# New Ensemble System Based on Mutual Entrainment

Y. Kobayashi Y. Miyake

Miyake Lab
Tokyo Institute of Technology
4259 Nagatsuta, Midori, Yokohama
E-mail: kobayasi@myk.dis.titech.ac.jp

**Abstract**

*Ensemble is a kind of cooperative musical performance between humans, and this kind of flexible cooperation can not be observed between human and music machine. Thus, to realize human-like ensemble, the analysis of actual performance between humans is necessary. In this research, we constructed a new ensemble system based on mutual entrainment, which re-creates the human ensemble process. By comparing among the ensembles of humans, previous system and our system, we clarify the importance of interaction between performers. Finally, we advocate Co-Creation and its process to expand the previous communication between human and machine.*

## 1. Introduction

Ensemble is a kind of cooperative musical performance between humans, and this kind of flexible cooperation can not be observed between human and music machine. For instance, KARAOKE plays its music independently from human's state [1]. Thus, human should adapt to the fixed music one-sidedly. On the other hand, mutual adaptation is widely observed in the ensemble between humans, and this kind of ensemble is usually much better than the other type.

Comparing these two types of music performance, it is thought that there is a significant difference in the interaction mechanism between players. In the ensemble between humans, mutual interaction of rhythms is widely observed, however, one-sided action is mainly used in man-machine communication. Therefore, to realize human-like ensemble system between human and musical machine, we have to focus our attention on the mechanism of mutual interaction between human and machine.

From these backgrounds, in this research, we try to realize human-like ensemble system by using mutual interaction between rhythms.

Many previous studies for artificial ensemble tried to overcome the problem of such one-sided action, and most of them aimed at the rapid response of machine which precisely follows human performance [2]-[5]. However, it

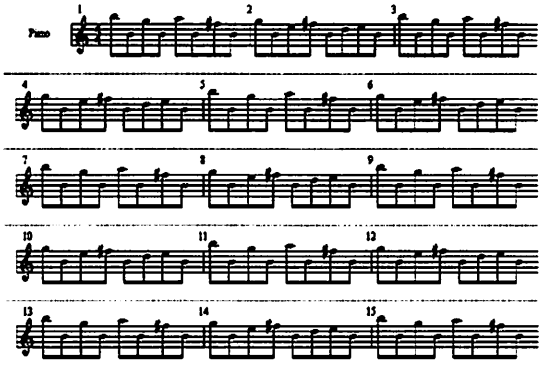

Figure 1 Score of experiment

is thought that the machine of such follow-up control is insufficient to realize ensemble between human and machine. This is because such kind of interaction can be regarded as an opposite one-sided action, i.e., interaction from human to machine. Furthermore, the present approaches in jam- session-system or jazz-session-system are also supposed to be insufficient to make mutual interaction [6]-[8]. These systems use two one-directional actions, and action from human to machine and from machine to human is realized alternately.

We think that simultaneous and mutual interaction between human and machine is important in ensemble and assumed that it is realized by using interaction of rhythms. However, at the present, such kind of artificial ensemble system has not been developed. In this research, to realize human-like ensemble, we learn from ensemble between human and human. In order to achieve such a purpose, in chapter 2, we analyze the interaction in performance between human and human. In chapter 3, such an interaction is modeled and ensemble system is constructed. In chapter 4, 5, experiment is done and system is evaluated.

## 2. Human performance

To realize human-like ensemble, the analysis of actual performance between human and human is neseccery. Then according to the results of analysis, model is made and system is constructed. To analyze such an ensemble, we attach importance to how they interact each other. An

Table 1 Results of hitting cycle and hitting lag

| | 1 | 2 | 3 | 4 | 5 | All |
|---|---|---|---|---|---|---|
| Average of hitting cycle (msec) | 214 | 194 | 191 | 195 | 191 | 197 |
| Average of hitting lag (msec) | 22.1 | 8.2 | 1.7 | 8.2 | 0.0 | 8.1 |
| Standard Deviation of hitting lag (msec) | 20.0 | 25.3 | 22.4 | 25.2 | 22.9 | 23.2 |

