ANALYZING USER INTERACTIONS WITH MUSIC INFORMATION RETRIEVAL SYSTEM: AN EYE-TRACKING APPROACH

Xiao Hu[1] **Ying Que**[1] **Noriko Kando**[2] **Wenwei Lian**[1]

[1]University of Hong Kong
{xiaoxhu, queying, lian0624}@hku.hk

[2]National Institute of Informatics, Japan
kando@nii.ac.jp

ABSTRACT

There has been little research considering eye movement as a measure when assessing user interactions with music information retrieval (MIR) systems, whereas many studies have adopted conventional user-centered measures such as user effectiveness and user perception. To bridge this research gap, this study investigates users' eye movement patterns and measures with two music retrieval tasks and two interface presentation modes. A user experiment was conducted with 16 participants whose eye movement and mouse click behaviors were recorded through professional eye trackers. Through analyzing visual patterns of eye gazes and movements as well as various metrics in prominent Areas of Interest (AOI), it is found that users' eye movement behaviors were related to task type. Besides, the results also disclosed that some eye movement metrics were related to both user effectiveness and user perception, and influenced by user characteristics. It is also found that some eye movement and user effectiveness metrics can be used to predict user perception. This study allows researchers to gain a deeper insight into user interactions with MIR systems from the perspective of eye movement measure.

1. INTRODUCTION

With the rapid development of Music Information Retrieval (MIR) as a field, user has been increasingly recognized as playing an essential or central role in the process of music retrieval [14][25]. Users' interactions with MIR systems have been drawing researchers' attention for the purposes of evaluating MIR techniques [13] and interfaces [5] from users' perspectives, as well as improving understanding of users [19].

Various approaches have been utilized to collect data of user interactions with MIR systems including listening histories [26], click streams in system logs [11], user surveys or diaries [32], etc. A series of user-centered metrics have then been proposed and used to measure users' interactions including those in user effectiveness (e.g., number of songs listened to) [11] and user perceptions (e.g., emotion, satisfaction) [32]. To facilitate this process, various tracking techniques have been employed through off-the-shelf software tools and/or self-development apps [11][32].

As an alternative approach to measuring users' interactions with various stimuli (e.g., computer systems, web interface, learning materials), eye movement has been used to explore the relationship between human's cognitive process and their behaviors [3]. It is believed that people's cognition can be reflected by their eye movement patterns [27][30]. Nowadays, fast-developing high-tech devices enable researchers to acquire eye tracking data in an accurate and reliable manner, such as eye tracking glasses and eye movement sensors. Compared to measures that rely on users' self-report, eye tracking data are objective and thus could help avoid possible response bias [4]. With current (wearable) technology, the process of collecting eye movement data is becoming less obtrusive. If relevant methodologies and techniques are appropriately applied to MIR studies, measurements of eye movement could be potentially helpful for analyzing and evaluating user interactions with MIR systems.

This study aims to explore the application of eye movement measure to investigating user interactions with a MIR system, through a user experiment involving different MIR tasks and interface modes. The relationships between eye movement measure and traditional measures such as user effectiveness and user perception are analyzed. Possible influence of user characteristics (e.g., gender) is also taken into account. The findings are expected to provide empirical evidence on exploiting eye movement data in studying user interactions in MIR.

2. RELATED WORK

2.1 User Interactions in MIR

Users' interactions with MIR systems have been mostly studied in the context of user-centered evaluation of MIR. Compared with the system-centered evaluation paradigm, user-centered evaluation is not as popular, but is increasingly adopted, with the recognition of users being the center of MIR [25]. It is advocated that user interactions provide a more direct way to understand how human perceive and use MIR systems, as serving the users is arguably the ultimate goal of MIR systems [14]. Another purpose of studying user interactions in MIR is for designing interactive systems where users' behavioral, psychological and physiological measures could serve as input as well as feedback to systems [15][11][32].

To date, a number of user-centered measures and metrics have been applied in studying user interactions in MIR. Conventional user-centered metrics including those of user effectiveness and user perception [16] have been adopted in MIR, such as number of songs found, task completion time and user satisfaction, etc. Besides, enjoyment has been recognized as an important aspect of user perception unique in MIR [18]. Along with the line of research considering the casual-leisure nature of music retrieval, new metrics were proposed for evaluating MIR, including novelty, aesthetics, and content quality [12].

