Tempo Prediction Model for Accompaniment System

Shizuka Wada **Yasuo Horiuchi** **Shingo Kuroiwa**

Chiba University

sheeson@chiba-u.jp hory@faculty.chiba-u.jp kuroiwa@faculty.chiba-u.jp

ABSTRACT

In this paper, we propose a tempo prediction model for an accompaniment system. To realize a system that acts as a human performer, it is necessary to clarify how performance control is done by human accompanists. In our previous study, we proposed a method for predicting the next beat time by using the previous history of two parameters: the time difference between the soloist and accompanist and the change of the beat duration of the accompanist. However, the study analyzed only simple music that consisted of notes with the same musical length; therefore, the model cannot be applied to general music with different note lengths. In this study, we analyze general ensemble recordings by virtuosi to investigate the effectiveness of prediction with the two parameters proposed in the previous study and a newly added parameter: the time length difference between the soloist and accompanist. Then, we propose a method for predicting the next beat time of the accompanist that is applicable to general music. The result of an evaluation experiment shows that this model can predict the next beat time with errors of 25 - 45 ms.

1. INTRODUCTION

The automatic accompaniment system [1-5], which performs accompaniment synchronizing with human soloists, is an important piece of research as one of the real-time interactive systems between human and computer. The system has both solo and accompaniment scores and controls the timing of the system's performance to cooperate with a soloist's performance. The previous pieces of research on the accompaniment systems focus mainly on the tracking of a soloist's performance for performance mistakes or large tempo changes and give little consideration to the quality of the musical performance of the accompaniment in regard to whether the accompaniment performance is adequate in comparison with human accompanists. Therefore, a method for controlling human accompanists has not been clarified yet.

As a previous study on the analysis of ensemble performances, Rasch [6] analyzed and discussed the asynchronization of trio ensembles by using directional microphones but did not mention the performer's control method. In some pieces of research [7-9], a model for accompaniment performance performed by using a simple etude with constant note length was proposed. However, general ensemble music has various notes with different note lengths, so the model cannot be applied. Dannenberg [10] proposed a linear tempo model of tapping with music such as jazz or rock music, which has a relatively constant tempo, but did not consider the interaction between human musicians. In the researches of conducting system ([11] etc.), the virtual orchestra follows the human conductor but there is no sound interaction between the conductor and the orchestra.

In this study, to improve the accompaniment control model of our previous study [9], we propose a method of accompaniment control that is applicable to general music. The learning of parameters for the control model uses performances of virtuosi recorded on CDs.

2. PREVIOUS STUDY [9]

In this section, we explain in detail about the accompaniment control model proposed in our previous study [9].

In the study, it was assumed that human performers played music on the basis of an internal clock corresponding to the beat, and in an ensemble, he/she played while adjusting their clock to synchronize with the clock of the ensemble partner. Therefore, we proposed a model for predicting the next beat time from the history of performances of all performers in an ensemble. In the previous study, ensemble performances given by human performers (graduates of a piano course from a college of music) were recorded. The recorded piece was a piano etude consisting of only eighth notes (played by only one hand for each performer). Analyzing the performance and mutual relationship in time, the model for predicting a human performer was estimated.

We define the history of the beat time of the soloist with $T_{n-1}, T_{n-2}, \ldots$ and the history of the beat time of the accompanist with $t_{n-1}, t_{n-2}, \ldots$ The model is intended to estimate the beat time of the accompanist in the future (t_n). In the prediction, we used two parameters: the time difference between the soloist and accompanist and the change of the beat duration of the accompanist.

The time difference between the soloist and accompanist is the difference of beat time between the two performers in each beat, defined as:

$$d_n = T_n - t_n \tag{1}$$

, where T_n is a soloist's onset time of the beat n, and t_n is an accompaniment's onset time of the beat n.