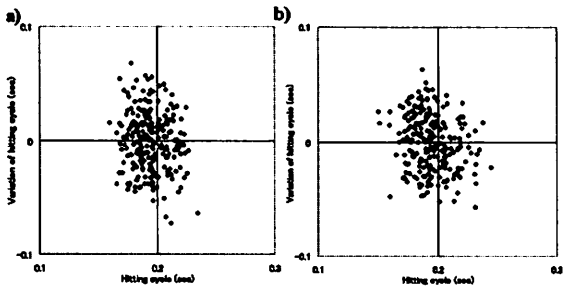

Figure 2 Correlation between hitting cycle and hitting cycle variation

Table 2 Correlation between hitting cycle and variation of hitting cycle

| | 1 | 2 | 3 | 4 | 5 | All |
|---|---|---|---|---|---|---|
| High→Low | -0.52 | -0.11 | 0.02 | -0.11 | -0.13 | -0.17 |
| Low→High | -0.37 | -0.21 | -0.11 | -0.22 | -0.25 | -0.13 |

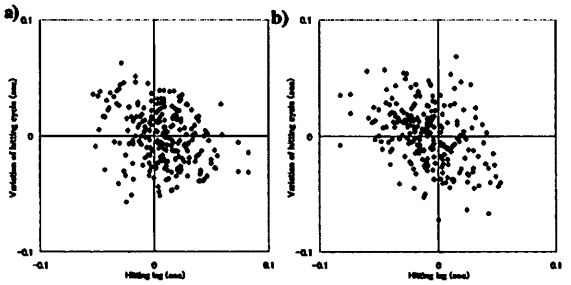

Figure 3 Correlation between hitting lag and variation of hitting cycle

Table 3 Correlation between hitting cycle and variation of hitting cycle

| | 1 | 2 | 3 | 4 | 5 | All | t-Test |
|---|---|---|---|---|---|---|---|
| High | -0.29 | -0.29 | 0.30 | -0.29 | -0.34 | -0.30 | * |
| Low | -0.42 | -0.38 | -0.33 | -0.38 | -0.39 | -0.38 | * |

experiment is done by ensemble between two persons. Performers use the same special score as **Fig.1**. The score is constructed only by an eighth note, so the intervals of notes are constant basically. The one performer performes faithfully to the score and the other performer performes 2 octove lower than the score on the same MIDI keybord. Performers are requested to cooperate each other but basically they are free. Experiment is done 5 times by changing performers, and time of key hitting is measured. Hitting period of one performer and hitting lag of two

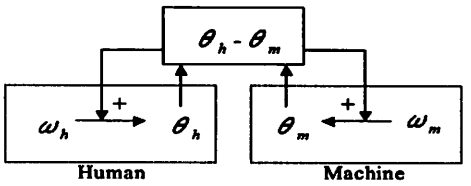


Figure 4 Architecture of interactive performance system

performers are calculated, and correlation between them are analyzed to clarify interactions between them.

**Table 1** shows the average of hitting cycle, average of hitting lag and standard deviation of hitting lag at each trial. It shows the hitting lag is much less than hitting cycle (less than 20% of the hitting cycle) that there are synchronization between performers. To accomplish such a synchronization, two kindes of method is expected. The one method is to use the hitting period of partners perfomance for index of synchronization, and the other one is to use the hitting lag for its index. To exame which one is the major index, we compare the influence value to the hitinng period variation by using correlation between them. The example(the first trial) of relation between hitting cycle of partner and hitting cycle variation of immidiate cycle is shown as **Fig.2**. In this graph, **Fig.2a** shows the influence from the higher part to the lower part, and **Fig.2b** shows the the influence from the lower part to the higher part. **Table 2** shows the correlation of them at each trial. In this table, the upper section indicates the influence from the higher part to the lower part, and the a lower section indicates the influence from the lower part to the higher part. From this table, the corrlation of them are not high and their value is nither not constant. Against this, the example(the first trial) of relation between hitting lag and hitting cycle variation is show in **Fig.3**. In this graph **Fig.3a** shows the influence to the higher part and **Fig.3b** shows the influence to the lower part. **Table 3** shows the correlation of them at each trial. In this table, the upper section indicates the influence to the higher part, and the a lower section indicates the influence to the lower part. Comparison from **Table 2**, its correlation value is constantlly higher and significantlly different(* means $P<0.01$ from t-Test) that, the existance of a feedback mechanism from the hitting lag to the hitting cycle variation is observed.