To collect and analyze the aforementioned measures and metrics, a range of methods have been employed. Traditional methods of user studies such as survey, interview and observations are included [18] and further developed to work with current technologies such as experience sampling with mobile apps [32]. Interactive behaviors are often recorded as system logs with the support of software tools that can track mouse clicks and keyword input [11]. More recently, with the development of wearable technology, studies started to collect users' physiological signals while they interact with MIR systems [22]. Nevertheless, to date there has been little research investigating user-MIR interactions with eye movement measure [29] which has been recognized as effective in capturing people's cognitive process and used in a number of domains.

2.2 Eye Movement Tracking and Cognitive Behaviors

It is recognized that eye movement can reflect human cognitive behaviors. In recent years, eye movement has been used in the domain of text retrieval, particularly on the influence of search interface design on users' search behaviors [6], as well as the relationships between document relevance and users' cognitive efforts [8]. Eye movement has also been widely employed in studies on people's reading behaviors such as identifying the sections focused on by readers during reading process [24]. In the education domain, there are many studies benefited from analyzing learners' eye movement such as computer programming [23], language learning [1], and music education [20].

In leveraging eye movement data, an important step is to find concrete measurements that can describe eye movement in an accurate and reliable manner. Fixation and saccade are two primary measures of eye movement, which are regarded as plausible evidences to detect cognitive processing during interactive tasks such as information retrieval [4]. Fixation indicates that a person stares at certain location and lasts for a period of time, presumably to process certain cognitive tasks, while saccade is a dynamic visit between two fixations [19]. Furthermore, investigators have often attempted to categorize eye movement behaviors into different patterns so as to interpret possible cognitive implications by comparing the patterns [4][8][24][30]. Relatedly, visualization of eye movement patterns such as heatmaps of eye gazes and plots of eye scan paths is also a viable and frequently adopted approach to making sense of eye movement data [4]. For instance, based on the positions of intensive fixations and the speed of saccades, authors of [30] found the correlation between eye fixation and participants' learning performance.

MIR also involves cognitively intensive activities where eye movement has great potential in studying user-system interactions. An existing research gap is that very few user studies have considered eye movement as a measurement of interactions between user and MIR systems, even though there are studies exploiting eye movement in music psychology [17] and music performance [31]. Therefore, this study aims to bridge the gap by providing empirical evidence on studying interactive MIR systems from an eye-tracking approach.

3. RESEARCH QUESTIONS

Aiming to showcase how eye movement measures can be used in MIR research, this study considers multiple constructs in typical MIR research settings: retrieval tasks, system interface modes, user-centered measures as well as user characteristics. Specifically, through conducting a user experiment involving these factors, this study focuses on three research questions as follows:

Q1. To what extent do different music retrieval tasks and system interfaces have effect on users' eye movement measure?

Q2. To what extent is eye movement measure related to user effectiveness and user perception measures?

Q3. To what extent is eye movement related to user characteristics (e.g., gender, music listening habit)?

To answer these research questions, a set of hypotheses are formulated. There are two hypotheses under Q1:

H1: eye movement measure differs in different task types.

H2: eye movement measure differs in different interface modes.

Under Q2, there are three hypotheses as well:

H3: eye movement measure is related to user effectiveness measures.

H4: eye movement measure is related to user perception measures.

H5: eye movement and user effectiveness measures can predict user perception.

Under Q3, one hypothesis is formulated:

H6: eye movement measure is related to user characteristics.

Answers to Q1 will have methodological implications on the extent to which eye movement can be used in MIR task/system design and evaluation. Answers to Q2 can reveal how eye movement measures are related to commonly used measures of user behaviors and whether they can be complimentary to one another in the context of MIR. In particular, the answer to whether or not user perception, as a subjective measure, is predictable by other objective measures can have the potential of facilitating MIR researchers and system designers to better understand users' perceptions through objective measures. User characteristics (e.g., background in music training) have been recognized as influential in MIR process[19][25], and Q3 is to investigate their relationship with eye movement measure.

4. METHORDS

To fulfill the goals of this study, a user experiment was conducted with a mood-aware MIR system. During the experiment, users' eye movement and other user-centered measures were collected. This section describes details of the system, tasks, experiment procedure and data analysis.