The change of the beat duration of the accompanist is a variation of an accompanist's inter onset interval (IOI) in each beat, defined as:

$$c_n = l_n - l_{n-1} \tag{2}$$

, where l_n is an accompanist's IOI, defined as:

$$l_n = t_n - t_{n-1} \tag{3}$$

To predict the next beat time of the accompaniment t_n, we proposed a model for predicting c_n from the history of these two parameters with Equation 4.

$$c_n = \sum_{i=1}^{p} \alpha_i\, d_{n-i} + \sum_{j=1}^{p} \beta_j\, c_{n-j} \tag{4}$$

, where α_i and β_i are the partial regression coefficients applied a multi-regression analysis on the ensemble data of human performers, and p is the number of the history of the parameters. Then, variable selection is conducted by using the stepwise method.

In the previous study [9], we analyzed a piece where there are notes on all beats; however, general music has situations where a note is not present on the beat. Therefore, in that case, the model cannot be applied.

In this study, we analyze general music and propose an accompaniment control model that can be applied for any score. Furthermore, to be able to analyze large amounts of general music, we analyze recorded CDs played by virtuosi.

3. PREDICTION PARAMETERS

3.1 Measurement of Beat Time

The previous study [6], in which the author used directional microphones placed in front of each performer, allowed the sound of each instrument to be analyzed separately. However, it is difficult to collect a lot of data with this method. Therefore, in this study, we analyzed commercial CDs played by virtuosi by using the method proposed in [12, 13]. First, we fitted the acoustic signal of the CDs to the score data in the MIDI format by using the tool proposed in [12], and then, we measured the onset time. However, some fitting errors and subtle errors occur with this automatic fitting, so we modified the fitting result by hand. By observing spectrograms and listening to acoustic signals [14] by using the tool shown on p.67 in [13], we determined the onset time of the notes of each beat more precisely.

3.2 Proposed Prediction Parameters

With the previous model [9], it was not possible to predict the timing of the next beat of the accompanist when there is no note that sounded at the same time with the beat of the accompaniment. In this study, if there is a solo note on the beat in this situation, we assume that the beat time of the accompanist is the same as that of the soloist, thus allowing the calculation of each parameter for the prediction of the next beat. Moreover, in this study, we added a variable, the time length difference between the soloist and accompanist, as the parameter for predicting the next beat time, and it is defined as:

$$ld_n = (T_n - T_{n-1}) - (t_n - t_{n-1}) \tag{5}$$

ld_n is equal to the deviation of the time difference as shown in Equation 6, and [7, 15] show that the correlation between this parameter (ld_n) and the tempo modification (c_n) is high.

$$\begin{aligned} ld_n &= (T_n - T_{n-1}) - (t_n - t_{n-1}) \\ &= (T_n - t_n) - (T_{n-1} - t_{n-1}) \\ &= d_n - d_{n-1} \end{aligned} \tag{6}$$

In this study, we propose a model for predicting the next beat time of the accompanist from these three parameters (d_{n-1}, c_{n-1}, and ld_{n-1}) in the ensemble data between the soloist and accompanist.

4. EXPERIMENT

4.1 Material

We used the forth movement from “Sonata in A major for violin and piano,” composed by C. Franck for the analysis. We use the arrangement version of this piece for flute and piano because our accompaniment system is developed for solo flute with piano accompaniment. We analyzed measures 1 - 36, 51 - 64, 87 - 98, and 185 - 235 (the Auftakt notes in measures 1, 51, and 185 were included in the analysis data). The pairs of musicians are shown in Table 1. The average of each accompanist's IOI, beats per minute (BPM), and the standard deviation are shown in Table 2.

4.2 Simple Regression Analysis

To examine the correlation between the predicted value c_n and the three predictor parameters (d_{n-1}, c_{n-1}, and ld_{n-1}), scatter diagrams are shown in Figure 1, in which the horizontal axes are the three predictor parameters, and the vertical axes are the predicted values. Also, the coefficients of correlation between each predictor parameter and predicted value are shown in Table 3. From Figure 1 and Table 3, these three parameters have a correlation with the predicted value c_n. As the result of a statistical test, the null hypothesis that there is no correlation was rejected at the 1% significance level for all parameters; therefore, these three predictor parameters are shown to be effective for predicting the next beat time.

