we are representing and tracking a tempo rather than individual beats, it allows the robot to selectively, or gradually, entrain its rhythm to perceived beats. We have used this adaptive method in some of our studies (e.g. sections 5.1 and 5.2), but there is a tradeoff between smoothness and responsiveness to changing rhythms. Given the unpredictability of human movement and the differences between various interactive situations, it is difficult to establish a general rule for how quickly the robot's tempo should adjust to changing rhythms. Since the physical human movements we have observed (and encouraged through our interaction design) occur in the range of .5Hz to 4Hz, an enlarged window would include at least 3 seconds of history. In practice, however, we found that people often change tempo abruptly (e.g. doubling their frequency), making a larger window insufficiently responsive to these changes. Furthermore, this method does not necessarily result in synchrony in phase, so while the tempo might be matched, subsequent beats may not co-occur with perceived beats, which is an important property of entrainment.

A third method of entrainment that we developed is based on the theory of coupled oscillators discussed in Chapter 2. In this approach, Keepon has an intrinsic rhythm determined by an oscillator whose phase point rotates around a circular limit cycle at some preset angular velocity. A perceived beat is treated as a force that perturbs the oscillator in a given direction. In one half of the cycle, this force moves the phase point forward; in the other half of the cycle, this force moves the phase point backward. In this way, the robot's intrinsic oscillator eventually synchronizes to the perceived beats through repeated phase adjustment. One limitation of this mechanism is the strict boundary conditions required for entrainment to occur; the frequency of the autonomous oscillator must initially already be rather close to that of the external force. We have addressed this problem by allowing the robot's intrinsic frequency (the phase point's angular velocity) to adjust according to the determined tempo of a perceived rhythm; however, we face the same problem as in the previous tempo-based system of selecting an appropriate window to match. Another limitation is that it takes time for the driven oscillator to entrain to the external rhythm. In our experience, for the types of behaviors involved in our dance- and game-based interactions, this delay (on the order of several seconds or phase cycles) made the robot even less responsive than in our tempo-based method. Although natural phenomena such as the blinking of fireflies, or even the clapping of crowds of people, exhibit a gradual entrainment that can be modeled by coupled oscillators, the rhythms involved in our social tasks of dancing and game-playing seemed to require a much quicker and more responsive type of synchronization. It is unclear to us (and it remains unanswered in the social science literature) whether and which types of social synchrony might be modeled in this manner.

To summarize, we have developed a reactive method, a tempo-tracking method, and an oscillator-coupling method to allow the robot to entrain to perceived rhythms. The reactive method requires, from a hardware and control perspective, a highly responsive physical behavior in order to create the appearance of simultaneous movement. Since we found that our robot was capable in this respect, we used this method in order to be able to rapidly entrain to abrupt changes in rhythm. Our tempo-tracking method and oscillator-coupling method are both designed to allow anticipation of future events by assuming rhythmic continuity, and they allow smoothing of erroneous data. However, they suffer from the difficulty of selecting an appropriate time window for considering historical data, and they are slow in adapting to changes in perceived rhythms. The oscillator-coupling method is physically, mathematically, and biologically motivated, and it has the advantage over the tempo-tracking method that it can intrinsically account for the amplitude of rhythmic perception and behavior (which we are not currently taking into account). It would furthermore be representationally satisfying for the robot to have its own intrinsic rhythmicity that is influenced by, and yet has some independence from, the environment. Although we have not used this method in our studies, we believe that with appropriate high-level and task-specific modeling of possible changes in rhythm, it could be used in technology for interactional synchrony.

4.3.2 Analysis

We have discussed how, on one level, the robot analyzes sensory input in order to perceive rhythms and to synchronize its behavior to them. On another level, we as researchers must analyze the resulting social interactions in order to understand whether and how synchrony is established, and what effects this synchrony has on other aspects of the interaction. Our analysis has involved both human coding of video data and machine processing of sensor data. We now present examples of the analytical methods we have used.

While performing research on a series of interactive robots—the Roboceptionist [Gockley *et al.*, 2005], GRACE [Michalowski *et al.*, 2007a], and a hand-puppet of Roillo [Michalowski *et al.*, 2006a]—we used traditional social science tools such as video coding, interview transcription, and textual analysis of videotaped interactions [Sabanovic *et al.*, 2006]. We have also used some of these methods in our work on rhythmic synchrony. For our initial pilot studies with Keepon and children dancing together, we videotaped several hours of interaction with hundreds of children [Michalowski *et al.*, 2007b] and used two methods in coding it. Initially, we employed a per-child coding scheme in which the coder simply watched the video at full-speed and paused it when a new child approached the robot. The coder then created a data entry for the child and made binary notations regarding conditions at the start of the interaction and, upon careful observation of the interaction, the occurrence of subsequent events. We then had another coder analyze the video using Noldus Observer software [Noldus, 2006], which allows fine-grained, continuous, time-based labeling of the robot's and participants' behaviors for statistical and temporal analysis. The software allows us to define a coding scheme consisting of events of interest (involving a discrete timestamp and specification of one or more interactors) as well as states (which additionally have duration). The coder then advances through the video frame-by-frame, selecting codes for micro-behaviors such as utterance, spatial movement, gesture, and gaze as performed by the interactors (people or robots). The resulting data is subjected to statistical analysis to describe the frequency of various types of events, and lag-sequential analyses are used to identify temporal relationships between antecedent and subsequent events. The results of these analyses are discussed in section 5.1.

In prior studies with the Roboceptionist [Sabanovic *et al.*, 2006], we aligned and merged the coded data from Noldus Observer with logged data from the robot's sensors and internal state, so that we could compare the robot's evaluation of what was happening in the world (during the interaction) with our own (after the fact). After our pilot studies with Keepon, we unfortunately lacked the human resources to perform further intensive video analysis of recorded interactions. However, we developed methods for logging Keepon's perception and behavior for later automated analysis. Our focus was on automating analysis of the rhythmic aspects of this recorded data, but future work can integrate these results with human-coded video analysis in order to draw conclusions about the relationship between rhythm and other aspects of the interaction. Using the sensors and techniques introduced in this work, we can assist human coders by automatically recognizing the rhythmic (and, by definition, repetitive) aspects of behavior so that they may focus on emotional expressions, vocalizations, and other social behaviors that are as yet very difficult to classify automatically.

In order to record various data during the robot's interactions, our system includes a multi-track sequencer that accepts aligned streams of beats, sensor data, and motor commands (from music, accelerometers, pressure sensors, and the robot) for later playback. This sequencer uses the Max/MSP `mtr` object, which records multiple tracks of time-coded messages, can save the entire collection to a file, and can play back the tracks of messages. These messages can be low-frequency (e.g. state changes) or high-frequency (e.g. raw sensor data). In analyzing this data, we can find co-occurrence of events, differences in activity between different conditions, relationships between tempos, or the degree of synchrony. For example, at any point in the data, we can start a timer on one event and determine the duration of time until another event; we can analyze the variation in accelerometer or pressure data to estimate overall activity level; or we can compare the tempo of the music to the tempo of the child's movement to determine whether one is faster, slower, or a multiple of the other.

Our estimation of a person's activity level is dependent on the sensory modality used. For example, on the WiiFit board, we obtain the norm of the 3-vector [*positionX*, *positionY*, *normalizedWeight*] from the sensor or from a recording of the data. This maps the 3D movement of the person into a 1D stream of movement data (at approximately 20Hz). The overall activity level is defined as the standard deviation of the movement data over the previous five seconds. It is higher when the person is engaged in higher-amplitude movements up-and-down or side-to-side on the board. We have used a similar technique in measuring overall activity level in the movement of the WiiMote. However, in the case of the WiiMote, we can (to date) only measure magnitude of accelerations (or the abruptness of changes in direction) rather than the magnitude of linear movement.

To measure the frequency of people's periodic movement from the sensor data—in order to compare it to the rhythm of music or the robot—we have experimented with a discrete Fourier transform (frequency domain representation) of our typical beat information (as obtained from the WiiFit board, but it would be equally applicable to data from the WiiMote). Unfortunately, we found in practice that this signal-processing technique was not applicable to the time scales we are concerned with. The DFT is appropriate when the frequencies to be examined are a small fraction of the sampling rate. The fundamental frequency of a DFT is the sampling rate divided by the number of samples per analyzed time slice. In our case, the frequencies we are concerned with (full-body periodic movement) can range from .5Hz to 4Hz. Since our sampling rate for WiiFit data is 20Hz, it follows that a time slice must use at least 40 samples (2 seconds of data) in order to show response in this frequency range. However, given the discretization of the resulting transform, this does not give us sufficiently high resolution to accurately pinpoint the recent dominant frequency in our range of interest. Furthermore, the long time window cannot adjust quickly enough to rapid changes in frequency, which often occur abruptly.

To address this problem, we have developed a mechanism that scans over the data and compares the current beat interval (the time elapsed between the current beat and the previous beat) to the last beat interval (the time elapsed between the previous two beats). If the current interval is greater than half and less than twice the previous interval, the current frequency is calculated from the mean of the two intervals. This policy has a number of good properties:

- It ensures minor smoothing (averaging over two recent intervals) of relatively constant tempos

- It ensures a rapid response (immediately responding) to minor changes in frequency
- It allows a rapid response (within two beats) and an immediate jump to significant changes in frequency (greater than twice, or less than half)
- It excludes spurious readings if they occur in isolation
- It gives a frequency reading of 0 within twice the previous interval if the beat stops

When comparing multiple rhythms (e.g., the robot's and the child's, or the child's and the music's), we can account for harmonics (integral multiples of frequencies) by considering half or twice a given frequency, plus or minus a buffer area, to be a matched frequency. In this way, we can evaluate periods of synchrony or asynchrony between two rhythms. Although we have used this method only in post-facto data analysis, it may also be applied to real-time behavioral control in the future. For example, the robot could compare the rhythms of the music and the child in the course of the interaction to perform behaviors intended to speed up or slow down the child's behavior. Additionally, this method of adjusting to changing rhythms may be incorporated into the tempo-tracking or oscillator-coupling methods described above.

Finally, in the game interactions described in Chapter 6, the interaction design itself incorporates a task whose performance can be explicitly measured (as the target is either struck or missed based on the performance of the participant). As we shall discuss, we can compare overall performance in different phases of the game, and we can evaluate the change in performance over time within and between these phases.

4.4 Summary

We have developed hardware and software for exhibiting rhythmic behavior on the robot Keepon, enabling it to generate oscillatory movements according to a given rhythm. We have adapted hardware and created software for perceiving human rhythmic behavior, including audio, video, accelerometers, and pressure sensors. And we have used various techniques for entraining to rhythms in real time on the robot, as well as for analyzing rhythms and other behaviors in our recorded video and sensor data from experiments.

In the following chapter, we discuss a series of experiments focused on dance-based interaction with musical rhythms.

Chapter 5

Musical Dance Interaction

Dance, as a human art form, can be seen as a representation of the *imitation* and *pattern*[1] that we find pleasurable in social interactions. Indeed, anthropology and social science literature seem to support this view, with the observation that analysis of everyday social interactions often reveals that comfort and rapport in an interaction depends on a subtle and fine-grained coordination [LaFrance, 1979]. Just as music has been proposed to have an actual (and not just metaphorical) relationship with vocal communication [Trevarthen and Malloch, 2000], we believe that the imitation and pattern found in dance are not unrelated to the "dance" of coordinated sound and movement between interactors [Hall, 1983].

What we call dancing, usually to the accompaniment of music, is a constrained form of social interaction that in general consists of more structured, regular, and exaggerated behavioral rhythms than those observed in conversational interaction. These structures and constraints can be useful in developing tools and processes to allow social robots to participate in interactional synchrony. For example, just as any social interaction is influenced by the temporal properties of environmental stimuli, music serves as a structured rhythmic backdrop for dance. Just as a speaker may try to inform, persuade, or question an audience or a conversational partner, a dancer can physically lead a group of dancers (at a distance) or a single partner (in close contact). And just as listeners provide feedback to speakers through nonverbal behavior, a follower in a dance can exhibit resistance or willingness to the leader. These analogies between the regular, structured nature of dance, on one hand, and the complex, unconstrained nature of conversational interaction, on the other hand, suggest that investigating rhythmic synchrony in dance-oriented human-robot social interactions may be informative in designing robots that can perceive and behave according to natural rhythms in more open-ended interactions. The regularity

---

[1] Aristotle's *Poetics* proposes these as the two innate motivations for creating and enjoying art.

of rhythm in the domain of music-based dance provides us with problems that are likely more tractable than those presented by unconstrained interaction, and dance provides an engaging context for experimentation with children. And even it if happens that the principles of rhythmic synchrony do not easily transfer between the two domains, robots capable of playful dance interaction with children may be useful in pedagogy, cognitive development research, and therapy, as discussed in section 2.1.3.

While developing the technology described in the previous chapter, we have also worked to evaluate the interactive properties of this technology. After we observed measurable differences in children's interactions with Roillo's simple rhythmic dance-like behavior (discussed in Chapter 3), we implemented an autonomous dancing system on Keepon and situated it in a variety of settings. In this chapter, we describe a series of studies that use dance as an organizing context for facilitating and studying rhythmic interaction between Keepon and children. First, we situated the robot in an unconstrained open-house setting in which we found that the robot's synchrony to music had a positive effect on the elicitation of rhythmic behavior by children (section 5.1). Next, we designed a strictly controlled experimental protocol to examine differences caused by the robot's rhythmic attention to music or to children's movement as perceived using an accelerometer (section 5.2). This experiment was not effective in eliciting the behaviors we wished to study, so we redesigned our protocol and used an older group of children. Finally, we conducted two experiments with children dancing on pressure sensors, in which we found that children could assume the roles of leader and follower in a dance and that this affected their rhythmicity (section 5.3).

5.1 "Open house" study

Our rhythmic robotic system was first physically demonstrated, and an exploratory study conducted, with children interacting with the robot in an open-ended dance task (fig. 5.1) [Michalowski *et al.*, 2007b]. The goal was to learn how children would physically and rhythmically interact with the robot. Although the context and stimuli were designed to afford dancing, the interactions remained unconstrained so that we would be able to observe a wide range of interaction styles. We were particularly interested in observing the effects of the robot's synchrony to music on children's interactive behavior: if children encountered the robot when it happened to be dancing in sync with the music, would they be more likely to dance themselves? Our analysis suggested that there was indeed a positive effect of the robot's synchrony on the occurrence of rhythmically interactive play. Through a behavioral analysis of videotaped interactions, we saw that music provided a powerful environmental cue for the elicitation of rhythmic behavior relative to that of the robot, and that the robot's responsiveness to people's behaviors positively affected their

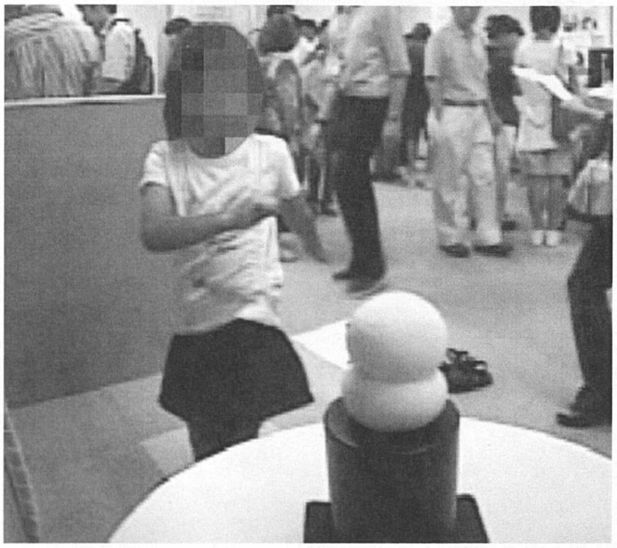


Figure 5.1: Keepon dancing with a child at the NICT open house.

engagement with it. We believe that children were capable of perceiving the robot's synchrony to environmental stimuli, and that this heightened its attractiveness and elicited physical playful interaction with the robot. We additionally discovered possible gender differences in the degree of rhythmic movement and touching behavior. Furthermore, we observed a number of expected and unexpected styles and modalities of interactive exploration and play that informed the next steps in our design.

5.1.1 Study design

Keepon was on display at the annual laboratory open house of Japan's National Institute of Information and Communications Technology (NICT) in July 2006. Members of the community, especially parents and children, were invited to see the various research projects undertaken at the Institute. Keepon was positioned atop a small table in a half-cubicle-partitioned area in a room with other robots on display. The floor on one side of the area, in front of Keepon, was tiled with carpeted panels to provide children with an affordance for dancing. A sign hung on one of the partitions encouraged children, in Japanese, to "dance with Keepon!" Selections from a collection of well-known Japanese children's songs played randomly from speakers behind a partition. Keepon was controlled by the same PC that played the music, and danced when a song was playing. During periods of silence (in between songs), Keepon stopped dancing.

Keepon was configured to extract rhythm from visual movement, rather than audio, with the goal that the robot would synchronize to the physical dancing of visiting children.

The impression the children had was that of a robot dancing to music, when in fact the robot was "deaf" to the music and could respond only to their movements. A small fixed camera was trained on the carpeted area for video capture and processing, and this feed was also recorded for later analysis. The on-board cameras could not be used as it would be extremely difficult to isolate environmental movement from the robot's own movement. A subwindow of the video stream (in the area children would stand in to be in front of Keepon, with the cubicle wall as a static background) was processed by an optical flow vision operator as described in section 4.2.2. Changes in this direction (with hysteresis) were considered to be "beats" of dominant movement in the environment (presumably that of a human dancing partner). The Max/MSP `sync~` object received this series of taps and returned the BPM of the last three approximately evenly-spaced taps. Keepon's dancing was set to this tempo, and remained at this tempo until a new tempo was detected. As a result, Keepon's tempo was influenced by movement in the environment. In practice, the perception of rhythm from optical flow was rather poor, and the visual scene was rather noisy, with children moving erratically, in and out of the camera's view, or in groups. The result is that the robot achieved synchrony with children's movement only occasionally. However, we did not consider Keepon's unpredictable and imperfectly synchronized behavior to be a liability, as it resulted in the robot being sometimes (indirectly) synchronized to the music, and sometimes unsynchronized (whether this occurred as a result of Keepon's correct or incorrect perception of children's synchronized or unsynchronized dancing). Keepon's general direction of attention was guided by an operator, so that gaze could be maintained towards children, but the rhythmic movement and dancing (centered around the direction of attention) was generated autonomously as described in Chapter 4.

We recorded visitors interacting with the robot and analyzed the video to observe general patterns of behavior and, specifically, rhythmic behaviors. From five hours of recorded video, we isolated the segments (a total of one hour and forty-eight minutes) in which people interacted with the robot, where interaction is defined as visual attention, tactile contact, or bodily movement in front of the robot. We had two coders analyze the video with two different methods: a per-child coding scheme and a time-based behavioral analysis.

5.1.2 Per-child coding

Initially, we analyzed the video following a coding scheme designed to determine the effect of Keepon's synchrony to music on the rhythmic behavior of children who interacted with it. Half of the video (69 minutes), with the interactions of 116 children, was coded in this way. Since Keepon's tempo was only loosely related to the actual rhythms of each

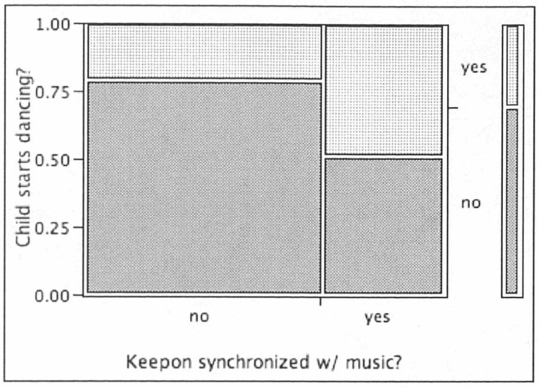


Figure 5.2: Whether children started dancing based on Keepon's initial synchrony with music (from binary per-child coding method). A significantly higher percentage of children began dancing when Keepon was initially synchronized to the music than when it was not. Overall proportions are on the right.

interacting child, and since the robot's tempo was maintained even after a child left until a new child appeared, Keepon was at times synchronized to the music when a new child approached it, and at times not. We were interested in the effect this synchrony would have on the subsequent interaction. For each child, a coder noted (a) whether Keepon was initially synchronized to the music as the child approached for interaction with the robot, and (b) whether the child then began to dance or perform clearly rhythmic behaviors synchronized with the robot or with the music. These codes were binary, in the sense that the coding method was only concerned with whether an initial condition was true and whether the child initiated rhythmic behavior, without regard to the duration of either the interaction or the period of rhythmic activity.

