“VirtualPhilharmony”: A Conducting System with Heuristics of Conducting an Orchestra

Takashi Baba
Kwansei Gakuin University
takashi-b@kwansei.ac.jp

Mitsuyo Hashida
Kwansei Gakuin University
hashida@kwansei.ac.jp

Haruhiro Katayose
Kwansei Gakuin University
katayose@kwansei.ac.jp

ABSTRACT
“VirtualPhilharmony” (V.P.) is a conducting interface that enables users to perform expressive music with conducting action. Several previously developed conducting interfaces do not satisfy users who have conducting experience because the feedback from the conducting action does not always correspond with a natural performance. The tempo scheduler, which is the main engine of a conducting system, must be improved. V.P. solves this problem by introducing heuristics of conducting an orchestra in detecting beats, applying rules regarding the tempo expression in a bar, etc. We confirmed with users that the system realized a high “following” performance and had musical persuasiveness.

Keywords
Conducting system, heuristics, sensor, template.

1. INTRODUCTION
Musical conducting, which sometimes involves directing an orchestra with as many as one hundred members, has been a sought-after profession since its professional status was established in the 19th century. A conductor’s manner of musical expression is quite different from that of an instrumentalist playing a musical instrument or a vocalist singing a song. Inspired by this unique manner of creative expression, composers and engineers in the field of computer science started to develop conducting interfaces in the late 1980s [1][2]. On the other hands, there are many amateurs who want to conduct an orchestra. Many of them are air-conductor, they shake a baton while listening to the great recordings of the maestros. Recently, conducting interfaces for entertainment purposes [3], for example, one developed by Wii Music (2008) [4], have started to be actively developed.

Recently, a new approach to music called active music listening has been advocated. One of the crucial merits of active listening is that it fosters listeners’ understanding of music by involving them in the music with more of a self-guided attitude. As a result, user expectations of the conducting systems and automatic accompaniment systems have been increasing.

The existing conducting systems were designed as a means of musical expression through computer-generated music or for entertainment. In either case, usability is one of the most important issues when designing such systems. Another important point to be considered when designing a conducting system is how well it can provide the user with the sensation of conducting a real orchestra. Systems that can create this sensation well enable professional musicians to give a more sophisticated performance and provide general users with high-end entertainment. Well-designed conducting systems are also applicable for educational purposes.

With the aim of providing players with the sensation of conducting a real orchestra, we have been developing a conducting system called “VirtualPhilharmony” (V.P.) in which the heuristics of conducting an orchestra are given importance. In this paper, we describe the conceptual basis and the design of V.P., focusing on the implementation of several gestural sensors and we explain “Concertmaster Function” that is designed by incorporating with the heuristics of conducting an orchestra to get the sensation of conducting a real orchestra.

In Section 2, we highlight related work on conducting systems, and describe the aim of V.P. In Section 3, we provide an overview of V.P., and describe the concertmaster function in Section 4. Section 5 contains an examination and evaluation of our results, and Section 6 contains our conclusions.

2. RELATED STUDIES AND AIM OF V.P.

2.1 Related Studies
One of the earliest studies done on conducting systems investigated the “Radio Baton” developed by Mathews in 1987 [1], which controls a MIDI (Musical Instrument Digital Interface) control signal using two batons with one’s hands on a sensor plate. Each baton is used for giving beats or changing the pitch and timbre of a music piece in real time. Although this system was not focused on conducting simulation, the idea of tempo control by user-given beats is similar to conducting.

Usa et al. developed the “Conducting Simulator,” which was aimed at generating a detailed recognition of conducting motions that followed the orthodox conducting method using an acceleration sensor and the hidden Markov model (HMM) [2].

“iFP” by Okudaira et al. is a performance interface for playing piano music by hand-shaking [5]. It extracts beat information from the hand-shaking and feeds it into a expressive performance template, where the expression is created from existing great piano performances.

“UBS Virtual Maestro” by Nakra et al. detects beats from the movement data of the accelerometer sensor in a Wii Remote and stretches the audio signal by a phase vocoder [6]. The big advantage of the system is that the performance templates were recorded by the Boston Symphony Orchestra; the multimedia performances with synchronized audio and video provide a more realistic conducting experience. The exhibition booth for the system demonstration is located in the lobby of the Boston Symphony Hall. From the point of view of the musical experience, conducting systems using audio signals are superior to those using MIDI signals. However, the audio conducting systems have to resolve a crucial problem regarding deterioration of the sound quality caused by the phase vocoder.

"Maestro Musik" [3] and "Wii Music" [4] are video games that simulate the conducting. The former uses a special baton device and the latter uses a Wii Remote, giving precedence to entertainment value rather than the actual sensation of conducting. As a result, the feeling of conducting is missing an element of reality and the user often fails to accurately conduct the performance from a musical standpoint because of the oversensitive feedback.

2.2 Heuristics of conducting an orchestra

As mentioned above, conducting systems have been made for the purpose of questing for musical expression in computer-generated music or entertainment. However, the existing conducing systems so for provide its users with little sensation conducting a real orchestra, from the experts' point of view.