From above results, it is considered that human performers use hitting lag rather than hitting cycle for the indicator of synchronization.

### 3. Model for Ensemble

We modeled the mechanism of above-mentioned by using mutual entrainment respect from forced entrainment. The mutual entrainment is the phenomenon that two oscilator interact each other and their rythms became the same [9]. Their rythm is regurated by natural frequency and phase. The mechanism of preceding chapter is able to substitute for the entrainment. The hitting cycle is able to substitute

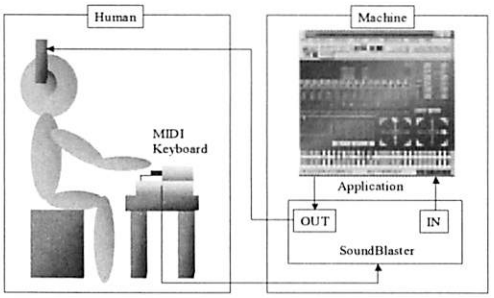


**Figure 5** MIDIdata flow

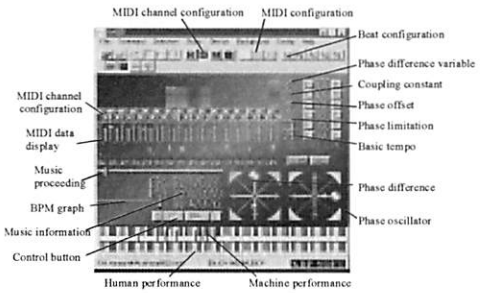


**Figure 6** Application

for a reciprocal of natural frequency and hitting lag is als able to substitute for phase difference.

As shown in **Fig.4**, human performs at the left side and machine performs at the right side. The tempo (BPM) of music performance in each player is defined based on the frequency $\omega$, and $\omega$ usually affect to the performance directly. Then, the temporal development of music performance is described by using the phase $\theta$. In ensemble process, the phase relationship ( $\theta_h$- $\theta_m$) is feed backed to each player. Thus $\theta$ changes not only by frequency $\omega$ but also by their phase relationship of performance ( $\theta_h$- $\theta_m$). In this condition, the harmonics between them can be defined by the temporal development of the relationship ( $\theta_h$- $\theta_m$).

As a detailed mechanism of entrainment between human and the machine, a coupled nonlinear oscillators model was used. The tempo of the machine's performance is controlled based on the frequency of the model oscillator. And they are expressed as follows,

$$\dot{\theta}_m = \omega_m + K_m \cdot \sin(\theta_h - \theta_m) \quad \text{.....(Machine)}$$

$$\dot{\theta}_h = \omega_h + K_h \cdot \sin(\theta_m - \theta_h) \quad \text{.....(Human)} \qquad (1)$$

where $\theta$ is the phase of each oscillator, $\omega$ is the intrinsic frequency, and K is the coupling constant. In this equation, the tempo of performance corresponds to $\omega$. The temporal development of music performance corresponds to $\theta$. The phase difference indicates the relationship between these two players. Coupling constant $K$ indicates the influence from the feedback. Forced entrainment is realized when $K$=0, and mutual entrainment is realized when $K$>0, respectively.

## 4. Experimental system and experimental method

Our system is constructed as cooperation between human player and MIDI (Musical Instruments Digital Interface) application designed on Windows as shown in **Fig.5** [10]. In this experimental system, human plays MIDI keyboard and its signal is transmitted to MIDI application. On the other hand, the sound signal from MIDI application is transmitted to human by using speaker.

