Dance Interaction with QRIO: A Case Study for Non-boring Interaction by using an Entrainment Ensemble Model

Fumihide Tanaka and Hirotaka Suzuki

Sony Intelligence Dynamics Laboratories, Inc.
3-14-13(-4F) Higashigotanda, Shinagawa-ku, Tokyo, 141-0022 Japan
E-mail: {boom, hirotaka}@idl.sony.co.jp

Abstract

This paper introduces our new project, Dance Interaction, where we will investigate design methodology or principles for realizing long-term interaction between a human and a robot. By exploiting the powerful hardware of an entertainment robot QRIO, we will tackle this pretty hard problem by a repeated try and test approach. As the first attempt, we try designing by an Entrainment Ensemble Model that will be presented in this paper. Though studies are just starting and have not matured yet, current progress and prospects for the near future will be addressed.

1 Introduction

1.1 Background

Entertainment robots have potential to help many people's life being further enjoyable, and their technologies have been making steady progress recently.

But still now, designing a robot that enables long-term interaction with his owner without him boring is a pretty hard task, presenting us an ultimate goal in this field. Losing interests, we might be able to say that it is a nature of us humans.

1.2 Our Approach

Knowing that, we will try to investigate how to design an interaction system that attracts human's interests as long as possible. As the goal seems to be too general including lots of directions, at first we will set a concrete application image to make our target clearer.

The target is Dance Interaction between a human and a robot where we will investigate continuing interaction by using both bodies. We believe that communication ability is crucial for building long-term interaction, and as one of the most primitive shape, we focus on body-based communication. Further, a dance is a universal activity for human beings that itself has special effects on our mind and body, continued long time in our history. These are the reason why we choose the Dance Interaction as our target challenge.

Section 2 will explain the challenge from the viewpoint of using a real robot. After that, in Section 3 we will present an Entrainment Ensemble Model that will be our first implementing idea in an example case with an entertainment robot QRIO described in Section 4. Finally we will summarize this paper in Section 6 after discussing future research plans in Section 5.

2 Dance Interaction Challenge between a Human and a Robot

2.1 Overview

As a benchmark test-bed for a series of studies, we are making an application where a human and a real humanoid-type robot facing each other do dancing together in real time and the real world.

Both dance (or just move) freely but basically their movements are based on imitation of each partner. Regarding the robot, it exploits its full of sensors such as vision or audio to detect its partner's movement, and then produce and output any corresponding motor-action sequence. The human also reacts inspired by the robotic dance, and interaction is formed by continuing the process. Ideally, both dance based on not only just imitation but also active creation of new dances, therefore the dance itself can be seen as evolving through the interaction process.

The goal of this challenge is very simple: Keep human's interest as long as possible. This is a bit rough goal but connotes affluent research topics below:

2.2 Research Topics

- **Principles for continuing long-term interaction:** We are interested in whether there is

0-7803-8570-5/04/$20.00 ©2004 IEEE.

any intrinsic key that keeps a human attracted to the interaction process itself. Attraction peculiar to the partner such as the cuteness of a robotic design might be an important factor. But what we'd like to address here mainly is not that but anything that is buried in the interaction framework or process. If there is any, then we can exploit the knowledge into the entertainment robot living with its owner through his daily life.

- **Mutual imitation in real time:** Imitation is no doubt a basis for building up communication between human beings or some other beings. There have been many studies also in the field of robotics that deal with this topic [1, 2, 3, 4, 5, 12]. Usually robotic imitation is based on its visual sensor information and it has been very difficult problem for years to make a noise-robust and real time working system. Especially, the real time response is crucial in our task domain (dance), and it claims a new approach.

- **Using a humanoid that has a physical body:** Through our many experiences of demonstrations or exhibitions outside, we feel that real humanoid's real time performance gives people a kind of special impact actually. It is clearly different from it by CG-animations or even video-movies of the same humanoid robot. To clarify the reason and investigate the role of embodyment and its presence are also our themes.

3 Entrainment Ensemble Model

3.1 Background

In this section we present a model that we consider suitable for describing the interaction process for our purpose: designing long-term interaction between a human and a robot. Before presenting the model, let us explain a basic strategy.

We consider that there are two concepts that should be basic units designing the interaction model. One is *sympathy* between the two (human and robot), and the other is *variation* to it. Our basic hypothesis is that by the repetition of the *sympathy* and *variation*, long-term interaction can be realized.