In case of real conducting, a conductor faces with members of an orchestra. The members do not always obey the directions of the conductor completely. Each of the members has her/his own musicality and intentions just like the conductor. Interaction between a conductor and the orchestra yields real musical performances. The biggest issue of the conducting systems is that few systems consider the viewpoint orchestra members' musicalities or the intentions when generating music performances.

iFP, mentioned above, is one of the few systems that deal with interaction between the player and the machine musicality. The scheduler of iFP mixes the player's intention of tempo and dynamics and that described in a expressive performance template. It schedule the timing of notes to be issued, using the predicted tempo given by the following equation

$$P_{n+1} = \alpha A_n + \beta B_n + \gamma C_n \quad (1)$$

where P_{n+1} is the predicted tempo of the next beat, A_n is the moving average of the last four beats from the performance history, B_n is the differentiation between the current and the last tempo, and C_n is the tempo that the template indicates. α, β and γ are the weighting coefficients for each variable $(0 \le \alpha,\beta,\gamma \le 1)$. In iFP, the balance of these weight parameters characterizes the performance taste that is interaction between the player and the template that strands for the virtual orchestra. But the weight parameters used in iFP are static regardless the music style and the tempo all the time. The weight parameters should be adjusted as they fit to the music style (E.g. a march, a waltz), and the template itself also should be adjusted in accordance with tempo dynamically.

The goal of V.P. is to provide its player with the sensation of conducting a real orchestra. We implemented a function called "Concertmaster Function" where, set of heuristics about musicalities or the intentions that orchestra members are accumulated for this goal. A human concertmaster conveys instructions of a conductor to orchestra members, and tells the orchestra members' will to the conductor. Generally, the leader of the first violin section of an orchestra takes charge of the concertmaster. In V.P., concert-master plays the role of adjustment of intention of its player and the template. To be concrete, heuristics such as, which of the past beats are crucial when predicting the future beat position, or how the data of a

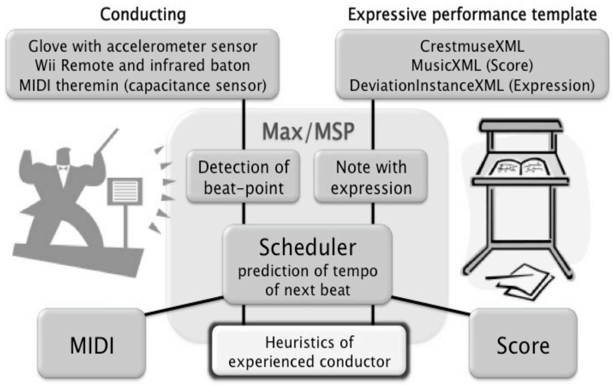

Figure 1. Overview of V.P.

template should be modified as they fit the tempo change are examined and applied through the analysis of virtuoso music performances.

3. SYSTEM DESIGN OF V.P.

V.P. is a conducting interface, the basic idea of which is derived from iFP. Figure 1 shows an overview of V.P. In V.P., like many other conducting systems, a tempo of a beat is detected by some gesture sensors from conducting motions that a player draws. It is a conducting interface based on beat tapping. It captures only one vertical dimension of a gesture of conducting.

V.P. is implemented on Max/MSP by Cycling'74. In V.P, a tempo of the next beat is predicted from the history of beat that the player has given. And an expressive performance template is used, as well as iFP. We introduce heuristics of conducting an orchestra, as the template works more naturally.

3.1 Gesture sensing device

In V.P., the local minimum point of the conducting motion captured by various sensors. This point is regarded as beat-point. Three methods were used in the acquisition of the beat-point to correspond to various conditions of the stage: 1) a glove interface using an accelerometer sensor, 2) a baton using an infrared sensor (Wii Remote), and 3) a capacitance sensor (MIDI Theremin).

3.1.1 Glove interface using an accelerometer sensor

The V.P. players attach an original glove equipped with an accelerometer sensor to their right hands. The LilyPad Arduino [7] is used for the accelerometer sensor and the microcontroller. The three-dimensional acceleration information with a sampling time of 10 ms is sent to the system through Bluetooth using a BlueSMiRF [8]. The gravity component is extracted from the three-dimensional acceleration data. This direction is labeled as the vertical component, and the local minimum values of the vertical component are considered beat-points. Figure 2-a shows pulling on the glove.

We had the option of using the Wii Remote, which is also equipped with an acceleration sensor, like UBS Virtual Maestro [4], however, the Wii Remote (160 g) is much heavier than an ordinary conductor's baton (10 g). We were concerned that the heaviness of the Wii Remote would worsen the player's conducting motion because the center of gravity of the Wii Remote is far from the player's hand. In addition, although the weight of the original glove is about 100 g, this weight is not a problem because the player uses a real baton, and its center of gravity is nearer the player's hand than it is with the Wii Remote.

There are two advantages to using the acceleration sensor. One is not being limited by the V.P. player's placement and direction and the other does not need to use caution when making the conducting motion. V.P. players can construct any conducting figure they like compared with the two ways described in more detail below that have large restrictions on position.

In contrast, the acceleration value increases and decreases significantly as a result of various factors other than the vertical movement, such as rotation. Therefore, beat tracking using acceleration sensors will introduce more mistakes than other sensing inputs.

3.1.2 Baton using infrared sensor (Wii Remote)

An infrared LED (TOSHIBA: TLN115A, 950 nm) is attached to a baton tip. The system traces the LED's motion in 2 dimensions using a CMOS sensor (infrared camera) with a built-in Wii Remote. The sampling time is 10 ms. The beat-point is extracted from the local minimum values of the vertical component (Figure 2-b).

