Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Understanding differences of eye movements patterns while reading musical scores between instructors and learners to design learner-centered teaching strategies

Katsuko T. NAKAHIRA[a*] & Muneo KITAJIMA[a]
[a]*Nagaoka University of Technology, Japan*
*katsuko@vos.nagaokaut.ac.jp

Abstract: While learning instruments playing, one engages in the following mental and motor activities; reading notes in a musical score, retrieving relevant information from long-term memory concerning how to play the notes, moving relevant body parts, and verifying his/her performance against the expected sound. The learner is required to coordinate these activities in the learning process. This paper assumes that how the process of coordination is carried out should depend on the degree of matching between the learner's skill and the musical score which should appear in the process of reading; visually encoding the to-be-played notes and verifying the sound from the instrument he/she has just played. This paper proposes a method for objectively measuring the degree of matching by using eye movements data while reading a musical score with hearing notes.

Keywords: Education, Eye-Tracking

1. Introduction

Learning instrument playing is one of popular activities that many people do in their lives. One learns instrument playing to become a skillful performer, to enjoy themselves as hobby, to be licensed as a musical teacher, etc. Musical scores are the materials these people would use while learning, and their contents would have a large impact on the course of learning. The learner would be willing to use a musical score that he/she estimates it matches his/her purpose of instrument learning, on the other hand, he/she would not use such a music score that is too difficult for him/her to play. This paper proposes a method for objectively measuring the degree of matching by using eye movements data while one is reading a musical score with hearing notes with some empirical data.

There are many works that deal with eye movements while reading music. Wristen (2005) argued that sight-reading is an integral part of the experience for all musicians, and argued the direction of its ability. In the review, sight-reading task divided into two large phases: preparation and performance, and performance. Especially, performance needs a number of cognitive demands. At least, to recognize the score, the basic elements for sight-reading is rhythm, melody, harmony, and context. After getting the knowledge of the elements, Lehmann and Ericsson (n.d.) showed to decipher a score at sight reading, the musicians have to recognize musical patterns, generate a large-scale performance, plan to govern performance of the piece as a whole, and learn to anticipate how the music continues. With the idea, several researches tried to describe the difference of sight-reading skills for music learners. For example, Kopiez and Galley (2002) tried to compare features of eye movements in musicians to a non-musician. Thus, the relation of eye movement, sight-reading, and the music learners' skills is going to explain. If the idea will be able to be realized practically, we can estimate easily the learners' ability for sight-reading, and expansively. For the perspective, we can easily imagine the application of the method to e-Learning. Before applying, we need to consider how to include the technique for e-Learning. In general, almost instructors thought that training instruments playing skill is incongruous with e-Learning. But during the several years, the causes of incongruence partly have been improved with some researches with the analysis of the relation between training elements and human knowledge or feeling.

Nakahira and Kitajima (2013) proposed an educational design for piano playing and singing that is capable of transferring skill related to not only explicit knowledge but also implicit knowledge. In order to explore specifications of such an educational design, we adopted a methodology called Cognitive Chrono-Ethnography (CCE), which was successfully applied to understand people ’s daily behavior, for studying how trainees interact with and utilize a proposed educational design in real educational settings, which would eventually lead to an appropriate specification of student-centered educational design for piano playing and singing skill transfer. They set the design consisted of the following three elements: (i) having the trainee watch a video of the model performance for the purpose of building mental image of the performance goal, (ii) having the trainee mimic the model performance while referring to guidance comments added on the musical score, and video-record the performance, and (iii) having the trainee submit the video and critically review his/her own performance. Using this, they designed student-centered design for education in skill transfer. Through the research, they suggested to identify critical parameters and understand the trainees’ behaviors while they used the student-centered educational design for piano playing and singing skill transfer.

2. Methodology

While a learner is practicing instrumental playing, he/she has to engage in the following four processes;

1. reads the musical score and associate it with his/her body movements,
2. sends signal from his/her brain to his/her body parts,
3. moves his/her body for playing,
4. checks accuracy of his/her playing.

In this paper, we focused on the first process of reading musical score and the fourth process of checking accuracy of their playing

2.1 Reading Musical Scores

In the following, we describe in detail the processes involved in reading musical scores.

