Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Effect of Visual Cues in Synchronization of Rhythmic Patterns

Sisi Sun[1], Trishul Mallikarjuna[2], Gil Weinberg[3]

Center for Music Technology, Georgia Institute of Technology, Atlanta, GA, U.S.A.
[1] eversisi@gatech.edu, [2] trishul.mallikarjuna@gatech.edu, [3] gilw@gatech.edu

ABSTRACT

We conducted a rhythmic pattern learning and synchronization experiment. During the experiment, each of 20 experiment subjects was learning 7 patterns in different level of difficulty from a drummer robot. They played all the patterns twice in 2 different visual conditions: being able to see, and not being able to see the robot's movement. 10 of the subjects could see the robot the first time they played the 7 patterns, and they then played the patterns the second time without seeing the robot. The other 10 played in the opposite order of visual conditions. We applied Dynamic Time Warping algorithm on the onset time values to find the best matches between the subjects' and robot's hits. Then we used 4-way Analysis of Variance with the factors: existence of visual cues, order of visual conditions, subjects, and onset times, to analyze their influence on the time difference between matching onsets. The average of onset time differences was treated as a measure of synchronization. The data showed that, in case of more difficult patterns, the average onset time difference had higher variance when there were no visual cues compared to when there were visual cues, while in case of easier patterns, the variance was not significant. Thus we infer that visual cues can influence synchronization in a task that requires learning of more difficult rhythmic patterns. We also inferred that subjects showed a tendency to learn new patterns faster with visual cues. We further observed that people tend to play in a lag with visual cues during the learning period, and play better after learning. However, more experimentation is needed to establish statistical significance of the last two effects.

INTRODUCTION

Synchronization is a fundamental task in ensemble music. Specifically, rhythmic synchronization has long been considered one of the most important elements in music. It is obvious that auditory cues and events, such as a mistake in one's playing, or a change in tempo, affect the level of synchronization between two or more musicians. Here we explore the role of visual cues in aiding a player's performance in terms of the level of synchronization.

Some related research has been done in such musical contexts. Luck and Sloboda (2008) studied the synchronization between a participant's tapping and the beats depicted by a point-light representation of a conductor. They examined the effect of various conducting movement rates and showed that the highest correlation between taps and the conductor occurred when there was an absolute acceleration of the trajectory of the conductor's hand. Hoffman and Weinberg (2010) used a robotic musician to study the effects of visual cues on synchronization. They demonstrated that when human musicians are able to see the robot, synchronization improves only when the robot musician is playing in an unpredictable tempo. Other works of research studying the effect of visual and auditory cues on synchronization that inspired us are by Repp (2006) and by Repp and Penel (2004).

Brian Blosser (2010) studied synchronization in an ensemble performance setting using different visual modalities. In his series of experiments, six drummers were divided into two groups according to their skill level: one consisting of experienced drummers, and the other consisting of novice drummers. In each group, each person would play as a leader once while the other drummers attempted to synchronize with this leader under two conditions: looking-out and looking-in. Under the former condition the drummers could not see each other, and under the latter the drummers could see each other. The experimental results showed that for experienced drummers visual cues helped synchronization, but for novices visual cues decreased synchronization, which contradicted Blosser's hypothesis.

Some improvements can be made to Blosser's experiments to address concerns for impreciseness and distraction. Firstly, the sample size of 6 people was too small. Any extreme outlier would lead to a large data offset. Secondly, the difficulty levels of the rhythmic motifs played were not fully controlled, because they were generated intuitively by the leader. Thirdly, three drummers playing simultaneously in one group might lead to interference; mistake by one follower might influence the other follower. Finally, the order of visual conditions was not controlled; the order was predominantly looking-out after looking-in, which might influence the accuracy as people might get more familiar with the piece after playing it once.

EXPERIMENT

A. Motivation and Hypothesis

The effectiveness of different visual modalities is an important issue in music ensembles and may lead to a better understanding of how musicians naturally synchronize. To measure the effects of visual modalities with fewer distractive factors a new experiment is necessary. Here we design a new experiment to more accurately measure these effects, which focuses on the synchronization of rhythmic motifs between a single musician and Haile, the robotic drummer.

We make two hypotheses: subjects should synchronize better with visual cues, and subjects may learn the patterns faster with visual cues. The latter is to say, the temporal duration to reach a certain level of synchronization will be shorter with the aid of visual cues.