Application software to operate the above experimental system is shown in **Fig.6**. This application performs the SMF format MIDI file designed on the Windows [11][12]. The event timing of the machine performance is determined by the phase defined in the coupled oscillator model. This application also measures the timing of human performance in the MIDI data in real time. Used music is the same as used in the human performance. The experiment was done in a silent room by using PCs (VAIO Pentium3 1GHz) and electric pianos (Roland RD-600). Human performs by watching music score and PC was located to the hide space.

### 4.1 Experiment for parameter setting.

The behavior of the model and this system is strongly related to the coupling constant K. So we have to search the best parameter K which works system as human like.

We changed the coupling constant K as different ratio of ω(0% of ω 5% of ω 10% of ω ....50% of ω). In each condition, human performer played with system two times (upper part and lower part). And recorded the temporal development of phase difference between human and machine. We also recorded the performance between humans and temporal development of phase difference between them. Five students who play piano more than 10 years were used as subjects. We calculated the average and the S.D. of phase difference in each performance and used as an indicator of performance, which indicates how the performance was human like.

Finally, we decided the unique coupling constant K to use in the evaluation experiment.

### 4.2 Experiment for evaluation

We compared three kinds of the ensemble styles. The first one is the ensemble between humans (Situation A). The second one is the ensemble between human and our system (Situation B). In the our system, the coupling constant K which decided in the above experiment was used. The last one is the ensemble between human and the recorder, which recorded the partners, solo performance (Situation C).

So for example, if human α and human β is placed in these situation, in the situation A, human a performs with human β directly. But in the situation B, each performer plays with each system independently. And in the situation C, human α performs with recorder which plays human β's solo performance and human β performs with recorder which plays human α's solo performance.

Four students who play piano more than 10 years were used as subjects. All of the 6 pair played in these three

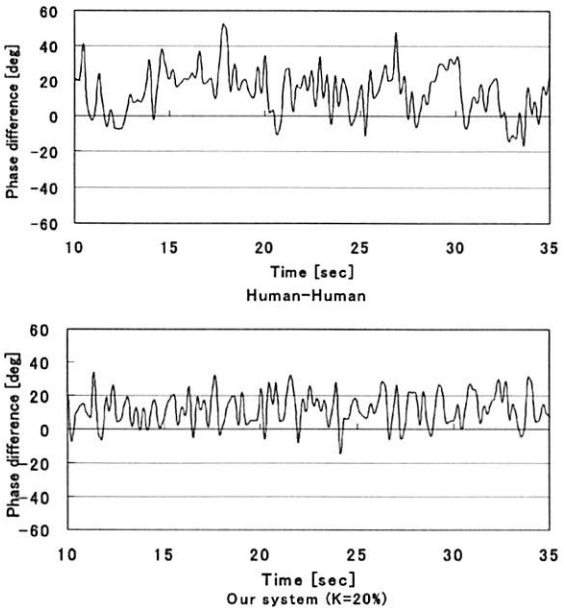


**Figure 7** Temporal development of phase difference

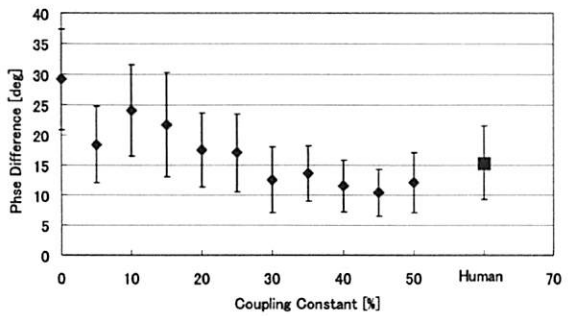


**Figure 8** Average of phase difference

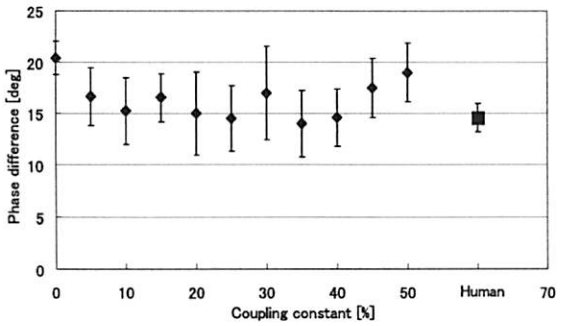


**Figure 9** S.D of phase difference

kinds of situation randomly. One session was constructed by three situations and in each pair two sessions, which exchanged the playing parts, were done. So all of the experiment is constructed by 12 sessions. In each session, the situations A,B and C was ordered randomly and human