About a third of the children (35 out of 116) who interacted with the robot performed some dance-like rhythmic behavior, from oscillatory hand movements to full-body physical exertion. The codes obtained from this analysis show that significantly more children started to dance with the robot when it was already dancing in synchrony with the music than when it was not ($\chi^2(1, N = 116) = 9.321, p = 0.0023$) (fig. 5.2). This suggested that the children were able to perceive properly synchronized movement, and that this recognition was a powerful motivator for turning passive interactions (such as observing and touching) into more active and engaged dancing interactions.

5.1.3 Time-based behavioral analysis

Encouraged by these results, we had a second coder analyze the video (in its entirety) using Noldus Observer software [Noldus, 2006], which allows fine-grained, continuous labeling of the robot's and participants' behaviors for temporal analysis. This coding method is different than that just described in that it considers events and states as durations of time rather than considering initial conditions or eventual behaviors of individual participants. The results are consequently presented as percentages of time rather than as a sample of children.

First, Keepon's synchrony with the music (independent of children's movement) was coded for the duration of the video. Keepon was more often not synchronized with the music, either because people were not moving rhythmically to the music or because Keepon failed to perceive if they were: for 38.6 minutes (37.9% of the time) Keepon danced in synchrony with the music; for 53.64 minutes (52.64% of the time) Keepon was out of synchrony with the music; and for the remainder of the time Keepon was not moving either due to silence between songs or because of minor technical interruptions in functioning. The beginnings of interaction sequences, in which children interacted with the robot individually, simultaneously in groups, or by taking turns, were evenly distributed over these durations: 38 interaction sequences began with Keepon synchronized to the music, 57 when not synchronized, and 17 when Keepon was not moving.

The behaviors of children who interacted with the robot, as well as their gender, were coded according to a scheme focusing on the rhythmicity of their movements and synchrony with the robot or the music. The scheme includes rhythmic dancing behavior such as bobbing, rocking back and forth, and full-body movements to the music; gesturing with the arms (synchronously or asynchronously with the music); touching the robot (tapping/patting just once, or repetitively; synchronously or asynchronously); and holding or stroking the robot.

We coded Keepon's interactions with 113 girls and 96 boys. Children of all ages interacted with the robot, from very young (approximately two years) to the mid-teens. While we were not able to accurately estimate age, it did not appear to have a significant effect on the behaviors we analyzed. The majority of interactions involved children interacting with Keepon in groups: only 25 children interacted with Keepon alone, 52 interaction sequences involved pairs of children, and 35 interactions involved 3 or more children (with some overlap, as children left or joined these groups).

Our main focus in coding, as before, was to see how the apparent synchrony of Keepon's behavior in its relation to the music affected the manner in which children approached an interaction with the robot. For the children, there were two major rhythmic affordances in the environment: one was the almost constantly moving Keepon, and

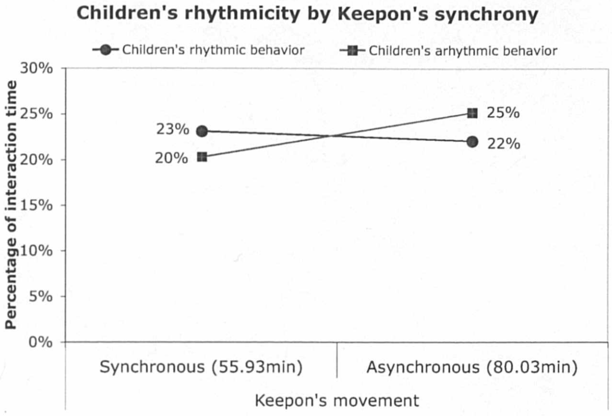


Figure 5.3: The percentage of interaction time children spent performing rhythmic or arhythmic behaviors according to Keepon's synchrony to music.

the other was the music. Although the rhythmic behaviors of other people in the environment can be expected to have an influence on children's behavior as well, we did not code people's interactions with each other as we were primarily interested in the development of interaction between the children and the robot.

We first examined the effect of Keepon's synchronous or asynchronous relationship with the music on the occurrence of corresponding rhythmic (bodily movement, gesturing, or touching the robot in a rhythmic manner in synchrony with the music) versus arhythmic (gesturing or touching the robot without apparent rhythm and not synchronous with the music) behavior by children. The coder did not distinguish the stimulus to which a child synchronized, but only whether the child initiated rhythmic dance-like behavior. We determined the amount of time children spent performing rhythmic and arhythmic behaviors toward the robot as percentages of the total time that all children spent interacting with the robot in each of these two conditions (55.93 minutes when Keepon was synchronous, and 80.03 minutes when Keepon was asynchronous). In this method of coding, by measuring percentages of time, we found support for the findings in our initial analysis. Figure 5.3 illustrates that the synchrony of the robot's behavior to music resulted in more time performing rhythmic behaviors relative to arhythmic behaviors.

Interestingly, the gender of the interacting children appears to have had an effect on the incidence of children's rhythmic behavior as well as on the relative amounts of movement and touching behaviors. Girls spent longer periods of time with Keepon (total 82.38 minutes, average .73 minutes) than boys did (total 54.66 minutes, average .57 minutes). Furthermore, girls spent a higher percentage of their total interaction time performing rhythmic rather than arhythmic behaviors, while for boys this trend is reversed (fig. 5.4, left). Finally, girls spent a higher percentage of their time touching Keepon, a

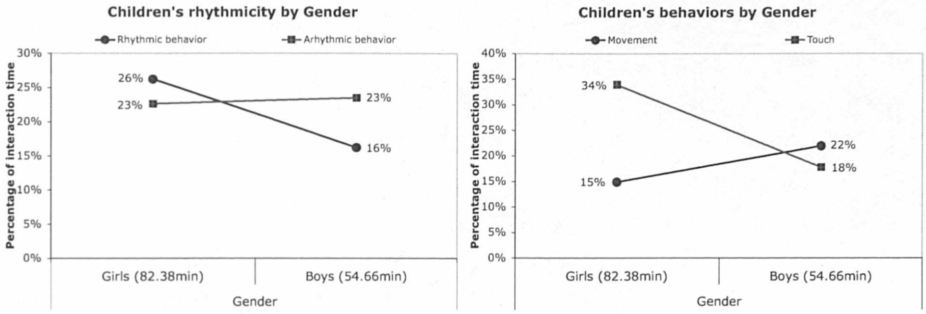


Figure 5.4: The percentage of interaction time children spent performing rhythmic vs. arhythmic behaviors (left), and movements vs. touching behaviors (right), according to gender.

phenomenon corroborated by our qualitative observations (fig. 5.4, right). This may be viewed in the context of the theory that girls tend to be empathetic while boys tend to be systematic [Baron-Cohen, 2003].

5.1.4 Qualitative observations

We also made several qualitative observations based on our experience during the study and our subsequent observation of the recorded video. Many children did not dance at all, and we suspect that this is because the music was oriented for rather young children and because the demonstration took place in a large public space with mostly strangers. Additionally, a cultural sense of embarrassment or modesty may have precluded the initiation of free "interpretive" dancing (without rigid structure or rules). As there was intentional ambiguity concerning what Keepon was responding to in the environment, we observed some confusion about what to do and the effects of one's behaviors on the robot's behavior. Given the difficulty of accurately sensing bodily movement using the simple method of optical flow, this confusion probably hindered the development of more involved interactions. In the cases where children were easily able to perceive the robot's responsiveness to their own rhythmic movement (perhaps a dozen cases), they exhibited a clearly recognizable enjoyment and satisfaction with the robot's performance.

Several of the children, and even some adults, began their interactions with Keepon by simply observing the robot for several seconds, and immediately began to nod or dance in rhythm with Keepon when the robot's rhythm happened to begin to synchronize with that of the music, even if only for a few cycles. It may be that the presence of music created an expectation that the robot should behave in a rhythmically appropriate manner, and that when this synchrony was observed, it provided a strong motivation to participate in the rhythmic activity.

While we did not have an opportunity to ask children how they viewed Keepon, during our observations we noticed them displaying various behaviors toward the robot in interaction styles that might be roughly categorized as "exploratory" or "caregiving." For example, some children seemed to explore the robot as a mechanical object: we observed them closely examining the robot's internal and external cameras (interacting specifically with the external camera), testing the floor tiles to see whether stepping on them would elicit a reaction from Keepon, and squeezing the robot to feel what was inside the body. Some children positioned their heads near the robot as if to listen for sounds during movement. Other children enagaged in petting and stroking behavior as one would expect to see in interactions with an animal.

Even though the children were instructed only to "dance with Keepon," several children tried hard to make Keepon "do the right thing" and synchronize its movements to the music. Some children tried to provoke reactions from Keepon by waving, stomping around, or hitting the robot in time to the music. A number of children exaggeratedly danced to the music when Keepon was not synchronized, and even petted or grabbed the robot rather strongly in order to try to get it to move correctly, while other children went to great effort to imitate Keepon's movements very closely, sometimes ignoring the musical rhythm altogether. We also observed children mimicking the robot's morphological constraints; children seemed to mirror the robot's capabilities by bobbing their heads and torsos or rocking side to side. Synchrony to the music rather than to the robot (when the two were not synchronized) was more prevalent in children's behavior. This may be because the robot's behavior was not sufficiently rhythmic, or because its movement was relatively less salient in the context, or because audio might be a stronger cue to rhythm than movement. However, in a later study (described in section 5.3), when we introduced leadership roles and intentionally made the robot deviate from the music, we found that the robot could indeed draw children away from an environmental rhythm.

Finally, we were distributing cell phone charms with an image of Keepon on them; several children made the charms "dance" in front of Keepon, often in rhythm with the music, as if Keepon had a special (perhaps communicative) relationship with the charms or would recognize a depiction of itself. This mode of interaction inspired our later development and use of accelerometer-based perception of physical movement.

5.1.5 Discussion

The range of interaction modes we observed with the robot made direct comparisons between children difficult, as they entered these open-ended interactions in an uncontrolled manner with very different sets of expectations, desires, and contexts for participation. However, as this was an exploratory study, the observations we made were useful in the development of future interactions and studies. In particular, the differences in human behavior based on Keepon's variable synchrony to environmental stimuli led us to later explicitly and experimentally control Keepon's rhythmic attention. Furthermore, the difference between children who attempted to show Keepon how to dance and children who tried to imitate the robot inspired, in later studies, our explicit introduction of leadership roles in the interaction design.

This study also demonstrated the challenges of developing perceptual technologies for detecting social and environmental rhythms. The difficulty of achieving autonomous rhythmic synchrony stemmed largely from a difficulty in perceiving what are often small changes in visual stimuli. The optical flow method used here required that the interaction partner remain in a rather constrained space, which often failed in these freeform interactions. To this end, we identified a number of directions for addressing the problem of rhythmic motion perception:

- We can look at how successfully we can detect oscillatory or repetitive movements in the environment from a camera using simple vision processes other than, or in addition to, the optical flow technique described here, such as face detection, skin color detection, feature tracking, etc. [Crick *et al.*, 2006] detected extremities in the spatial movement of the skin color of a "conductor's" hand, and we can imagine doing the same with skin blobs detected from the head or hands.

- In order to obtain more accurate information about human body movement during social interaction, we can look for repetitive or rhythmic movements in motion capture data. Alternatively, the [Ogawa and Watanabe, 2001] robotic telepresence system (which allows two remote humans to communicate with each other through a pair of robots whose behaviors are controlled by models of speech/motion correlation obtained from human speakers and listeners) measures head, arm, and body motions using magnetic sensors, which provide similar data to motion capture data. Unfortunately, both of these methods are rather invasive and uncomfortable, so we would prefer to use passive sensing.

- As many children spent a large proportion of their time with the robot touching it, and some in fact tried to force the robot to move in a certain way, it is clear that proprioceptive recognition of rhythmic input might be a powerful means of detecting rhythmic movement, eliciting rhythmic play, satisfying the children's expectations of tactile contingency, and most importantly, communicating rhythms to the children through a salient tactile (and not just visual) channel. Keepon's motors can be queried to provide a measure of the torque being applied. If this can be separated from the torque induced by Keepon's current movement, it would be possible to extract rhythm from any repetitions or regularities in externally applied force. Alternatively, or additionally, we might detect deformation in the body surface by embedding tactile sensors in the skin. Unfortunately, time precluded such investigations.

- We had selected dancing as the task for our social interactions because the rhythmic properties of audio and movement are salient and regular enough to be easily sensed. Similarly, music production or performance, such as hitting a drum, is a task that would allow explicit rhythmic input and even the use of specific sensors for this purpose (e.g., an electrically instrumented drum). We have demonstrated Keepon dancing rhythmically to the beating of a modified drum fitted with a microphone [Michalowski and Kozima, 2007a]. The movement of some other instrumented object or avatar, such as the cell phone charms observed here, is another possibility that we later explored with the use of accelerometers (see section 5.2).

- While the upper-body movements in children's freeform dance can be rather erratic, there appeared to be a more regular vertical movement that might be exploited with the use of simple pressure sensors in the floor. In fact, a few participants mistakenly believed that the floor was in this way instrumented, similar to input devices for currently popular musical video games. We eventually moved to this perception method (in section 5.3 and Chapter 6).

Perceiving rhythms in these different modalities is important not just for establishing rhythmic synchrony more effectively, but also for clearly and powerfully exhibiting responsiveness and contingency to the many different ways that people in unconstrained interactive situations might engage with the robot.

The dance-oriented human-robot interactions between Keepon and a child would at first appear to be dyadic. However, with music as a stable and unchanging[2] rhythmic presence, it might instead be considered a *triadic* interaction, with the music serving as a very salient and powerful time-giver. Alternatively, the music can be seen as a part of the environmental context in which the dyadic pair operates. Children could (and did) synchronize with Keepon, with the music, with both, or with neither (though we were not able to code these differences explicitly), and Keepon could synchronize with the music only indirectly, by sensing movement in the environment. However, people may hold a cultural model in which they expect that if Keepon can hear the music, it will dance appropriately. Differences in their expectations and behaviors, then, may be indicative of their model of Keepon's awareness of its environment.

---

[2] This need not be so—the music might also be made to change tempo according to rhythmic movements of Keepon or the child.

Furthermore, while we did not analyze for differences between dyadic interactions and group interactions, we believe that a group context affected children's behavior—causing them either to hang back as a bystander or to be more active in competing for the robot's attention. To control for these effects, our later studies were generally constrained to one-on-one interactions.

Our initial method of coding was based on a per-child binary labeling of the robot's synchrony to music and the rhythmicity of children's resulting behavior. We found a statistically significant effect of the former on the latter, which was supported by our subsequent time-interval-based method. This suggests that children were able to perceive the synchronicity of the robot's behavior, and that this increased their engagement and participation in the activity. Our analysis in this study did not account for various synchronous modes, due to the uncontrolled nature of the robot's behavior, the rapidly shifting rhythms of the various stimuli and interactors involved, and the difficulty (and time-intensiveness) of manually coding periods of synchrony. This demonstrates the importance of carefully designing both the interactive scenario and the analytical methods used to evaluate it. In later studies, we developed additional perceptual techniques to allow the robot to more effectively synchronize to people's rhythmic movements, designed experimental protocols to control the robot's synchrony to entrain to particular perceived rhythmic behaviors by people, and logged rhythmic information received from sensors and sent to actuators in order to identify the effects of synchrony, not only to music, but also to the movements of interaction partners.

5.2 Controlled rhythmic attention study

While our goal is to develop technological artifacts that can participate in rhythmically coordinated social interactions, it is also necessary to consider appropriate methodologies for the evaluation of such technology. In research with human beings, quantitative measures are most useful when data can be directly compared across participants or scenarios in controlled experimental settings. Modern psychological experiments of this nature, in order to minimize the number of confounding factors, typically attempt to control differences between participants' experiences and to clearly define these differences in terms of independent and dependent variables.

This section describes our experience in developing and revising such an experimental paradigm. We conducted a study on the effects of the robot's synchrony to either environmental or social rhythms on dance-oriented play with children [Michalowski and Kozima, 2007c]. This study was intended to evaluate and improve our perceptual and behavioral system. Even more importantly, our goal was to develop an experimental paradigm appropriate for our research. While this study was an attempt to examine rhythmicity in

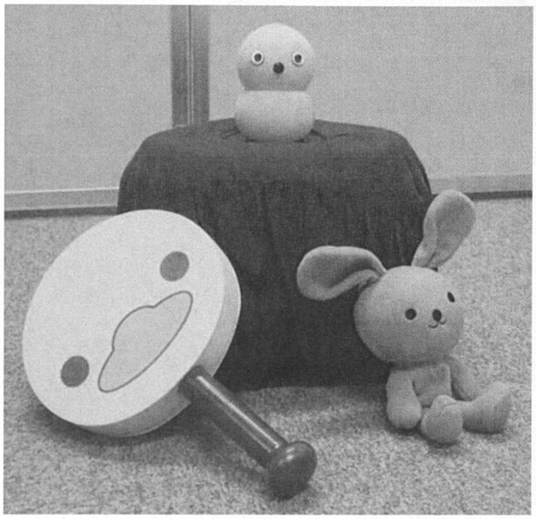


Figure 5.5: Keepon and the two toys used with the accelerometer: a stuffed rabbit (169g, 290mm long) and a soft paddle (137g, 320mm long).

a controlled manner, our experience led us to re-evaluate our approach. We found that our strictly controlled experimental paradigm and constrained interactions precluded the full expression (and therefore observation) of interactive behaviors sufficiently rich and natural for the study of the complex yet fundamental rhythmic properties of social interaction. Children were insufficiently comfortable to engage in the desired rhythmic behavior, possibly due to the robot's novelty, the children's age, the laboratory setting, and the controlled scenario. Although we were unable to draw statistically significant conclusions given the small number of participants and high variability in observed behaviors, our experience allowed us to create more effective protocols for subsequent studies.

5.2.1 Experimental design

Our initial formulation of the experimental procedure was designed to examine, in a controlled manner, the active or passive roles that a child and a robot might take in a rhythmic social interaction. After our difficulties in sensing human movement through vision in the previous experiment, we developed a system for detecting rhythmic movement through the use of an accelerometer (as described in Chapter 4, implanted in one of two toys: a stuffed rabbit and a foam paddle, fig. 5.5). Similarly to the visual system, our software looked for three evenly-spaced beats in order to determine a tempo of the rhythm. The robot was configured to be able to synchronize its dancing behaviors either to music or to the movement of the toy, to present stimuli in a controlled manner, and to record sensor data for later analysis and comparison with the rhythms of the musical stimuli and the robot's movements. By controlling for other variables, we tried to identify the condition in which the movements of the child would be most closely synchronized to the music. One one hand, the robot might reinforce the music's rhythm and lead the child to dance with it; on the other hand, if a child recognizes his or her influence over the robot's dance, this might encourage more explicit demonstration to the robot.

Our procedure was as follows: A facilitator (not the experimenter) brings the child and caregiver into a play area (1.9m by 3.12m, wall height 1.5m) with Keepon on the floor about 50cm from the middle of one of the shorter walls. The facilitator retrieves the accelerometer-equipped rabbit (and, in later trials, the paddle) from behind a wall of the play area, and Keepon establishes gaze following (under the guidance of a teleoperator) toward the toy. The facilitator demonstrates Keepon's contingency and maintenance of attention to the movement of the toy. The facilitator places the rabbit next to Keepon and Keepon gazes at it. As soon as the child picks up the toy, music begins to play and Keepon begins to dance. Keepon dances only when the child is holding and moving the toy (automatically sensed by the accelerometer). Meanwhile, the teleoperator maintains Keepon's attention toward the toy. The only difference between conditions is that, in the first condition, the rhythm of Keepon's dance is synchronized with that of the music; in the second condition, the rhythm is derived from the detection of direction change points in the movement of the accelerometer. After two songs (approximately 4 min.), an indefinite period of free play is allowed, with rhythmic synchrony to the movement of the toy (for all children).

This manipulation aimed to identify the influence of a robot's behaviors on a child's when the robot reinforces an environmental rhythm (in a sense trying to "guide" the child's movements) versus when it responds to the child's movements ("following" the rhythm of the child). The recorded rhythms could be compared quantitatively, in terms of synchronization, duration of movement, and amount of movement, and the responses and attitudes of children in the two conditions could be compared qualitatively. Our aim for the study was to target differences between rhythmic attention to one of two modalities.