4.3 Multiple Regression Analysis

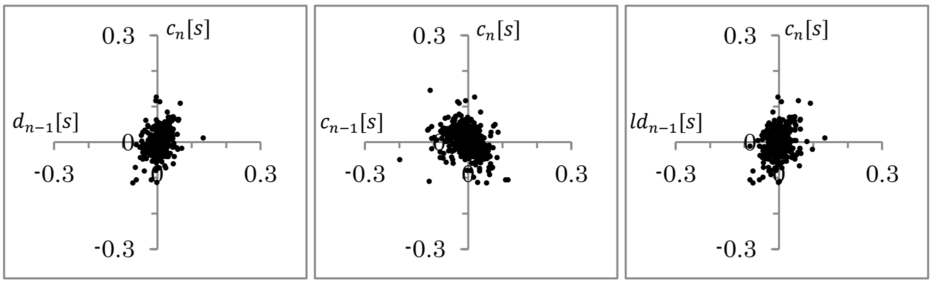
Although the model for predicting c_n can use these three parameters, the possible combinations of the three parameters are different depending on the presence/absence of a note on the beat. The possible combinations and the frequency in analyzed data are shown in Table 4. Within possible seven combinations, we can use three combinations of parameters: $[d_{n-1}, c_{n-1}, ld_{n-1}]$, $[c_{n-1}, ld_{n-1}]$, and only $[c_{n-1}]$, which have a sufficient amount of data to perform multiple regression analysis.

	Soloist (flute)	Accompanist (piano)
Duo A	Emmanuel Pahud	Eric Le Sage
Duo B	James Galway	Martha Argerich
Duo C	Jean-Pierre Rampal	Pierre Barbizet
Duo D	William Bennett	Clifford Benson

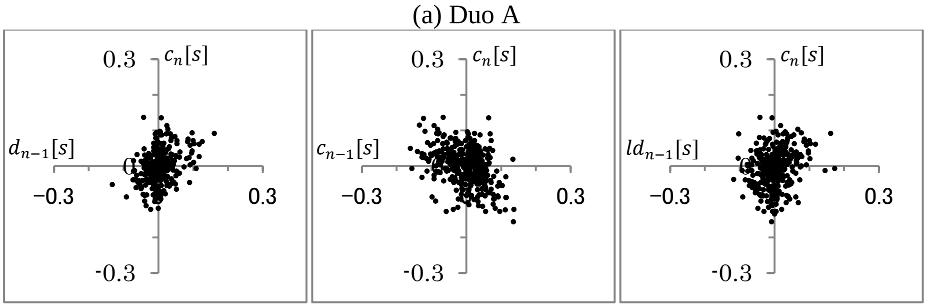
Table 1. Musician information of analyzed CDs

	Duo A	Duo B	Duo C	Duo D
Accompanist's IOI [s]	0.37	0.38	0.32	0.41
Tempo [bpm]	164	160	189	147
Standard deviation of accompanist's IOI [s]	0.040	0.063	0.036	0.054

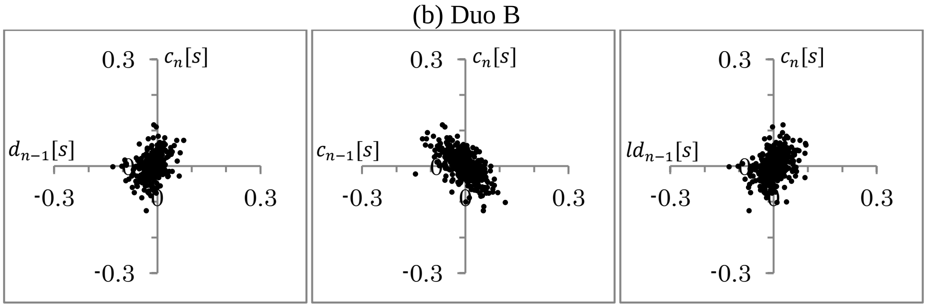
Table 2. Average of accompanist's IOI, tempo, and standard deviation



