RECOGNITION OF PHYSICAL MOTION PATTERN USING STOCHASTIC PETRI NETS FOR SOUND SYNTHESIS

Munetatsu Go   *Naotoshi Osaka*

Tokyo Denki University
School of Science and Technology for Future Life
2-2 Kanda-nishiki-cho, Chiyodaku, Tokyo, 101-8457, Japan
go@srl.im.dendai.ac.jp, osaka@im.dendai.ac.jp

ABSTRACT

A number of dance or simply gestural movement analysis and performance systems have been developed over the years. Recent developments have extended functionality for recognition and performance systems. Most of the recognition procedure is based on the Hidden Markov Model (HMM). The Hidden Stochastic Petri Net (HSPN) is a general model for the purpose of estimation of multiple concurrent process events, which have synchronizations with other processes, given physical observation that does not necessarily correspond to the event. The recognition procedure for an HSPN is done by transforming a given Petri net into an HMM. The HSPN also solves the problem of giving a concrete meaning to each state of the obtained HMM. This paper reports the theoretical study of an HSPN's implementation in a MAX/MSP environment for dance movement analysis and recognition algorithm.

1. INTRODUCTION

Several attempts have been made to associate physical motion to sounds in media art works, including Atau Tanaka's BioMuse [1], as a live performance, and *Very Nervous System* by David Rokeby [2], as an installation, and Eyesweb [3], used as a support system for music in a dancing drama, *Rew* by Herve Robbe [4]. The authors explore a physical motion recognition system and sound control system using its output for the purpose of multimedia content creation.

In ordinary dances, figure skating, or rhythmic gymnastics, where both physical motion and music are performed simultaneously, music leads physical motion. The other way around is also possible; sound generation can be controlled by physical motion. This is the same relation as that of the physical motions of a musical instrument performance and the sounds generated by them. Sound control technology by physical motion adds a new artistic role to dancing, and enables new channels for sound representation.   With this as motivation, we are developing an interactive system that enables sound generation from physical motion recognition. In this paper we propose a new algorithm, Hidden Stochastic Petri Net (HSPN), which is a Petri net based "grammar description" of physical motion. The algorithm is an advanced version of HMM (Hidden Markov Model) [5] and expanded to HMM. Back-ground and resolution algorithm are introduced in the subsequent sections.

2.   BASIC DESIGN CONCEPT OF THE SYSTEM

Figure 1 depicts a block diagram of the system. In order to realize a real time process in overall system, it has the following features.
i)   Motion capturing is done using video camera,
ii) OSC (Open Sound Control) protocol [6] is used as an interface to communicate among system in order to assure real time processing.
iii) Eyesweb [3] is used for physical motion future extraction.
iv) Max/MSP is used as an audio programming environment.
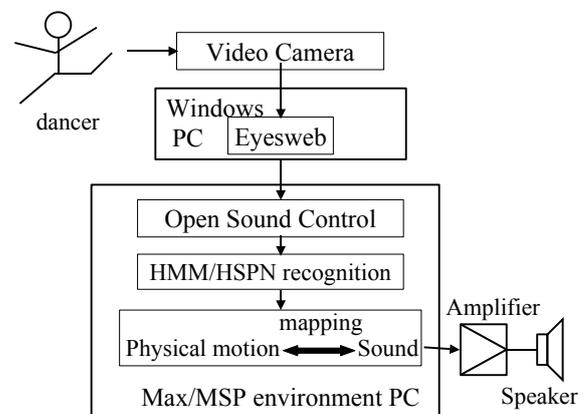v) HMM recognition and HSPN recognition are implemented in Max/MSP using newly coded external objects.

Figure 1. Block diagram of the system

2.1. Physical motion capturing

There are two common paths to capture physical motion: image processing from the video output and sensory data processing. Image processing can be done in real time using Max/MSP/Jitter [7]. A wireless sensor system is developed at IRCAM and sold from La Kitchen Inc. and its precision is guaranteed. Our ideal concept is to use both image processing and sensory data processing, compensating for lacking data if only one of them were used. For the time being, we first take on the problem of image processing.

2.2. Data communication

OSC (Open Sound Control) [6] is used as a communication protocol, which has many advantages over MIDI in real time computer music communication. Communication libraries using OSC are already implemented in various software and programming languages. Therefore, using OSC is an effective choice for the systems in which many application systems and hardware are combined. OSC has already been implemented in both Max/MSP and Eyesweb.

2.3. Physical motion analysis and recognition

From a motion captured system, we expect fundamental physical features such as displacement, velocity and acceleration of important points of moving body. Eyesweb divides the silhouette into regions corresponding to a portion of human body. Human skeleton is reconstructed from the coordinates of the centroid pixel of each region, and two dimensional coordinate data are derived in real time including head, Center of Gravity of the body from the video image input.