In the first condition, Keepon is synchronized with the music's rhythm; we might say that Keepon is following the environmental rhythm and trying to lead the child's rhythms. In the second condition, Keepon should follow (or synchronize to) the child's rhythm. The goal of this experiment was to determine which condition results in a closer coordination between the three rhythms. There are two possible outcomes. If Keepon following the music results in greater coordination than Keepon following the child, it may suggest that Keepon's physical expression and reinforcement of audible environmental rhythms is salient and encourages the child to follow this rhythm as well, while Keepon following the child encourages the child to explore his or her influence over the robot without regard

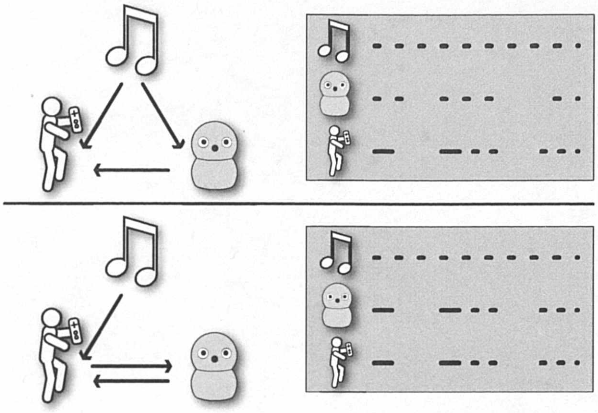


Figure 5.6: The proposed "flow" of rhythmic influence between the music, the robot, and the child in the two experimental conditions: Keepon synchronizing to the music (top) and to the toy (bottom). Large gaps represent periods in which the child is moving the toy; short and long bars represent fast and slow rhythms.

to the music. If Keepon following the child results in greater coordination than Keepon following the music, it may suggest that the child's perception of Keepon's rhythmic contingency to his or her own encourages the child to more clearly follow the music so that the two may dance together, while Keepon following the music results in lower engagement or in attempts to make Keepon deviate from the music.

There are a few important points to note about this procedure:

- A teleoperator maintains Keepon's attention toward the toy for the duration of the experiment, to encourage manipulation of the toy as the channel for interaction with Keepon.
- During the presentation of musical stimuli, and during free play, Keepon's expression of rhythmic dancing behavior is in both conditions temporally contingent on movement of the toy (whereas rhythmic contingency differs between the conditions).
- There are (at least) three important rhythms in effect during this type of interaction: the environmental rhythm of the music, the rhythm of Keepon's dancing, and the rhythm of the child's movement of the toy. All three rhythms are recorded for later comparison and analysis.

Fig. 5.6 illustrates the proposed relationship between rhythms in the two conditions.

The recorded rhythm of the child's movement of the toy was to be analyzed for quantitative information about three variables: first, the number of movements made by the child (e.g., in fig. 5.6, the child changes the toy's direction seven times); second,

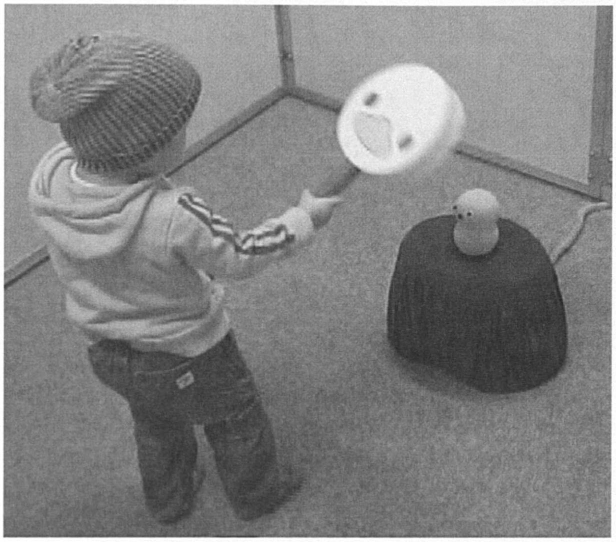


Figure 5.7: Keepon dancing with a child.

the duration of movement by the child (e.g., in fig. 5.6, the total number of seconds in the three periods of movement); third, the degree of synchrony between the child's movements and the music (i.e., the correspondence between tempos in the two rhythms). These values, in addition to the qualitative analysis of recorded video, can provide a numerical indication of the difference between experimental conditions. We expected to find more movement with longer duration when Keepon synchronized to the children, and a higher degree of synchrony to the music when Keepon reinforced the rhythm of the music.

5.2.2 Conducting (and revising) the experiment

We recruited eight volunteer parents to bring their 3-year-old children (34.69 ± 0.736 months) to a child study laboratory at Kyushu University in Japan in February 2007. We had access to a pool of participants in this age group, and we believed that it was an appropriate age for encouraging rhythmic play with toys and music. As the experiment was explained to parents and consent forms filled out, children played in a staging area with young lab assistants to make them feel comfortable. Next, the experimental sessions lasted about twenty minutes each. Finally, a debriefing session in the staging area was used to explain the system and the reasons for the study.

After three sessions, we found that our choice of toy had an impact on the sensor data we obtained. The shape and weight distribution of the toy was such that children tended to hold the rabbit by the neck and spin it rapidly back and forth by rotating their wrists. This movement was both difficult to sense (being rotation rather than translation around the center of the accelerometer) and too rapid to provide meaningful rhythmic input to the system. Therefore, for the remaining sessions, we selected a different toy: a foam paddle with a plastic handle and the face of a yellow "chick"; this could plausibly be called Keepon's "mother" by the facilitator, was larger than the rabbit, and provided an affordance for holding and waving rather than spinning. For four of the remaining five children, the difference with this toy was indeed the elicitation of a different quality of movement, and sensor data that reflected the clearer rhythmicity of this movement.

We closely followed the experimental procedure described above for the first three sessions (Keepon synchronized to the toy for the first two sessions, and to the music for the third) and noted a common shyness or reluctance to engage with the robot, which is typically a very interesting artifact for children in the classroom settings in which we normally observe it. We suspected that the facilitator's limited introductory behaviors with the robot, and his subsequent withdrawal from the interaction (all intended for controllability), resulted in the lack of a scaffold, or context, for the children to begin their exploratory play. The short demonstration of Keepon's contingency to the location of the toy was hindered by a competition between the robot and the toy for the child's attention; since the robot was moving under its own power, children focused on it as a novel sort of artifact. This novelty meant nearly no engagement during the music, and generally only an exploration of contingency after some time during the post-music free play period.

Therefore, we decided to introduce a short introductory free play period before the music, with Keepon only maintaining attention to the toy, to remove some of the element of novelty and surprise during the music. This lasted between 30 seconds and 1 minute, based on the judgment of the facilitator. For the fourth and fifth participants, the change resulted in a higher degree of comfort and uninhibited play during the musical period. However, we found that the introduction of music itself seemed to be a novel enough stimulus that it continued to elicit an initial period of uncertainty or renewed observation for any difference in the robot's reaction with the music, not necessarily providing the active movement of the toy necessary to contingently trigger Keepon's dancing. We also saw the fourth and fifth participants rhythmically hit Keepon's head with the paddle, and this was sufficiently reinforced by Keepon's responsive rhythmic behavior that it was repeated. However, this behavior was not performed with music or for sufficient duration to provide meaningful quantitative rhythmic data.

Consequently, we further modified the procedure to add music to the free play period in the hopes that the exploratory behavior would, with more time, lead to desired rhythmic behavior with the music. After the introductory attention-contingent free play, a song was added in which the robot's rhythm was synchronized with that of the toy for all children. Following a thirty-second pause, two songs played in which the robot danced in synchrony with the music. Following another pause, a fourth song played with toy-synchronized dancing once again. For the limited number of children we had seen, we decided that any differences between conditions were being overwhelmed by individual differences between participants, so a within-subjects design might be more appropriate than between-subjects.

The sixth participant was extremely shy and unresponsive to almost any of the robot's behaviors or the facilitator's encouragement. The seventh and eighth participants, however, performed a high amount of rhythmic behavior, particularly after becoming comfortable with the robot's attentional contingency and the presence of music. For these last two participants, we further deviated from the procedure in making Keepon much more social, in the sense of frequently shifting attention between the child and the toy, and this seemed to be a powerful cue for motivating use of the toy. In the case of the eighth participant, a five-year-old brother was permitted to participate in the interaction, and we found that his company and facilitation encouraged exactly the kind of rhythmic behavior we had hoped to see, without seeming to bias or detract from the free expression of these behaviors or their synchrony with the robot or the music. Running around Keepon was seen by this final pair of children as the most interesting way of eliciting movement and attentional contingency. During one period, when the final participant actively moved the paddle, we observed a mixture of attention-directing and rhythmic stimuli from the child: she made large movements of the paddle and held it in particular locations for Keepon to gaze at, interspersed with rapid repetitive waving of the paddle to elicit dancing. It is our opinion that she recognized that these were qualitatively different types of stimuli that elicited qualitatively different behaviors from the robot.

5.2.3 Discussion

Although we did not observe as much rhythmic behavior from most of the children as we had hoped, we came closer to an understanding of the types of scenarios that are (or are not) conducive to such behavior. We identified several types of problems that may be responsible for the low rate of rhythmic engagement in this study: technical problems of perception and entrainment, constraints inherent in the lab setting and controlled protocol, and children's comfort with the robot as determined by age and familiarity.

From a technological perspective, this study was intended to experiment with a new method of perceiving rhythmic behavior. We had investigated the use of vision and sound, and now we incorporated an accelerometer in the detection of physical rhythms. While having direct access to the movement of the toys resulted in cleaner data for rhythmic perception, we encountered difficulties in getting children to move the toys in the translational rather than rotational manner required by the accelerometers. We therefore switched to the use of pressure sensors in later studies (which did not require the manipulation of a toy). With respect to entrainment, we observed that correspondence in tempo or frequency alone was not as effective in eliciting engagement as phase-matching (i.e., the co-occurence of beats). We believe that children's motivation to explore is related to their perception of controllability or reactivity in the robot, and the most basic form of this controllability is the recognition of contingency. When the robot matched a beat with a movement of the child, or attended to the toy's spatial location, it was a salient instant in time. On the other hand, synchrony in tempo alone has duration and requires sustained observation. Even when rhythmic contingency was recognized by children, it was usually recognized as contingency in the *onset* of oscillatory movement more than in the establishment or maintenance of synchrony in frequency. For our later studies, therefore, we switched to an event-based rather than tempo-based mechanism (as described in Chapter 4), in which the robot immediately responded to detected beats. It may also be helpful to to saturate the interest in instantaneous contingency (through introductory periods of familiarization with the robot) in order to encourage children's later attention to more continuous behaviors.

From a perspective of setting and protocol, we found that strict control of the situation (precise timing of stimulus introduction, limited introductory period, minimal facilitator participation, and an unfamiliar, sterile, anxiety-inducing laboratory setting) overwhelmed or significantly delayed the exhibition of natural rhythmic interaction. We believe that a looser, more naturalistic environment (and less strictly adhered-to experimental procedure) is helpful in promoting rich interactions that can yield useful and interesting data for studying social rhythmic phenomena. Rhythmicity is a *holistic* property of social interaction, and a reductionist approach that attempts to cut this phenomenon into pieces (to see the effects of changes in one set of parameters on another set of parameters) looks only at the tip of an iceberg. Furthermore, enforcing a scenario that emphasizes a small number of piecewise relationships may actually stifle the holistic emergence of a phenomenon like social rhythmicity. It is out of the scope of our work to perform psychological experiments that prove the existence of a social phenomenon, as it is not feasible to run enough subjects through strictly controlled experiments to find statistically significant effects over the noise of individual variability.[3] Rather, we are starting with an accepted principle of social interaction and we aim to demonstrate the occurrence of this phenomenon in human-robot social interaction. Conducting field stud-

---

[3] When we cannot demonstrate effects of rhythmic synchrony on the quality of controlled experimental interactions, it is unfortunately difficult to determine whether it is because our technology is inadequate (in perception or behavior), if our architecture or formulation of the relevant principles is incorrect, or if our experimental methodology is insufficient for eliminating confounds/variability and illuminating common features of the interaction.

ies [Lofland and Lofland, 1995] in a more naturalistic environment (and with a less strictly adhered-to experimental procedure) can alleviate conditions for failure by promoting rich interactions that yield useful and interesting data for studying social rhythmic phenomena by drawing generalizations about the common features of the interactions. Of course, condition-blindness should be enforced so that the facilitator does not bias the outcome of the experiment, and there is a danger of introducing biases if analysis must be tailored to individual participants. Ideally, we would be able to facilitate and observe unconstrained interactions, record and perform video coding on behavioral aspects of engagement, and record and align sensor data to these codes to relate them to the rhythmic properties of movement. Since we lacked the resources for intensive behavioral analysis, we continued to work to develop controlled studies that were more flexible in their protocol and facilitation.

From a perspective of participant comfort, we observed that the novelty of the robot seemed to inhibit the free and natural expression of rhythmic behavior. For example, children were usually testing the robot's contingency or watching to see what it would do next, rather than assuming the basic social contingency they would expect in another human or in a pet. We expected (and, indeed, had experienced) that children familiar with Keepon generally acted very differently, and in a rhythmically more interesting manner, than did the children in this study. Indeed, by expanding the introductory period, we found that children became more comfortable and behaved more actively, and we included an extended introduction in our next studies. Once children are familiar and comfortable enough in a particular situation, it is easier to introduce or encourage the desired modes of play. Although this prior exposure might be considered a bias, we believe that familiarity and comfort with the robot are helpful for the expression and observation of the behaviors in which we are interested—if the study cannot be performed in a familiar setting (as in the Roillo study) or in an unconstrained manner (as in the open house study). Individual differences between children require that interactions under study incorporate adequate facilitation and that they be allowed to develop and unfold in a dynamic and flexible way. Children exhibited a wide range of interactive modes, possibly stemming from prior exposure or pre-existing ideas about robots, styles of play at home, comfort around strangers, and so on. This unpredictability and variability make it difficult to create well-defined procedures and consistent scenarios between participants in such a study. The facilitator, as we have observed, should establish a sufficient context for social interaction between the child and the robot. To do this effectively, he or she should be allowed to adapt the procedure to account for the individual personality and disposition of the child, while maintaining as much consistency between participant experiences as possible. Furthermore, given our observation of the five-year old brother who was more comfortable engaging with the robot, we moved to working with older children in our subsequent studies.

The main utility of this study was to develop and refine our technology and experimental protocol. In the following studies, we used: a perceptual technology (pressure sensors) that gives clear and clean data on children's behavior; an entrainment method that is more responsive with respect to matching phase; an experimental protocol that constrains participants only enough that they provide data to our sensors in a consistent manner to the backdrop of a consistent environmental stimulus; and facilitation sufficient for introducing (older) children to the robot and making them comfortable during the experiment.

5.3 Rhythmic leadership in dance

Based on our experiences in the previous study, we moved to a rhythmic perception system involving pressure sensors, revised our experimental protocol, and obtained access to an older population of participants. In this section, we present two experiments in which Keepon danced with children to music and synchronized either to the music or to their dancing behaviors [Michalowski *et al.*, 2009b]. Our objective, as before, was to identify differences between participants' rhythmicity and engagement between these conditions. A familiarization session was designed to promote greater comfort in the children, the school setting was familiar to the children, and the protocol was somewhat looser to allow natural interaction between the children and the facilitator, so our design was much more successful in eliciting the desired rhythmic behaviors (fig. 5.8). Furthermore, in the second experiment, we explicitly introduced and manipulated the roles of leader or follower. We found that children can assume these roles, that followers indeed tend to synchronize with the robot's movements, and that the role of follower causes children to more closely follow a musical rhythm than the role of leader.

5.3.1 Experimental design

We designed an experimental procedure for studying the effects of rhythmic synchrony on social engagement. The procedure involved Keepon dancing with individual children while music was playing—a triadic interaction in which child and robot were influenced by an external environmental rhythm. In the experimental conditions, similar to the previous study, Keepon danced in synchrony either to the rhythm of the music or to the rhythm of the child's movement, and we measured children's movements and responses. Our participants were 5-year-old students in the kindergarten of the Children's School at Carnegie Mellon University. Prior to the experiments, the teachers provided a ranking of each child's "musicality" as either high, medium, or low. When the participants (20 in

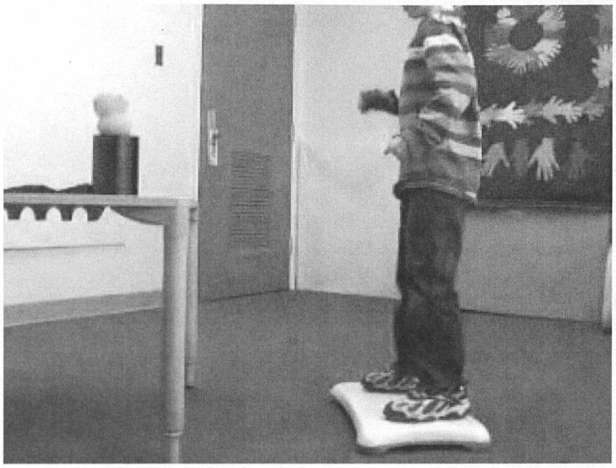


Figure 5.8: Keepon dancing with a child on a pressure-sensor board.

the first experiment, 21 in the second experiment) were randomly divided into two groups for different experimental conditions, the groups were balanced through pre-stratification (or "blocking") so that our randomization would be less likely to bias one of the groups with more musically-inclined children. This is an accepted design for random assignment when using small groups of participants with a known "nuisance" factor that might be a source of variability [NIST/SEMATECH, 2009].

5.3.1.1 Introduction & familiarization

In order to promote the children's comfort with Keepon and the experimenter prior to the experiments, we conducted a familiarization period of two freeform play sessions.

The experimenter first brought Keepon into the kindergarten two days before the experiment for an introduction. He introduced himself and and the robot, saying, "I've brought my friend Keepon to meet you. He can see you through the cameras in his eyes and he can hear through his nose." The experimenter told the children that Keepon would be visiting for a few days and that they would be able to play with it. For the purposes of realistic interaction, we incorporated an element of teleoperation whereby the experimenter controlled the overall direction of attention of the robot through a WiiMote (fig. 4.6) in order to maintain appropriate gaze. While controlling Keepon, the experimenter invited the children to gather around the robot. Keepon looked around at the children and made emotional movements (e.g. bouncing, rocking) in response to their actions or queries. The children were asked not to touch the robot so as to encourage (and create an expectation of) distal rather than tactile interaction.

The next day, the experimenter brought Keepon again for a free-form dancing session during "circle time," a period when the entire class gathers together, seated on the floor, and receives instruction, stories, questions, or direction from the teacher. The teacher played a few common songs (a normal activity for the children) while Keepon, in the front of the room, danced to the music along with the children.

5.3.1.2 General protocol

On the days of the experiments, Keepon was set up on a table in a small room off the main classroom (fig. 5.8). The WiiFit board (fig. 4.6) was placed in front of the table, and speakers near the table were used to play music. Children (having previously been assigned to a condition) were pulled from class in random order, with some shuffling taking place to accommodate individual children being absent or occupied. The experimenter brought one child at a time into the room, re-introducing Keepon to them. Keepon's attention was appropriately maintained (under the experimenter's WiiMote control) on the entering child from the time he or she entered the room, shifting to the board when waiting for someone to stand on it. The experimenter briefly demonstrated how to stand on the board (to which Keepon responded with an excited bounce). The experimenter got off, explaining that the child should also stand on the board and begin dancing with Keepon when the music starts playing. The experimenter then sat to the side of the room, minimizing his movements and vocalizations, for the duration of the song. Keepon danced either to the music or to the child's movements, but in both cases the amplitude of Keepon's dancing was determined by the child's energy level (determined by the standard deviation of weight on the board). This contingency to energy level was meant to promote recognition that Keepon would dance only if they danced; if they did not move, neither did Keepon. The only experimental difference between the conditions was the rhythm to which it danced.

In these experiments, we perceived children's rhythmic bodily movement through the WiiFit board. Keepon's dancing was constrained to a style that mainly emphasized the bouncing behavior, with minimal turning or nodding, and some amount of posture mirroring by leaning side-to-side in response to lateral movement of the child's center of gravity. As discussed in Chapter 4, the most salient rhythmic movement—bouncing—executes with minimal latency (approximately 100ms) and for our purposes may be considered simultaneous when triggered in response to a detected beat.

At the end of the first song, the experimenter thanked the child (saying, "great job!") and asked if he or she would prefer to dance to another song or return to the classroom (a measure of retention to gauge interest in the interaction). A second song was played if the child expressed desire. For the duration of the experiment, the data from the WiiFit board was recorded along with commands to the robot and the rhythm of the music.