First, the learner gets notes from the musical score. In this process, he/she visually acquires each note as a code of sound though his/her eyes, and the visual information is sent to his/her brain. The signal of note information is used to retrieve a set of information associated with the note from long-term memory, including the sound pitch, the position of instrument to strike a note, and so on. After retrieving the relevant information, the learner moves his/her body to hit the note. Upon hitting it, the instrument makes some sound. The learner recognizes the sound he/she has just made as a signal, and compares it with the information having been retrieved from long-term memory. If there is no mismatch, he/she recognizes the hitting was correctly carried out, and the learner proceeds to get a next note. If not, he/she recognizes the hitting is incorrectly performed, and the learner retries to play the correct note. We can identify two critical processes that should affect the performance of learners; The first is a visual process which is the process of getting information necessary for hitting the note, including such information as posture of fingers, how to move fingers, sound to be heard, etc. The second is the accuracy check, which is done by comparing the pitch of sound he/she has just played and the corresponding information retrieved from long-term memory. This is essentially the process of matching the auditory image that has been put in working memory by playing the note with the one having been loaded in working memory prior to playing it.

The first process of visual information processing and the second process of mental process of auditory information matching have to be carried out in parallel and in synchronous with the action of instrument playing. We assume that the degree of difficulty in carrying out these processes coherently should appear in the eye movements of the learner. If the learner feels the difficulty level of the musical score just matches his/her skill level, he/she would smoothly get the note information from the musical score. As a result, his/her eye movements would show coherent patterns, which would be observed while one is reading sentences in his/her favorite books. On the other hand, however, if he/she feels the musical score too easy or too difficult, he/she may feel uncomfortable about getting the note information in his/her favorite pace. His/her eye movements would show incoherent patterns, i.e., abrupt leaps, abnormally long fixations at specific places, etc.

2.2 A Model of Eye Movements While Reading a Musical Score

Figure 1 shows a model of musical score reading. Suppose a situation in which a musical score is placed in front of the learner. We are interested in the two datasets; one is the series of eye-fixations on the musical score, and the other is the series of note-playing events. Both datasets can be regarded as a function of time, and therefore we can investigate the mutual relationships by arranging them in the time dimension and trying to discover patterns that would be associated with the degree of matching of the musical score and the learner's skill level. In the figure, the horizontal line shows the eye position in the horizontal dimension, and the vertical line shows the elapsed time since the onset of the first note, which has a linear relationship with the time course of note-playing events. We denote the time course of note-playing events as a set $\boldsymbol{Q}_{\text{note-playing}}$ (sound sequences in Figure 1) and eye fixation events as a set $\boldsymbol{P}_{\text{fixation}}$ (pattern A – C in Figure 1):

$$\boldsymbol{Q}_{\text{note-playing}} = \{ (x(t_j),\ y(t_j)) \mid j = 1, 2, \ldots\ldots \},$$
$$\boldsymbol{P}_{\text{fixation}} = \{ (\xi(t_k), \eta(t_k)) \mid k = 1, 2, \ldots\ldots \}$$

where $(x(t_j),\ y(t_j))$ represents the (x, y) coordinate of the j - th musical score image, and $(\xi(t_k), \eta(t_k))$ represents the (ξ, η) coordinate of the k - th fixation recorded at $t = t_k$. Using these descriptions, we assume three possible patterns in the relationships between $\boldsymbol{P}_{\text{fixation}}$ and $\boldsymbol{Q}_{\text{note-playing}}$;

- **Pattern A:** a large set of (ξ, η) coordinates in $\boldsymbol{Q}_{\text{note-playing}}$ corresponds with the one in $\boldsymbol{P}_{\text{fixation}}$, meaning that the learner reads the musical score as a large chunk while he/she hears the stream of sounds of notes being played with the instrument,
- **Pattern B:** a smaller set of (ξ, η) coordinates in $\boldsymbol{Q}_{\text{note-playing}}$ corresponds with the one in $\boldsymbol{P}_{\text{fixation}}$, meaning that the learner reads less frequently than Pattern A learner does,
- **Pattern C:** there is little correspondence between (ξ, η) coordinates in the set $\boldsymbol{P}_{\text{fixation}}$ and (x, y) coordinates in the set $\boldsymbol{Q}_{\text{note-playing}}$.