Direct mapping of these primitive physical parameters and sound is a common method of control. However, another possibility lies in higher level mapping, that is, choreography pattern to sound or sound class. Higher level mapping does not mean to throw primitive mapping away, but rather, it provides multi-layer mapping. Fig. 2 depicts an example of multi-layer mapping between physical motion and instrumental sound. Choreographic pattern determines the instrument and lower level information controls the expression of the music performance, or prosody in speech case.

In general, speech has its own correspondent symbol or letters. Similarly, since dancing consists of sequences of discrete physical motion, and choreography of some dancing such as classical ballet has its descriptive system such as Labanotation [8]. The difference of linguistic and choreographic symbol is that dancing is
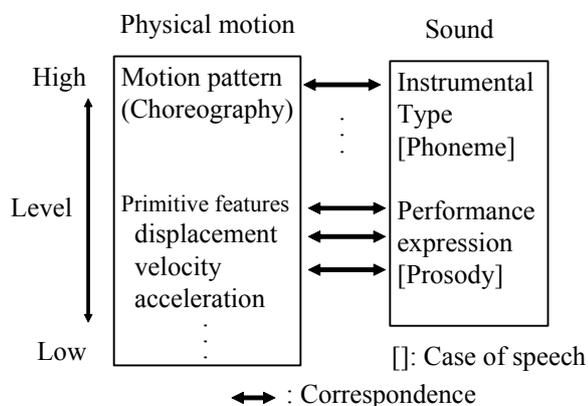

Fig. 2 Correspondence between physical motion and Sound

multi channels and their sequences are all in parallel, giving influences to each other. We first pose a physical motion to symbol mapping problem. In the higher levels this is equivalent to consider the motion pattern to sound pattern mapping such as choreography to phoneme mapping. At the lower level, it needs to be discussed whether or not it is better to adopt an intermediate symbol, since analogue data are treated in both sides.

We adopt HMM to recognize simple physical motion patterns. This is extended to HSPN which can incorporate the parallel event sequence. All the processing is implemented in a Max/MSP/Jitter environment. Some new externals were added. These components provide the means for real-time recognition processing.

3. MODELING PHYSICAL MOTION USING HSPN

3.1. Outline of HSPN

Hidden Stochastic Petri Net (HSPN) is a model in which event descriptions are done using Petri net, and finally converted to HMM via Stochastic Petri Net (SPN.) Petri nets are well-suited to describe parallel event sequences, running concurrently which partly synchronize with each other. In other words, HSPN is an extended algorithm of HMM, specifically interpreted as a multi-channel coupled HMM which incorporates synchronization of inter-channel states. "Hidden," similar to that in HMM, means that events described by places are not directly observed. For example, an event "right arm up" is not directly observed but is interpreted after the recognition process of physical observation.

In general HMM, such as that which is applied to speech recognition, the state is not only hidden, but the meaning of state as an event is also unclear. However, we can consider HMM in which states are so assigned as to be more correspondent to normal or specific event and not abstract. HSPN is an extension of the latter type of HMM with multiple channels.

Originally HSPN was developed for the purpose of modeling conversational turn taking of two conversants by the second author [9], and it was applied to the minimal class of the problem, that is, two channels. Moreover, real-time processing was not investigated yet. The problem being faced here is dealt with by multi-channel processing in real time by implementing external objects into Max/MSP environment, and is enabled to model physical motion via Petri net.

3.2. Petri net

Petri net [10] is specified by a finite set pf places, P, (drawn as circles) and a finite number of transitions, T, (drawn as bars), along with an input function and an output function, each of which associates a set of places with a transition (drawn as a set of directed arcs). Places

contain zero or more tokens (drawn as dots). A transition is enabled whenever there is at least one token in each of its input places. An enabled transition can fire by removing one token from each input place and depositing one token in each output place. A marking is defined by a vector, the elements of which specify the number of tokens in each place. The reachability tree is the set of markings that are reachable from a given initial marking.

3.3. Features of Petri net used in HSPN

Similar to HMM, HSPN has a transition probability and an output probability of physical observation, both attached to transition of the Petri net being used.

The Petri net in general has wider ability than Markov model in expressing system behavior. However, the Petri net used in HSPN is a restricted one that has always only one token per channel so that each channel is in either one of defined places.

This restriction assures a safe and bounded network, giving the same descriptive ability as a Markov chain. This level of description is as low as that of a Petri net and equivalent to regular grammar if applied to grammar. Since the network is safe, marking set becomes finite, and reachability graph is easily derived. This reach- ability enables the expansion to a Markov chain. Each marking corresponds to the state of Markov chain.