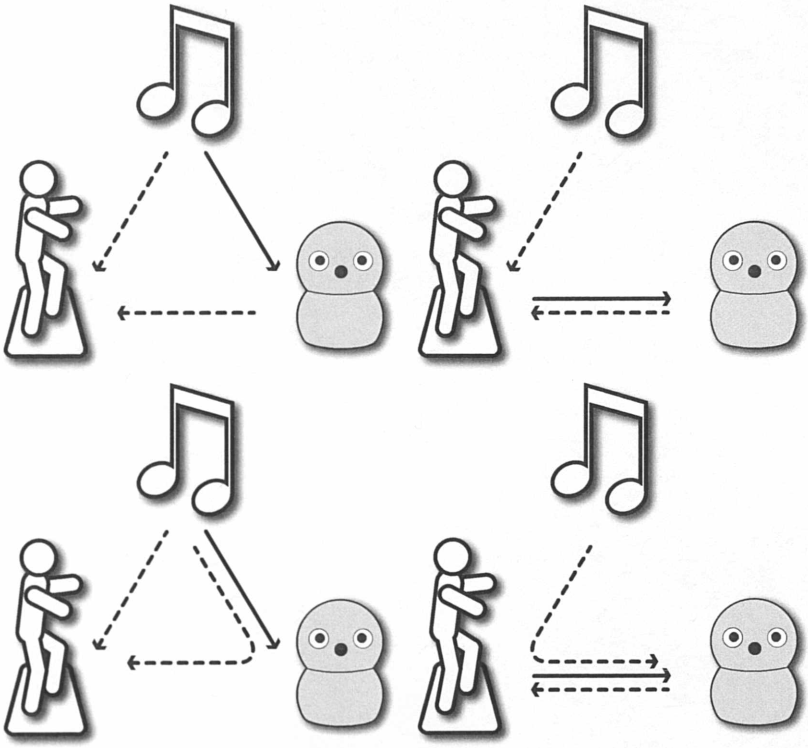


Figure 5.9: The designs of Experiment 1 (top) and Experiment 2 (bottom), with the experimental conditions of Keepon dancing to the rhythm of the music (left) or to the rhythm of the child's movement (right). Solid arrows indicate a strong coupling, i.e., the stimulus that directly controls Keepon's rhythm. Dashed arrows indicate a weak coupling, i.e., the rhythms by which the child is perceptually influenced. Curved arrows indicate a verbally established relationship, i.e., one partner demonstrating the musical rhythm to the other.

We performed two experiments, with minor changes to our technology, differences in the musical rhythm, and slightly different protocols between them.

5.3.1.3 Experiment 1: Robot attending to music or movement

In the first experiment, the child was told, "Keepon can see how we dance when we stand on the white board, like this. Now it's your turn to stand on the board, and I'm going to play a song. Go ahead and dance with Keepon!" The experimenter tried to use similar words with every child, but allowed some variation to remain naturally conversational. The experimenter was not aware of the participant's group assignment prior to beginning each trial.

The children were divided into two conditions: Keepon dancing in synchrony with the rhythm of the music, or Keepon dancing in synchrony with the child's movement on the WiiFit. The design is illustrated in fig. 5.9, top. We measured the degree of synchrony between the child's rhythm and that of the music. One hypothesis was that children would follow the music more closely when Keepon synchronized to (and reinforced the rhythm of) the music; that they would exhibit a higher average energy level when Keepon synchronized to their movement; and that they would be more interested in continuing to dance with Keepon for another song when Keepon synchronized to their movement.

The music had been provided by the teachers in order to be familiar to the children, but it is unknown whether (or how recently) they had heard it.

5.3.1.4 Experiment 2: Modeling or imitating a varying rhythm

The second experiment was conducted 2.5 months after the first experiment, so that the children's previous experiences with the robot would not unduly bias their second trials. Between the two experiments, we also made improvements to Keepon's perception of rhythm from the WiiFit (adding the ability to handle lateral periodic movement; increasing the weight normalization window from 1 second to 2 seconds; and earlier detection of bounce onset).

The song used in this experiment was played by the teachers during circle time the day before the experiment so that the children would be more immediately familiar with the music. This song differed from the music in the first experiment in that the tempo changed over the course of the song, which consisted of alternating fast and slow sections. The children were once again divided into two conditions: with Keepon dancing in synchrony with the rhythm of the music or in synchrony with the child's movement. However, in this experiment, we made the roles more explicit to the children. The children in the first condition were told, "Keepon is going to show you how to dance to this song, so try to dance like he does." The children in the second condition were told, "Keepon doesn't know this song, so try to show him how to dance to it." In this way, we reinforced Keepon's behavior with an expectation in the children that they should either follow or lead Keepon in the dance. The design is illustrated in fig. 5.9, bottom.

In the first condition, during the final slow section of the song, we had Keepon dance (for short periods) at thrice and then at twice the frequency of the music. By creating in the robot a physical rhythmic stimulus at odds with the auditory rhythmic stimulus, we hoped to determine whether children tended to follow the robot or the music. Our main hypotheses were the same as in Experiment 1, with the additional question of whether children in the first condition were more likely to synchronize with the music in the role of leader or follower.

5.3.2 Results

Our familiarization strategy was successful insofar as all the children in the classroom engaged with the robot and seemed to enjoy themselves, very likely due to the comfort and mutual encouragement of the group setting. The familiarization, the change in age group (from 3 years in section 5.2 to 5 years in these experiments), and the school setting also helped us to achieve a sufficient comfort level during the experimental trials themselves, as all children (except one during Experiment 2) stood and moved on the WiiFit for at least one song per session.

5.3.2.1 Experiment 1

In this experiment, contrary to our hypothesis, we found no differences between the two conditions with respect to children's energy level ($F(1,18) = 0.0307, p = 0.8628$, fig. 5.10, top left). We found trends, not statistically significant, in average frequency (marginally higher when Keepon synchronized to the child, $F(1,18) = 2.5997, p = 0.1243$, fig. 5.10, top right) and duration of synchrony with the music (marginally higher when Keepon synchronized to the child, $F(1,18) = 1.9071, p = 0.1842$, fig. 5.11, left).

However, we did find a significant difference in retention—the children's preference to continue dancing to a second song. Contrary to our hypothesis, we found significantly higher retention when Keepon had been dancing to the music than when it had been dancing to their movements ($\chi^2(1, N = 20) = 13.33, p = 0.003$). Perhaps this suggests that the interaction was more engaging when Keepon danced to the music because this was more "appropriate" or "intelligent" behavior than pure imitation. However, this significant result was not replicated in our second experiment.

5.3.2.2 Experiment 2

In this experiment, we found some trends that are worth noting. Children danced with a higher average energy level ($F(1,16) = 0.6270, p = 0.4400$, fig.5.10, bottom left) and higher average frequency ($F(1,16) = 0.6210, p = 0.4422$, fig.5.10, bottom right) when Keepon was dancing to their movements than when it was dancing to the music. Although these trends were not statistically significant, they suggest that the robot's rhythmic contingency to their movement may have resulted in more engagement or exploration, particularly since the frequency measure in Experiment 1 exhibited the same trend.

We did not replicate the significant difference in retention between the conditions. There was, in fact, a within-subjects correlation between children who asked to dance to a second song in the two experiments (although they had been randomly re-assigned to different conditions). We might attribute this result to an artifact of our group assignment

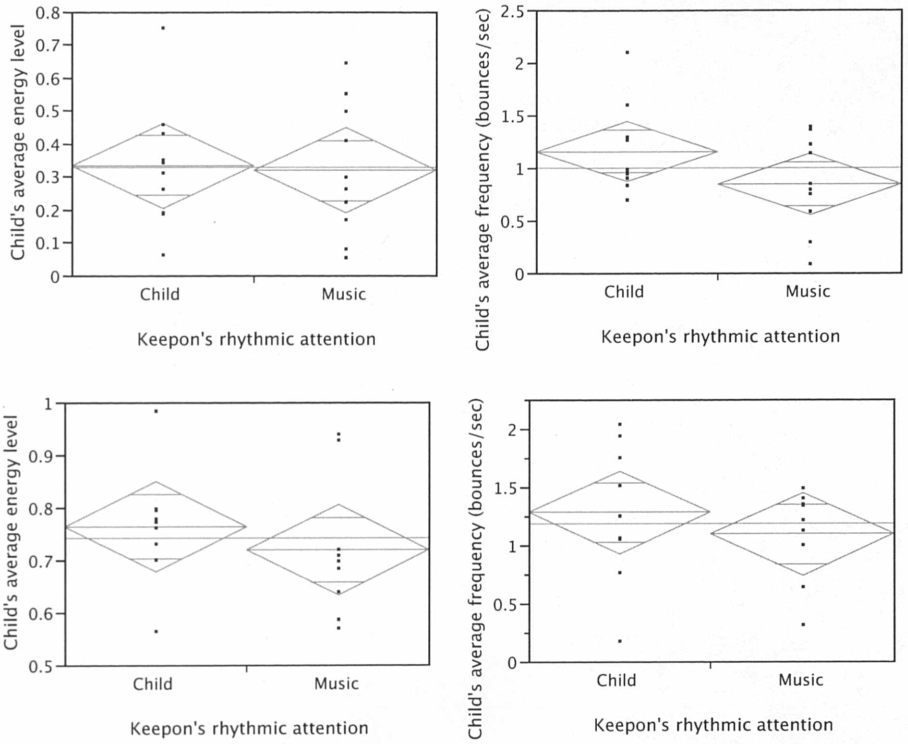


Figure 5.10: Activity-level measures of energy level (left) and average frequency (right) in Experiment 1 (top) and Experiment 2 (bottom).

or to a memory effect between the two experiments (i.e. if participants remembered their first experiences with Keepon).

We did find a statistically significant effect of Keepon's rhythmic attention on children's synchrony to music. When Keepon was dancing to the music, children spent significantly more time in synchrony with the music than when Keepon was dancing to them ($F(1,16) = 8.7768, p = 0.0092$, fig. 5.11, right). This effect is even stronger considering that Keepon actually deviated from the music's rhythm for several seconds during the second half of the song (in the condition where it was dancing in synchrony with the music). Indeed, in the condition where children followed Keepon, they spent significantly more time synchronized with Keepon than with the music during the period of Keepon's deviation (as determined by a matched-pair t-test, $t(8) = 4.8103, p = 0.0013$)—the second half of the child data in fig. 5.12 illustrates an example of this. This suggests that children indeed followed Keepon's rhythm in the condition where we asked them to do so, and that by following the robot, they more closely followed the music than those children

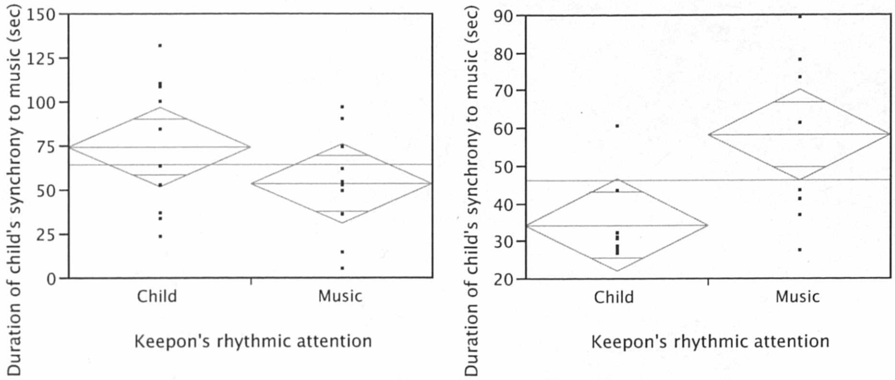


Figure 5.11: Experiment 1 (left): Children spent slightly more time synchronized with the music when Keepon was synchronized to them than when it was synchronized to the music. Experiment 2 (right): Children spent significantly more time synchronized with the music when Keepon was synchronized to the music than when it was synchronized to them (and they were instructed either to follow or to lead).

who were asked to demonstrate to the robot.

One child in the second experiment was very shy and did not participate in the activity, and two more children had to be excluded from the results after the experimenter incorrectly told them whether they should be leading or following the robot. However, one of these children correctly pointed out during the experiment that the robot was not following her, as it was expected to; this validates that the children could perceive the relationship between the robot's dancing and their own.

5.3.3 Discussion

We have presented a dance-based experimental protocol for studying the effects of rhythmic attention on engagement and synchrony, and the results of two experiments we have carried out under this protocol. It is worth drawing attention to the fact that our group familiarization sessions served to make the children comfortable with both Keepon and the experimenter and set up an understanding or expectation that Keepon is aware of its environment, has intentions and emotions, and has both an ability and a strong motivation to dance. We believe that such a protocol is a reliable way of making a group of children comfortable without biasing them before conducting "lab-like" studies with individual children on a one-on-one basis.

We believe that our first experiment was unsuccessful in revealing differences between the two conditions in part because of technical reasons, e.g., Keepon's perception of rhythmic movement was inaccurate when compared to our evaluations of children's actual

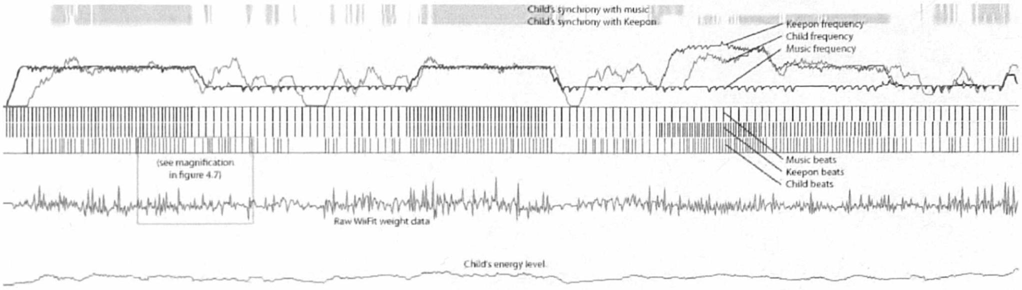


Figure 5.12: An illustrated song session for one example child in Experiment 2, condition 2 (Keepon dancing to the music). A number of data streams are represented in parallel, starting at the top: The first row of grey bars marks periods of synchrony between the child and the music; the second row of grey bars marks periods of synchrony between the child and Keepon (in this case, different only when Keepon deviates from the musical rhythm in the second half of the song). Below this, the black line marks the music's frequency; the blue line marks Keepon's frequency; and the red line marks the child's frequency. Below this, the black ticks show the music's beats; the blue ticks show Keepon's beats; and the red ticks show the child's beats. Below this, the red line shows raw weight data from the WiiFit. Below this, the red line shows the energy level of the child's movement on the WiiFit. Duration is 2.38 min.

dancing behaviors.

Another reason the first experiment failed to find differences may be that the song we had selected had a consistent rhythm that either became boring to the children or did not sufficiently make tempo a salient property of the interaction. For the second experiment, we selected music that had clear changes in tempo (and even referred to speed in the lyrics) so that we could more clearly see the effects of changing tempo on the interaction.

Finally, our protocol in the first experiment may not have provided sufficient context for the children. By making the roles (of "leader" and "follower") explicit in the second experiment, children had a clearer goal in the interaction.

It is still unclear how best to measure engagement in such interactions. Asking children's preference to dance to a second song seems to be a good measure, but we failed to replicate a significant result from the first experiment in the second.

We are encouraged by the result that Keepon's attention to the rhythm of the music, and establishing the role of "follower" in the children, indeed made them follow the rhythm shared by Keepon and the music. The fact that children spent less time in synchrony with the music when they were "demonstrating" to Keepon may be the result of further problems with Keepon's perception and imitation of the children's rhythm, since we expected that children in this condition would try harder to demonstrate the correct rhythm. While the regularity of musical rhythms can be a useful constraint, we would also like to develop dyadic interactions (closer to real dialogue) that involve turn-taking and do not involve music, since the ultimate goal is to create interactions that involve shifting roles of leader and follower between robot and human.

5.4 Summary

We have presented a robot that is able to perceive, using a variety of sensors, rhythmic properties of the environment and to generate synchronized rhythmic movement. We first situated the system in unconstrained interactions with hundreds of children in which dancing was the target activity. Our analysis of these interactions suggested that synchrony in the robot's behavior had a motivating effect on people's rhythmic behavior.

Next, we created an improved accelerometer-based perception system and designed a strictly controlled experimental paradigm in order to explore differences in the effects of the robot's rhythmic attention. The difficulties we faced in this study led us to design an experiment that was better suited to encouraging the kind of behavior we wished to study. In addition to the use of pressure sensors and an older participant population, our revisions to the experimental protocol—namely, an extensive introduction or familiarization session and a less constrained but still consistent trial protocol—were important in obtaining data that was able to be compared between participants to obtain quantitative measures of behavioral differences. In two experiments, we found: that children preferred to continue the activity when Keepon synchronized with the music; that they may have been more active when Keepon synchronized to them; that children spent more time in synchrony with the music when they were following Keepon's lead; and that they spent more time in synchrony with Keepon than with the music when Keepon deviated from the music.

Such dance-based interactions are useful in allowing us to develop rhythmic perceptual and behavioral technology. The constraints of musical structure and regularity, and the exaggeration of movement in dance, make it easier to measure, compare, and analyze rhythmic behavior. However, dance also carries a specific cultural context that may limit the portability of our findings to unconstrained social interaction, and it fails to provide a context of goals, intentions, and motivations that are a feature of many of our social interactions. The game-oriented experimental design presented in the following chapter is meant to broaden the scope of our work by examining interaction in a different social domain.

Chapter 6

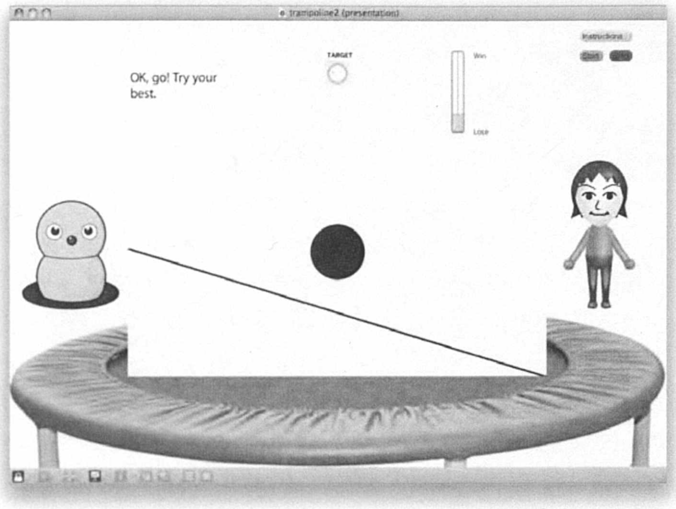
Cooperative Game Interaction

The previous chapter described a series of studies in which Keepon's interactions with children were open-ended, dance-oriented, aesthetically motivated, and mediated by music. The focus of those studies was the relationship between Keepon's rhythmic attention and the roles of "leader" or "follower" in the interactions. We have discussed various limitations of this context and our desire to expand our inquiry to different type of interaction.

In this chapter, we examine the relationship between rhythm and leadership roles from another angle. We describe a video game we have designed to study the effects of rhythmic leadership on task performance. The human participant must cooperate with Keepon by bouncing in synchrony on a virtual "trampoline" in order to score points. Throughout the game, the robot shifts its physical and rhythmic attention, there are changes in access to game state information, and players exchange roles of leader or follower. We describe an experiment in which we used this game to explore the effects of rhythmic synchrony and these changing roles on quantitative measures of task performance. We found that participants could indeed assume the roles of leader or follower in a rhythmic interaction, that participants performed better when they were following the lead of the robot, and that the particular transitions between different roles had specific effects on the time participants took to become accustomed to a new role.

6.1 Game design

After our prior work involving music and dance, we desired an interactive context that involved rhythmic behavior, had a measurable, quantifiable objective, and required synchrony in order to succeed. We considered several real-life athletic and cooperative physical activities that would be amenable to representation as a video game. We desired



OK, go! Try your best.

Figure 6.1: The interface for the trampoline game. This depicts a moment in which the human is bouncing (his or her side of the trampoline is at its lowest point), and Keepon is halfway between two bounces (its side of the trampoline is at its highest point). Since the players are out of phase with one another, the ball would not bounce high enough to hit the target.

that Keepon and a participant would have to cooperate in order to score points, and that the game would involve simultaneous rhythmic behavior as well as turn-taking. The central conceit of the game we eventually created is that both players are standing on a trampoline, one on either side. A ball is bouncing up and down in the middle of the trampoline. Above the ball is a target, and points are scored when the ball hits the target. The ball bounces high enough to hit the target only when the surface of the trampoline is oscillating in a horizontally level attitude, that is, when the two players are bouncing in synchrony. Furthermore, the ball changes size/mass over the course of the game, which changes its rate of bouncing and therefore determines the rate at which the players must bounce. We implemented the game in Max/MSP, which allowed us to easily create an animated depiction of the game elements (ball and trampoline surface), interface elements, and connections to our existing robot control and perceptual systems. The interface is illustrated in figure 6.1.

In our game interface, the surface of the trampoline is depicted as a line connecting two endpoints, each of which represents the height of the trampoline's surface under one of the players. Keepon's side is on the left and the person's side is on the right. Two icons representing the players hover above their respective sides of the trampoline. Each player's side of the trampoline starts at its lowest point, begins moving upward when its corresponding player "bounces" (in Keepon's case, when it compresses; in the human's case, when he or she rapidly bends his or her knees), eventually reaches its peak, and then falls back to the bottom. Each endpoint's motion is described by a parabola, such that it rapidly ascends, lingers at its apex, and rapidly descends in a physically realistic manner. The period of the endpoint's complete parabolic cycle is determined by the time interval between the previous two jumps. That is, when a player is bouncing at a steady or slowly changing tempo, his or her side of the trampoline will return to its lowest point at approximately the same time as his or her next bounce. If the person bounces before it reaches its lowest point, it begins a new period and immediately starts moving up. In this way, the rate or tempo of the trampoline's oscillation matches that of the player's bouncing.