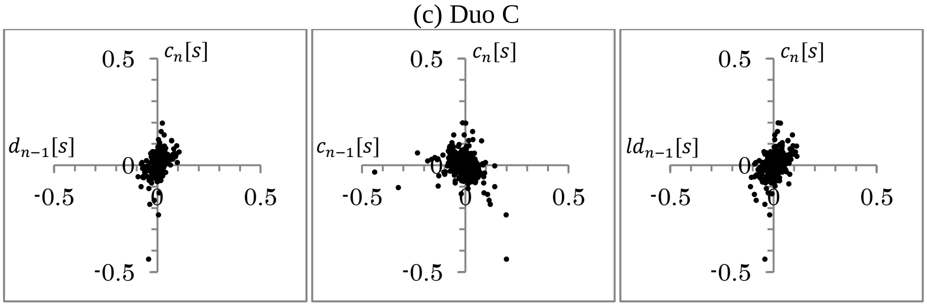
(a) Duo A



(b) Duo B



(c) Duo C



(d) Duo D

Figure 1. Scatter diagrams of each accompanist

	Duo A	Duo B	Duo C	Duo D
d_{n-1}	0.28	0.34	0.32	0.41
c_{n-1}	-0.35	-0.39	-0.60	-0.31
ld_{n-1}	0.30	0.24	0.38	0.41

Table 3. Coefficient of correlation in each parameter

	$d_{n-1}, c_{n-1}, ld_{n-1}$	d_{n-1}, c_{n-1}	c_{n-1}, ld_{n-1}	d_{n-1}, ld_{n-1}	d_{n-1}	c_{n-1}	ld_{n-1}
Number of data	294	0	55	2	1	55	0

Table 4. Combinations and frequency of parameters

The models for predicting the three combinations are shown in Equations 7 - 9, and we change the equation depending on the possible combination of parameters.

$$c_n = \alpha_3 d_{n-1} + \beta_3 c_{n-1} + \gamma_3 ld_{n-1} + e_3 \tag{7}$$

$$c_n = \beta_2 c_{n-1} + \gamma_2 ld_{n-1} + e_2 \tag{8}$$

$$c_n = \beta_1 c_{n-1} + e_1 \tag{9}$$

We perform the multiple regression analysis for these three combinations and calculate the partial regression coefficient and the constant term for each performer pair. We performed two types of experiments; one was a closed test where all data were used for learning and testing, and the other one was an open test where the first half of the data (measures 1 - 36, 51 - 64, 87 - 98) was used for learning and the last half of the data (measures 185 - 235) were used for testing because this piece has a similar phrase structure in the first half and the last half. The partial regression coefficient and constant term calculated by the multiple regression analysis are shown in Table 5. In Table 6, we apply the coefficients in Table 5 to Equations 7 - 9 and evaluate the prediction error for each combination and each pair.

4.4 Proposed Prediction Method for General Music

To make a practical model, we would like to propose a prediction method that uses all data by combining Equations 7 - 9. However, in some cases, it is not possible to predict because there is no solo or accompaniment note that is required to apply a model that combines these equations. Even in this case, the accompaniment system must continue to play, so it is necessary to predict the next beat time. In any case, the system has the histories of t_n because our accompaniment system determines the beat time sequentially while continuing to play. If there are t_n of three beats past, on the basis of Equation 9, we can predict c_n with Equation 10.

$$c_n = \beta_1((t_{n-1} - t_{n-2}) - (t_{n-2} - t_{n-3})) + e_1 \tag{10}$$

In Equation 10, when the notes on t_n, t_{n-1}or t_{n-2} are not present, apply the onset time of the solo if the onset time of solo T_n, T_{n-1}or T_{n-2} is present. When neither is present, calculate by using t_n obtained sequentially from the past.