B. Stimuli

The stimuli consist of 7 drum patterns described by Povel and Essen (1985). Each pattern is in 16 beats (240ms per beat in this experiment) and repeats itself for 1 minute. The duration and onset sequences of the patterns are shown in Table 1. All patterns contain the same total number of hits and the same number of hits of each duration, but at different temporal positions in a bar. We assume that all patterns are new to all the subjects. The difficulty level increases with each pattern as described by Povel and Essen (1985).

Table 1. Rhythmic patterns used in experiment

Pattern	Duration Sequence	Onset Sequence
1	1 1 1 1 3 1 2 2 4	■■■■■□□■■□■□■□□□
2	1 1 2 1 1 2 1 3 4	■■■□■■■□■■□□■□□□
3	1 1 2 1 3 1 2 1 4	■■■□■■□□■■□■■□□□
4	1 2 1 1 1 2 1 3 4	■■□■■■■□■■□□■□□□
5	1 1 1 1 2 1 2 3 4	■■■■■□■■□■□□■□□□
6	1 1 1 2 2 3 1 1 4	■■■■□■□■□□■■■□□□
7	1 1 1 2 1 1 3 2 4	■■■■□■■■□□■□■□□□

To have more control in this experiment, we use GTCMT's robot Haile (/ˈheɪli/) (Figure 1) to generate the stimuli. Compared to a human, the drummer robot's playing is measurably more accurate. The only uncontrollable factor is the mechanical friction and delay; however the robot is still more precise than a human player. The up-and-down movement of Haile's left arm resembles that of a human's (in terms of rate) and is used in this experiment (Weinberg and Driscoll, 2007).

Figure 1. Drummer robot Haile in GTCMT

C. Subjects

20 subjects with different levels of music background attended this experiment. We randomly divided the subjects into 2 groups, Group A and Group B. The two groups played the drum patterns in different visual cue orders. We purposefully did not consider the skill level in music or percussion when dividing the groups, because a measure of skill is somewhat subjective and different standards may have different criteria for measuring it. Therefore it can be assumed that the primary difference between the two groups is the visual cue order.

D. Procedure

For Group A, subjects were asked to follow the 7 patterns played by Haile twice. During the first round they could see Haile's movement and during the second round Haile was occluded with a large white board so that the participants would play without making visual contact with the robot. Subjects in Group B followed the same procedure, but in reverse order. They played the 7 patterns without visual cues the first time and then with the visual cues the second time.

E. Equipment

To detect and measure Haile's strike onsets, a piezoelectric sensor element was placed on the surface of the drum to detect vibration. When the surface of the drum is struck the piezo element generated a voltage between its two electric poles due to the change in shape. An Arduino microcontroller system was used as an AC/DC converter, which read the piezo output voltage value and transformed it into digital form. We programmed the Arduino control software to print a time measurement in milliseconds when the voltage exceeded a certain threshold; this threshold could be adjusted to change the level of sensitivity in different experiment environments. Additionally, a microphone was used to record the acoustic sound of the drum as a reference wave file.

For human subjects, we used an electronic drum pad as the input device to collect onset data as MIDI data, which was connected to ProTools software for recording.

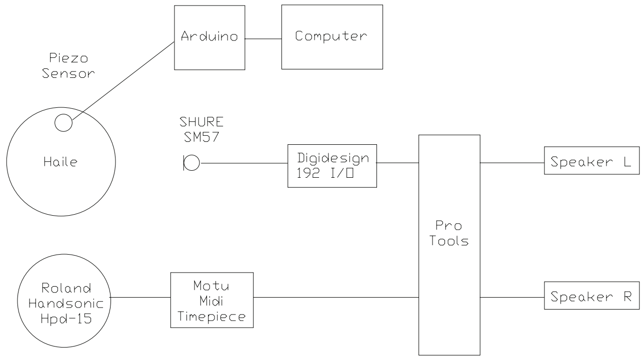

Figure 2.Recording System

DATA ANALYSIS

A. Noise Data Removal

We used a threshold of 200ms as a minimum interval between two of Haile's onsets, following from the tempo used for the patterns corresponding to 240ms per beat. This reduced the possibility of interpreting a single strike with large vibrations as multiple onsets; if the time interval between two detections was less than 200ms, the second one was ignored and the next detected onset was compared with the first. We found 200ms to be a suitable threshold for eliminating noise and false onsets. This threshold is small enough to allow for onset shifts due to robot's mechanical frictions; at the same time it's big enough to eliminate most of the double triggered onsets.