The ball bouncing on the trampoline follows a motion trajectory that is also parabolic, like those of the sides of the trampoline. The ball's speed, or the period of the ball's parabolic trajectory, is determined by the ball's size: fast when the ball is small (and has low perceived mass), and slow when the ball is large (and has high perceived mass). The temporal and physical start point of the trajectory is determined by the ball's impact with the trampoline surface. Upon being launched at the start of the game, the ball falls until it crosses the midpoint of the line representing the trampoline surface, at which time it begins its upward trajectory. The ball's period varies from .75 seconds between bounces (when the ball is at its smallest) to 1.5 seconds between bounces (when the ball is at its largest). Over the course of the game, the ball size changes according to the following repetitive sequence: 1 second/bounce for 5 seconds; decrease to .75 seconds/bounce over 5 seconds; increase to 1.5 seconds/bounce over 10 seconds; decrease to 1 second/bounce over 7 seconds. Therefore, over a period of 27 seconds, the ball shrinks and grows in a piecewise linear manner.

The target (which is considered to be "struck" when the ball overlaps it) is positioned just below the height reached by the ball after having struck the trampoline surface when at its maximal height. Therefore, the optimal frequency at which the players should bounce is that of the ball's period (as determined by its size), and the optimal phase (or moment at which the players should bounce) is when the ball reaches its apex. In other words, the person should be exactly out of phase with the ball. If the person bounces twice as fast as (but still in phase with) the robot, they would still succeed in striking the target; likewise, if they bounce at half the frequency, they would strike the target half the time.

In this game, we perceived participants' rhythmic bodily movement using the WiiFit (fig. 6.2). As described in section 4.2.4, we detect "bounces" by interpolating between these four pressure sensor values, smoothing and normalizing the data around the person's average weight, and detecting increasing zero-crossings in the weight data (as illustrated

Figure 6.2: The Nintendo Wii Balance Board, used for rhythmic perception.

in fig. 4.7). The bounce is recognized as the person begins to apply downward force and before they reach their lowest (or heaviest) point.

Meanwhile, Keepon has full access to the game state and is capable of bouncing such that, if the robot were in control of the entire trampoline, the surface would be positioned to strike the ball at the optimal time to achieve maximum height. However, since the two partners act independently on the two sides, and trampoline needs to be level for the surface's midpoint to reach its highest point, it is necessary for both players to bounce in a coordinated rhythm for the ball to strike the target.

In this experiment, we constrained Keepon's rhythmic movement to bouncing, a natural counterpart to the movement that participants would perform. This movement is very salient both visually (due to the rapid contraction and widening of the robot's body) and acoustically (due to the sound made by the motor responsible for compressing the body). Furthermore, bounce commands are executed with minimal latency and for our purposes may be considered simultaneous when triggered in response to a detected beat. In this experiment, the robot's attention was directed either towards the participant or towards the computer display, as described below. These two locations were determined by the experimenter prior to the experiment, based on the positioning of the robot, the display, the WiiFit, and the average height of the participants.

A score bar in the interface, which is initially half-full at the start of the game, grows when the ball hits the target and shrinks when the ball misses the target. "Winning" or "losing" corresponds with the bar becoming full or empty. The rate of change of this bar is not constant; that is, earlier hits and misses result in less rapid change than later hits and misses, in order to enforce a minimum playing time (of approximately 1 minute) for very good or very bad players, while ending the game within a reasonable amount of time (approximately 3 minutes) for players who were roughly balanced in hits and misses.

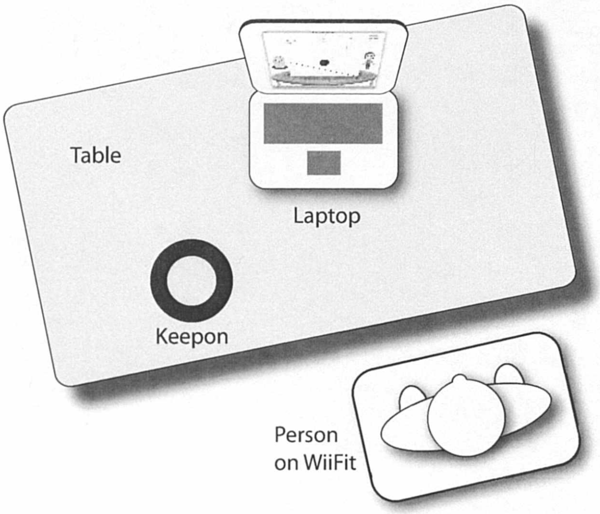


Figure 6.3: The configuration of Keepon, child, and laptop in the trampoline game.

6.2 Experimental Design

We designed an experimental procedure for studying the effect of different roles and selective availability of information on task performance in the trampoline game. Our participants were 35 middle- and high-school students, between 11 and 18 years of age, who were attending the C-MITES and Andrew's Leap summer programs at Carnegie Mellon University. IRB approval was obtained for the study, and we requested that parents sign consent forms for the students' participation.

The game software ran on a laptop on a table in a room near the students' classroom. Keepon was positioned to the left of the laptop in such a way that it could look at the laptop screen. The WiiFit was positioned on the floor in front of and to the right of the laptop, and the participants were instructed to stand on the board. The robot and the human thus made a triangle with the laptop and shared their attention toward it (fig. 6.3). The two players were also closest to their respective sides of the trampoline interface displayed on the screen.

A video camera captured a view of Keepon, the participant, and the laptop screen (fig. 6.4).

6.2.1 Introduction & training

Participants were brought into the room one at a time by the experimenter, who explained that we were designing a video game and that we would like the students to try it. The

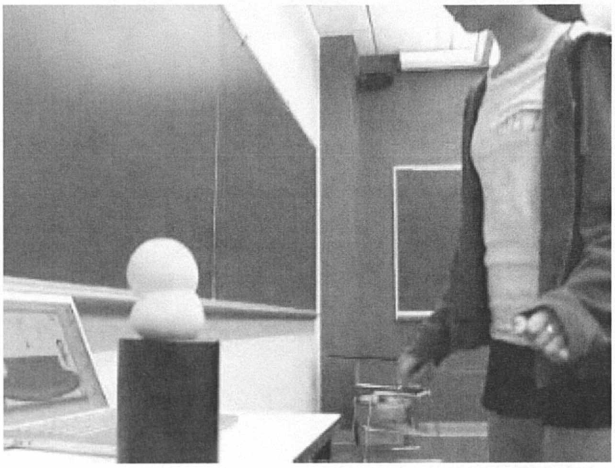


Figure 6.4: Keepon and a child playing our trampoline game.

experimenter introduced Keepon (which was looking at the student) and explained the premise of the game: that the participant should imagine being on a trampoline with Keepon and that the goal would be for them to cooperate in causing a ball to strike a target. A brief instruction period, lasting approximately one minute, consisted of a step-by-step explanation of the relevant game elements:

1. Keepon looked at the screen and proceeded to bounce, controlling the entire trampoline surface (with only Keepon's icon visible) and making the ball hit the target perfectly. The experimenter explained that Keepon's bouncing was causing the trampoline to move, and that when the robot bounced at the correct time, the ball bounced high enough to hit the target.

2. Keepon looked at the experimenter, who stood on the WiiFit and proceeded to bounce as Keepon had. At this point, the experimenter was controlling the entire trampoline surface (and only the human icon was visible). The experimenter demonstrated the most effective way of bouncing, that is, bending and then straightening the knees. The experimenter descended from the board and invited the participant to stand on it and practice until he or she hit the target a few times.

3. The experimenter switched the game to the mode where each player controls one side of the trampoline, and explained that it is necessary for both players to "work together" to make the ball hit the target. Keepon looked at the screen and bounced perfectly, but the experimenter did not explicitly state that the robot was bouncing perfectly. The experimenter let the participant practice until he or she successfully cooperated with Keepon to hit the target a few times.

4. A gray box covered the view of the trampoline. The experimenter explained that the view of the trampoline would sometimes be blocked in this way, and that the participant should then follow Keepon's lead in playing the game.

5. The occluding box disappeared, and Keepon turned to look at the participant. Keepon bounced whenever the participant bounced. The experimenter explained that Keepon would occasionally turn towards the participant in this way, and that at this time the robot would be following his or her lead.

6. The ball shrunk and grew, changing its speed (and rate of bouncing, as described above) as it changed size. The experimenter explained that the ball would change size and speed in this manner over the course of the game.

7. The score bar grew and shrank with hits and misses, and the experimenter explained that the game would end when it became full or empty.

8. The experimenter notified the participant that the game would start, and encouraged him or her to do their best.

6.2.2 Conditions

The experiment was a within-subjects study in which each participant, over the course of a single game, was alternately confronted with one of three different modes of play:

- *Together*: Keepon looked at the screen and bounced "correctly," that is, at the ball's apex. The trampoline surface was visible to the participant. In other words, the players acted independently and each had full access to the game state. This was step 3 in the training session, although Keepon's rhythmic attention or correctness had not been made explicit.

- *KeeponLead*: Keepon looked at the screen and bounced correctly, but the participant's view of the trampoline surface was blocked by the gray box. In other words, the players acted independently, but only Keepon had full access to the game state. This was step 4 in the training session, when participants had been told that they should follow Keepon's lead.

- *KeeponFollow*: Keepon looked at the participant and bounced when he or she did. The trampoline surface was visible to the participant. In other words, Keepon's behavior was dependent on the person's, and only the person had access to the game state. This was step 5 in the training session, when participants had been told that Keepon would follow their lead.

These modes were presented to the participants in "epochs" or time periods of length 13 seconds. For the first 15 participants, the game controller cycled between them in the order *Together*, *KeeponLead*, *KeeponFollow*. For the remaining 20 participants, our design evolved such that the game controller cycled between the modes in random order. Meanwhile, as described above, the size and speed of the ball changed with a 27-second period. This difference in period lengths ensured that the ball was of unpredictably different sizes during different epochs in the game.

6.2.3 Hypotheses

We expected that there would be performance differences between the different modes of the game, even though task performance is entirely dependent on the participant's performance (as Keepon behaves either optimally or under the person's control), and the relevant information about the appropriate time to bounce (at the apex of the ball's trajectory) is available to the participant at all times. The only differences between the three modes are the visibility of the trampoline surface and Keepon's rhythmic attention (whether it is bouncing according to the ball position or the participant's bouncing) as communicated by Keepon's gaze. It should be noted that the experimenter, after some practice with the game, can play it perfectly in any mode, but it is (by design) difficult for novices.

Specifically, we expected that the game would be easiest in the *Together* mode, when the person is fully aware of what is happening in the game (including learning the effects of their own movements on the trampoline surface) and can follow either the ball position or Keepon's movement. Next, we expected good performance in the *KeeponLead* mode, when Keepon bounces at the correct time and provides a salient cue for participants to follow, but the participants do not have access to direct feedback from the movement of the trampoline surface to learn the effects of their own movement. Finally, in the *KeeponFollow* mode, we expected that participants would have a harder time determining the appropriate time to bounce based on the movements of the trampoline surface and the ball.

Second, we expected that there would be changes in performance within an epoch. That is, a switch to a new epoch can be expected to initially throw a participant off until he or she becomes accustomed to it, so later performance within an epoch would be greater than earlier performance—an increasing learning curve. We expected to see this improvement in all three modes, but that it would not be as rapid in the KeeponFollow mode since it is more difficult.

Third, we expected that there would be differences between the transitions from one epoch to another. The mode of the previous epoch could have an effect on the current

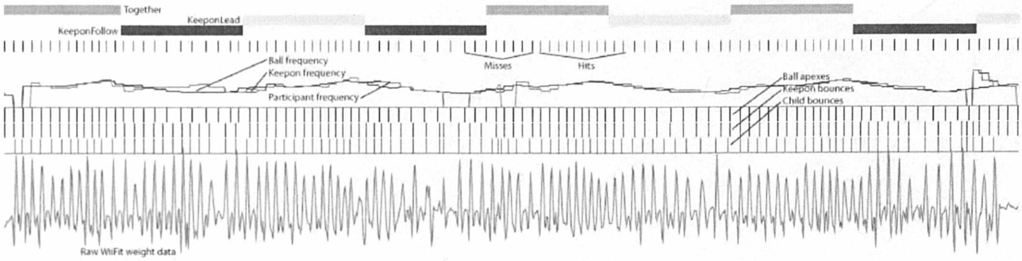


Figure 6.5: An illustrated game session for one example participant. A number of data streams are represented in parallel, starting at the top: green bars mark epochs in the Together mode; yellow bars mark epochs in the KeeponLead mode; and blue bars mark epochs in the KeeponFollow mode. Below this, black ticks show misses, and red ticks show hits, between the ball and the target. Below this, the black line marks the frequency of the ball's bouncing; the blue line marks the frequency of Keepon's bouncing; and the red line marks the participant's frequency. Below this, the black ticks show the ball's apexes; the blue ticks show Keepon's bounces; and the red ticks show the participant's bounces. Below this, the red line shows raw weight data from the WiiFit. Duration is 1.8 min.

epoch because the participant could be expected, in the beginning of an epoch, to continue the rhythm that he or she was using in the previous epoch. The disruption to the participant's rhythm may be differently affected by the changes in available information, in Keepon's attention, or in both.

Finally, we expected that there would be overall improvement in performance for particular modes or transitions over the course of the entire game.

6.3 Results

Games were, on average, about two minutes long (M=123.9 seconds, SD=40.7 seconds), but ranged in length from 55 seconds to 196 seconds (or, in terms of epochs: M=8.8 epochs, SD=3.04 epochs, ranging from 4 to 15 epochs).

Figure 6.5 illustrates the recorded data (ball movement, Keepon's bounces, and WiiFit weight data) from a sample game. The participant follows the rhythm of the ball rather closely, but hits the target more frequently in the Together and KeeponLead modes, which is consistent with other participants, as we shall now discuss.

In processing the recorded information, we divided the participants' games into individual epochs (303 total for 35 participants). For each epoch we recorded:

- the mode of the epoch
- the number of times the player has previously seen an epoch of the same mode
- the mode of the previous epoch (i.e., the transition)

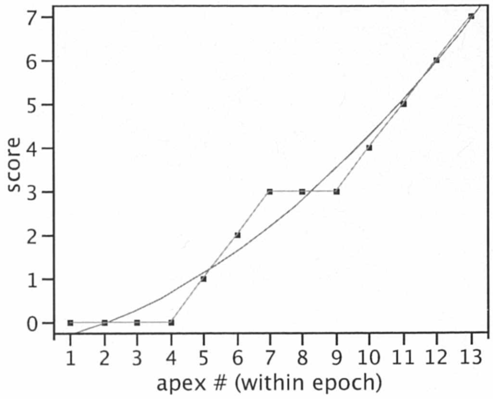


Figure 6.6: A sample epoch (in the *Together* mode, following a *KeeponFollow* mode). The straight dotted line represents the cumulative score function. The upward sloping curved line represents the learning curve.

- the number of times the player has previously seen this transition

Next, for each epoch, we noted the participant's record of target hits and misses for each apex of the ball (approximately 13 per epoch) in sequential order. Given such a sequence, we generate a plot of the player's cumulative score function in that epoch: after each apex, the score either remains the same in the case of a miss or increases by a point in the case of a hit. An example epoch is illustrated in fig. 6.6. We aligned all epochs on this event-by-event basis (i.e., so that the first apex in each epoch are considered together, then the second, and so on). In this way, for each epoch, we can calculate:

- the overall performance: the number of hits as a percentage of total apexes in the epoch, analogous to the first derivative of the cumulative score function
- the learning curve: the change in performance over the epoch, as the second derivative of a least-squares quadratic polynomial fit to the cumulative score function (if positive, performance improves over the epoch; if negative, performance decreases)

First, we compared overall performance between the three modes (fig. 6.7, top). A one-way ANOVA showed a statistically significant difference between the mean performance of the three modes ($F(2, 300) = 8.58, p = 0.0002$). A Student's pair-wise t-test showed that this significance came from the fact that performance in both the *Together* mode ($M = 42.64\%, SE = 1.93$) and in the *KeeponLead* mode ($M = 46.42\%, SE = 2.00$) was significantly higher than in the *KeeponFollow* mode ($M = 35.01\%, SE = 1.98$) ($p < .01$). The fact that the *KeeponFollow* mode resulted in the worst performance was in line with our hypothesis. However, the difference between the *Together* mode and the *KeeponLead*

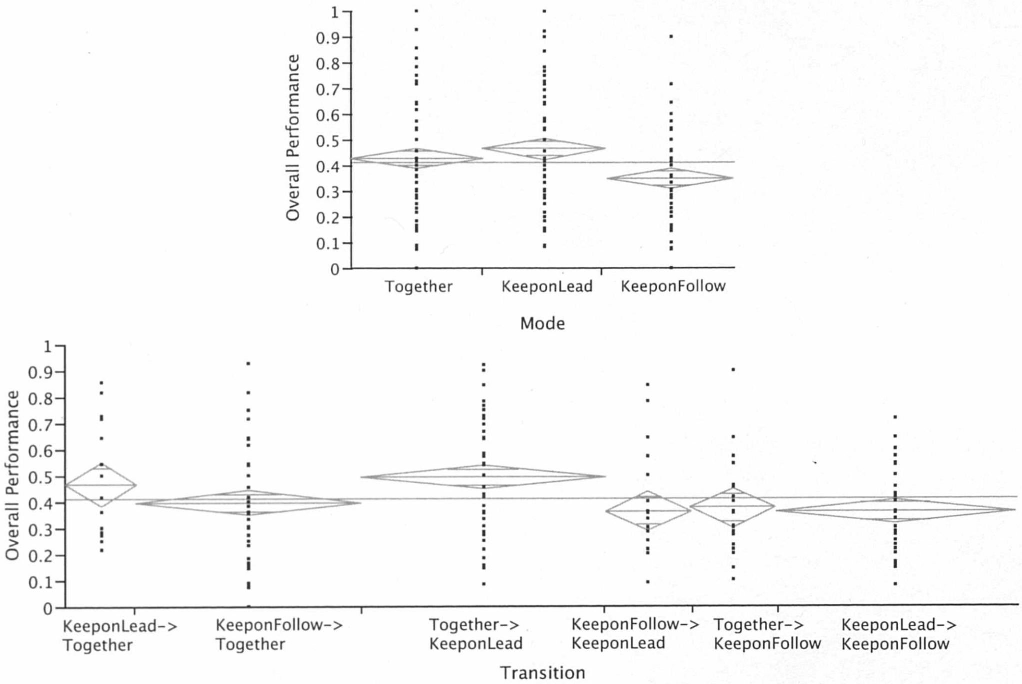


Figure 6.7: Top, overall performance for the three modes. Bottom, overall performance separated by the previous mode.

mode was not statistically significant, and performance was marginally better in the KeeponLead mode ($p = .174$), contrary to our expectations. This result suggests that it was easier to follow Keepon than to lead, and that the similarity between the two following modes (when Keepon's rhythm is dictated by the ball) dominated the difference in game state information between the two modes (the trampoline's visibility). The marginal difference suggests that participants might actually have been slightly distracted by the motion of the trampoline's surface (which is rapid and is the product of two independently acting systems), and that this reduced their performance.

Next, we separated out these performance measures based on the mode of the previous epoch (fig. 6.7, bottom). A one-way ANOVA showed significant differences between the mean performance in each of the six transitions ($F(5, 262) = 4.51, p = 0.0006$). For the *KeeponLead* mode, there was significantly higher performance if the previous epoch was *Together* ($M = 49.29\%, SE = 2.25$) than if the previous epoch was *KeeponFollow* ($M = 36.21\%, SE = 3.81$) ($p = .003$). This is reasonable, given that in the *Together* → *KeeponLead* transition, Keepon continues to provide the correct rhythm, which the player would already be used to, and the unnecessary distraction of the trampoline is

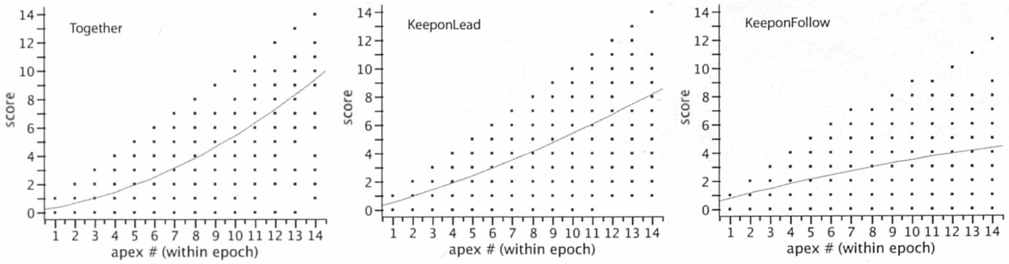


Figure 6.8: Average learning curves for the three modes. (Dots represents all apex-score data points for all participants.)

removed from view. For the *Together* mode, there was marginally better performance if the previous epoch was *KeeponLead* ($M = 46.72\%, SE = 4.18$) than if it was *KeeponFollow* ($M = 39.61\%, SE = 2.34$) ($p = .139$). Again, Keepon's rhythmic leadership remains constant in the *KeeponLead* → *Together* transition, but the appearance of the trampoline might have a distracting effect. For the *KeeponFollow* mode, there was no difference based on whether the previous epoch was *Together* ($M = 37.55\%, SE = 3.81$)or *KeeponLead* ($M = 36.23\%, SE = 2.28$); since Keepon's changes in attention and rhythm are so salient, the change in game state visibility has minimal effect.