4.1 The System

Moodydb is an online MIR system that supports searching and browsing music by mood [13]. Based on spectrum features extracted from the music audio, this system classifies music into five mood categories, namely *passionate*, *cheerful*, *bittersweet*, *silly/quirky* and *aggressive* [10]. There were 750 songs in the system when the study was conducted, with 738 Western popular pieces (from the U.S. and the U.K.) and 12 Chinese popular songs. The pieces were unevenly distributed across the mood categories with *bittersweet* being the largest class (226 songs) and *silly/quirky* being the smallest (35 songs). The rest categories contained 141 – 184 songs. When a user input singer's name or song's title in query box, it will prompt a set of songs that match the textual query. After user chooses one of the songs as a seed song, the system will retrieve a group of songs with similar mood to the seed song and display them as recommended songs to the user.

The system interface is presented in two different ways; one is in a traditional list-based layout and the other is visualization of album covers based on nested figures as in the treemapping display method [28]. As shown in Figure 1, the List layout ranks the recommended songs from the top of the screen down according to how similar the moods of the recommended songs are to that of the seed song. The Visual layout shown in Figure 2 uses the size of the album covers to represent the degree of similarity: the larger size an album image is in, the more similar the song is to the seed song in term of music mood.

4.2 Tasks and Topics

There were two music information retrieval tasks designed for this study: searching and browsing. As for searching, the participant was given a seed song (e.g., Irreplaceable by *Beyoncé*) to start with, and then make use of the system as a search engine to find other songs whose moods are similar to that of seed song. As for browsing, the participant was required to browse music in a given (seed) mood (e.g., *bittersweet*) and find songs in the mood. For each task, there were two topics that had different seed songs (for the searching task) or different seed moods (for the browsing task), making a total of four topics. All participants conducted both tasks (searching and browsing) with both interface modes (List and Visual), making a 2 * 2 experiment design. All participants conducted the same topics (same seed songs /moods) but the task and interface mode combinations were ordered in a Graco-Latin square arrangement to counterbalance possible sequence effect.

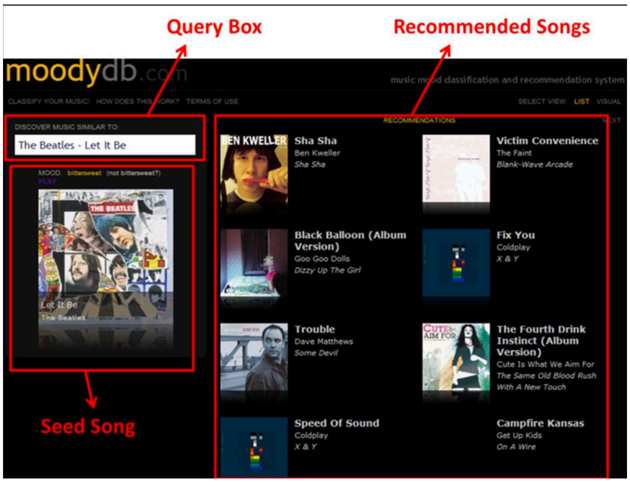

Figure 1. List-based interface mode in the system

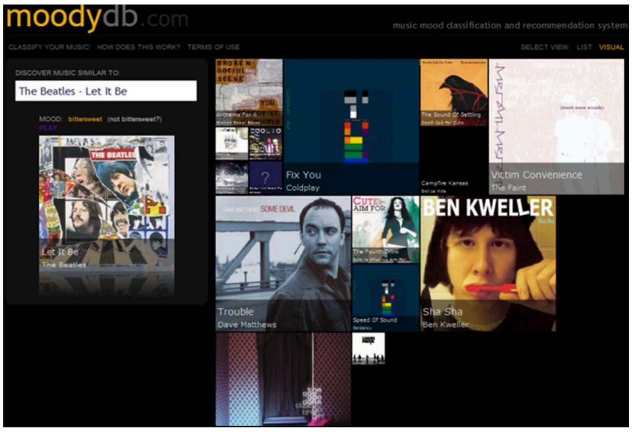

Figure 2. Visual-based interface mode in the system

4.3 Procedure

The experiment started with a pre-session questionnaire about demographic information, music knowledge, and general music information behaviors of the participants. After that, a research assistant demonstrated how the system could be used and the participants used the system for several minutes to familiarize with the system. An eye tracking sensor, Tobii T60, was used to collect data of participants' eye movements. Calibration was done with the eye tracker before a participant started working on each topic, with the assistance of a researcher. After completing all four topics (which covers both tasks and interface modes), participants were inquired about the feelings of the experiment through a post questionnaire. The experiment lasted for approximately one hour, and each participant was paid a nominal remuneration upon completing the experiment.