Using this method, performance prediction becomes possible in every beat. We conducted the closed and open experiments by using the partial regression coefficients in Table 5. The results of calculating prediction error for the beat notes of all of the experimental data are shown in Table 7.

4.5 Discussion

As a result of the evaluation experiment, the prediction error was about 25 - 45 ms for any pair from Table 7. According to [6], the perception of the gap of the onset timing of two sounds on playing instruments is more difficult than on psychological stimulation experiments because of the effect of pitch and timbre and a longer rise time of musical notes than stimulation, and therefore, it is considered that errors of about 25 - 45 ms mean rather high quality for a prediction model for an accompaniment system. For reference, the average of time differences between the soloist and the accompanist are shown in Table 7. It looks that the prediction errors are comparable with the time differences between them.

In the open experiment, the situations where the prediction error was more than 100 ms were about 2% in all data. Figure 2(a) is the score excerpted from measures 184 - 186. Our model can predict the note from the third beat of measure 185 by using beat time (t_n) of the preceding three notes, and the prediction error of this note is more than 100 ms for Duo A, Duo B, and Duo D because the Auftakt note of measure 185 was played very long and the prediction model did not work well. The error of Duo C was also relatively large. Figure 2(b) is the score excerpted from measures 219 - 221. In this situation, the prediction error was more than 100 ms for all musicians. The large error occurred at the second, third, and fourth beats of Duo A, the second and third beats of Duo B, the third beat of Duo C, and the first, second, and third beats of Duo D. These are the situations where the tempo slowed down by ritardando, and after that, the tempo accelerated by animato, so these large tempo changes made it difficult to apply the prediction model. The proposed model cannot handle these situations because it models the situation where the tempo is nearly constant, but it would be possible to handle this situation by adding a parameter of the tempo change expected from the score.

	Duo A	Duo B	Duo C	Duo D
α_3	0.18	0.39	0.23	0.51
β_3	-0.43	-0.37	-0.54	-0.49
γ_3	0.21	0.018	0.096	0.16
e_3	0.0020	0.0	0.0030	0.0040
β_2	-0.030	-0.69	-0.66	-0.15
γ_2	0.026	-0.72	0.010	0.14
e_2	-0.012	-0.025	0.0010	-0.020
β_1	-0.22	-0.19	-0.42	0.080
e_1	0.011	0.013	-0.0030	0.0020

(a) Closed experiment

	Duo A	Duo B	Duo C	Duo D
α_3	0.23	0.32	0.28	0.55
β_3	-0.39	-0.33	-0.61	-0.39
γ_3	0.22	0.097	0.14	0.14
e_3	0.0020	0.0010	0.0030	0.0060
β_2	-0.18	-0.65	-0.66	-0.40
γ_2	0.29	-0.36	-0.069	0.12
e_2	-0.010	-0.030	0.0040	-0.029
β_1	-0.30	-0.24	-0.46	-0.079
e_1	0.0070	0.011	0.0010	0.000017

(b) Open experiment

Table 5. Multiple regression analysis result (partial regression coefficient and constant term)

	$d_{n-1}, c_{n-1}, ld_{n-1}$				c_{n-1}, ld_{n-1}				c_{n-1}			
	Duo A	Duo B	Duo C	Duo D	Duo A	Duo B	Duo C	Duo D	Duo A	Duo B	Duo C	Duo D
Closed	0.030	0.040	0.027	0.043	0.040	0.052	0.028	0.056	0.036	0.049	0.034	0.034
Open	0.029	0.041	0.027	0.044	0.051	0.052	0.034	0.070	0.047	0.054	0.042	0.041

Table 6. Prediction error [s] of each combination

		Duo A	Duo B	Duo C	Duo D
Prediction error	Closed	0.032	0.041	0.028	0.043
	Open	0.033	0.043	0.029	0.044
Average difference		0.026	0.038	0.031	0.030

Table 7. Prediction errors [s] of evaluation experiment and average of time differences [s] between the soloist and the accompanist

(a) Measures 184 - 186

(b) Measures 219 - 221

Figure 2. Excerpt of score

In addition, errors of more than 100 ms occurred in other situations of each of the two notes of Duo B and Duo D. In these situations, it is considered that the accompanists performed the notes with tempo rubato or large agogik because of the different interpretation of each performer.