Next, we looked at the change in performance within epochs (i.e., the learning curves) by fitting quadratic polynomial curves to the cumulative score function of each epoch (fig. 6.8). The second derivative of these curves is a measure of improvement within an epoch. A one-way ANOVA showed significant differences between the mean improvement in each of the three modes ($F(2, 300) = 3.89, p = 0.0215$). A Student's pair-wise t-test showed that improvement in the *Together* mode ($M = .0088, SE = .0027$) was significantly greater than in both the *KeeponFollow* ($M = -.0015, SE = .0028$) and *KeeponLead* ($M = .0006, SE = .0028$) modes ($p = .009$ and $p = .038$, respectively). This suggests that in the *Together* mode, players required some amount of time to recognize Keepon's correct bouncing, and to understand and react to what was happening on the game interface. The difference between improvement in epochs of the *KeeponLead* and *KeeponFollow* modes was not significant. However, the *KeeponLead* mode had an improvement close to zero, suggesting consistent performance over the course of an epoch given the straightforward task of immediately matching and following Keepon's rhythm. The *KeeponFollow* mode actually resulted in negative improvement (a decrease in performance); this might be explained by initially successful rhythms (from a previous *Together* or *KeeponLead* epoch) being sustained into the beginning of the epoch, followed by a subsequent deterioration in performance due to the difficulty of playing without Keepon's assistance. Fig. 6.9, top, illustrates these differences.

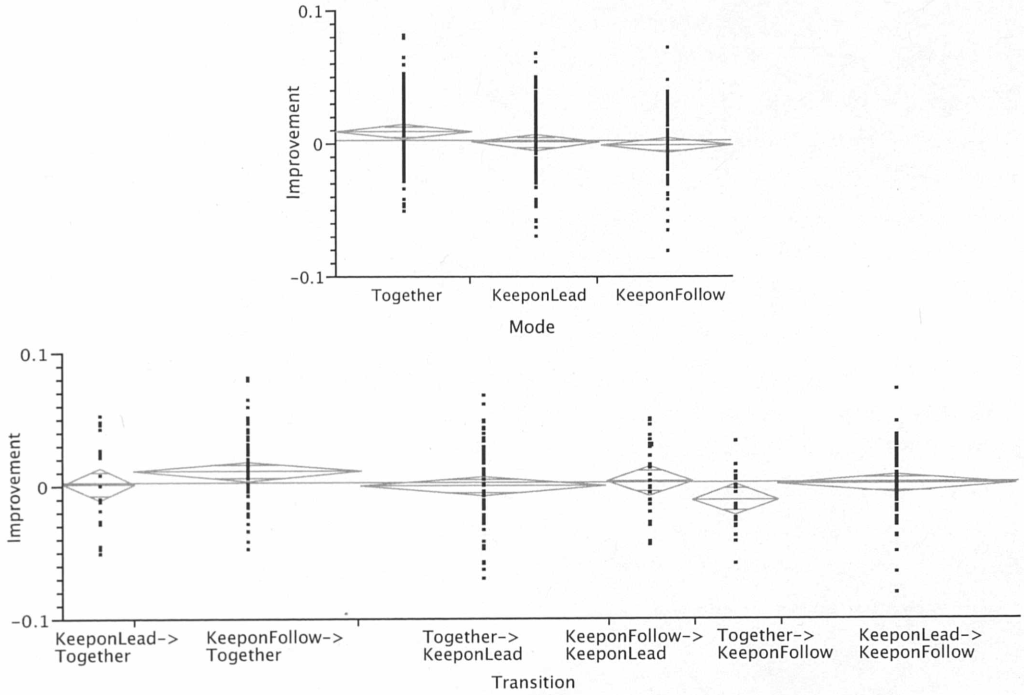


Figure 6.9: Top, within-epoch improvement for the three modes. Bottom, within-epoch improvement separated by the previous mode.

Next, we separated out these within-epoch improvement measures based on the mode of the previous epoch (fig. 6.10). A one-way ANOVA showed significant differences between the mean improvement after each of the six transitions ($F(5, 262) = 2.28, p = 0.0474$). However, the actual differences between the relevant pairs were insignificant, although we can look for suggestions in the trends. For the *KeeponFollow* mode, the improvement was marginally higher following a *KeeponLead* mode ($M = .0009, SE = .0035$) than following a *Together* mode ($M = -.0108, SE = .0058$) ($p = .0876$). This may be that the change in both Keepon's rhythmic attention (to following the person) and trampoline visibility required more time to get used to than a change in Keepon's rhythmic attention alone. For the *Together* mode, there was slightly greater improvement following a *KeeponFollow* mode ($M = .0107, SE = .0036$) than following a *KeeponLead* mode ($M = .0011, SE = .0064$) ($p = .1921$). This suggests that the rhythmic inertia was stronger than the maintenance of full game state access. For the *KeeponLead* mode, there was no difference following a *KeeponFollow* mode ($M = .0034, SE = .0058$) than following a *Together* mode ($M = -.0005, SE = .0034$) ($p = .5602$), consistent with the notion that it is easiest to immediately start following Keepon's lead when there is no additional distraction. Fig. 6.9, bottom, illustrates these differences.

A few other transition comparisons can help us to separate the effects of change in Keepon's rhythmic attention from the effects of change in trampoline visibility. The improvement in *KeeponFollow* → *Together* is slightly but insignificantly greater than that in *KeeponFollow* → *KeeponLead* ($p = .2927$), suggesting that when the person transitions from a leading to a following role, it takes a longer time to meet Keepon's rhythm when there is the visual distraction of the trampoline than when it is invisible. It may also be that in the *KeeponFollow* → *Together* transition it is less clear that Keepon has switched to the correct rhythm, and it takes more time for the participant to recognize this.

Additionally, the improvement in *KeeponLead* → *Together* is virtually the same as the improvement in *KeeponLead* → *KeeponFollow* ($p = .9774$), which means that getting used to the appearance of the previously hidden trampoline is not affected by whether the person is staying in a following role or transitioning into a leading role.

Finally, we analyzed changes in performance or learning curves over the course of the game, as each mode was seen a second, third, and fourth time (fig. 6.11). While we did not see significant changes in the learning curves, we noticed an increase in performance, and then a decrease, in the *KeeponLead* and *KeeponFollow* modes, perhaps suggesting an initial practice effect followed by confusion or boredom. The opposite happened in the *Together* mode, with a decrease followed by an increase. However, the changes were not statistically significant, and these analyses do not control for absolute time playing the game so far (which varied due to the ordering of the epochs.) We believe that the short duration of the game precluded any significant practice effect.

Nearly all participants reported enjoying the game when asked what they thought, but we expect that there was a typical bias towards satisfying the experimenter. More interestingly, when asked which mode was easier, most participants reported that it was easier to lead, even though their performance was better when following. This suggests a possible preference to be in control, or a confidence in their own ability relative to the robot's. Such considerations might play a role in designing behaviors for role transitions, such as exhibiting appropriate acknowledgment when requesting control as well as when ceding it.

6.4 Summary

We have presented a game-based experimental protocol for studying the effects of rhythmic attention and role switching on task performance and the results of an experiment we have carried out under this protocol. We believe that such a protocol is a good way of creating a constrained yet enjoyable rhythmic interaction that can allow us to explore the use of rhythm in the negotiation of social roles.

The results of this experiment point to several interesting conclusions. We found (consistent with our dance-oriented findings) that it was significantly easier for participants to follow a robot's rhythm than to figure out the game directly from the state as displayed on the screen, which is a promising direction of research for physical human-robot collaborative tasks. We might ask whether we would see the same effects if Keepon were animated on the screen rather than embodied. We believe that people interact qualitatively differently with physical artifacts that share a space with them than with images displayed on a screen, and that processing an animated version of Keepon would be more cognitively demanding (even if its movement would be spatially closer to the game depiction). One reason for this is auditory: not only is Keepon moving in the participant's direct or peripheral vision, but its motors are making noises that serve as rhythmic cues. Indeed, several participants volunteered that it was helpful for them to listen to Keepon even when they were focused on watching the game. Keepon's production of real-life visual and auditory cues was likely useful to participants because of, not despite, their transmission over different perceptual and cognitive channels than what was happening on the screen, and because of the social stance that people can be expected to take towards a robot like Keepon.

We also found that some of the transitions, or shifting leadership roles, had a significant effect on the "learning curve" or rate of improvement in subsequent phases of the interaction. This is important because it demonstrates that the negotiation of role-switching and the transfer of control is not immediate, but that there may be a delay that is highly dependent on the nature of the transition. In this experiment, we can attribute these delays to biological and cognitive limitations related to attention, distraction, and behavioral "intertia." On the other hand, the role transitions in this experiment were externally imposed rather than initiated by the participant (in which case we might expect that the initiator would not exhibit this inertia). In any case, it will be important that interactive robots, and researchers studying human-robot interaction, understand and take into account the temporal effects of these specifically human limitations—for example, to expect and account for delays in response when a robot requests acknowledgment, assistance, or an object from a person.

Furthermore, the effect of visibility or invisibility of the trampoline surface was dominated by Keepon's rhythmic attention, whether it was in a shift from leading to following or the other way around. There is also a tendency for participants to continue following a particular rhythm even after a change in game state, and we have found that this inertia is affected by the information available to the participant. All our results suggest that the visual presence of the trampoline surface increased participants' cognitive load: either it generally diminished their performance when present, or increased the amount of time necessary to adjust to its presence.

There are several changes we might make to this game based on our experiences. First, we would like to make it longer; despite our attempt to regulate game length by varying the rate at which the score bar changed, there was great variance in the length of games between participants, and most games were not long enough to observe an overall improvement through practice. In order for the game to be longer, it needs to be more interesting, as participants seemed to get bored towards the end of longer games. This could be achieved through better game interface design and attractive reinforcement of successful play, although this was beyond the scope of our work. We also might allow the ball to change size/mass more randomly or rapidly, although it would be important to carefully isolate these changes from the role transitions in post-analysis.

While the regularity of rhythms in this game is a useful constraint at this point, we would also like to develop dyadic interactions (closer to real dialogue) that do not involve an external rhythm (determined, in this case, by the ball). Our goals in future work involve creating interactions that involve more naturally shifting roles of leader and follower between robot and human. One interesting addition to this game would be to allow participants to choose when to switch between modes. This can be encouraged by selectively blocking access to other parts of the game state (not just the trampoline surface, but the ball or even the score as well) in order to encourage actively handing control to the robot, or degrading the robot's performance at certain points in order to encourage the taking of control from the robot. We expect that such transitions would be stimulated when the expected improvement in performance would outweigh the cognitive load of switching roles. Such modifications to the game would illuminate differences between self-initiated and externally imposed role transitions, and would make the game even more relevant to general questions of turn-taking and role negotiation in open-ended social interaction.

The data we have collected (and can collect) using this type of experiment is rich for further exploration. For example, when we analyzed performance differences between modes and transitions, and changes in within-epoch performance between modes and transitions, we treated success as binary: either the ball struck or missed the target. We have not yet accounted for how spatially close the ball has come to the target, nor have we taken the related measure of how temporally close the player has come to the optimal time to bounce. Studying these changes over the course of a game, between different modes, and after role transitions would illuminate more fine-grained temporal aspects of how people perceive and adjust to changing rhythms. A further innovation would be to combine such a study with gaze analysis, in order to relate people's changes in temporal rhythmic behavior to their shifting visual attention.

We are making progress in demonstrating the importance of rhythmic synchrony in social interaction by showing that simple synchrony to different social and environmental rhythms has a real effect on performance. We believe that these properties of non-verbal communication comprise a critical area of research for any robots that are expected to interact with people naturally and comfortably.

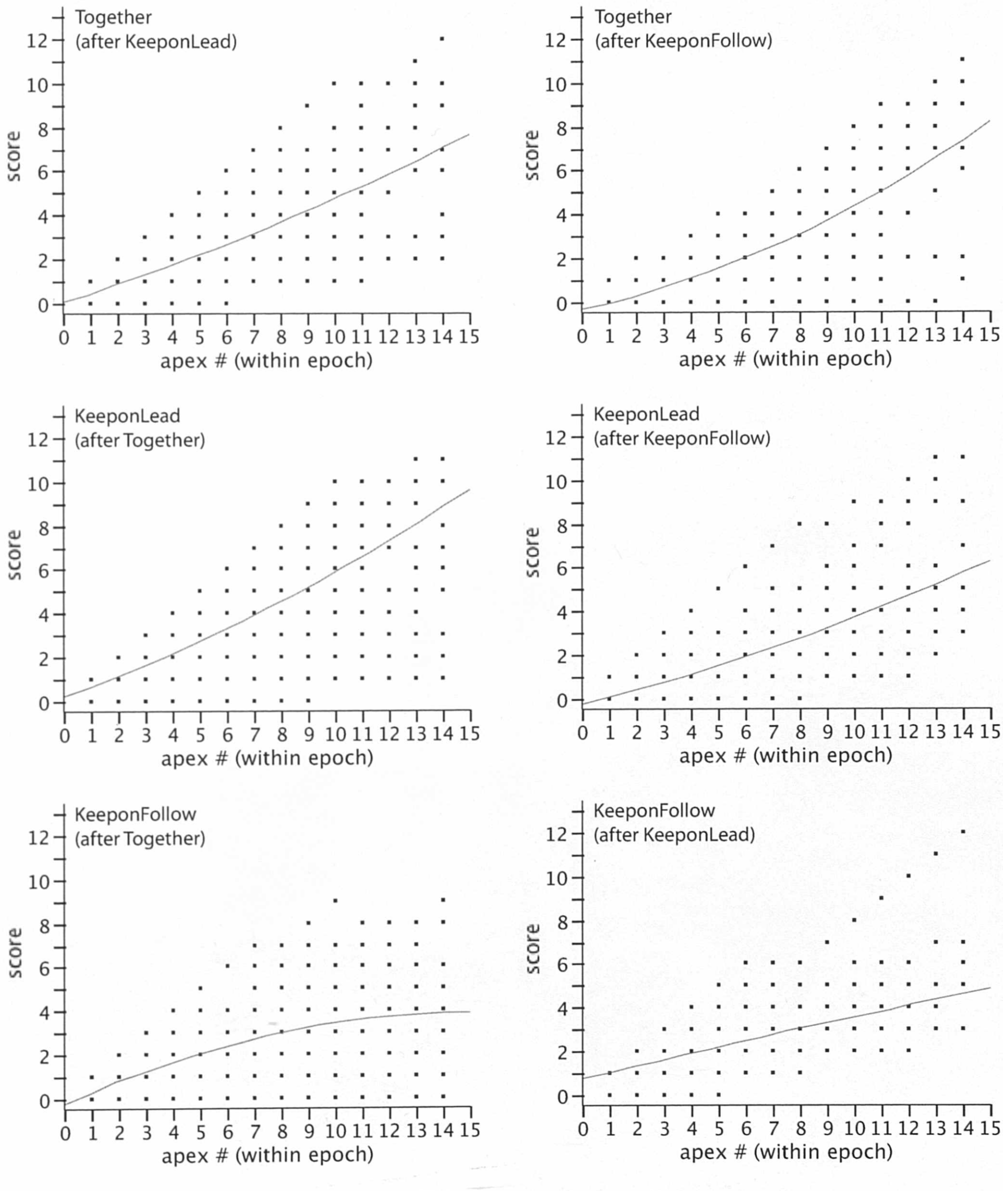


Figure 6.10: Average learning curves for the three modes, separated by previous mode.

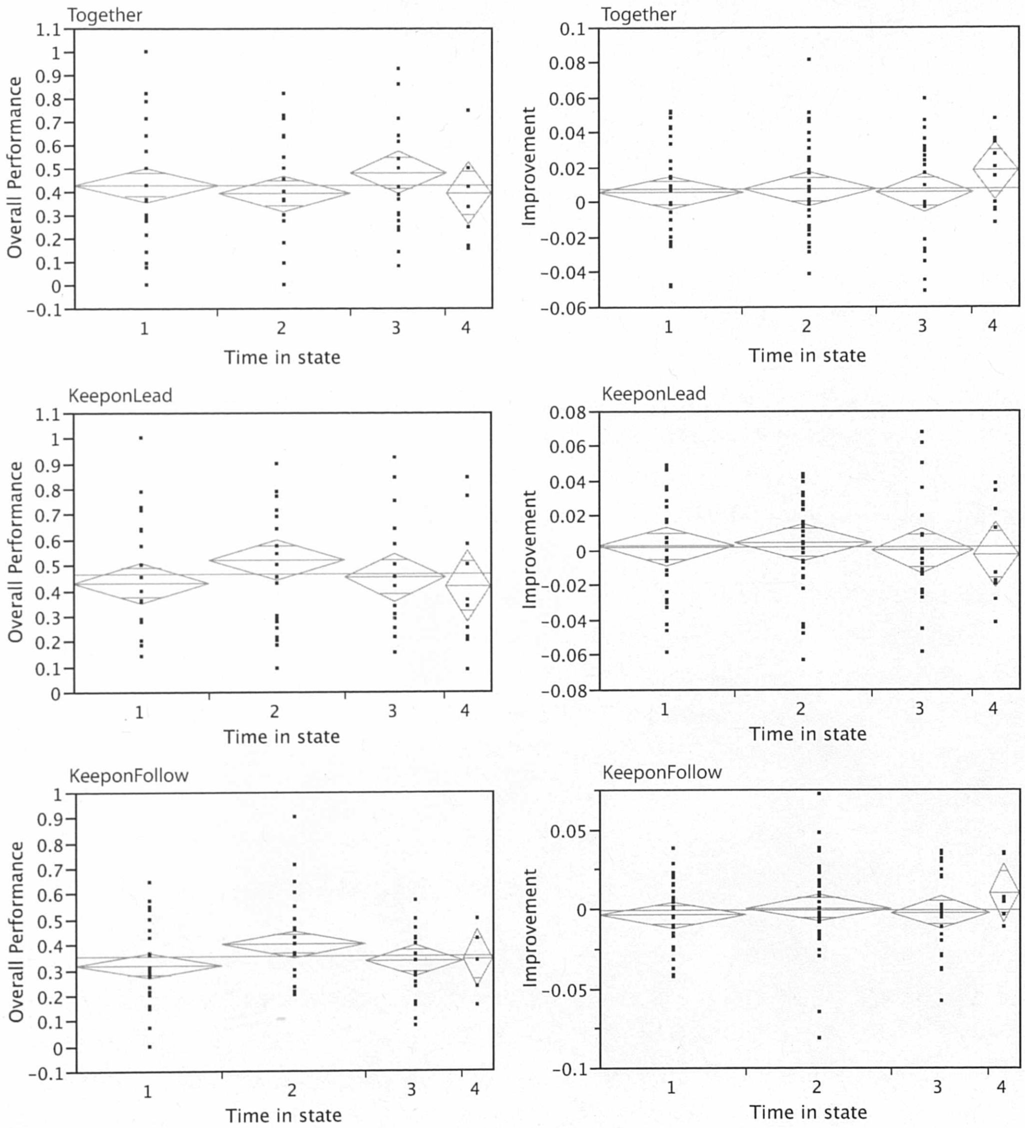


Figure 6.11: Changes in performance and improvement (learning curves) as modes were experienced multiple times.

Chapter 7

Conclusion

In this thesis, we have presented our motivation for exploring the development of rhythmic intelligence in interactive robots, proposed a framework for building this type of intelligence, designed technology to enable human-robot rhythmic synchrony, and conducted a series of experiments in which we observed and studied simplified examples of synchronous interaction.

We believe that the coordination or synchrony of subtle, nonverbal auditory and physical behaviors will be as much a prerequisite for comfortable, natural social interaction between humans and robots as we know it is between humans and other humans. Our work is intended as a starting point for developing the required technology and studying the properties of these social principles as they might apply to unconstrained human-robot interaction.