4.4 Measures

The following user-centered measures are used to answer the three research questions raised in this study.

User effectiveness was used to assess user performance in the experiment. It included four metrics: (a) Number of songs found to fulfill each topic; (b) Number of songs played during the process of working on each topic; (c) Completion time of each topic (measured in minute); (d) Number of mouse clicks during the process of working on each topic. The first metric came from answers submitted by the participants while others were recorded by Tobii 60.

User perception was used to understand users' feelings after they completed each topic. It included three metrics: (a) Task easiness; (b) Preference on the given seed song; (c) Satisfaction with songs found. Measured by a 7-point Likert scale, the perception degree ranges from the most negative to the most positive. Take task easiness as an example: point 1 means the user felt the topic not easy at all, while point 7 means he/she felt the topic very easy.

Eye movement was described statistically with five metrics: (a) Fixation Duration; (b) Fixation Count; (c) Visit Duration; (d) Total Visits Duration; (e) Visit Count. Metrics related to duration such as fixation duration and visit duration is measured by second. It is noteworthy that fixation and visit are two different concepts; the former indicates that participants' eyes fixate on one point, while the latter measures a process of saccade (i.e. moving the eye gaze from one point to another).

The eye movement metrics are calculated in defined Areas of Interest (AOI). In this study, the interface of system is divided into four AOIs based on their functions in the music retrieval process. They are (a) Search Box; (b) Seed Song; (c) Recommended Songs; (d) Player, which are shown in Figure 3. It is noteworthy that a Visit Duration in an AOI is defined as the interval of time between the first fixation inside the AOI and the first subsequent fixation outside the AOI. Besides, while the Visit Duration measures the duration of each individual visit within an AOI, Total Visit Duration measures the sum of duration of all the visits within an AOI. In addition to the metrics in each AOI, we also take the total across all AOIs into account. Statistics of the eye movement metrics in each AOI are generated by the Tobii Studio Analyzer.

4.5 Data Analysis

Hypotheses proposed under the research questions are tested using corresponding statistical tests. H1 and H2 compare metrics between two tasks and two interface modes respectively, for which pair wised t-tests are applied. H3 and H4 investigate relationships between metrics, and thus correlation analysis is used, including Pearson's correlation (for numerical metrics) and Spearman's correlation (for ordinal metrics). Linear regression is used to test H5 which concerns the predictive power of the metrics to user perception. Last but not least, H6 is tested through point-biserial correlation analysis (for binary metrics) and Spearman's correlation analysis. For cases where multiple tests are conducted, Bonferroni corrections are conducted to control type I errors [9].

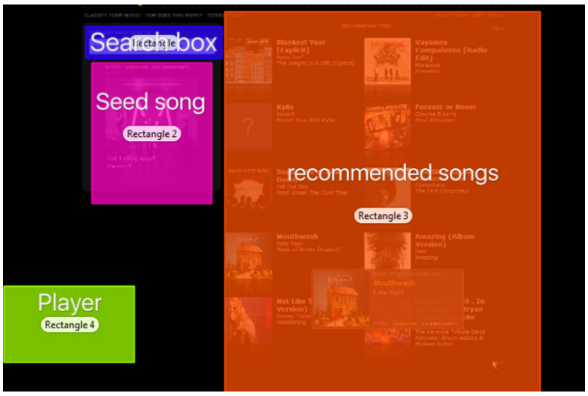

Figure 3. Four AOIs in this study

Besides, visualization as a data analysis method was used in this study as well. There are a number of visualization types in eye movement analysis, such as gaze plot and heat map. In this study, we used heat map to compare eye movement in two interface modes. Examples of heat map are shown in Figures 4 and 5 where aggregated time of fixations on the screen are illustrated by different colors. Red represents the most fixations or longest time period, while green is the least and shortest. The varying colors infer that the levels are in between.