These musical deviations cannot be treated in the proposed method, but they are essential in music performance and have musical characteristics [6]. Human performers have shared musical knowledge, and in general, they rehearse an ensemble before a concert or recording and share their musical expression. In the future, the accompaniment system has to share this musical knowledge with the soloist through rehearsal.

5. CONCLUSION

In this paper, we proposed a tempo prediction model for an accompaniment system. Improving the model of our previous study [9], the proposed method can be applied to general music including various notes. The results of evaluation experiment show that this model can predict the next beat time of an accompanist with errors of 25 - 45 ms, and this model could be implemented in the accompaniment system easily. Because the musical deviations (such as tempo rubato or agogik) are included in the analysis data, it is considered that the deviations caused some prediction errors. General performances have these musical deviations; therefore, in the future, we have to implement musical knowledge in the accompaniment system or introduce a mechanism for learning the deviation through rehearsal with a soloist.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26240025.

6. REFERENCES

[1] R. B. Dannenberg, “An On-Line Algorithm for Real-Time Accompaniment,” Proceedings of International Computer Music Conference, pp. 193-198, 1984.

[2] B. Vercoe, “The synthetic performer in the context of live performance,” Proceedings of International Computer Music Conference, pp. 199-200, 1984.

[3] Y. Horiuchi, H. Tanaka, “A Computer Accompaniment System with Independence,” Proceedings of International Computer Music Conference, pp. 418-420, 1993.

[4] C. Raphael, “Music plus one: A system for flexible and expressive musical accompaniment,” Proceedings of International Computer Music Conference, 2001.

[5] T. Nakamura, E. Nakamura, S. Sagayama, “Acoustic Score Following to Musical Performance with Errors and Arbitrary Repeats and Skips for Automatic Accompaniment,” Proceedings of Sound and Music Computing, pp. 299-304, 2013.

[6] R. A. Rasch, “Synchronization in Performed Ensemble Music,” Acustica, vol. 43, pp.121-131, 1979.

[7] Y. Horiuchi, K. Sakamoto, A. Ichikawa, “Estimating the Model of Human Performance in Ensemble,” Information Processing Society of Japan, vol.43, no.2, pp.260-267, 2002 (in Japanese).

[8] Y. Kobayashi, Y. Miyake, "New ensemble system based on mutual entrainment," Proc. of 12th IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN2003), Silicon Valley, U.S.A., No.235, pp.1-6, 2003.

[9] Y. Horiuchi, K. Sakamoto, A. Ichikawa, “Estimating the Model of Human Ensemble,” Information Processing Society of Japan, vol.45, no.3, pp.690-697, 2004 (in Japanese).

[10] R. B. Dannenberg, S. Mohan, “Characterizing Tempo Change in Musical Performances,” Proceedings of International Computer Music Conference, pp. 650-656, 2011.

[11] T. Baba, M. Hashida, H. Katayose, “VirtualPhilharmony: A Conducting System with Heuristics of Conducting an Orchestra,” Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), pp. 263-270, 2010.

[12] N. Takamune, H. Kameoka, M. Tsuchiya, S. Sagayama, “Auxiliary function approach for optimal MIDI-fitting to audio signals,” The Acoustical Society of Japan, 2012 Autumn Meeting, pp. 937-940, 2012 (in Japanese).

[13] H. Kameoka, “Statistical Approach to Multipitch Analysis,” Ph.D. Thesis, The University of Tokyo, 2007.

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

[15] M. Nojima, H. Shimo, Y. Miyake, "Subjective timing control in synchronized motion of humans: A basic study for human-robot interaction," In H. Asama, H. Kurokawa, J. Ota & K. Sekiyama (Eds.) Distributed Autonomous Robotic Systems 8, Springer-Verlag, Tokyo, pp.135-147, 2009.