3.4. Design of HSPN and its conversion flow to HMM

The flow of the resolution of the problem is denoted briefly as:
(1) Petri net -> TPN -> SPN -> HSPN
(2) HSPN -> semi Markov model -> HMM

Construction of the full system notation starting from Petri net is shown in (1), while (2) presents a conversion for solving HSPN. The Petri net here also has the properties of transition probability attached to a trans-
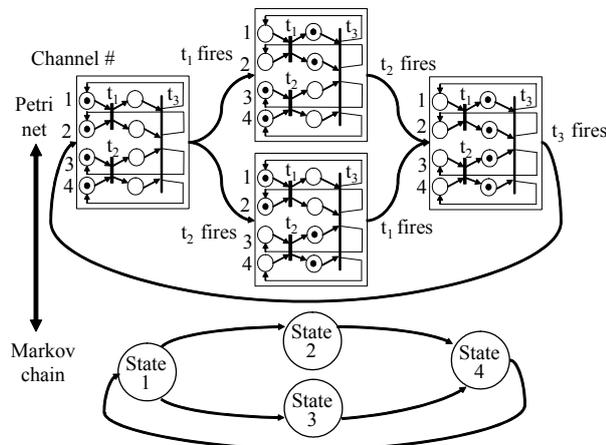
ition and output probability of physical observation attached to a transition as well.

An HMM network configuration is heuristically defined as a left to right model. These states are abstract and do not necessarily correspond to specific events. On the other hand, places correspond to specific events of physical motion, and the network configuration is manually written including their sequences and synchronization. Automatic learning of network configuration is a topic for future study and not considered here.

In the original Petri net, firings were done in zero time. However, delay should be considered in real event occurrence and inter-event synchronization, it is represented as "special" places. Moreover, temporal element and physical observation are attached to places to simulate a real physical motion. This is called a TPN (Timed Petri Net).

Stochastic Petri Net (SPN) is a typical TPN proposed by Natkin et al. [11]. In SPN a transition possesses the time factor, and it is restricted in that only one transition can be fired from enabled transitions at a time. A TPN is converted to an SPN (Stochastic Petri Net) to make transition possess the temporal factor.

Once an HSPN is defined, it can be expanded to a Markov chain from a reachablity tree, since the class of Petri nets are equivalent to Markov chains stated in 3.3. The Markov model in which time is attached to a state is called semi Markov model. Then HSPN is converted to HMM in which semi Markov model is used. Once HMM is defined, we can use an arithmetic property of HMM since it provides a learning algorithm of physical observation probability and transition probability using likelihood evaluation, and recognition algorithm, solving event sequence estimation problem, once the network configuration is defined.
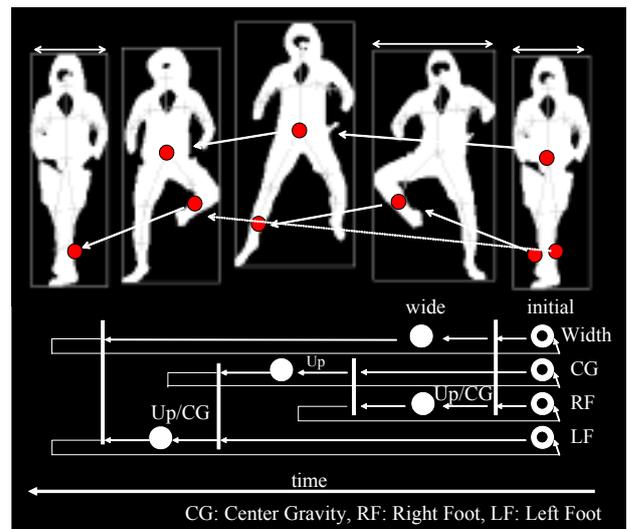
Fig. 3 depicts an example of conversion from Petri net

Fig. 3 State transition diagram of Markov chain derived from reachability of HSPN

Fig. 4. Petri net description of *pas de chat*

to Markov chain. In an initial marking, there are two possibilities of firing: t1 or t2. Either pass ends up with t3 firing, and this is represented equivalently in Markov chain. Correspondence between marking and the state can easily be seen.

4. IMPLEMENTATION OF HSPN

4.1. Description example of physical motion

By adopting Eyesweb as a physical motion analysis tool, the centroid coordinate of each physical portion is acquired. A specific motion can be represented by characteristic states of physical portions. Physical portions in question are assigned as channels of HSPN, and characteristic states for each channel are represented by places. Transition is used between events.