Let us present an example along dance interaction: First, a human presents some kinds of movements to a robot that is inactive then, expecting that it would react to his action. Then, the robot gradually follows his movements improving its imitation accuracy. When the two synchronize as accurate as to some extents, the human feels very well, and these correspond to the phase of the *sympathy*. But continuing the same situation for a long time makes the human boring. Human beings lose interests for situations whose behavior is completely predictable and unchanged. Therefore the situation should be changed such like presenting a new dance motion from the robot after a period of the *sympathy*. This corresponds to the phase of *variation*.

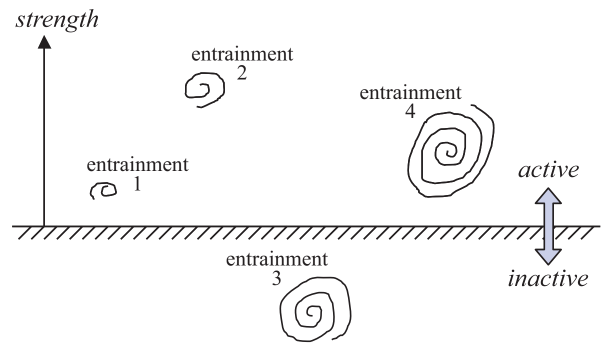

Figure 1: The concept illustration of an Entrainment Ensemble Model (EEM).

The example might be seen as simple and natural, but we consider it contains some essential points.

- How the *sympathy* is designed?
- How the *variation* is designed?
- How they (*sympathy* and *variation*) are controlled in an interaction system?

In this section we explain the former two topics. Regarding the third point, various approaches can be considered, and in this paper we will introduce some naive approaches in Section 4. Detailed works in this point will be presented in the near future.

3.2 Framework

Our basic idea designing sympathy is that there should be included a number of entrainment-factors at the same time in it. An entrainment factor corresponds to any kind of a concrete pair of a relationship between a human and a robot. For example, if we design a robot that shakes his hands by the speed of a moving-region in its visual field, then this relationship is an entrainment factor. If we add another function to the robot such as jumping in the rhythm of clapping by its owner, then there are two entrainment factors in this system.

We consider designing multiple entrainment factors in an interaction system. Figure 1 illustrates one interaction scene where there are four entrainment factors designed in it. We can see that one of them is inactive now, and active ones have their own strength of exhibition. On making a system actually, an entrainment factor can be realized by any computational approach such as a naive rule-base, many kinds of oscillators [6, 7], a recurrent neural network [8], and so on. We will show our example later in Section 4.

Now we call this model with multiple entrainment factors as Entrainment Ensemble Model (EEM). Once we design sympathy along with EEM, we can consider the ways of variation in two levels:

variation in a local level: This works within each entrainment factor respectively. We can design many ways of variation independently such as giving a random noise to outputs of a recurrent neural network representing one entrainment factor.

variation in a global level: This works across multiple entrainment factors. By changing the number of them that are active then, and (or) by changing the order of strength by controlling the value in each entrainment factor, global variation can be designed.

Practically, the concrete ways of designing variation are different in every target task domain. In this section, what we'd like to explain is the framework of EEM, which has potential for modeling complex relationships in interaction and designing variation flexibly. To be able to realize long-term interaction, complexity in it is crucial in some way to keep human's interest. We are trying to handle the complexity by using multiple entrainment factors explained above.

4 An Example Case

4.1 QRIO: A Small Biped Entertainment Robot [9, 10, 11]

We will use Sony's humanoid robot QRIO. Figure 2 illustrates its appearance and Table 1 describes its basic specification. It is a stand-alone autonomous robot with three CPUs: the first one is used for audio recognition and text-to-speech synthesis, the second one is for visual recognition, short and long term memory, and a behavior control architecture, and finally the third one is spent in its motion control. In addition to these, remote PCs can be exploited as remote-brains by using its embedded wireless-LAN system.

After years of research and developmental efforts, now QRIO's ability spreads over very wide range: walking, running, jumping, playing soccer, throwing a ball, swinging a putter, singing songs, recognizing

Figure 2: Sony Entertainment Robot 'QRIO' (: it is not a product but a test prototype).