7.1 Technical contributions

Our proposed framework for rhythmic intelligence, described in Chapter 3, consists of a hierarchy of components to allow robots to participate in coordinated, synchronous interactions. First, robots should have *rhythmic attention*: perceptual capacities for detecting rhythms in multiple modalities and selecting those that are applicable to the interactive goal. Second, robots should be able to *entrain* to these rhythms: first comparing internal and external rhythms to understand the differences between them, and then taking action either to match an external rhythm or to try to influence it. Third, robots should be able to actively establish synchrony or asynchrony in accordance with the *interaction state*: to adapt their rhythmic behavior according to changing roles, goals, and situations. Our work has operated under simplifications to this framework; for example, rhythmic attention was pre-selected according to experimental conditions; we used simple methods of entrainment by matching either frequency or phase; and interaction state and leadership roles were controlled according to our simplified (dance- and game-oriented) interactions. However, we believe that this framework itself is flexible and general enough to accommodate rich expansion and adaptation to more complex interactions and interactive systems.

We have developed and integrated several technologies for the perception, generation, and synchronization of rhythmic behavior in a robot. These are described in Chapter 4. Our software is written in Max/MSP, a programming environment well-suited to such work due to its native tools and support for timing constructs, signal processing, datalogging, interface design, and hardware control.

Perceptually, we have used audio, video, accelerometers, and pressure sensors to detect environmental and physical stimuli. For each of these sensors, we have developed techniques for detecting periodic activities. We found that accelerometers and pressure sensors, in the form of the inexpensive and easy-to-use WiiMote and WiiFit game interfaces, were ideal devices for providing clean information about human movement while affording and encouraging playful interaction. We believe that pressure sensors in particular are a promising technology for detecting gross human bodily movement in studies of proximal human-robot interaction.

Behaviorally, we have implemented an architecture for rhythmic expression on the 4-DOF robot Keepon. Given a stream of "beats," our software drives each DOF in an oscillatory manner centered around the current direction of attention. A wide variety of dance "styles" are possible by changing several parameters for each DOF. Dancing is therefore an autonomous behavior that can be composed with other high-level commands such as the control of a teleoperator.

From a perspective of entrainment, we have explored three methods for matching a robot's rhythm to those in perceived stimuli: a direct reactive method that immediately responds to triggering by perceived or pre-labeled rhythms; a tempo-based method that looks for consistent spacing between beats and sets a metronome to the detected frequency; and a model of coupled oscillators that uses incoming beats to perturb a rotating phase point.

In parallel with our technology development, we have worked to design and refine experimental protocols for conducting studies in rhythmic human-robot interaction. This iterative process took us from early experimentation with a puppet (fig. 3.2) to unconstrained dance-oriented interaction in a public setting (section 5.1). When we later attempted to conduct a strictly controlled experimental protocol, we encountered difficulties in eliciting comfortable rhythmic behaviors (section 5.2). Our eventual dance-oriented protocol incorporated adequate familiarization and relatively flexible facilitation while still creating consistent participant experiences that could be compared across participants, conditions, and leadership roles (section 5.3). Our game-oriented interaction described in Chapter 6 was a within-subjects design that examined role-switching in a task with quantifiable performance and transitions between leadership and game state access. We believe that both the dance and video game concepts are flexible and promising experimental paradigms for studying rhythm, turn-taking, engagement, and non-verbal communication in human-robot interaction.

7.2 Experimental contributions

In our early pilot studies with Roillo, we found that the puppet's contingency to children resulted in more rhythmic behavior from the children than when it danced blindly. After implementing rhythmic perceptual and behavioral technology on the robot Keepon and situating it in an open house setting, we found that the robot's initial synchrony to music had a strong effect on children's likelihood to engage in rhythmic behavior. We also found that girls engaged in more rhythmic and touching behavior than did boys.

In our experiments on rhythmic leadership in dance play, we found trends that general activity (energy level and average frequency) were higher when the robot synchronized to the child than when it synchronized to the music. We also found that children could assume either the role of leader or follower in demonstrating to or imitating the robot. Children spent significantly more time synchronized with the music when Keepon was synchronized to the music (and they were told to follow it) than when it was synchronized to them (and they were told to demonstrate to it). Moreover, when children were in a following role and Keepon deviated from the music, they were significantly more synchronized to the robot than to the music (thus following its lead).

Our game design was motivated by questions about the effects of the robot's attention, participants' access to world information, and particular transitions between leadership roles on human performance in the task. It is of course expected that performance would be best when the robotic partner is behaving optimally, but we found that even when that was the case, participants actually performed worse when they had full access to the game state than when they only had Keepon to follow. We also obtained interesting results concerning the transitions between modes of the game. We found a rhythmic "inertia" that resulted in entrained rhythms being maintained even after transitions occurred. We also found that changes in the robot's rhythm dominated changes in the participants' access to game state. Moreover, the rate of improvement (or "learning curve") was different between different transitions. Our results suggest that the visual presence of the trampoline surface increased participants' cognitive load: either it generally diminished their performance when present, or increased the amount of time necessary to adjust to its presence.

Our game results are interesting because they demonstrate that rhythmic behavior resists change, and that transitions in turn-taking or role switching must account for these delays in human response. Furthermore, they suggest that, in performing a joint task, there is a benefit to rhythmic synchrony in that it can reduce the cognitive load of the participants. By entraining to each other, partners can to a degree relieve each other of attending and responding to environmental changes by themselves. These results point to interesting properties of social rhythmic synchrony that impact not just how we develop this technology, but also suggest interesting avenues of research in general human social behavior. We believe that the human behavioral tendencies we are identifying in these constrained rhythmic tasks can be instructive and suggestive of what we might expect to find in unconstrained interactions in the future.

Our hypothesis was that a robot's rhythmic synchrony to environmental or human rhythms, even if not consciously perceived by people, would have measurable effects on their behavior. Our experiments have validated this claim in a several ways. Although it was more challenging to demonstrate effects on "engagement" through measures such as retention and overall activity level (rather than more intensive behavioral analysis, for which we unfortunately did not have the resources), we did find support for the notion that rhythmicity supports various roles, such as leader and follower. We also hypothesized, and demonstrated, that people could assume these roles in interactions with a robot, and that the role of follower typically resulted in higher fidelity to the relevant environmental rhythms (music or task-related cues).

7.3 Public exposure

Keepon's dancing behavior has been featured in several videos that have become popular on the internet. Although not experimental in nature, the public's experience with and reactions to these videos, as mediated by the internet, is a form of human-robot interaction that is different from the co-present and embodied interaction normally studied by HRI researchers. Yet it serves as an instructive example to researchers interested in creating compelling and attractive social robots.

Keepon's earliest capability for rhythmic behavior was originally demonstrated in a video (fig. 7.1, top [Michalowski and Kozima, 2007b]) of the robot dancing to the song *I Turn My Camera On* by Spoon, a rock band from Austin, TX. In under one month, the video had accumulated over one million online views. We were then invited by WIRED Magazine to create a professional follow-up video to the song *Don't You Evah* with Spoon in Tokyo (fig. 7.1, bottom [Kozima *et al.*, 2008]). We later collaborated with Carnegie Mellon University and Daniel Wilson on two promotional music videos [Michalowski *et al.*, 2008], and with the KAIST PES Design Lab on a story in which Keepon participates in traditional Korean "Pungmulnori" dancing [Michalowski *et al.*, 2009a].

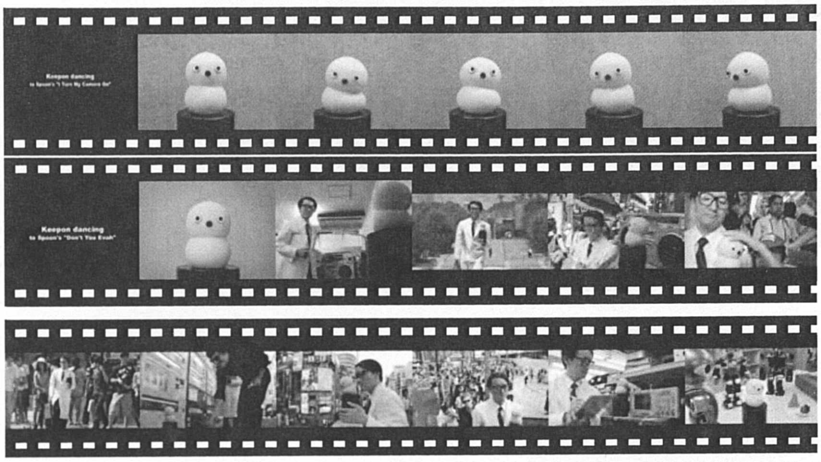


Figure 7.1: Frames from "Keepon dancing to Spoon's *I Turn My Camera On*" (top row) and "Keepon dancing to Spoon's *Don't You Evah*" (second & third rows)

These were not the first videos of robots, dancing robots, or even of Keepon that were available online, but they received wide publicity. The thousands of comments posted by viewers on these videos are nearly unanimously positive in their comments not just about Keepon's appearance but also about the attractiveness of its dancing ability (sometimes in comparison to their own). Dozens of viewers have found the videos compelling enough that they have created response videos. These are interesting from a sociological perspective, as they range from video "remixes" (in which the musical track is replaced), to "imitations" using foam or marshmallow depictions of Keepon, to full human enactments of the robot's dancing.

Keepon's popularity has given us the opportunity to exhibit it through a variety of channels, including WIRED's NextFest, NBC's *Today Show*, and Pittsburgh's WTAE. In the latter filming, we brought Keepon to a local facility for interaction with two teenage boys with autism. We configured the robot for teleoperated interaction in an unconstrained setting and had an interesting experience with the therapeutic potential of robotic rhythmic synchrony. We noticed that one of the boys was "stimming," or engaging in self-stimulating repetitive wrist movements that resembled the tapping of an invisible drum. The teleoperator caused Keepon to dance in synchrony with the boy's hand. When the boy noticed this, he stopped and smiled. The teleoperator responded contingently through the robot. The boy then continued stimming in an exploratory manner, testing and recognizing Keepon's rhythmic contingency to his behavior. Within minutes of beginning this interaction, the boy had caressed and kissed the robot, to the surprise of the assembled therapists.

Our colleagues in the field of movement therapy have discussed the use of rhythmic synchrony in their own work with autism, in which they might synchronize with a rocking child in order to establish a channel for engagement and communication. This suggests that, while we have been considering dance with a robot as a form of play that magnifies the rhythmic qualities of interaction (and is therefore useful for developing the relevant technologies), it is also a potentially important application in its own right.

7.4 Limitations

Our work is intended as an early step in the direction of making socially interactive robots behave in a more temporally appropriate ways and using such robots to study how people regulate their interactive behaviors through rhythm. As such, there are limitations in the sophistication of the technology we have worked with, in the complexity of the interactions we have facilitated and observed, and in the conclusions we may draw about interaction more generally.

In terms of perceptual technology, our use of accelerometers and pressure sensors for measuring human movement is advantageous in the simplicity of deployment and in the cleanness of the resulting data. However, these sensors are limited in their sensitivity, in their dimensionality, and in their descriptiveness of human behavior. Ideally, a high-resolution motion capture system would provide us with full-body and full-face data for interacting partners. However, in addition to the cost of such systems and the complexity of processing the resulting data, it is likely that such intrusive measurement would preclude comfortable, natural social behavior. With further development of non-intrusive vision-based human activity measurement, it should be easier to detect rhythmic movements in any part of an interactor's body.

In terms of behavioral technology, our use of Keepon's simple appearance and only four degrees of freedom was motivated by the ease of controlling the robot in a convincing manner in simple interactions. For the large, full-body movements involved in our dance-and game-based interactions, these simple behaviors were sufficient. However, in less constrained, more open-ended conversational interactions, periodic behaviors such as blinking, nodding, breathing, and tapping (not to mention the tempo of speech) are more subtle and require more degrees of freedom, perhaps with greater morphological correspondence between person and robot. Although we believe that a simple robot such as Keepon can be useful in studying relationships between rhythm, turn-taking, and leadership in simple or constrained interactions, it may be necessary to use more degrees of freedom for more complex social interactions.

Another limitation of this work is the difference in size between Keepon and a human being. The laws of physics dictate that the agility of animals' movement (relative to their size) generally varies inversely with their mass. For example, the frequency of periodic behaviors (e.g. walking gaits, chewing) is normally faster in mice than in humans, and faster in humans than in elephants. We believe that there may have been an analogous tension between the "comfortable" range of frequencies in bouncing behaviors between people and Keepon; bouncing continuously and as quickly and vigorously as Keepon can be tiring. Using a robot closer to human size might result in different and more socially relevant rhythmic relationships.

From a perspective of entrainment, one point that should be noted is our observation that frequency correspondence alone did not elicit responses as strongly as phase correspondence, or simultaneous movement. A purely reactive system cannot, theoretically, be synchronous in this way, due to delays in sensing, processing, and acting (particularly for biological systems—witness the difficulty that even our reflexes have dodging hits in the "hand-slap game"). It may be argued that one of the reasons for the rhythmic organization of social behavior is that it enables synchrony and simultaneous movement through *anticipation* of imminent activity. Interacting partners might be seen to have a responsibility to anticipate the behaviors of the other and to "respect" the other's anticipation of one's own behavior by continuing rhythmic behavior. It may be that social biological systems, due to very slow neural reaction times, require anticipatory rhythmic perception in order to achieve simultaneous movement. However, these perceptual mechanisms are also highly sophisticated in perceiving and even anticipating not just tempo but changes in tempo. Our artificial systems, on the other hand, are to date limited in their perceptual and representational capabilities with respect to these rhythms. While our tempo-tracking and oscillator-based methods aim to be anticipatory in this way, it was difficult for them to adapt to rapidly changing rhythms in a contingent manner—this brings the interactors' actions out of phase and precludes simultaneous movement. On the other hand, our robotic system's behavior (particularly bouncing) is extremely rapid, and nearly simultaneous action can be achieved in a reactive manner. Therefore, in order to study interactions in which simultaneous movement was possible even in the presence of rapidly changing human rhythmic behavior, we elected to study the robot operating under reactive control. While we believe that this was adequate for our limited purpose, a rhythmically intelligent system should ultimately synchronize in both frequency and phase in an anticipatory manner.

Another limitation of our work, for reasons of participant recruitment, derives from the ages of our participants. Our participants in the dance-based interactions were generally young children, between the ages of 3 and 6. Within that age range, we found that older children were typically more willing to engage in dance play with the robot. Yet there was still a wide range of comfort and rhythmic ability in our participant pool, which may have made it difficult to obtain consistency within groups. Studies of this type may therefore benefit from experimentation with even older children. Our game-based interactions were conducted with high-school students due to the more complex nature of the task, but we encountered greater reluctance to move freely (perhaps due to embarrassment) and greater desire to satisfy the experimenter in post-interviews. It could be productive to attempt such dance-, and game-based studies with children between these groups, around the ages of 8 to 12, when understanding of and fluency in social cues is probably more firm than in younger children.

Experimentally, our studies aimed to create a context for interactions that would be natural and comfortable, yet constrained to the extent that behaviors could be easily measured and compared across subjects. Our selections of dance and video gaming afforded environmental stimuli that served as a rhythmic backdrop and an external reference to which the robot's and person's behavior could be related. Natural social interaction, or unconstrained dialogue, can certainly share such a relationship to external stimuli (e.g. conversation in a busy public space vs. in a quiet room), but the interactants generally have greater freedom and control over the rhythms they exhibit. This is an important limitation of our work. Although the strict environmental constraints made it tractable to design applicable technologies and to collect and analyze real data from rhythmic human-robot interaction, future work must creatively expand the richness of experimental contexts as well as the flexibility of the technology to participate in more natural interaction.

7.5 Future work

Our dance- and game-oriented experimental paradigms are clearly simplified forms of social interaction, with regular and exaggerated rhythms that facilitated our technology development, perception, and experimentation. However, our long-term goal is to develop systems capable of naturally participating in the flow of unconstrained dialogue. It is notable that social scientists have not, to date, been able to operationalize observed interactional synchrony quantitatively by, for example, analyzing the nodding of two interactors through the lens of a coupled oscillator model. Likewise, the synchronous behaviors of even simple natural phenomena studied by physicists and biologists (e.g. firefly flashing) lie at the frontier of our mathematical understanding of synchrony, and the subtleties of human social behavior are clearly more complex than can be represented by such simple pulse trains.

Therefore, we must build experimental stepping-stones between work such as ours—with its major simplifications—and general interaction. Of course, the dance and video game paradigms are still rich with potential for expanding and developing increased sophistication. For example, we have discussed the introduction of human initiative in switching leadership roles. The performance of the robot could be manipulated to encourage the participant to take control, if he or she is performing better than the robot, or to cede control, if the robot is performing better. Initially, the human might be given a switch or verbal command[1] to initiate these transitions, and this would allow us to determine the performance conditions under which a participant is willing to experience the cognitive load of a transition. We might alternatively design dance-based studies in which the robot signals to its partner that it would like to demonstrate a particular way of dancing or would like to follow the partner's example.

However, we would eventually like to find more natural ways of taking and ceding control. We believe that "requests for transfer" might be detected rhythmically (e.g. in the form of pauses or significant changes in frequency), through gaze direction, or through explicit verbal communication. It would be interesting to design studies comparing these various methods for making transfer apparent to a person.

We believe that dance, as well as goal-oriented or game-like tasks, can continue to serve as a context for these more sophisticated negotiations of rhythm and roles. Additional playful forms of interaction, such as "Simon says"-like imitation games, could be designed without external environmental rhythms while still facilitating the expression of rhythmically patterned behavior. The point of following such a progression of tasks is to gradually increase the complexity of the interaction while maintaining enough control and consistency between trials to be able to compare groups and experimental conditions.

Nevertheless, we can also work from the other direction: integrating these concepts into work on higher-level dialogue-capable robotic systems. For example, if a realistic humanoid robot sends an oscillatory "noise" to its actuators in order to perform constant lifelike movement, the frequency of these movements should probably be related to ambient environmental rhythms. Similarly, humanoid conversational agents that can perceive or generate nodding behaviors should adjust the frequency of their speech or nodding to that of the speech or nodding of a human partner. Further progress could involve detecting the end of a human vocalization through rhythmic cues rather than waiting for a particular token or period of silence. Much of this future research must take place in the fields of linguistics [Auer *et al.*, 1999] and natural language processing (recognition as well as generation), but the role of roboticists in this endeavor must be to carefully design physical cues that will appropriately synchronize with, and in some cases even direct, the linguistic trajectory of the interaction. In contrast to the constrained interactions we have worked with, evaluation of these types of systems will require more sensitive measurements of human behavior as well as suitable interviews to measure subjective impressions.

---

[1] We informally tried this with a few participants in the game experiment after their trials. However, they usually picked and stayed with whatever mode they had thought was easiest, as we did not deteriorate the robot's performance.

It would also be interesting to explore the role of ontological expectations in the perception of a social robot's use of rhythm. That is, we are using a clearly social robot (vaguely anthropomorphic or zoomorphic) rather than a simpler physical object like a ball, so given appropriate facilitation, we can expect that people enter the interaction with a social stance. While we are interested in the effects of rhythmicity on sociality, it is also interesting to consider the effects of sociality on the perception of synchrony. For example, we might change the context within which such dance activities take place, such as using a mirror to focus the child's attention on the relationship between self-movement and the movement of the robot.

Although we have focused on dyadic interaction, group interactions should be further explored. As suggested by our experiences in the pilot study described in section 5.1, there will likely be rhythmic differences between interacting directly with a robot, versus observing the interaction as a bystander. And these roles can of course change frequently over the course of a group interaction, as each interactor switches attention between multiple partners (addressing them either individually or generally). We expect that even bystanders will exhibit rhythmic synchronicity not just passively, but also actively, in order to signal continued participation or to attract direct attention.

We must also continue to improve our technology, particularly in perception and entrainment. While our reactive method allowed us to create phase-synchronized simultaneous movement, it is not clear that it is the optimal way to entrain to human movement. Some of our results (suggesting that children were more engaged when Keepon was synchronized to the music than when it was synchronized to them) may be explained by the "naturalness" of Keepon's behavior; while it was easy for Keepon to entrain to a known musical rhythm, it is possible that Keepon did not look as natural when it was following the children's movement. Many of our experiments involved the extraction of beats from perceptual data, and there were always some false positives or false negatives. This disconnect and the resultant strange behavior may be one reason why Keepon did not stimulate higher engagement when it synchronized to human movement. Since perception will always be noisy, we believe that it would be advantageous for the robot to behave under its own intrinsic (and somewhat regular) rhythm, as supported by the oscillator-based method we have described.

