ACMC2015 - *MAKE!*

Proceedings of the Annual Conference of the Australasian Computer Music Association

Hosted by the Faculty of Arts and Social Sciences
University of Technology Sydney

Wednesday 18th- Saturday 21st November 2015

Proceedings of the 2015 Conference of the Australasian Computer Music Association
ISSN 1448-7780

Proceedings of the 2015 Annual Conference of the Australasian Computer Music Association Sydney Australia

18th – 21st November 2015

Keynote Speaker:
Associate Professor Cat Hope

Paper Review Committee:
Roger Alsop
Stephen Barrass
Ross Bencina
Rodney Berry
Daniel Blinkhorn
Andrew Brown
Noel Burgess
Warren Burt
Paul Doornbusch
Christian Haines
Lawrence Harvey
David Hirst
Cat Hope
David Kim-Boyle
Eve Klein
Somaya Langley
Terumi Narushima
Michael Norris
Tim Opie
Lindsay Vickery
Ian Whalley

Music Review Panel:
Jon Drummond
Donna Hewitt
Sophea Lerner

Conference Organisers:
Jon Drummond
Donna Hewitt
Sophea Lerner
Ian Stevenson

Published by:
The Australasian Computer Music Association

http://acma.asn.au
November 2015

ISSN 1448-7780

Proceedings edited by Ian Stevenson

All correspondence with authors should be sent directly to the authors.

General correspondence for ACMA should be sent to info@acma.asn.au

The peer review process applied to full papers published in these proceedings involved impartial and independent assessment and review of the papers before publication, conducted by independent, qualified experts as required by section 6.5 of the 2015 Higher Education Research Data Collection Specifications of the Australian Government Department of Education and Training.

TABLE OF CONTENTS

FULL PAPERS

MUSIC SCREEN-READING:
indicative results from two pilot studies

Lindsay Vickery
Edith Cowan University
WAAPA

Talisha Goh
Edith Cowan University
WAAPA

ABSTRACT

This paper discusses two pilot studies conducted by Lindsay Vickery and Talisha Goh, *Screening the Score: Exploring the Potentials and Limitations of Presenting Music Notation on the iPad* and *Sight-Reading Polyphonic Musical Textures: a Pilot Eye-Movement Study.* Vickery's experiment sought to investigate the activity of the eyes of musicians while performing a variety of notations and score presentation types from a screen. Goh's experiment explored sight-reading polyphonic keyboard music containing two, three, four and five voices, at a comfortable pace and with a click-track beat.

1. INTRODUCTION

Two pilot studies - *Screening the Score: Exploring the Potentials and Limitations of Presenting Music Notation on the iPad* (Lindsay Vickery) and *Sight-Reading Polyphonic Musical Textures: a Pilot Eye-Movement Study* (Talisha Goh) - were conducted examining the activity of the eyes of musicians while performing a variety of notation and score presentation types, and the effect of click-track speed on reading increasingly polyphonic music.

The goals of Vickery's study were to use eye-tracking technology to develop a methodology for investigating the effect of reading broad range of musical notation and presentation types from a computer or notepad screen and to identify potential anomalies for study more controlled examination. 11 subjects performing on a range of musical instruments (flute, clarinet, bass clarinet, violin, viola, piano, guitar, cimbalom and percussion) were tested, reading traditional notation, sonographic representation of sound, semantic graphic notation (in which non-standard symbols are used to convey precise information), non-semantic graphical (in which graphical symbols are interpreted by the performer with no pre-defined meaning) and text (Fig. 1.)

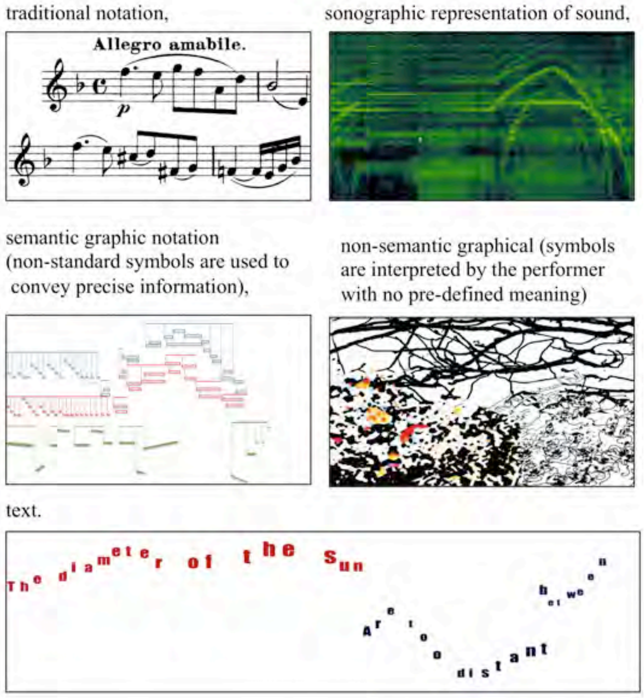

Figure 1. Traditional notation, sonographic representation of sound, semantic graphic notation, non-semantic graphical and text.