Beat detection using the location information can more accurately acquire data than using the accelerometer. However, there are many restricting factors when using infrared sensors, including outside light, obstacles between the LED and the camera, the range of the camera, and the irradiation angle of the LED. As a result, the detection mistakes will be greater than with other methods.

3.1.3 Capacitance sensor (MIDI Theremin)

An electrostatic capacitance sensor (the MIDI Theremin, see Figure 2-c), one of the main interfaces since iFP, is a device that captures data on the distance from the sensor to the hand through the change in electrostatic capacitance, converting the distance into a MIDI pitch-bend signal, which then controls the pitch in the sound source. The sampling time is 11 ms and the local minimum values of the pitch-bend signals are treated as beat-points. In Nakra's paper [6], it is described that the sampling time needs less than 10 ms. It is necessary for the extraction of the beat-point precisely. However, in the case of MIDI Theremin, 11 ms is the minimum sampling time. And, it is the capacitance sensor. Because it captures the distance, it can extract the beat-point more precisely than the other sensers.

The MIDI Theremin is the most stable device of the three sensors because it makes fewer mistakes in the recognition and detection of beat-points. However, the available range of the measurement is narrow (about 20 cm from the antenna), increasing the restrictions on user location compared with the other sensors.

3.2 Expressive performance template

V.P. utilizes an expressive performance template, as well as the score data. V.P. uses CrestMuseXML (CMX) for reading this detailed information about the score. CMX is a group of common data formats used to treat various types of information about music [9].
It deals with MusicXML, which describes score information, and DeviationInstanceXML (DIXML), which describes deviation information. The deviation information used in V.P. is 1) attack-deviation for all notes, 2) release-deviation for all notes, 3) volume-deviation for all notes and 4) tempo-deviation for all beats. The MusicXML data are available from various commercial music production software packages such as "Finale." The DIXML data are obtained from expressive performance data in SMF format using our original software.

Figure 2. Gesture Sensors

(a) Glove with accelerometer, (b) Baton with LED and Wii Remote, (c) MIDI Theremin, from the left

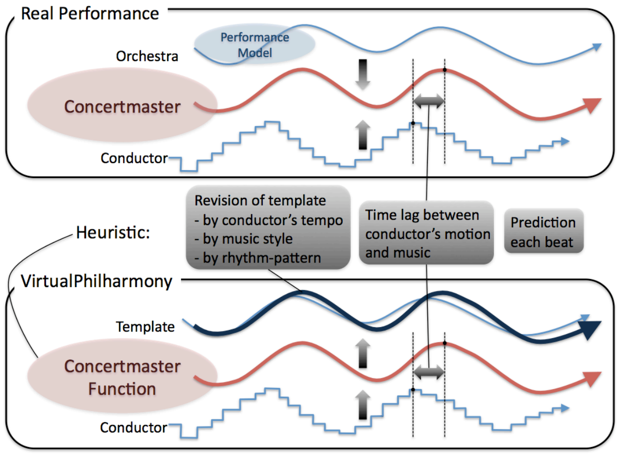

Figure 3. Real performance and V.P.

3.3 Concertmaster Function

Figure 3 shows the role of the concertmaster in real orchestra and V.P. As mentioned at the section 2.2, the concertmaster of V.P. plays a role of coordinating the intention of the conductor (player) and that of the virtual orchestra that is derived from the performance template. The scheduler of V.P. estimates the player's tempo of the next beat from the history of beat positions that the player has given, and calculates the scheduler tempo of the next beat, using the estimated player's tempo and the tempo described in the template.

In real performance, each member has a performance model: how the music should be performed. The model is constructed in his/her head based on the member's own musicality or various information about the music. The information includes the music style, the beat time and the tempo suitable etc. And, the performance is created by changing the model along to the tempo shown by the conductor. The concertmaster's role is combining the members with the conductor smoothly.

In V.P., because there is no orchestra, the performance model does not exist too. One of roles of the expressive performance template is a substitution of this model. Because the expressive performance in the template is created based on the performance model of the orchestra members. Then, the template should not be static, but dynamic. In iFP, because the template was static, there were cases where the player failed in the performance when the player's tempo was different from the template's tempo. In V.P., the expressive performance template is dynamic. It is revised by the music style or the prediction tempo similarly to the performance model of real orchestra. Hereby flexibility is added to the template. The role of a concertmaster in V.P. is combining the template with the player (conductor) smoothly. We describe the detail of the concertmaster function in Section 4.

3.4 Integration tempo scheduler

In most existing conducting systems, tempo is handled as a discrete values on each beat, which causes imperfection in scheduling notes within a beat, such as in case of *ritardando* and *accelerando*. In musical performance, a tempo change should not be regarded as a discrete change in each beat, but rather as a continuous change. To simulate this, an integration scheduler was introduced. The tempo change in a beat is expressed by approximating the continuous tempo change to the straight line. The present tempo T_n and the prediction tempo P_{n+1} are connected by a straight line. Figure 4 shows a case where a beat is divided into twelve equal parts. In case of a quarter per a beat, one interval corresponds to a triplet of a demisemiquaver.