Stepping back from the problem of creating rhythmically intelligent artificial agents, it would be beneficial to bring the social science research up to date using tools and techniques such as those we have presented in this work. The type of observational research done by Condon et al. (described in Chapter 2) can be performed more efficiently and accurately using accelerometers, pressure sensors, or other technologies such as motion capture—in interactions between two or more people—and the resulting data can be analyzed automatically to understand frequencies, relationships between events, activity level, and the degree to which these behaviors can be explained by a model such as oscillator coupling. The interactions might be regular rhythmic tasks, constrained by music or dance, or they might be open-ended interactions in which the partners converse and take turns. We hope that our techniques and methods might inspire social scientists, and not just roboticists, to pursue further research in quantitatively measuring and modeling social rhythmicity.[2] Another interesting direction would be to study rhythmic properties of interactions between people and pets or other animals. This would have the advantage of serving as a form of non-verbal interaction with agents that are expected to behave entirely differently from people, but in a temporally appropriate and relatable manner. Such research would also shed light on the question of appropriate movement speed as it relates to interactions between creatures of different sizes.

Finally, as we have discussed, we believe that our techniques and tools can be used to study and even assist those whose disabilities manifest themselves in impaired social interaction skills. While humans are typically very adept at social interaction, problems occur in the course of human cognitive and physical development that result in difficulties with this type of social intelligence. Furthermore, these disorders are occasionally treated by therapeutic methods involving music and dance that specifically address the rhythmic difficulties [Fledderjohn and Sewickley, 1993; Cruz and Berrol, 2004]. Therefore, by designing interactive robots that can appropriately perceive and behave according to the principles of interactional synchrony, we can not only achieve more natural human-robot social interaction, but we will also open the possibility of applying this technology in educational and therapeutic settings. For example, a dancing robot might be used by clinicians to study a child's behavioral patterns and identify rhythmic abnormalities that could be useful in diagnosis. We believe this socially assistive approach will be effective because (a) some people tend to be more comfortable interacting with an artifact than with a person, (b) a robot can be more available than a human therapist, and (c) a robot can enable more controlled and consistent interactions. Such techniques, coupled with the guidance of experienced professionals, may have significant impact on the fields of movement and dance therapy.

7.6 Summary

As discussed in Chapter 2, social scientists have highlighted the importance of interpersonal coordination in human social interaction. However, the relevant principles have been rather underexplored in robots. Our goal has been to introduce, to the field of social robotics, the idea that rhythmic synchrony is important and to offer suggestions

---

[2] The rhythmic properties of non-verbal and verbal interaction likely have interesting cultural differences that may be quantitatively identifiable.

on how to organize this line of research.

In parallel with the three concepts of *interaction rhythm*, *simultaneous movement*, and *smooth meshing* in interpersonal coordination [Bernieri and Rosenthal, 1991], we have proposed a framework in Chapter 3 for allowing robots to participate in rhythmically coordinated interactions: *rhythmic attention*, which handles rhythms in multiple sensory modalities; *entrainment*, which allows the robot to compare and match its rhythmic behavior to external rhythms; and *interaction state*, which uses task-, role-, and state-specific information to direct the robot's rhythmic behavior in an intentional manner.

Our long-term goal is to develop technologies that allow robots to establish rhythmic synchrony in open-ended social interactions. The major challenges are: perceiving human behavioral rhythms more accurately, developing more sophisticated cognitive or representational mechanisms for rhythmic entrainment, and experimentally studying the properties of human-robot rhythmic synchrony. Each of these three tasks are, at this point, difficult in the context of unconstrained dialogue. Issues of emotional expressiveness, content, natural language use, etc. make it difficult to isolate the purely rhythmic components of the interaction, not to mention that the complex rhythms involved are poorly operationalized and understood even by social scientists.

Against this backdrop, we believe it is worthwhile to develop and study such technology in the context of simplified and constrained interactions. We have studied dance-oriented interactions constrained by music, as well as similarly constrained game-oriented interactions, which provide regularity and encourage easily measurable behaviors (i.e. bouncing). An analogous human activity is the use by lumberjacks of a "two-man saw," in which a rhythm is negotiated by the partners feeling each other's forces. The activity is a very simple interaction, and the ability to participate in it of course does not immediately transfer to open-ended social interaction—yet the underlying principles, and the perceptual and representational technologies that we develop for such tasks, are potentially useful (and tractable) stepping-stones in the broader enterprise of developing rhythmic social intelligence. Rhythmically intelligent robots promise not only to entertain and assist us, but also to contribute to our understanding and appreciation of the rich rhythms that guide us in our lifelong dance with each other.

Bibliography

[Andry *et al.*, 2001] Pierre Andry, Philippe Gaussier, Sorin Moga, Jean Paul Banquet, and Jacqueline Nadel. Learning and communication via imitation: An autonomous robot perspective. *IEEE Transactions on Systems, Man, and Cybernetics*, 31(5), Sept 2001.

[Aristotle, circa 330 BC] Aristotle. *Poetics.* circa 330 BC.

[Auer *et al.*, 1999] Peter Auer, Elizabeth Couper-Kuhlen, and Frank Müller. *Language in time: the rhythm and tempo of spoken interaction.* Oxford University Press, 1999.

[Bailenson and Yee, 2005] J.N. Bailenson and N. Yee. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, 16:814–819, 2005.

[Bandura, 1977] A. Bandura. *Social Learning Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1977.

[Baron-Cohen, 2003] Simon Baron-Cohen. *The Essential Difference: The Truth about the Male and Female Brain.* Perseus Books Group, 2003.

[Bartenieff and Lewis, 1980] I. Bartenieff and P. Lewis. *Body Movement; Coping with the Environment.* Gordon and Breach, New York, 1980.

[Bernieri and Rosenthal, 1991] F.J. Bernieri and R. Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. In R.S. Feldman and B. Rime, editors, *Fundamentals of nonverbal behavior*, pages 401–432. Cambridge University Press, 1991.

[Blacking, 1983] John Blacking. Movement and meaning: Dance in social anthropological perspective. *Dance Research: The Journal of the Society for Dance Research*, 1(1):89–99, 1983.

[Breazeal and Scassellati, 2002] C. Breazeal and B. Scassellati. Robots that imitate humans. *Trends in Cognitive Science*, 6:481–487, 2002.

[Breazeal *et al.*, 2004] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 2004.

[Buck and Buck, 1968] John Buck and Elisabeth Buck. Mechanism of rhythmic synchronous flashing of fireflies. *Science*, 159:1319–1327, 1968.

[Burgoon *et al.*, 1995] Judee K. Burgoon, Lesa A. Stern, and Leesa Dillman. *Interpersonal Adaptation: Dyadic Interaction Patterns.* Cambridge University Press, New York, 1995.

[Chapple, 1982] Eliot D. Chapple. Movement and sound: The musical language of body rhythms in interaction. In Martha Davis, editor, *Interaction Rhythms: Periodicity in Communicative Behavior*, pages 31–51. Human Sciences Press, New York, 1982.

[Chi *et al.*, 2000] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 173–182, 2000.

[Clark and Brennan, 1991] Herbert H. Clark and Susan E. Brennan. Grounding in communication. In L. B. Resnick, R. M. Levine, and S. D. Teasley, editors, *Perspectives on socially shared cognition*, pages 127–149. APA, 1991.

[Condon and Ogston, 1966] William S. Condon and W.D. Ogston. Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 143:338–347, 1966.

[Condon and Ogston, 1967] William S. Condon and W.D. Ogston. A segmentation of behavior. *Journal of Psychiatric Research*, 5:221–235, 1967.

[Condon and Ogston, 1971] William S. Condon and W.D. Ogston. Speech and body motion synchrony or the speaker-hearer. In D.L. Horton and J.J. Jenkins, editors, *Perception of Language*, pages 150–173. Charles E. Merrill, Columbus, OH, 1971.

[Condon and Sander, 1974a] William S. Condon and Louis W. Sander. Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science*, 183(4120):99–101, 1974.

[Condon and Sander, 1974b] William S. Condon and Louis W. Sander. Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, 45(2):456–462, June 1974.

[Condon, 1986] William S. Condon. Communication: Rhythm and structure. In James R. Evans and Manfred Clynes, editors, *Rhythms in Psychological, Linguistic and Musical Processes.* Charles C. Thomas Publisher, Springfield, 1986.

[Crick *et al.*, 2006] Christopher Crick, Matthew Munz, and Brian Scassellati. Synchronization in social tasks: Robotic drumming. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006)*, University of Hertfordshire, United Kingdom, September 2006.

[Cruz and Berrol, 2004] Robyn F. Cruz and C.F. Berrol, editors. *Dance/movement therapists in action: A working guide to research options.* Charles C. Thomas, Springfield, IL, 2004.

[Cycling '74, 2006] Cycling '74. Cycling '74 Max/MSP: A graphical programming environment for music, audio, and multimedia., 2006. URL: http://www.cycling74.com/products/maxmsp.

[Dautenhahn, 1999] Kerstin Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. if the International Cognitive Technology Conference*, 1999.

[DiCarlo and Maunsell, 2005] James J. DiCarlo and John H. R. Maunsell. Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *Journal of Neurophysiology*, 93:2974–2986, 2005.

[Fledderjohn and Sewickley, 1993] Heidi Fledderjohn and Judith Sewickley. *An Annotated Bibliography of Dance/Movement Therapy, 1940-1990.* American Dance Therapy Association, Columbia, MD, 1993.

[Gentry *et al.*, 2003] Sommer Gentry, Steven Wall, Ian Oakley, and Roderick Murray-Smith. Got rhythm? haptic-only lead and follow dancing. In *Proceedings of Eurohaptics*, 2003.

[Gockley *et al.*, 2005] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz, and Jue Wang. Designing robots for long-term social interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, pages 2199–2204, Edmonton, Alberta, August 2005. IEEE.

[Griesbeck, 1996] Christian Griesbeck. Introduction to labanotation, 1996. URL: http://user.uni-frankfurt.de/g̃riesbec/LABANE.HTML.

[Hall, 1983] Edward T. Hall. *The Dance of Life: The Other Dimension of Time.* Anchor, 1983.

[Hatfield *et al.*, 1994] Elaine Hatfield, John T. Cacioppo, and Richard L. Rapson. *Emotional Contagion*. Cambridge University Press, New York, 1994.

[Ijspeert, 2008] Auke Jan Ijspeert. Central pattern generators for locomotion control in animals and robots: A review. *Neural Networks*, 21:642–653, May 2008.

[Intel, 2006] Intel. Intel Open Source Computer Vision Library, 2006. URL: http://www.intel.com/technology/computing/opencv.

[Jehan, 2004] Tristan Jehan. Msp externals: beat , 2004.

[Kalish, 1968] Beth Kalish. Body movement therapy for autistic children. *Proceedings of the 3rd annual conference of the American Dance Therapy Association*, pages 49–59, 1968.

[Kang *et al.*, 2008] Sin-Hwa Kang, Jonathan Gratch, Ning Wang, and James H. Watt. Does the contingency of agents' nonverbal feedback affect users' social anxiety? In *Proceedings of AAMAS 2008*, pages 120–127, 2008.

[Kendon, 1970] Adam Kendon. Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32:100–125, 1970.

[Kendon, 1990] Adam Kendon. *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press, 1990.

[Kosuge, 2010] Kazuhiro Kosuge. Dance partner robot: an engineering approach to human-robot interaction. In *HRI '10: Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 201–201, New York, NY, USA, 2010. ACM.

[Kozima and Nakagawa, 2006] Hideki Kozima and Cocoro Nakagawa. Social robots for children: Pactice in communication-care. In *IEEE International Workshop on Advanced Motion Control (AMC2006)*, pages 768–773, Istanbul, 2006.

[Kozima *et al.*, 2008] Hideki Kozima, Marek P. Michalowski, Jeff Nichols, and Melanie Cornwell. Hri caught on film 2: Keepon dancing to spoon's "don't you evah". In Christoph Bartneck, editor, *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008)*, page 386, Amsterdam, March 2008. ACM. URL: http://youtube.com/watch?v=nPdP1jBfxzo.

[Kozima *et al.*, 2009] Hideki Kozima, Marek P. Michalowski, and Cocoro Nakagawa. Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, 1, January 2009.

[Laban, 1980] Rudolf Laban. *The Mastery of Movement, 4th edition revised by Lisa Ullmann.* MacDonald and Evans, London, 1980.

[LaFrance, 1979] Marianne LaFrance. Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, 42(1):66–70, March 1979.

[Lamb and Watson, 1979] W. Lamb and E. Watson. *Body Code: The Meaning in Movement.* Routledge and Kegan Paul, London, 1979.

[Large and Kolen, 1994] E.W. Large and J.F. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6:177–208, 1994.

[Lesley University, 2006] Lesley University. Expressive therapies, 2006. URL: http://www.lesley.edu/library/guides/research/expressive_therapies.html.

[Lofland and Lofland, 1995] John Lofland and Lyn H. Lofland. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* Wadsworth, 1995.

[Marder and Bucher, 2001] Eve Marder and Dirk Bucher. Central pattern generators and the control of rhythmic movements. *Current Biology*, 11:R986–R996, 2001.

[Michaels *et al.*, 1987] D.C. Michaels, E.P. Matyas, and J. Jalife. Mechanisms of sinoatrial pacemaker synchronization: A new hypothesis. *Circulation Research*, 61:704–714, 1987.

[Michalowski and Kozima, 2007a] Marek P. Michalowski and Hideki Kozima. Keepon dancing to a drum, 2007. URL: http://www.youtube.com/watch?v=6VKE_VoXzNI.

[Michalowski and Kozima, 2007b] Marek P. Michalowski and Hideki Kozima. Keepon dancing to spoon's "i turn my camera on", 2007. URL: http://youtube.com/watch?v=3g-yrjh58ms.

[Michalowski and Kozima, 2007c] Marek P. Michalowski and Hideki Kozima. Methodological issues in facilitating rhythmic play with robots. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)*, pages 95–100, Jeju, Korea, August 2007. IEEE.

[Michalowski *et al.*, 2006a] Marek P. Michalowski, Selma Sabanovic, and Philipp Michel. Roillo: Creating a social robot for playrooms. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006)*, pages 587–592, Hatfield, UK, September 2006. IEEE.

[Michalowski *et al.*, 2006b] Marek P. Michalowski, Selma Sabanovic, and Reid Simmons. A spatial model of engagement for a social robot. In *Proceedings of the 9th International Workshop on Advanced Motion Control (AMC 2006)*, pages 762–767, Istanbul, March 2006. IEEE.

[Michalowski *et al.*, 2007a] Marek P. Michalowski, Selma Sabanovic, Carl DiSalvo, Didac Busquets, Laura M. Hiatt, Nik A. Melchior, and Reid Simmons. Socially distributed perception: Grace plays social tag at aaai 2005. *Autonomous Robots*, 22(4):385–397, May 2007.

[Michalowski *et al.*, 2007b] Marek P. Michalowski, Selma Sabanovic, and Hideki Kozima. A dancing robot for rhythmic social interaction. In *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2007)*, pages 89–96, Washington, DC, March 2007. ACM/IEEE.

[Michalowski *et al.*, 2008] Marek P. Michalowski, Hideki Kozima, and Daniel Wilson. Keepon: Friend or foe? & keepon auditioning, 2008. URL: http://cmu.edu/robou.

[Michalowski *et al.*, 2009a] Marek P. Michalowski, Jaewook Kim, Bomi Kim, Sonya S. Kwak, and Hideki Kozima. Keepon goes seoul-searching. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2009)*, pages 197–198, La Jolla, CA, March 2009. ACM. URL: http://youtube.com/watch?v=XwqfWR2KPd0.

[Michalowski *et al.*, 2009b] Marek P. Michalowski, Reid Simmons, and Hideki Kozima. Rhythmic attention in child-robot dance play. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2009)*, pages 816–821, Toyama, Japan, September 2009. IEEE.

[Michaud and Theberge-Turmel, 2002] F. Michaud and C. Theberge-Turmel. Mobile robotic toys and autism. In Kerstin Dautenhahn, Alan Bond, Lola Canamero, and Bruce Edmonds, editors, *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, pages 125–132. Kluwer Academic Publishers, 2002.

[Naguib *et al.*, 1999] Marc Naguib, Claudia Fichtel, and Dietmar Todt. Nightingales respond more strongly to vocal leaders of simulated dyadic interactions. *Proc Biol Sci*, 266(1419):537–542, March 22 1999.

[Néda *et al.*, 2000] Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási. Self-organizing processes: The sound of many hands clapping. *Nature*, 403:849–850, February 2000.

[NIST/SEMATECH, 2009] NIST/SEMATECH. e-handbook of statistical methods: Randomized block designs, 2009.

[Noldus, 2006] Noldus. Noldus Information Technology: The Observer, 2006. URL: http://www.noldus.com/products/observer.

[Ogawa and Watanabe, 2001] Hiroki Ogawa and Tomio Watanabe. Interrobot: speech-driven embodied interaction robot. *Advanced Robotics*, 15:371–377, 2001.

[Patel *et al.*, 2005] A.D. Patel, J.R. Iversen, Y. Chen, and B. Repp. The influence of metricality and modality on synchronization with a beat. *Exptl. Brain Research*, 163:226–238, 2005.

[Penny, 1997] Simon Penny. Embodied cultural agents: at the intersection of robotics, cognitive science, and interactive art. In *AAAI Socially Intelligent Agents Symposium*, 1997.

[Pikovsky *et al.*, 2002] Arkady Pikovsky, Michael Rosenblum, and Jurgen Kurths. *Synchronization: A Universal Concept in Nonlinear Science.* Cambridge University Press, 2002.

[Reidsma *et al.*, 2008] Dennis Reidsma, Anton Nijholt, and Pieter Bos. Temporal interaction between an artificial orchestra conductor and human musicians. *Computers in Entertainment*, 6:1–22, 2008.

[Restivo, 2002] Sal Restivo. Romancing the robots: Social robots and society. In *Robot as Partner: An Exploration of Social Robots Workshop, International Conference on Intelligent Robots and Systems*, 2002.

[Rider and Eagle Jr., 1986] Mark S. Rider and Charles T. Eagle Jr. Rhythmic entrainment as a mechanism for learning in music therapy. In James R. Evans and Manfred Clynes, editors, *Rhythms in Psychological, Linguistic and Musical Processes.* Charles C. Thomas Publisher, Springfield, 1986.

[Robins *et al.*, 2008] Ben Robins, Kerstin Dautenhahn, Rene te Boekhorst, and Chrystopher L. Nehaniv. Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinesics. In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 17–24, New York, NY, USA, 2008. ACM.

[Sabanovic *et al.*, 2006] Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. Robots in the wild: Observing human-robot social interaction outside the lab. In *Proceedings of the 9th International Workshop on Advanced Motion Control (AMC 2006)*, pages 596–601, Istanbul, March 2006. IEEE.

[Scheflen, 1964] A.E. Scheflen. The significance of posture in communication systems. *Psychiatry*, 27:316–331, 1964.

[Scheirer, 1998] Eric Scheirer. Tempo and beat analysis of acoustic music signals. *Journal of the Acoustic Society of America*, 103, 1998.

[Sidner *et al.*, 2005] Candace L. Sidner, Christopher Lee, Cory Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, May 2005.

[Stern and McClintock, 1998] Kathleen Stern and Martha K McClintock. Regulation of ovulation by human pheromones. *Nature*, 392:177–179, 1998.

[Strogatz *et al.*, 2005] Steven H. Strogatz, Daniel M. Abrams, Allan McRobie, Bruno Eckhardt, and Edward Ott. Theoretical mechanics: Crowd synchrony on the millennium bridge. *Nature*, 438:43–44, November 2005.

[Strogatz, 2003] Steven H. Strogatz. *Sync: How Order Emerges from Chaos in the Universe, Nature, and Daily Life.* Hyperion, 2003.

[Swaminathan *et al.*, 2009] Dilip Swaminathan, Harvey Thornburg, Jessica Mumford, Stjepan Rajko, Jodi James, Todd Ingalls, Ellen Campana, Gang Qian, Pavithra Sampath, and Bo Peng. A dynamic bayesian approach to computational laban shape quality analysis. *Advances in Human-Computer Interaction*, 2009:1–17, 2009.

[Tanaka *et al.*, 2006] F. Tanaka, J. R. Movellan, B. Fortenberry, and K. Aisaka. Daily hri evaluation at a classroom environment: Reports from dance interaction experiments. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction*, pages 3–9, Salt Lake City, Utah, 2006.

[Trevarthen and Malloch, 2000] Colwyn Trevarthen and Stephen N. Malloch. The dance of wellbeing: Defining the musical therapeutic effect. *Nordic Journal of Music Therapy*, 9:3–17, 2000.

[Weinberg and Driscoll, 2006] Gil Weinberg and Scott Driscoll. Robot-human interaction with an anthropomorphic percussionist. In *Proceedings of CHI2006*, pages 1229–1232, Montreal, Canada, 2006.