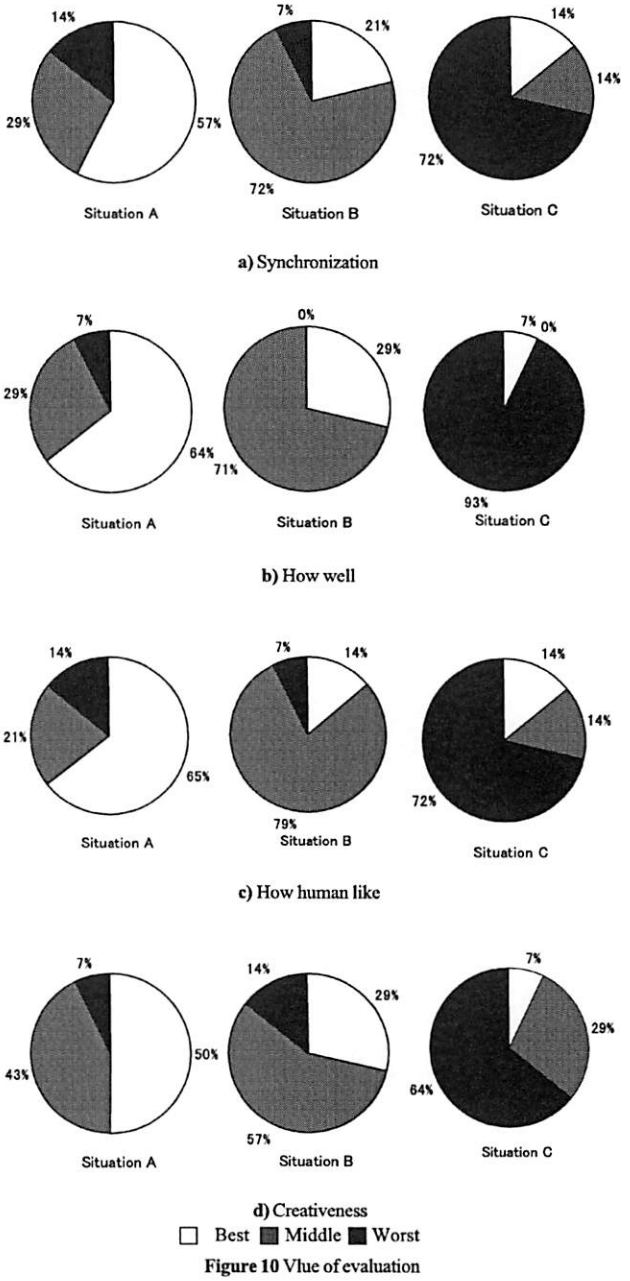


**Figure 10** Vlue of evaluation

performer can't know the order of A,B and C. Through all of performance, each human performer separated in distant places and can't see the other performer. So if the situation was A, human performer ensembles each other but, if the situation was B or C, each human performer only played with their machine and no interaction between humans. In such a blind test, we asked performs to evaluate and compare their situations in each session. The evaluation criterion is as following.

a): Which situation was human like?

b): How well did you ensemble?

c): How well did you synchronized with your partner.?

d): Which situation was more creative?

From these criterions, each performer compared and evaluated each situation as best one, worst one, and middle one.

Besides, in these experiment, only timing of key hutting was send to the other performer. So, the other musical information such as the velocities, the durations, and the miss touches are cut off by the filter from score information.