3.5 Display of score

V.P. shows the musical score on the display because of the importance of the score in conducting. It utilizes a scroll-bar that shows the present location in the score for the player who cannot read the score. The automatic page-turning function is realized by extracting the page information from the MusicXML file regarding the location where the score needs to be turned over (Figure 5). Moreover, The player can write a memo in the score, and can see it when he/she conducts some pieces.

4. CONCERTMASTER FUNCTION

There are three functions in the concertmaster function.

a. Revision of the template depending on the predicted tempo

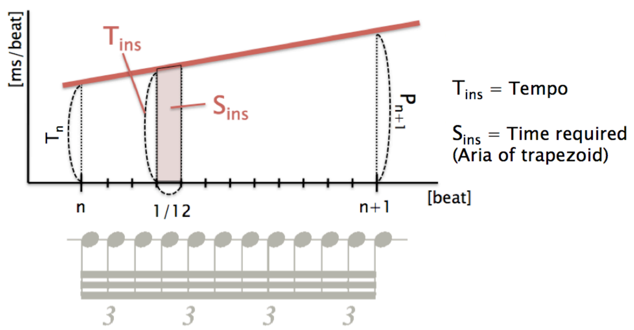

Figure 4. Integration Scheduler

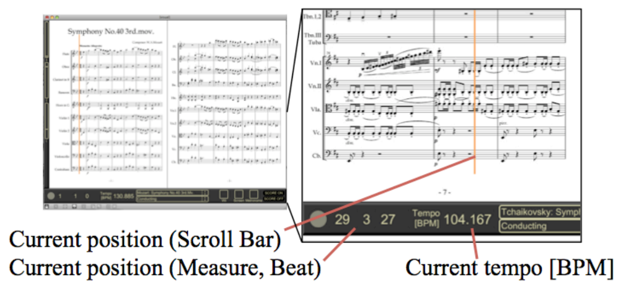

Figure 5. Display of score

In iFP, because the expressive performance template was static, when a tempo that player gave was different from the template tempo, a performance often failed. To cancel this problem, the template is revised by the prediction tempo in V.P. There are three revisions in V.P.

a-1. Revision of the template tempo per beat:

When the prediction tempo is different from the template tempo, the template tempo is revised to the prediction tempo per beat.

a-2. Revision of the tempo in a bar by the music style:

There are cases where the tempo of each beat in a bar is not constant depending on the music style. We investigate how each beat changes depending on the tempo of music in case of Viennese Waltz.

a-3. Revision of some rhythm-pattern in a beat:

When there are some specific rhythm-patterns in a beat, the performance technique of the rhythm-patterns change with a tempo change. We investigate how the rhythm-pattern changes depending on the tempo of music in case of the rhythm-pattern of dotted note.

b. Tempo prediction of next beat based on beat time and music style of a tune

Beat time and music style of a tune have great influences on tempo prediction of next beat. Then we examine how long past beats affect the tempo prediction, based on analyses of real tunes. We investigate correlations between past beats and next beat, and find the optimum degree of the prediction and the coefficient of the prediction by using the linear prediction.

c. Adjustment of time lag between conducting motion and music

In real performance, it is often the case that the beat of the music is later than the beat-point of the conducting motion (local minimum point). In V.P., the player can adjust the time lag.

4.1 Revision of the template depending on the predictive tempo

4.1.1 Revision of the template tempo per beat

The template tempo is revised by using the equation

$$U_m{}' = U_m + \alpha(P_{n+1} - U_m)\ (m \geq n+1) \tag{2}$$

where U_m is the template tempo before the revision, $U_m{}'$ is the one after the revision, P_{n+1} is the prediction tempo, α is the weighting coefficients $(0 \leq \alpha \leq 1)$.

4.1.2 Revision of the tempo in a bar by the music style

Viennese Waltz has the rhythm-pattern as shown in Figure 6-a. t_1 to t_3 are the times required of each note. Because it is dance music, its tempo is deeply related to human dance movement. Because dancers turn on the second beat in a Viennese waltz, t_2 lengthens as previously stated (t_1 shortens). Moreover, the ratio of $t_1 : t_2 : t_3$ changes depending on the tempo of music. We investigate these features by analyzing real tunes.

We analyze two pieces as shown in Table 1-A. The Blue Danube is a sample piece in which change of tempo is a little.

On the other hand, Die Fledermaus Ouvertüre is a sample piece in which change of tempo is bigger than The Blue Danube.

About these pieces, we get rate of a time required of each beat in a bar ($\frac{t_1}{t_M}, \frac{t_2}{t_M}, \frac{t_3}{t_M}$). $t_M (= t_1 + t_2 + t_3)$ is the time required of one bar. We show how these ratios are changed by the tempo in Figure 7-a and 7-b. When all of the three rates are 0.333, the three beats are performed equally (1:1:1).

In case of The Blue Danube, a feature of Viennese Waltz is seen remarkably. The second beats are longer than the other beats as shown Figure 7-a. The transition of the ratio of the tempo of three beats in a bar is approximated by lines with the least square method.

$$L_{a1} = y = 0.0019x + 0.1888 \tag{3}$$

$$L_{a2} = y = -0.0008x + 0.4067 \tag{4}$$

$$L_{a3} = y = -0.0011x + 0.4045 \tag{5}$$

These three equations show that as the tempo is faster, the tempo of the three beats are more equal. For example, the ratio of the tempo of the three beats in a bar is 0.3028:0.3587:0.3385 when the tempo is 60[BPM] and 0.3385:0.3275:0.322 when 75[BPM].