We apply an HSPN to basic choreography of classical ballet. Fig. 4 depicts an example of "Pas de chat". In the upper part of the figure, transition image frames are shown acquired from Eyesweb. In order to describe "pas de chat" both feet (RF and LF), Center of Gravity (CG) and width of the silhouette (Width) are chosen as characteristic channels. The first transition enables wide silhouette along with RF (or one foot) rise toward the direction of CF. The next transition makes CF rise since the dancer jumps and RF fall from CV for the preparation of landing. The third transition makes CG fall for landing with LF rise. The final transition brings all the channels back to the initial states. This example has no alternative other than 1 through 4 transition sequence and has four state serial sequences when expanded to Markov model.

4.2. Implementation of HSPN and HMM

We are still under way in construction of HSPN. Discrete HMM is implemented as an external object under Macintosh OS X, using Xcode and Max/MSP Software Development Kit. Users can specify the number of states, output labels and maximum observation events and model type (Ergodic/Left-to-Right) as argument of patches. In the left inlet observation sequence is input in real time, while in the right switching of learning/ recognition is done. Learned HMM can be stored and read as a text file.

In order to support observation input sequences, vector quantization using LBG (Linde- Buzo – Gray) algorithm [12] is also implemented. This object has both learning and execution functions as well as an HMM object. The system was applied to sound installation work [13], where a few simple moving patterns such as walking right/left and jumping are recognized.

5. CONCLUSION

In order to support new framework of physical motion controlled sound/music synthesis, we propose a new algorithm, HSPN, which recognizes choreographical pattern from physical movement. The system is being implemented into the Max/MSP environment. In HSPN physical motion or choreography is represented as multi-channel event sequences with synchronization among channels, using Petri net. This is finally expanded to HMM, from which the arithmetic is completely available to solve the problem of HSPN. We are still in the process of brushing up the system and planning to apply the system to classic ballet and works in which sound and music are controlled by a dancer.

6. REFERENCES

[1] A. Tanaka, "Musical Technical Issues in Using Interactive Instrument Technology", In Proc. ICMC, pp. 124-126, 1993.

[2] D. O'Sullivan, T. Igoe, "Physical Computing: Sensing and Controlling the Physical World with Computers", p240, 2004.

[3] A. Camurri et al., "EyesWeb—Toward Gesture and Affect Recognition in Interactive Dance and Music Systems," Computer Music J., vol. 24, no. 1, pp. 57-60, Spring 2000.

[4] A. Cera, H. Robbe, "Rew." - Music and Dance performance for 2 dancers and electronics. Premiered in Lisbon, 2003.

[5] S. Young et al., "The HTK Book", 2002.

[6] M. Wright et al., "Open Sound Control: State of the Art 2003", pp153-159, NIME03, Montreal, 2003.

[7] M. Puckette, Combining Event and Signal Processing in the MAX Graphical Programming Environment, Computer Music Journal, 1991.

[8] A. Hutchinson, "Labanotation: The System of Analyzing and Recording Movement" – 4th edition, 2004.

[9] Naotoshi Osaka, "Conversational Turn-taking Model Using Petri Net," ICSLP'90, pp.1297-1300, 1990.

[10] J. L. Peterson, "Petri net theory and the modeling of systems," Prentice-Hall, Englewood Cliffs, N. J., 1981.

[11] S. Natkin, "Les reseaux de Petri stochastiques et leur application a l'evaluation des systemes informatiques," These de Doct. Ing., CNAM, Paris, 1980.

[12] Y. Linde, A. Buzo, R.M. Gray, "An algorithm for vector quantizer design", IEEE Transactions on Communication, COM-28: pp.84-95, 1980.

[13] M. Go and H. Misu, "Physical Intimacy," Inter-college Media art exhibition, organized by Kyoto Seika University, Dec. 2006.

Physical modeling synthesis is a very different approach to creating sound that can yield unique results that go well beyond the idea of emulating acoustic systems. Virtual instruments like Sculpture are perfect for sound design applications thanks to an extensive degree of control and specificity. Mutable Instrument's Eurorack synth voice, Elements, allows for the modulation of parameters via CV (control voltage), which can create unusual hybrid sonorities when connected to analog or digital oscillators and envelopes. While the parameters in physical modeling synthesis may seem at first forei Using trajectories of action patterns reduces limitations of human motion rather than a method such as HMM. We use an omnidirectional image sensor for... Recognition of physical motion pattern using stochastic petri nets for sound synthesis. January 2007. Munetatsu Go. Relaxed soundness has been introduced previously as a more suitable requirement. The paper presents how methods from controller synthesis for Petri nets can be used to automatically make this type of models sound. For this reason we adopt the idea of controllability for Petri net workflow models. Making Workflow Models Sound Using Petri Net Controller Synthesis. 9 years 1 months ago.