CPU	64 bit RISC processor (x3)
Main Memory	64MB DRAM (x3)
Operating System	Aperios (a Sony real time OS)
Control Architecture	OPEN-R
Supplying Media	Memory Stick
Joint Degree of Freedom	Neck: 4DOF, Body: 2DOF, Arms: 5DOF(x2), Legs: 6DOF(x2), Hands with 5 fingers (x2)
Distance Sensor	Infrared system distance sensor: Head: 1, Hand: 1(x2)
Acceleration Sensor	Trunk: X,Y,Z / 3 axes, Foot: X,Y / 2 axes (x2)
AngularRate Sensor	Trunk: X,Y,Z / 3 axes
Foot Sensor	Foot: 4(x2)
Thermo Sensor	Hand: 1(x2), Foot: 1(x2), Head: 1, Body: 1, Actuator: 22
Touch Sensor	Head: 4, Hand: 1(x2), Shoulder: 1(x2)
Pinch Detection Sensor	Overall: 18
Grip Switch	Back: 1
Image Input	110,000 color pixels CCD camera (x2)
Audio IN/OUT	Microphone(x7) / Speaker(x1)
Input/Output	PC card slot (Type II), Memory Stick slot
LED display	Eye (4096 colors each), Ears (1 color, 16 gradation), Power (2 colors)
Weight	Approximately 7kg (with battery)
Dimension (H x W x D)	Approximately 580 x 270 x 190 mm

Table 1: The hardware specification of QRIO.

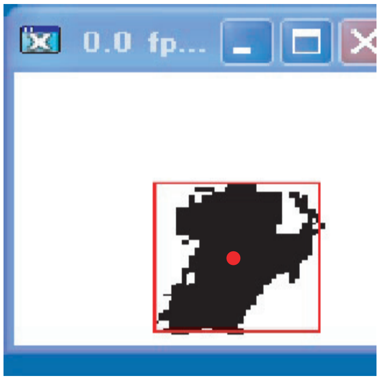

Figure 3: A moving region in QRIO's visual field and a block-region surrounding it.

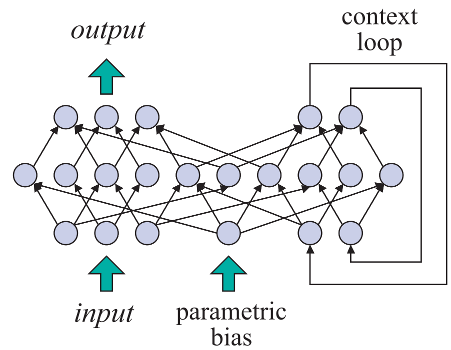

Figure 4: Recurrent neural network model with parametric bias [8].

humans by vision and audio, making a conversation (dialogue), and so on. Please refer to [9, 10, 11] for its detailed description.

4.2 Visual Processing for Rough Imitation

In our dance application, QRIO basically imitates (or just reacts to) a human's movement by exploiting its sensor information. We are planning to use not only visual sensors but also auditory or tactile ones, but in this paper let us explain our first attempt of using only vision system.

Robotic imitation has been investigated for years, and there have been presented several approaches recently [1, 2, 3, 4, 5, 12]. But still now, we can say that it is very hard to let a robot imitates human's gestures in real time and the real world by using only embedded vision system without setting any visual markers.

After considering various directions, we started our project with a policy of "*Rough but Robust Imitation*". As a first attempt, we focus on moving regions in a visual field of QRIO. Thanks to its FPGAs-embedded vision system, QRIO can obtain frame-difference and disparity (distance) information very first in real time. Therefore we can calculate a block-region (Figure 3) that encloses the nearest moving region by using basic visual processing technologies. By using the block-region with corresponding disparity value, QRIO can know its dance partner's rough shape in real time, robustly.

4.3 Data Structure for Representing an Entrainment Factor

As we described in Section 3, EEM uses multiple entrainment factor in parallel, and each representation is arbitrarily decided. In this section let us introduce our first attempt of implementing the model to QRIO.

We will use two kinds of data structure: one is a naive rule-base and the other is a recurrent network with parametric bias (RNNPB). In our application domain of dance interaction, performance as an application for user's entertainment is also required. Therefore we positively use designing rule-bases though we should recognize their roles carefully in evaluation.

RNNPB (Figure 4) is originally proposed by Tani [8] and we have interests to its potential. Actually, our colleagues have succeeded and reported its application to imitate human's simple gestures with QRIO [12]. They showed that by using one RNNPB, multiple gestures can be memorized into it in generalized way at a dynamically structure level. Once they are learned, each gesture can be extracted by using a parametric bias value as a set of time sequence data.

Now we are thinking of exploiting RNNPB as describing one entrainment factor, and consider designing an interaction system with EEM.