In case of Die Fledermaus Ouvertüre, the variance of the ratios is very large as shown in Figure 7-B. Because this sample has extreme change of tempo including *fermata* and *ritardando*.

From these results, in V.P., the rhythm-patterns of Viennese Waltz are revised by using the lines $L_{a1} - L_{a3}$ depending on the prediction tempo. It is the Viennese Waltz Mode. This mode is applied to the case of Die Fledermaus Ouvertüre. Though it cannot reenact complicated tempo change, can create an atmosphere of Viennese Waltz.

By the way, in case of Viennese Waltz, one bar is conducted with one stroke, like drawing a circle, generally [10]. But, in V.P., the prediction schedulers that have been described up to now assume that the beat-point of the conducting motion and the beat of music have a one-to-one correspondence; it is not possible to conduct two or more beats together by shaking one stroke (one beat that conductor shakes). In the Viennese Waltz Mode, the player shakes three beats by one stroke.

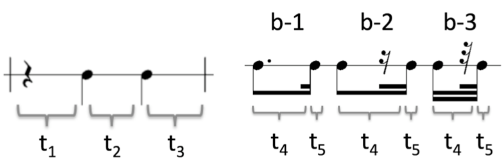

Figure 6. a. Viennese Waltz b. rhythm of dotted note

Table 1. Subject pieces for the analysis

A) Viennese Waltz	J. Strauss II: The Blue Danube 4th waltz (3 beats)
	J. Strauss II: Die Fledermaus Ouvertüre Waltz (3 beats, Bar: 316-350)
B) Rhythm of dotted note	A. Dvořák: Humoresque No. 7 Main theme (2 beats, Bar: 1-8)
	D. Shostakovich: Symphony No. 11 4th mv. 1st theme (4 beats, Bar: 6-17)

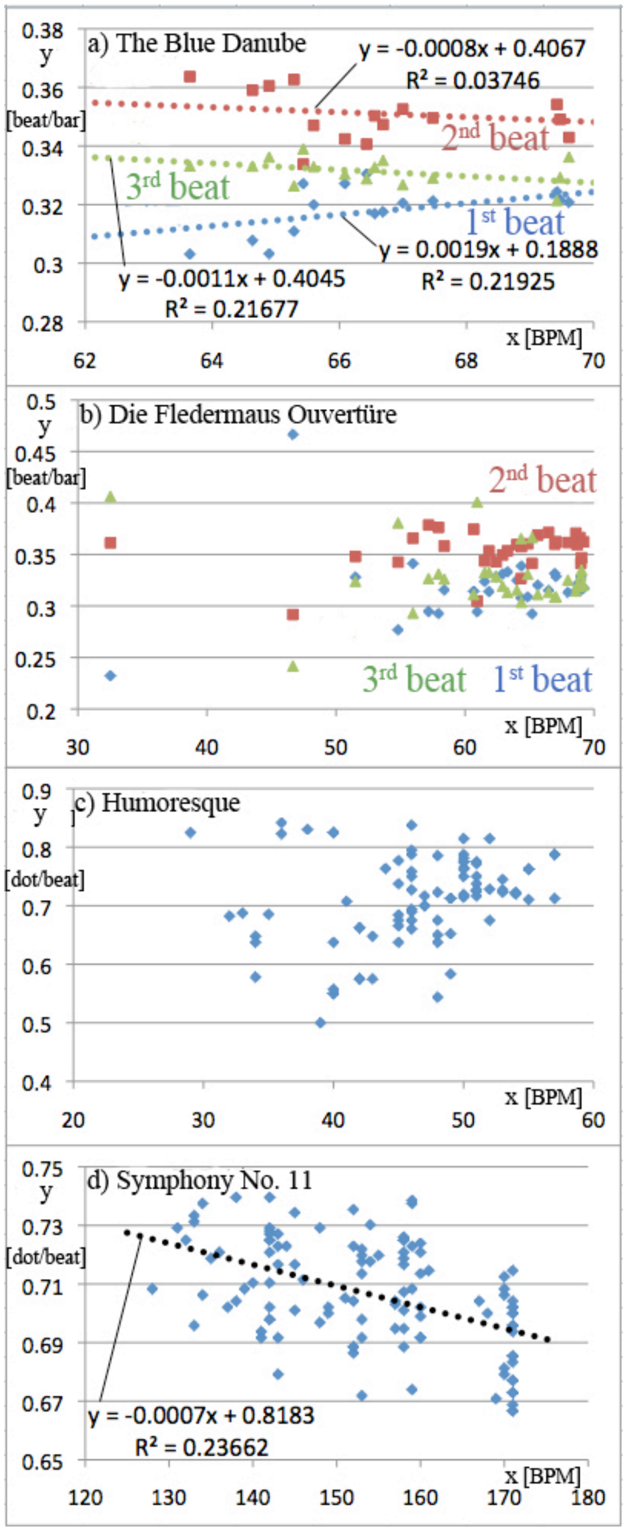

Figure 7. Rhythm-patterns and tempo

4.1.3 Revision of some rhythm-patterns in a beat

In V.P., we handle the rhythm-pattern of the dotted notes as shown in Figure 6-b. t_4 and t_5 are the times required of each note. The ratio of $t_4 : t_5$ changes depending on the tempo of music. We investigate these features by analyzing real tunes. Though these rhythms exist innumerably in various patterns in music, V.P. treats only the case where one beat (one stroke) is constituted the two notes.