Using the three patterns, we can estimate the learner's status of instrument playing skill. When a learner has sufficient ability to read musical scores and hear the stream of sounds of notes in parallel, the learner would show the behavior, so called "shadow hearing" of the musical score. In this case, be treated as a subset of $\boldsymbol{Q}_{\text{note-playing}}$, that corresponds to Pattern A. A skilled instrumental player would show a stream of eye fixations similar to $\boldsymbol{P}_{\text{fixation}}$ even if there is no sounds of notes to hear. $\boldsymbol{Q}_{\text{note-playing}}$, when exits, might be used as cues for keeping timing, but it would not be necessary for those who can do "shadow hearing." Less skilled players would show less coherent relationship between $\boldsymbol{Q}_{\text{note-playing}}$ and $\boldsymbol{P}_{\text{fixation}}$ than the one shown by the skilled players as described above. We assume that there should be two cases that would be distinguishable in terms of the relationship between $\boldsymbol{Q}_{\text{note-playing}}$ and $\boldsymbol{P}_{\text{fixation}}$. The first case is characterized by unstable $\boldsymbol{P}_{\text{fixation}}$ without musical sounds, but stable $\boldsymbol{P}_{\text{fixation}}$ with $\boldsymbol{Q}_{\text{note-playing}}$; in other words, musical sounds are inevitable for them to exhibit consistent eye movement patterns. Musical sounds help the learner search the appropriate notes from learner's long-term memory. We assume that this case corresponds to Pattern B. The second case is characterized by a reverse relationship between $\boldsymbol{Q}_{\text{note-playing}}$ and $\boldsymbol{P}_{\text{fixation}}$, namely,

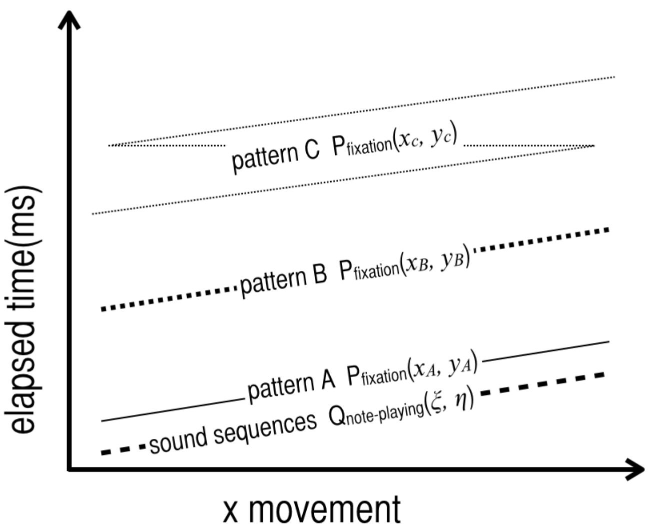

Figure 1. A model of musical score reading.

stable $\boldsymbol{P}_{\mathrm{fixation}}$ without musical sounds, but unstable $\boldsymbol{P}_{\mathrm{fixation}}$ with $\boldsymbol{Q}_{\mathrm{note\text{-}playing}}$. In this case, the introduction of musical sounds would cause extra cognitive burden on the learners that make difficult for them to search notes from learners' long-term memory. We assume that this case corresponds to Pattern C

3. Experiment

Applying the methodology, we make a hypothesis for eye tracking of reading music and the trajectory. If the leaners groups' ability is almost similar, the learners' reading music split some patterns;

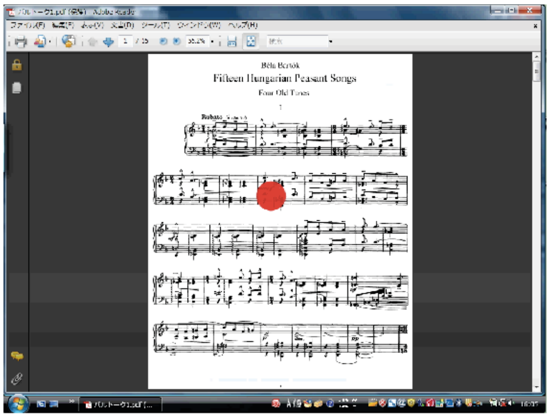

Figure 2. A screenshot for the experiment. A large red circle displayed on the monitor indicates the recorded participant's gaze point. The musical score was displayed in the entire area of the monitor.

(1) independently with or without musical sound, the learners' eye tracking pattern has no difference,
(2) dependently with or without musical sound, the learners' eye tracking pattern has some difference, dividing into learners' ability of reading musics.