## 5. Result

### 5.1 Result of parameter setting.

Typical pattern of temporal development of phase difference is shown in **Fig.7**. From these graphs, we are able to calcurate the average and S.D. of phase difference in one session.

Result of the average of phase difference is shown in **Fig.8**. From this graph it is shown that coupling constant K is strongly affect to the synchronization of performance and there is negative correlation between the coupling constant K and the average of phase difference. In all condition, the value of average of phase difference is positive, that it means the performance of human is forward to the performance of machine. From this graph, the nearest value of average of phase difference to humans is achieved between 20%-30% of value of coupling constant K.

Result of the S.D. of phase difference is shown in **Fig.9**. From this graph it is shown that coupling constant K is strongly affect to the synchronization of performance and we can presume the best coupling constant K, which achieves minimum S.D. of phase difference, which represents the best synchronization between performers. From this graph, the nearest value of S.D. of phase difference to humans is achieved between 20%-40% of value of coupling constant K.

From these results, the suitable coupling constant K for this system is expected between 20%-30%. So we have to choose static value from this range. From this range, we choose 20% with following reason. If coupling constant K become bigger, this system would be unstable because system is more influenced by human performer. So, to work this system stable, smaller K is better for ensemble.

### 5.2 Result of evaluation

Result of the evaluation in each criterion is shown in **Fig.10**.In the entire criterion, situation A marked well than the others and situation C marked worst. And in the entire criterion, there is significant difference between situation B and situation C ($P<0.01$ ANOVA). Also there is significant difference between situation A and situation C. Between situation A and B, only in the criterion a), there is tendency of difference ($P<0.1$), and in the other criterion, there is little difference ($0.1<P<0.5$). So from these results, it is able to say that situation A is the best of three situations, and situation B is improved from situation C and approached to situation A.

## 6. Discussion

From the results, situation A marked highest value and situation C marked lowest value. The situation A is the performance between humans, so we can confirm that the performance between humans is the best from the results. In our experiments, the only the musical information of playing timing is used and the other information such as velocities and durations are cut off by the filter. So the difference between these situations is only the difference of tempo and its temporal development. It is important that such a simple information causes to the difference of subjective feelings.

Furthermore it is important that the situation C marked lowest value in all the criterions. In the situation C, the performance is the record of real humans and it is equal to ordinary system as KARAOKE or CD medias. However, the evaluations were worst in all the criterions. Especially in the criterion a), its value was worst and evaluated that it was not human like. Situation C is the playback of the performance of the real humans, so it would be evaluated to most human like. But from these results, it is clarified that human like ensemble is not achieved only by the static characteristic. The reason why there is a difference between the situation A and the situation C, one of the clue is the evaluation result of the situation B.

In the situation B, there is an interaction between the human and the machine by using our model. The baseline of our system is static tempo, so there is no data and characteristics of human performance such as situation C. However, the situation B marked higher evaluation of criterion a) and the others by interact with the human performer in real time. Especially, there is a significant difference between situation B and situation C, so the interaction between performers will be one of the important elements. In present ordinary machine, how to playback the human performance faithfully would be the major interest. But from the results, it was shown that how to interact with the human performers is important element to achieve human like ensemble system.

Furthermore, as in the criterion d), we asked performers about the creativeness. At first, we expected that there is no significant difference of creativeness between the situations because we used very simple music as shown in Fig.1. But against to our expectations, there were differences. We expect the reason of these differences is caused by one of the improvisation. The improvised performance is accomplished only by through the real time performance and its interaction. These variations are not able to input in the machine beforehand, because these characteristics are basically depend on human performers. We define this kind of creativeness, which generated from real time interaction as Co-Creation, and we will analyze such the Co-Creation process and apply to present machines to realize better communication.