We analyze two pieces as shown in Table 1-B. Humoresque has a slow tempo, and it is a sample in which change of tempo is big (the rhythm-pattern is Figure 6-b-3). Shostakovich's Symphony No. 11 has a rapid tempo, and it is a sample in which change of tempo is little (the rhythm-pattern is Figure 6-b-2).

About these pieces, we get rate of a time required of a doted note in a beat ($\frac{t_4}{t_B}$). $t_B(=t_4+t_5)$ is the time required of one beat. We show in Figure 7-c and 7-d. When the rate is 0.75, the two notes are performed in the ratio of 3:1.

In case of Humoresque, the variance of the ratios is very large as shown in Figure 7-c. Because various expressions are added, the ratios are never constant.

In case of Shostakovich's Symphony No. 11, when the tempo quicken, the ratios of the dotted notes approach from 3:1 (0.75) to 2:1 (0.66) as shown in Figure 7-d. Because it becomes difficult to play fine notes. The transition of the ratio of each beat is approximated by a line with the least square method.

$$L_b = y = -0.0007x + 0.8183 \quad (6)$$

From these results, in V.P., when tempo is 100 [BPM] and over, the rhythm-patterns of dotted notes are revised by using the line L_b depending on the prediction tempo. For example, the ratio of the tempo of the two notes in a beat is 0.7483:0.2517 when the tempo is 100[BPM] and 0.6783:0.3217 when 200[BPM].

4.2 Tempo prediction of next beat by viewpoint of orchestra

First, we investigate correlation between the prediction tempo of next beat P_{n+1} and the past beats in real tunes, to examine that how long does past beats affect the tempo prediction. N is a number of past reference. Prior to every even beat

$$P_{n+1} = \sum_{k=0}^{N} a_k T_{n-k} \quad (7)$$

a_k (Linear predictive coding: LPC), the average ε of the square root of the square error margin for the P_{n+1} and T_{n+1} of the actual measurement value, is minimized and found for each beat.

$$\varepsilon = E\left[\sqrt{\left|T_{n+1} - P_{n+1}\right|^2}\right] \quad (8)$$

Eight pieces that have some typical tempo-change features were used for the analysis (Table 1).

Category A contains pieces in which each beat is played in time.

Table 2. Subject pieces for the analysis

A) Constant tempo	J. Strauss I: Radetzky March Main theme (2 beats, Bar: 5-20)
	P.I. Tchaikovsky: Marche Slave Main theme (4 beats, Bar: 5-12)
B) Frequent auf-tact	W.A. Mozart: Symphony No. 40 3rd mv. main theme (3 beats, Bar: 62-77)
	J. Brahms: Symphony No. 1 4th mv. 1st theme (4 beats, Bar: 1-43)
C) Unsettled tempo	P.I. Tchaikovsky: Symphony No.6 1st mv. 2nd theme (4 beats, Bar: 89-100)
	J. Brahms: Hungarian Dance No. 1 Main theme (2 beats, Bar: 73-80)
D) Waltz-style	J. Strauss II: The Blue Danube 4th waltz (3 beats)
	P.I. Tchaikovsky: Waltz of the Flowers Main theme (3 beats, Bar: 34-53)