We used a similar minimum interval threshold method for removing noise and false onsets from the subjects' MIDI data. Noise data could be generated in one of two ways: the elasticity of the drum skin might cause a double hit or the simultaneous strike of two sections of the segmented MIDI drum pad (this would cause two MIDI notes to be recorded at the same time). To account for these possible false onsets we set a threshold of a minimum 10ms interval.

B. Onset Sequence Matching

In order to measure the synchronization between the robot and human, it is necessary to find each of their corresponding onset pairs. To do this we regard the piezo onset sequence (played by the robot) as a reference and try to match the MIDI sequence (played by human) with it. The simplest method is to find the nearest one in time. Though simple, this algorithm may cause multiple MIDI onsets to be matched with a single piezo onset or multiple piezo onsets to be matched to a single MIDI onset. This method also does not guarantee accurate matching between the two corresponding sequences of onsets.

To solve the matching problem, we devised a better algorithm based on Dynamic Time Warping Algorithm. DTW is a well-known technique to find an optimal alignment between two given (time-dependent) sequences under certain matching restrictions (Müller, 2007).

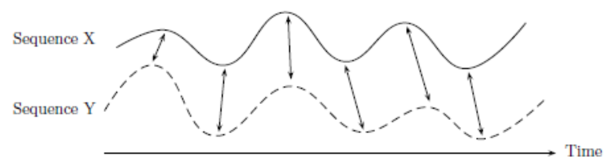

Figure 3.Dynamic Time Warping

To determine an optimal path, defined by the points of matching between the two sequences X & Y pictured along the two axes of a matrix grid, one could test every possible warping path between X and Y, but it would result in very high computational time complexity that is exponential in the lengths *N* and *M*. To reduce the time complexity, recursive computation and dynamic programming can be applied. A new matrix named accumulated cost matrix is defined:

$$D(n,m) = DTW(X(1{:}n), Y(1{:}m))$$

with the cost:

$$c(x_n, y_m)$$

The resulting accumulated cost matrix D can then be computed by the recursion:

$$D(n,m) = \min\{D(n-1,m-1), D(n-2,m-1), D(n-1,m-2)\} + c(x_n, y_m)$$

for $1<n\le N$ and $1<m\le M$

Instead of finding all possible routes through the grid that satisfy the constraints, this algorithm works by keeping track of the cost of the best path to each point in the grid. During the computational process populating the accumulated cost matrix, any path can potentially be the lowest cost path. But the optimal path can be traced back once the entire matrix is populated.

Several constraints are applied in the process. There cannot be multiple MIDI onsets being matched to single piezo onset or one MIDI onset being matched to multiple piezo onsets.

Further, if a MIDI onset *m* has been matched to a piezo onset *n*, the MIDI onsets after *m* are not allowed to be matched to the piezo onsets before *n*; thus, a cross-match is avoided to maintain temporal order.

We also set $c(x_n, y_m) = 500\text{ms}$ when there is no match, as a penalty. Based on our observation and analysis of the experiment data, a subject is not likely to try to make a MIDI onset that is more than 500ms away from the piezo onset that he/she intends to follow. The subjects either make the onset closer to the piezo onset or miss it altogether. Based on the observation, we set a window with length of $500\text{ms} * 2$ and only onsets inside the window could be selected as possible matches. This could increase the computation speed.

This algorithm has some advantages comparing to simply finding the nearest onset. Figure 4 shows us a possible situation. The upper part shows piezo onsets from 40s to 45s, and the lower one shows the MIDI onsets during the same period. For the 3rd red onset R3, we can tell by observation that the subject intended to catch the 3rd piezo onset B3, but played a little late. R4 is supposed to be matched with B4. With the simple algorithm, R3 will be matched to both B3 and B4 since it is the closest one to both of them. If only one pair is to be selected, B4 would be matched to R3, thus making the wrong match and abandoning B3 and R4 onsets. However, with the DTW-based algorithm, matches are made correctly.