To test the hypothesis, we conducted an experiment. At the experiment, we focus the ability of collaboration between visually and auditory. Due to this, we adopt the model performance for musical sound which means the musical sound does not stop due to the players 'mistake. If the learner achieve a mastery of the musical score, he/she can match $\boldsymbol{Q}_{\mathrm{note\text{-}playing}}$ pattern with their $\boldsymbol{P}_{\mathrm{fixation}}$, if not, he/she cannot match. And the $\boldsymbol{P}_{\mathrm{fixation}}$ pattern will change whether the experiment include with or without musical sound. We chose participants, expected to have uniform learners' level, from a pre-school teacher education institution which has a class for basic piano/singing course. Fourteen students and two instructors participated in the experiment. All students had basic piano playing skills. We prepared a one page musical score whose level was at the middle class for piano instrument, including some chordal notes. Although the score had five lines, we used first two lines for acquiring eye tracking data in order to avoid recording errors. It included nine bars with about 30 note heads. Each musical score was displayed on a 22 inch LCD monitor, with a resolution of 1024 x 768 pixels. The distance between the monitor and an participant was about 30cm. We used a Tobii Eye Tracker to record participants' eye movement with sampling rate of 50Hz. Figure 2 shows a screenshot from the experiment. The experiment was conducted according to the following flow:

First, the experimenter explained the outline of the experiment, then instructed the important points that the participant was expected to follow consisting of the following four items (the participant is indicated as "you" in the following instruction);

(1) When you read the musical score, please keep your head in the region about 30cm from the monitor's surface;
(2) There are two conditions. One is "read musical score without sound" condition and the other is "read musical score with sound" condition.
(3) When you are in the "read musical score without sound" condition, please imagine the melody from the musical score, and try to synchronize the position of notes you read and the sound pitches you image;
(4) When you are in the "read musical score with sound" condition, please synchronize the sound pitches you hear and the position of the note heads you read from the musical score;

After calibrating the Tobii eye tracker, the experimenter presented the musical score on the monitor with full screen, followed by sending a starting sign to the participant. Between the two conditions, a short break was given to the participant.

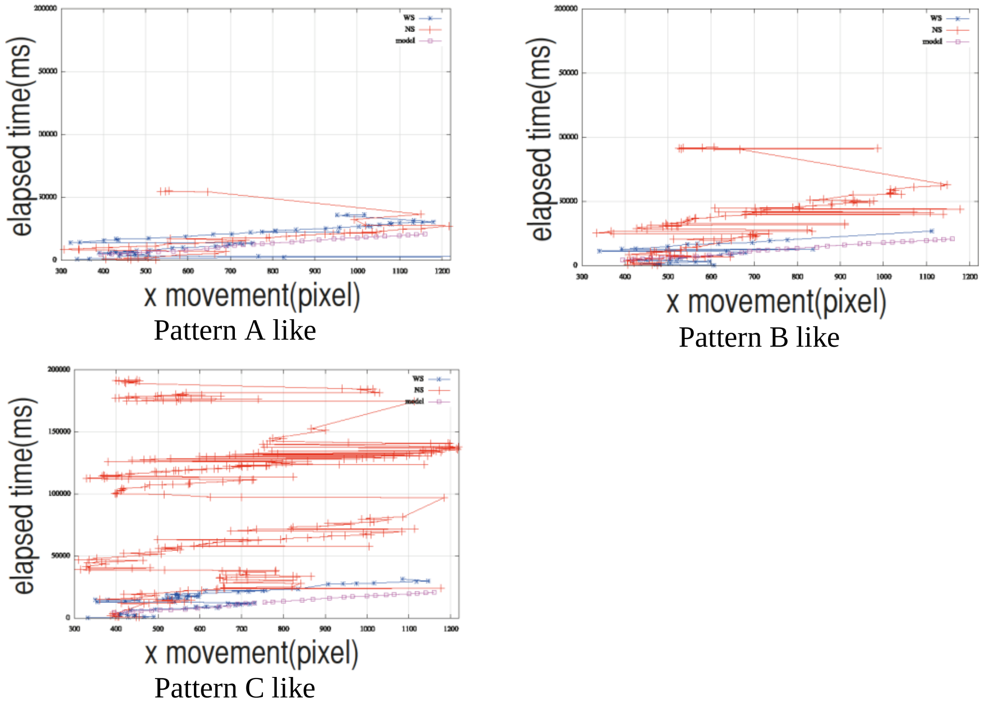

Figure 3. Typical trajectories of participants' eye tracking patterns. □ represents $\boldsymbol{Q}_{\text{note-playing}}$, + represents $\boldsymbol{P}_{\text{fixation}}$ for reading soundless musical score, * represents $\boldsymbol{P}_{\text{fixation}}$ with sound reading music.