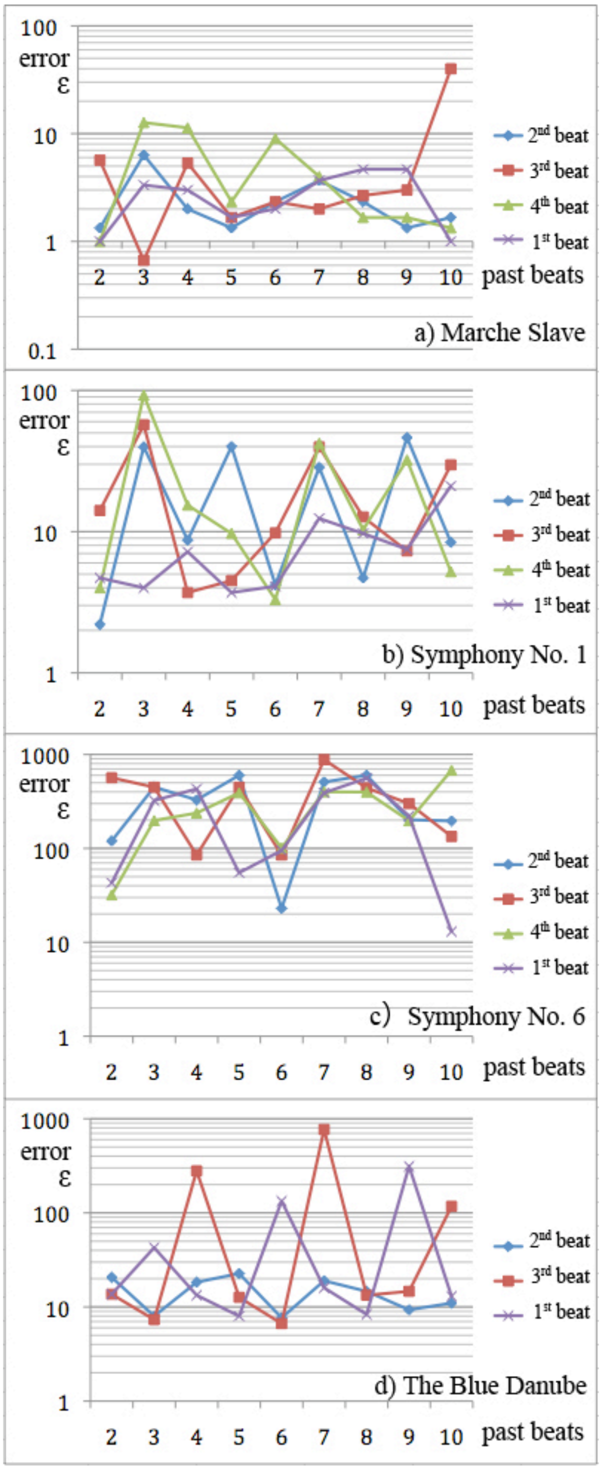

Figure 8. Past beats and the prediction error

Category B contains pieces in which phrases often begin with auf-tact (It means that a start of a phrase shifts to previous from a bar). In the pieces given in Table 1, the last beat in a bar is longer than the other beats.

Category C contains pieces in which change of tempo is extreme. Tchaikovsky's Symphony is a part of *espressivo* (expressively), and Hungarian Dance is a Csárdás (Hungarian Volk Dance with extreme tempo change).

Category D contains waltz-style pieces. These pieces are dance music, so their tempo is deeply related to human dance movement. For example, when dancers turn on the second beat in a Viennese waltz, the second beat lengthens as previously stated (the first beat shortens). Because this feature is seen only in case of Viennese Waltz, we compare with the Russian Waltz.

For these eight tunes, the tempo of each beat was estimated by human listening (ten recordings per tune for a total of eighty recordings).

For example, for the cases of (a) "Marche Slave", (b) Brahms' Symphony No. 1, and (c) Tchaikovsky's Symphony No. 6, (d) "The Blue Danube." Figure 8 shows the transition of the error ε when the number of past references is changed from two to ten (for an average of ten recordings). When the value of ε is large, the prediction of the tempo of the next beat becomes difficult.

a) The error ε is low. Therefore, to keep tempo constant, we use moving average for the algorithm of the prediction, the number of past reference N is fixed with the number of the beat time (if a tune is 4 beat, N is 4).

b) When the first beat is predicted, the error ε is less than the others. When the number of past references is an odd number, the error increases; when it is an even number, the error decreases. Therefore, when the first beat is predicted, N is fixed with the number of the beat time. And when the other beats are predicted, V.P. refers to prior to 2, 4, 6 beats. Each LPC is used for coefficient of the prediction.

c) The error ε is extreme high. Moving average is used, and N is fixed with the number of the beat time. The sensory threshold is provided, and if a sudden tempo change more than the threshold happens, $N = 1$ $(P_{n+1} = T_n)$.

d) Although the error ε is uneven based on beat position or the number of past references, periodicity is seen in it. When the second beat that is the feature in the Viennese waltz is predicted, the error is small. When the second beat is predicted, N is fixed with the number of the beat time. In case of the first beat, it predicted except prior to 3 beats, 6 beats. In case of the third beat, it predicted except prior to 4 beats, 7 beats. Each LPC is used for coefficient of the prediction. This information is included in the Viennese Waltz Mode.

The same analyses were done about other tunes; we got the similar result above about Category A, B, C. In Category D, Russian Waltz lengthens the third beat than the second beat, and this tendency is the same as Category B.

4.3 Adjustment of time lag between conducting motion and musical performance

A time lag between a beat-point of a conducting motion and a beat in musical performance is dependent on the tempo; in general, the time lag lengthens in the case of music with a slow tempo and shortens in the case of music with a rapid tempo. In V.P., a parameter that can adjust for the amount of time lag is prepared.

$$Timelag = \frac{60000\alpha}{P_{n+1}} \tag{9}$$

where $Timelag$ [ms] is amount of the time lag, P_{n+1} is the prediction tempo [BPM], α is the weighting coefficients $(0 \le \alpha \le 1)$.

5. EVALUATION AND EXAMINATION

The goal of V.P. is to provide its players with a sensation of conducting a real orchestra, and we introduced the concertmaster function, which is equipped with heuristics of conducting an orchestra. In this section, we examine the validity of introducing heuristics for increasing the sensation of conducting a real orchestra. First, we show you an interview with a professional conductor who played the conducting with V.P. Next, we introduce the feedback of those who played the V.P. Then we discuss the possibility and the future works of V.P.

5.1 Interview with an expert conductor

We conducted an interview with Mr. Hoshina, H. who is a conductor and a composer, also one of the most famous supervisors of directing brass band in Japan, for evaluating V.P. After explaining the outline of V.P, we asked him to conduct Tchaikovsky's Symphony No. 6 and the Blue Danube using V.P. Hoshina's comments are summarized as follows:

1). Concertmaster function with the heuristics provides better sensation of conducting than without heuristics.

2). Tempo following and delay time, when conducting Viennese Waltz performance reproduce the sensation of conducting a real orchestra very well.