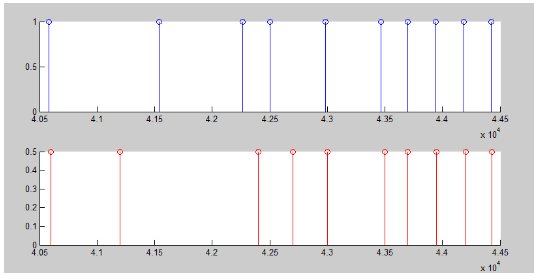

Figure 4. One segment of piezo and MIDI onset sequences

The results of matching were graphically verified and were found to be as intended. Figure 5 is the first 10 seconds' matching results for the first subject, with visual cues, and for pattern 1. Both x and y axes are time. Green lines are subject's onsets, and blue lines are piezo onsets. The red dots mark the matches determined. The black line is where y=x, which gives a reference for perfect synchronization. The figure also shows some noisy subject onsets, which would later be filtered off.

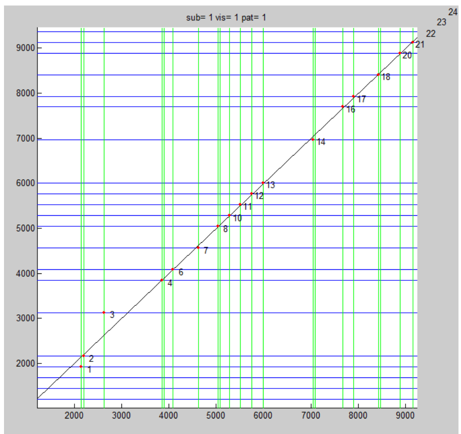

Figure 5.Matching result of first 10 second for first subject

C. Analysis on time difference between matched onset pair

After getting all the matching pairs, statistical analysis was conducted on the time difference (subject's onset time - piezo onset time) between each matching pair. We used absolute time difference to analyse how well they are synchronized and used the original time difference to analyse the prediction/lead-lag effect during learning. We applied 3-way Analysis of Variance (ANOVA) (Hogg, 1987) on the two versions of time difference, with respect to the parameters: visual cue order (experiment group), visual cue on/off, and pattern. The ANOVA analysis was applied to not only the whole 1 minute of each pattern's playtime, but also to the first 10s, 10-20s, 20-30s and 30-60s, in order to be able to observe the performance differences in different learning periods.

The analysis of data from the experiment showed the following effects of visual cues:

(1) Overall synchronization accuracy: the analysis on absolute time difference shows that two out of seven patterns, (4 and 7), which we labeled as medium and high difficulties, showed significant difference in the synchronization metric between the visual on/off conditions (Figure 6), and the metric for these patterns was worse in the case of visual off condition. Differences were not significant with easier patterns. We infer that visual cues aid synchronization in case of more difficult patterns.

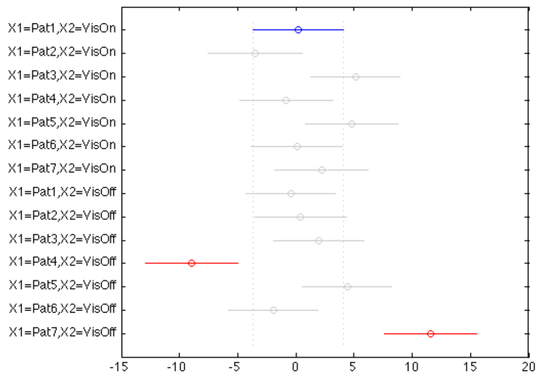

Figure 6. Comparison of synchronization metric for the 7 patterns under Visual On/Off conditions

Analyzing the data of the relatively steady latter half of the sessions, from 30th to 60th second, we also observed the trend wherein the mean synchronization metric was slightly better for visual-on condition compared to that for visual-off.

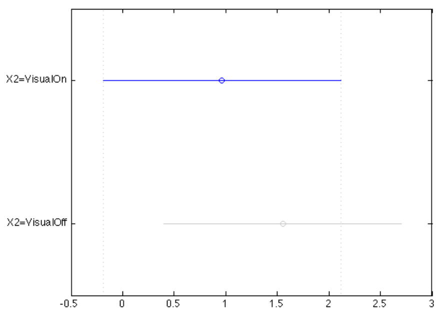

Figure 7. Comparison of synchronization metric for Visual On/Off conditions for the steady part of the sessions

(2) Learning effect: analyzing the absolute time difference data of the relatively steady latter half of the sessions, from 30th to 60th second, we also found that the subjects group that played the pattern with visual cues first before playing them without the cues had a slightly better synchronization metric than the other group that was exposed in the reverse order (Figure 8). While more experimentation is required to establish this with statistical significance, looking at the general trend, we infer that visual cues helped subjects in learning the patterns faster.