4. Analysis

After the experiment, we analyzed the recorded eye movement data. We use first two lines for analysis. At first, we tried to get rid of noises from the data from the *i*-th participant's $\boldsymbol{P}_{\text{fixation}}$. In this research, we make smoothing the dataset $\boldsymbol{P}_{\text{fixation}}$, namely calculated average for $\xi_i(t_k)$ and $\eta_i(t_k)$ coordinate per each 10 points. We regard the smoothed $\boldsymbol{P}_{\text{fixation}}$ as point of gazes. Next step of the analysis is to calculate stationary points for smoothed $\boldsymbol{P}_{\text{fixation}}$. In this analysis, we set the human central area of vision θ_c as 2 degree (--23 pixels). If the *i*th participant's two adjacent gaze points distance is longer than θ_c, we judged that the stationary point had moved. If the distance is shorter than θ_c, we regarded that the stationary point had not moved. Through this process, we can derive the *i* - th participant's stationary points series.

Figure 3 shows the typical trajectories of participants' eye tracking patterns. Square symbol represents $\boldsymbol{Q}_{\text{note-playing}}$, plus symbol represents $\boldsymbol{P}_{\text{fixation}}$ for reading soundless musical score, asterisk symbol represents $\boldsymbol{P}_{\text{fixation}}$ with sound reading music. From the figures, we find the learners have several difference of $\boldsymbol{Q}_{\text{note-playing}}$ and $\boldsymbol{P}_{\text{fixation}}$ with/without sound reading music. From the figures, we find almost learners' trajectories of participants' eye tracking will be distinguished pattern A, B, or C.

5. Discussion

Through the analysis, we considered the learners' types. There are some possibilities for selecting musical score to play appropriate for their playing skill levels. When people want to play a musical instrument, they need to engage in elemental-level training to treat instrument. In the process, learners cannot select the etude or melody music target, because they have least skills for playing instrument. But when they move their skill level status from the elemental level to middle levels, they have possibilities of selecting musical scores due to their variety of playing skills. Though their selections could become diverse, sometimes they tend to choose such musical scores just only too easy or too difficult. In the case of hobby training, there would be no problem in specific selections. But if the learners need or want to improve their playing instrument skills, they have to select musical scores at a challenging level.

6. Conclusion and Future Works

In this paper, we proposed the method for estimating learners' skills to fill in the gaps of their skills and the requirement skill from musical score. Focusing on the processes of reading musical score and of checking accuracy of their playing, we set the hypothesis about the relation between eye movement, auditory checking accuracy for pitch, and reading music. For checking the relation, we set an experiment of eye tracking for reading music, with or without musical sound. We found that the learners state could be divided in four categories by using the features measured by dh and dv, showing differences between sound time sequences and eye positions in the horizontal dimension with or without musical sound. With these categories, we proposed a method of easy estimation for learners' skill and musical scores which fits learners' skill for future

Acknowledgements

The study described in this paper has been partially funded by the Scientific Research Expense Foundation C Representative: Muneo Kitajima (24531274).

References

Kopiez, R., & Galley, N. (2002). The musicians' glance: A pilot study comparing eye movement parameters in musicians and non-musicians. In *Proeedings of the 7th international conference on music perception and cognition* (pp. 683–686). Adelaide: Causal Productions.

Lehmann, A. C., & Ericsson, K. A. (n.d.). Performance without preparation: Structure and acquisition of expert accompanying and sight-reading performance. *Psychomusicology, 15*, 1-29.

Nakahira, K. T., & Kitajima, M. (2013). Development of a student centered educational design for piano playing and singing skills. In A. S. Mashat, H. M. Fardoun, & J. A. Gallud (Eds.), *Proceedings of IDEE 2013 - 2nd international workshop on interaction design in educational environments* (p. 146-152).

San Agustin, J., Skovsgaard, H., Hansen, J. P., & Hansen, D. W. (2009). Low-cost gaze interaction: Ready to deliver the promises. In *Chi '09 extended abstracts on human factors in computing systems* (pp. 4453–4458). New York, NY, USA: ACM.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D.W., et al. (2010). Evaluation of a lowcost open-source gaze tracker. In *Proceedings of the 2010 symposium on eye-tracking research 38; applications* (pp. 77–80). New York, NY, USA: ACM.

Wristen, B. (2005). Cognition and motor execution in piano sight-reading: A review of literature. *Applications of Research in Music Education*, *24*(1), 44-56.