4.4 Dance Interaction with QRIO: A Preliminary Application

Based on technologies explained so far, we made a preliminary prototype of a dance interaction application with QRIO (Figure 5). Now we're testing various kinds of entrainment factor sets, an example of which with three factors is like this:

1. Focusing on the horizontal rhythm of block-region movement. By tracking the central point of the block-region, a rhythmic wave (sequence value data) can be obtained. It is transformed to QRIO's shoulder pitch angle information by passing a rule-base system and a RNNPB that learned several rhythmic waves' dynamics previously.

Figure 5: Dance Interaction with QRIO. (This is a scene where QRIO trys to imitate human's motion and build a sympathy phase.)

2. Similar to 1, the square measure of the block-region is tracked, and another rhythmic wave is calculated here also. This time, it is transformed to QRIO's shoulder roll angle information by using a pre-designed rule-base.

3. Disparity information (the distance to the object) is tracked, and the third rhythmic wave is calculated. It is transformed to QRIO's trunk pitch angle information by using another rule-base.

By controlling the number (strength) of entrainment factors that are active, QRIO can present variation. The controlling scheme right now is pre-designed or just random, but here we can exploit various approaches such as estimation techniques of user's emotion by using facial or voice information [13].

Up to now, no creation ability is given to the robot. But our preliminary test showed that even the naive implementation could produce complex and unpredictable movements, presenting a possibility of attracting the partner human's interest. It is a typical example of exploiting environmental complexity in a good way for the application.

5 Future Works

Now we are planning for evaluation experiments especially in various settings for entrainment factors. Besides them, we will try below matters in the near future:

- Input a number of dance primitive motions with the help of professional dancers.
- Use PC-clusters as a remote-brain of QRIO to speed up the processes of RNNPB.
- Output original movements spontaneously from QRIO based on real time learning mechanism.

6 Summary

This paper described our challenge for non-boring interaction between a human and a robot, attracting the human's interest as long as possible. After introducing our benchmark challenge, Dance Interaction, we explained our first attempt of using an Entrainment Ensemble Model, and a preliminary example with a humanoid robot, QRIO.

Researches are ongoing and lots of works are to be done. We hope to expand this project with the helps of many people including those from psychological fields, professional dancers, choreographers, and so on.

References

[1] S. Schaal, "Is Imitation Learning the Route to Humanoid Robots?", *Trends in Cognitive Science*, Vol.3 No.6 pp.233–242, 1999.

[2] G. Cheng, A. Nagakubo and Y. Kuniyoshi, "Continuous Humanoid Interaction: An Integrated Perspective - Gaining Adaptivity, Redundancy, Flexibility - In One.", *Robotics and Autonomous Systems*, Vol.37 No.2-3 pp.161–183, 2001.

[3] T. Inamura, and et al., "Imitation and Primitive Symbol Acquisition of Humanoids by the Integrated Mimesis Loop.", *2001 IEEE International Conference on Robotics and Automation*, 2001.

[4] C. Breazeal and B. Scassellati, "Robots that Imitate Humans.", *Trends in Cognitive Science*, Vol.6 No.11 pp.481–487, 2002.

[5] A. Billard, and et al., "Discovering Optimal Imitation Strategies.", *Robotics and Autonomous Systems*, Vol.47 No.2-3 2004.

[6] M. Williamson, "Designing Rhythmic Motions using Neural Oscillators.", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1999.

[7] A. Ijspeert, J. Nakanishi and S. Schaal, "Learning Rhythmic Movements by Demonstration using Nonlinear Oscillators.", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[8] J. Tani, "Learning to Generate Articulated Behavior through the Bottom-up and the Top-down Interaction Processes.", *Neural Networks*, Vol.16 No.1 pp.11–23, 2003.

[9] M. Fujita, and et al., "SDR-4X II: A Small Humanoid as an Entertainer in Home Environment.", *International Symposium of Robotics Research*, 2003.

[10] T. Ishida, and et al., "Development of Mechanical System for a Small Biped Entertainment Robot.", *IEEE International Workshop on Robot and Human Interactive Communication*, 2003.

[11] Y. Kuroki, and et al., "A Small Biped Entertainment Robot Exploring Human-Robot Interactive Applications.", *IEEE International Workshop on Robot and Human Interactive Communication*, 2003.

[12] M. Ito and J. Tani, "On-line Imitative Interaction with a Humanoid Robot Using a Mirror Neuron Model.", *IEEE 2004 International Conference on Robotics and Automation*, 2004.

[13] J. Movellan and M. Bartlett, "The Next Generation of Automatic Facial Expression Measurement.", In: P. Ekman (eds.) *What the Face Reveals*, Oxford University Press, 2003.