From the results, it is clarified that our system is able to approach to humans, but in all the criterions humans is better than ours. The reason why there are differences, one of the possibilities is the simplicity of the model we used. Human performers to play more complicated and we have

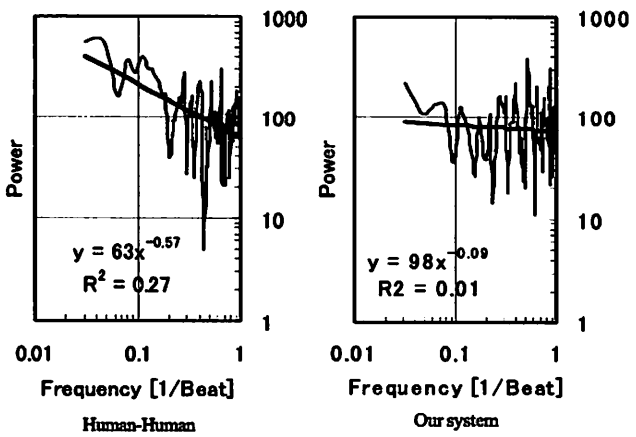


Figure 11 Power spectrum of phse difference

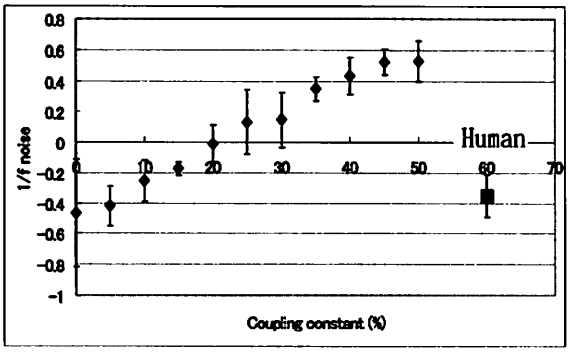


Figure 12 Tendency of 1/f noise

to improve our model. One of the clue is the power spectrum of phase difference as shown in **Fig.11**. There is a difference of spectrum pattern between humans ensemble and our system. As show in **Fig.12**, as coupling constant K increases as tendency of 1/f noise increases. This mean, lower frequency of temporal development of phase difference is decreased by the action of model. However the result of humans is more similar to 1/f noise, and lower frequency is remained. This dynamical difference is also able to confirm from the temporal development of phase difference as shown in **Fig.7**. This difference is exactly caused by the simplicity of our model, and the model, which uses longer span of the interaction, will be need.

## 7. Conclusion

We analyzed performance of humans and proposed

## Referemces

[1] K.Yaguchi,"The introduction of music composing by Mac,"SaitoCo.,1993.

[2] Y.Nagashima,S.Hahimoto,Y.Hiraga,K.Handa,"The Computer and Music,"Kyoritsu Shuppan Co.,1998.

[3] R.B.Dannenberg, "An On-Line Algorithm for Real-Time Accompaniment,"Proc. of ICMC,pp.275-283, 1984.

[4] J.J.Bloch and R.B.Dannenberg, "Real-Time Computer Accompaniment of Keyboard Performances,"Proc. of ICMC,pp.279-289,1985.

[5] B.Veroce, "The synthetic performer in the context of live performance,"Proc. of ICMC, pp.199-200, 1984.

[6] R.Rowe, "Interactive Music Systems Machine Listening and Composing," The MIT Press, 1993.

[7] S.Waki,H.Kato,N.Saiwaki,S.Iguchi,"-JASPER-"Proc. ofIEICE,Vol.35,No.7,pp.106-107,1995.

[8] M.Nishijima and K.Watanabe, "Interactive music composer based on neural networks," Proc. of ICMC, pp.53-56,1992.

[9] Y.Kuramoto,K.Kawasaki,M.Yamada,S.Kai, M.Sasamoto,"Pattern Formation,"Asakura Shobo Co.,1991.

[10] Y,Ohkura,"An elementary knowledge of Sound and Music,"Kokushokankokai,1999.

[11] Richard J. Simon, Tony Davis, Joyn Eaton, R. Murray Goertz, "Windows API Bible3,"Shoei Co.,1998.

[12] J.Arai,"SMF Refference book,"Ritto Music Co.,1996.

## Acknowledgement

There was a great contribution of Smith Andrew Mcmanama in the experiments. As a token of his gratitude, we admire him here.