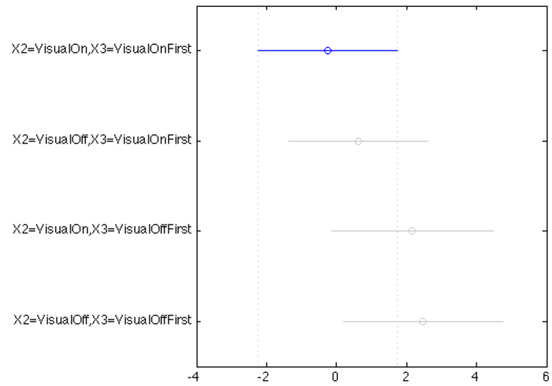

Figure 8. Comparison of synchronization metric for the Visual On/Off conditions, for the two subjects groups differing in the order of visual condition

(3) Lead and lag: the analysis on original time difference shows that subjects played in a bigger lag when playing with visual cue at first 10 seconds, and subjects played better after the first 10 seconds (Figure 9). The overall performance for 1 minute, with presence of visual and without, didn't show significant difference. This means people take some time on seeing the visual cues and then play the onset during the learning period, but it doesn't influence the overall accuracy. They play better with visual cues after learning period.

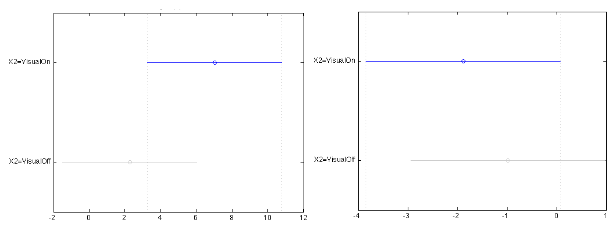

Figure 9. Comparison of synchronization metric for the Visual On/Off conditions for first 10 seconds (left) and 10-20 seconds (right)

CONCLUSION

When learning a rhythmic pattern at a medium tempo, in our experimental setup, visual cues were found to be helpful in facilitating synchronized performance only in case of more difficult patterns, when subjects were first introduced to new patterns. Subjects also showed a tendency to learn new patterns faster with visual cues. They further tended to play in lag with visual cues in the early learning period, but to play better later after learning with visual cues. The latter two effects however were not statistically significant.

ACKNOWLEDGMENT

We thank Brain Blosser from GTCMT for spearheading the previous versions of the experiment. We thank Andrea L Thomaz and Michael J Gielniak from GT RIM Lab for their assistance with initial experiments. Thanks also to Mason and Aaron for proof-reading. We thank all the professors and classmates at GTCMT for either participating in the experiment or for their precious suggestions.

REFERENCES

Blosser, B., Mallikarjuna, T. & Weinberg, G. (2010). *The Effect of Visual Modality on Musical Ensemble Synchronization*. Georgia Tech Center for Music Technology. Retrieved April 2012, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.182.4779.

Hogg, R. V., and J. Ledolter (1987). *Engineering Statistics*. New York: MacMillan.

Hoffman, G. & Weinberg G. (2010). Synchronization in Human-Robot Musicianship. *Proceedings of The 19th International Symposium on Robot and Human Interactive Communication (RO-MAN 10)* (pp. 718-724). Viarggio, Italy.

Lim, A., Mizumoto, T., Cahier, L., Otsuka, T., Takahashi, T., Komatani, K., Ogata, T. & Okuno, H.G. (2010). Robot musical accompaniment: integrating audio and visual cues for real-time synchronization with a human flutist. *Proceedings of The IEEE/RSJ International Conference on Intelligent Robots and* Systems *(IROS)* (pp. 1964-1969).

Luck G. & Sloboda J. (2008). Exploring the spatio-temporal properties of simple conducting gestures using a synchronization task. *Music Perception, Vol. 25, Issue 3*.

Müller, M. (2007). Dynamic Time Warping. In *Information Retrieval for Music and Motion* (pp. 69-84). Springer.

Povel, D. & Essens, P. (1985). Perception of Temporal Patterns. *Music Perception, Vol. 2, No. 4* (pp. 411-440).

Repp, B. H. (2006). Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology, Vol. 2, No. 2*-3 (pp. 163-181).

Repp, B. H. & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research* (pp. 252-270).

Weinberg G. & Driscoll, S. (2007). The Interactive Robotic Percussionist: New Developments In Form, Mechanics, Perception And Interaction Design". *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2007) (pp. 97-104). Arlington, VA.