3). When I tried sudden tempo down in conducting Tchaikovsky's Symphony No. 6., the V.P. did not follow properly.

4). I'd like to let V.P. be used by amateurs to train conducting orchestras.

As for 3), it is impossible to predict the conducting motion of the player perfectly. Especially in case of tempo down, systems based on a predictive scheduler are obliged to issue the notes on the next beat. To cope with this problem, it is necessary to capture the other data than the beat-point from the conducting motion. We are considering to implement a function to capture conductor's sign from her/his gestures, using knowledge of the possible cases given beforehand.

5.2 Feedback from Demonstrations

V.P. has been demonstrated at Japanese domestic symposiums, EC 2009, Rencon 2009, and Interaction 2010, and the international CrestMuse Workshop 2009, so far. We obtained various variable comments of the participants, including some experts in computer music field (Figure 9).

Approximately 120 people, including R. Dannenberg, and C. Raphael enjoyed playing with V.P. and gave us favorable comments.

Comments about the system improvement are mainly regarding the narrow range of gesture sensor when using the infrared camera. We are going to use a wide-angle lends for improving this point.

Figure 9. Demonstration in CrestMuse Workshop 2009

Twenty percent of the respondent had experiences of some musical instruments. Around half of them commented that they wanted to use both hands, when conducting. In real conducting, a conductor uses both hands.

Main request of those who had little experience in playing the instruments is to provide a CG character that animates synchronized with the music instead of displaying a score. It is difficult to read musical scores for musical beginners. A CG character animated with music will increase fun of playing with V.P. and it can be used as an indicator of the gesture sensors. The latest version of V.P. provides a simple CG character responding to this request.

5.3 Discussion

The heuristics implemented in V.P. are 1) revision of the template tempo per beat, 2) revision of the tempo in a bar by the music style and 3) revision of some rhythm-pattern in a beat. We discuss effectiveness of introducing these heuristics through observation of conducting three music pieces.

In the case of Brahms's Symphony No. 1 (Category B in Table 2), the tempo of the predicted beat is highly correlated with the tempo of the beat prior to the every even beats as shown in figure 8-B. This piece is four beats. In case of four beats, the first and the third beats are given accents, and the second and the fourth beats are not given clear accents. The prediction error of the first beat is small. This may be related with this piece has an auf-tact structure. The tempi of the fourth beat slowed down, as we had expected when we conducted this piece. This is one of the evidences that the scheduler with the heuristics works well.

As shown in 5.1, sudden tempo down, which is commonly seen in the performance of Tchaikovsky's Symphony No. 6 (Category C in Table 2), may confuse the system. We verified revising the template in accordance with the tempo lessens the error, compared with the processing without revising the template. As a future work, we would like to examine the effectiveness of the countermeasure described in 5.1.

When playing the Blue Danube among the all pieces in Table 2, we obtained the most realistic sensation of conducting. We regard that there are two reasons. First, when calculating the tempo of the next beat, priority is given to the prior beats, which are correlated with the next beat to be predicted (see 4.2). Second, the rhythm-pattern of Viennese Waltz template is revised in accordance with the tempo when the piece is played (see 4.1.2).

And in all cases of conducting the pieces, the time lag value between conducting motion and musical performance affects sensation of conducting. The time lag 0 ms is unnatural especially for those experienced in conducting. The time lag between 100 ms and 300 ms is the realistic for those experienced in conducting.

6. CONCLUSION AND FUTURE WORK

In this paper, we described the conducting system "VirtualPhilharmony", which includes heuristics of conducting an orchestra.

To obtain a genuine conducting feeling, we improved the previous system (iFP) in terms of tempo scheduling, and we introduced heuristics of conducting an orchestra. Furthermore, we added a glove interface and an infrared sensing baton as well as the MIDI Theremin to provide V.P. players with various methods of input to simulate conducting with a baton.

V.P. contributes to the development of high-end entertainment, musical education and the practice of real conducting. We are still improving V.P. to create a more genuine and natural conducting performance, including the use of left-hand gesturing. In the future, we will develop V.P. into a conducting educational tool for children.

7. REFERENCES

[1] ICMA Video Review, Vol. 1, 1991 (90 minutes) Review of Research in Computers and Music.

[2] Usa, S. and Mochida, Y., "A Multi-modal Conducting Simulator," In Proc. Int. Computer Music Conf. (ICMC'98), pp. 25--32 (1998).

[3] Katayose, H. and Okudaira, K.: iFP A Music Interface Using an Expressive Performance Template, Entertainment Computing 2004, Lecture Notes of Computer Science, Vol. 3166, pp.529-540 Springer (2004.9).

[4] Nakra, T., Ivanov, Y., Smaragdis, P., Ault, C.: The USB Virtual Maestro: an Interactive Conducting System, pp.250-255, NIME2009 (2009).

[5] http://www.gae.co.jp/game/ms/maestro_top.htm

[6] http://www.nintendo.co.jp/wii/r64j/

[7] http://web.media.mit.edu/~leah/LilyPad/

[8] http://www.sparkfun.com/commerce/categories.php

[9] Hashida, M., Matsui, T. & Katayose, H. A New Music Database Describing Deviation Information of Performance Expressions, International Conference of Music Information Retrieval (ISMIR), pp.489-494, 2008.

[10] New Grove Dictionary of Music and Musicians