5. RESULTS AND DISCUSSIONS

5.1 Participants

A combination of purposive and convenient sample methods was recruited in the user experiment, including 16 Japanese undergraduate and graduate students (8 female). Their ages ranged from 19 to 50, with a mean of 22.9 and a median of 21. The standard deviation of the age was 7.37. They were from a range of majors including Science, Medicine, Economics, Law and Humanities. A majority of them (11) were able to play an instrument while 5 of them could not. About half of them (7) searched for music fairly frequently, at least several times a week. While the rest of them searched music at least one a month. In term of how often they listened to music, 6 of them rated weekly, 7 daily, and 3 multiple times a day.

5.2 Influences of Task Types or Interface Modes on Eye Movement

This section aims to answer Q1 and test H1and H2. Table 1 presents means and standard deviation (in parenthesis) of eye movement metrics with significant difference in corresponding AOIs between the browsing and searching tasks, detected by paired t-test after Bonferroni correction [7]. As Table 1 shows, during the searching task, participants had longer fixation duration and more fixation count in areas of seed song as well as more visit count in total AOIs. Therefore, our hypothesis H1 is partially supported.

It is not surprising that there were more eye movement in the seed song area during searching task as users needed to compare the seed song and retrieval results, while there was no need for such comparison during the browsing task. The difference on visit count in total AOIs could possibly be attributable to the fact that comparing seed songs and retrieved songs would need a user to move eye gazes between the two AOIs which in turn generated more visits.

Measure	Browsing	Searching	*p* value
FixDur in SeedSong	16.22 (10.91)	27.66 (15.33)	.001*
FixCnt in Seedsong	69.56 (39.37)	109.72 (59.86)	.001*
VisCnt in Total	114.22 (48.98)	152.69 (69.50)	.000*

Table 1. Eye movement measures with significant differences between browsing tasks and searching tasks

The paired t-test on interface mode did not generate significant results after Bonferroni correction, and thus H2 is not supported. This can be explained by the fact (and limitation) that the difference between the two interface modes is actually within an AOI, the recommended song AOI (Figure 3). This calls for alternative methods to compare the two interface modes.

To compare eye movement in a qualitative manner, we present the heat maps of the two interface modes in Figures 4 and 5 respectively. As is shown in Figure 4, fixations from the list-based interface mode disperse along with the positions of the recommended songs. In contrast, in visual-based interface mode (Figure 5), fixations of participants nearly concentrate on the center of recommended songs area. In addition, although, by convention, the List mode ranks highly relevant results up in the list, Figures 4 shows that participants did not only focus on the top rows. For the Visual mode, while more relevant results (with bigger album images) could appear anywhere in the recommended song area, participants' attention seem to have focused on the middle only (Figure 5). These observations are somewhat anti-intuitive and worthy of future investigation.

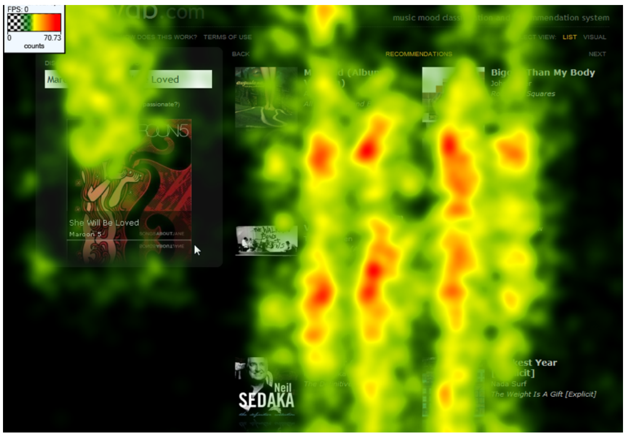

Figure 4. Eye fixation heat map of the List interface mode

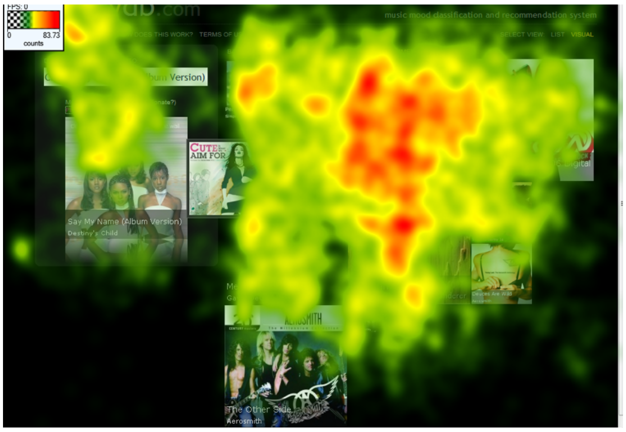

Figure 5. Eye fixation heat map of the Visual interface mode

5.3 Relationships among Eye Movement, User Effectiveness and User Perception Measures

This section aims to answer Q2 and tests H3 to H5. To test H3, Pearson's correlation coefficients were calculated for user effectiveness metrics on interval scales: number of songs found in each topic, number of songs played in each topic and completion time of each task. Significant results after Bonferroni corrections are shown in Table 2.

Effectiveness	Eye Movement	Coefficient	*p* value
Completion Time	TotVisDur in RecomSong	.416	$.001^{*}$
	TotVisDur in Total	.476	$.000^{*}$

Table 2. Correlation between eye movement and user effectiveness (Pearson)

As we can see in Table 2, completion time has significant correlations with total visit duration in recommended songs and in total AOIs, and the correlations were moderately positive (coefficient > 0.3). In other words, when users spent more time moving their eyes within the recommended song areas, the more time they would need to complete the tasks. In contrast, other two effectiveness metrics have no significant correlation with eye movement. Thus, our Hypothesis H3 is partially supported.

To test H4, Spearman's correlation coefficients were calculated for user perception metrics on ordinal scales: task easiness; satisfaction with songs found; preference on seed song. The significant results after Bonferroni correction are shown in Table 3.

Perception	Eye Movement	Coefficient	*p* value
Satisfaction with Songs Found	VisDur in Total	.396	$.001^{*}$
	VisCnt in SeedSong	-.403	$.001^{*}$

Table 3. Correlation between eye movement and user perception (Spearman)

As is shown in Table 3, satisfaction with songs found has significantly moderate correlation with visit duration in total AOIs and visit count in seed song (coefficient > 0.3). Interestingly, the correlations were positive and negative respectively. In other words, the more time users spent in moving their eyes around the AOIs, the more satisfied they would be with the songs they found. However, the more times their eyes visited the seed song area, the less satisfied they would be with the found songs. One possible explanation of the former relationship could be from the perspective of enjoyment [18]: satisfied users presumably enjoyed the task and thus they spent more time looking around on the system interface. For the latter relationship, visiting the seed song area might have been an *effort* one had to pay in order to complete the task (e.g., comparing the seed song to a recommended song), and thus repeated visits to this area would entail more efforts and less satisfaction. As there is no significant correlation between the other two perception metrics (task easiness, preference on seed song) and eye movement, our hypotheses H4 is partially supported.

To test hypothesis H5, we ran a linear regression analysis on predicting user perception from eye movement and user effectiveness. To save space, only significant variables in the prediction models are reported (Table 4). As we can see in Table 4, several eye movement metrics have statistically significance in predicting participants' satisfaction with songs found (one user perception metric). In addition, it is noteworthy that the number of songs found, a user effectiveness measure, was significant in predicting all the three metrics of user perceptions. Thus, our hypothesis H5 is partially supported. This finding that users' eye movement and user effectiveness can contribute to predicting user perception measure has methodological implications in that eye movement and user effectiveness measures can be captured automatically in unobtrusive and objective manners. This can potentially provide novel and reliable methods for detecting user perceptions which have to rely on self-report to collect in traditional methods.

Perception	Metrics	Coefficient Beta	*t*	*p*
Task Easiness	# of songs found	.443	2.400	.021*
Satisfy with Songs Found	VisDur in Player	.533	2.308	.026
	VisDur in Total	.774	3.012	.004
	VisCnt in Search Box	-.984	-2.346	.024
	VisCnt in RecomSong	.718	2.127	.040
	# of songs found	.510	3.505	.001
Like Seed Song	# of songs found	.453	2.572	.014

*: significant at $p < 0.05$ level; $R^2 = .334, .586, .394$ for the three predictive models respectively

Table 4. Regression analysis on user perception

5.4 Relationship between Eye Movement Measure and User Characteristics

To answer Q3 and test H6, we calculated point-biserial correlation coefficients for user characteristics metrics on binary scales: gender (Male=1, Female=2), and being able to play music instrument (Y=1, N=2). We also ran Spearman's correlation on two ordinal variables, frequency of listening to music and frequency of searching for music. Results with significant correlation after Bonferroni correction are shown in Table 5.

Characteristics	Eye Movement	Coefficient	*p* value
Listening Freq	VisCnt in Total	.436	.001*
	VisCnt in Seed-Song	.485	.001*

Table 5. Correlation between eye movement and user characteristics

As shown in Table 5, frequency of listening to music (one user characteristic) has significant and moderately positive correlation with visit count in Total AOIs and that in the seed song area. The more often users listened to music, the more times their eyes visited the seed song area and all AOIs. As visit count of the eyes is related to attention [4], This result is possibly related to the enthusiasm that these users had in music. Nevertheless, other three user characteristics have no significant correlation with eye movement. Thus, our hypothesis H6 is partially supported.

6. CONCLUSION AND FUTURE WORK

This is an empirical study on using eye tracking method to analyze user interactions with MIR systems. We analyzed eye movement metrics in four AOIs and most of our hypotheses were partially supported by the results. Specifically, some eye movement metrics differed between different retrieval tasks. Furthermore, some eye movement metrics were related to user effective and user perception measures. It is potentially useful in research methodology that some metrics of eye movement and user effectiveness can be used to predict user perception. There was no significant difference on eye movement metrics on the different interface modes, which calls for alternative analysis methods beyond AOI-based ones. Also, through comparing visualized heat maps, we found that fixations of participants may not follow patterns found in text retrieval, which warrants further investigation.

This study aims to stimulate and encourage more work to utilize eye tracking in MIR research. There is much room for future work in investigating eye movement with different MIR tasks, use cases and user groups. It could be paradigm shifting if more findings in future work support deriving users' subjective perception (i.e. satisfaction, perceived difficulty) from unobtrusive and objective measures including eye movement.

7. ACKNOWLEDGEMENT

This study is supported by the JSPS KAKENHI Grant (No. JP16H01756), National Natural Science Foundation of China (No. 61703357), and a General Research Fund from the Research Grants Council of the Hong Kong S. A. R., China (No. HKU 176070).

8. REFERENCES

[1] Berzak, Y., Katz, B., & Levy, R. (2018). Assessing language proficiency from eye movements in reading. In Proceedings of NAACL-HLT (pp. 1986-1996).

[2] Bojko A. (2009). Informative or misleading? Heatmaps deconstructed. In Jacko J. A. (Ed.), Human-Computer Interaction, Part I (pp 30-39). Heidelberg, Berlin: Springer.

[3] Chandra, S., Sharma, G., Malhotra, S., Jha, D. & Mittal, A. P. (2016). Eye tracking based human computer interaction: Applications and their uses. International Conference on Man & Machine Interfacing.

[4] Cole, M. J., Gwizdka, J., Bierig, R., Belkin, N. J., Liu, J., Liu, C. & Zhang, X. (2010, August). Linking search tasks with low-level eye movement patterns. In Proceedings of the 28th Annual European Conference on Cognitive Ergonomics (pp. 109-116). ACM.

[5] Dias, R., Pinto, J. & Fonseca, M. J. (2014, September). Interactive Visualization for Music Rediscovery and Serendipity. In Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014-Sand, Sea and Sky-Holiday HCI (pp. 183-188). BCS.

[6] Farzan, R. & Brusilovsky, P. (2009). Social navigation support for information seeking: If you build it, will they come?. International Conference on User Modeling. Springer-Verlag.

[7] Goeman, J. J. & Solari, A. (2014). "Multiple Hypothesis Testing in Genomics". Statistics in Medicine. 33 (11): 1946–1978.

[8] Gwizdka, J. (2014). Characterizing relevance with eye-tracking measures. Information Interaction in Context Symposium. ACM.

[9] Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6, 65-70.

[10] Hu, X. & Downie, J. S. (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. In Proceedings of the 8th Annual Conference of the International Society for Music Information Retrieval (ISMIR), Vienna, Austria (pp.67-72).

[11] Hu, X. & Kando, N. (2012). User-centered measures vs. system effectiveness in finding similar songs. In Proceedings of the 13th Annual Conference of the International Society for Music Information Retrieval (ISMIR), Porto, Portugal (pp. 331-336).

[12] Hu, X. & Kando, N. (2014). Evaluation of music search in casual-leisure situations. In Workshop on Searching for Fun, the Information Interaction in Context conference (IIiX), Regensburg, Germany.

[13] Hu, X., Kando, N. & Yuan, X. (2011). User evaluation of an interactive music information retrieval system. In Proceedings of the 5th Workshop on Human-Computer Interaction and Information Retrieval (HCIR), Mountain View, CA.

[14] Hu, X., Lee, J., Bainbridge, D., Choi, K., Organisciak, P. & Downie, J. S. (2017). The MIREX Grand Challenge: A framework of holistic user experience evaluation in music information retrieval, Journal of the Association for Information Science and Technology, 68(1), 97–112.

[15] Hu, X., Li, F. & Ng, T. D. J. (2018). On the Relationships between Music-induced emotion and physiological signals. In Proceedings of the 19th Annual Conference of the International Society for Music Information Retrieval (ISMIR), Paris, France (pp. 362-369).

[16] Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval, 3(1-2): 1-224.

[17] Lange, E. B., Zweck, F., & Sinn, P. (2017). Microsaccade-rate indicates absorption by music listening. Consciousness and cognition, 55, 59-78.

[18] Laplante, A. & Downie, J. S. (2011). The utilitarian and hedonic outcomes of music information-seeking in everyday life. Library & Information Science Research, 33(3), 202-210.

[19] Lee, J. H. & Price, R. (2016). User experience with commercial music services: An empirical exploration. Journal of the Association for Information Science and Technology, 67(4), 800-811.

[20] Madell, J. & Hébert, Sylvie. (2008). Eye movements and music reading: where do we look next?. *Music Perception: An Interdisciplinary Journal, 26*(2), 157-170.

[21] Moro, R., Daraz, J. & Bielikova, M. (2014). Visualization of Gaze Tracking Data for UX Testing on the Web. CEUR Workshop Proceedings. 1210.

[22] Murakami, M., Sakamoto, T. & Kato, T. (2018, July). Music Retrieval and Recommendation Based on Musical Tempo. In International Conference on Applied Human Factors and Ergonomics (pp. 362-367). Springer, Cham.

[23] Papavlasopoulou, S., Sharma, K. & Giannakos, M. N. (2018). How do you feel about learning to code? Investigating the effect of children’s attitudes towards coding using eye-tracking. International Journal of Child-Computer Interaction, 17, 50-60.

[24] Sevcech, J. & Bielikova, M. (2014). User's Interest Detection through Eye Tracking for Related Documents Retrieval. International Workshop on Semantic & Social Media Adaptation & Personalization. IEEE.

[25] Schedl, M., Flexer, A. & Urbano, J. (2013). The neglected user in music information retrieval research. Journal of Intelligent Information Systems, 41(3): 523-539.

[26] Schedl, M. (2016, June). The lfm-1b dataset for music retrieval and recommendation. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (pp. 103-110). ACM.

[27] Sharma, K., Alavi, H. S., Jermann, P. & Dillenbourg, P. (2016, April). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 417-421). ACM.

[28] Shneiderman, B. (1992). Tree visualization with tree-maps: A 2-d space filling approach. In ACM Transaction on graphics, 11(1): 92–99.

[29] Stober, S. & Nürnberger, A. (2010, June). MusicGalaxy: a multi-focus zoomable interface for multi-facet exploration of music collections. In International Symposium on Computer Music Modeling and Retrieval (pp. 273-302). Springer, Berlin, Heidelberg.

[30] The, B. & Mavrikis, M. (2016, April). A study on eye fixation patterns of students in higher education using an online learning system. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 408-416). ACM.

[31] Vandemoortele, S., Feyaerts, K., Reybrouck, M., De Bievre, G., Brône, G. & De Baets, T. (2018). Gazing at the partner in musical trios: a mobile eye-tracking study. Journal of Eye Movement Research, 11(2), 6.

[32] Yang, Y. H. & Teng, Y. C. (2015). Quantitative study of music listening behavior in a smartphone context. ACM Transactions on Interactive Intelligent Systems (TiiS), 5(3), 14.