A range of screen-score presentation modes were also examined: traditional music score, scrolling score (moving from right to left across the screen), 2D scrolling score (moving both right to left and up and down across the screen), rhizomatic score (moving along predetermined 2D pathways) and scatter display (in which notation is presented in different parts of the screen) (Fig. 2.)

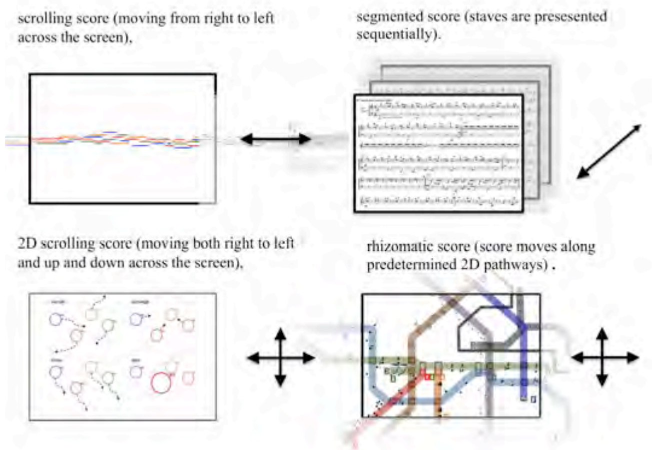

Figure 2. Traditional music score, scrolling score, 2D scrolling score, rhizomatic score and scatter display.

 ISSN 1448-7780

The goals of Goh's study were to examine the effect of increasing the polyphonic density of music notation and performing in conjunction with a click-track, upon performer's eye movements. In the experiment the eye movements of an expert musician were examined while sight-reading polyphonic keyboard music containing two, three, four and five voices, at a comfortable pace and with a click-track beat during three encounters with the music. The first (A) and fifth (E) examples are shown in Fig. 3 and Fig. 4 respectively. The number of fixations per second of performance, fixation duration, eye-hand span and position of fixations were assessed.

Figure 3. Excerpt A: Bass Voice

Figure 4. Excerpt E: Soprano, Alto, Tenor 1, Tenor 2 and Bass Voices

The technological setup for this study was supplied by Edith Cowan University's *Office of Research and Administration*, and consisted of a laptop computer connected to a large widescreen monitor (1920x1080 resolution; picture size 1900x900 pixels) and a web camera. The computer ran the program Tobii Studio v.3, which was designed to set up eye-tracking experiments and record and analyse the data acquired by a Tobii X2-30 eye-tracker was connected to the computer by USB (Tobii X2-30, 2014).

In addition to recording raw-data exportable in .xls format, the tracker visually represents fixation and saccade activity in real-time via video and allows the export of images of the cumulative plotting of fixation and saccade – "gaze-plot" - activity in relation to the score and "heat-maps" frequently viewed regions of the screen (Fig. 5).

Vickery's experiments were conducted in the Music Auditorium of the Western Australian Academy of Performing Arts (WAAPA). The musician's read from either image files of traditional notation or .AVI video files of "animated" scores presented on screen at a comfortable reading distance.

Goh's experiments were conducted in a designated room at Edith Cowan University. The computer screen with the eye-tracker was positioned in front of a digital keyboard with 88 weighted keys. It was also connected to a small damper foot pedal, again of comparable size to that of a piano.

The slides used for the experiment were generated using the Sibelius 7 manuscripts. These were converted into picture files that were compatible with Tobii Studio, and the music was displayed at a size comparable to a paper score. The slides were designed to display the excerpts in alphabetical order so that a voice was added on each successive excerpt.

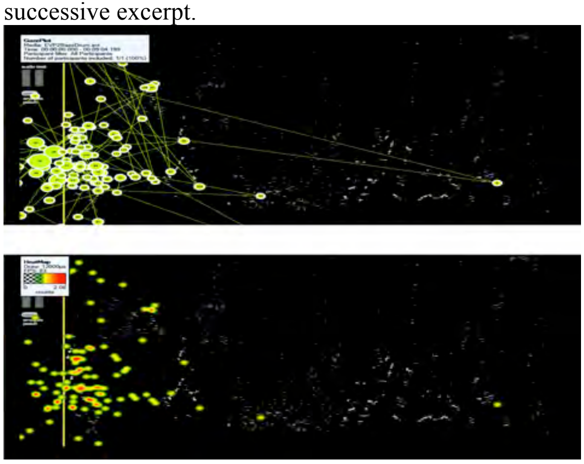

Figure 5. Gaze Plot (above) and Heat Map (below) of a Sonographic Score produced by the X2-30 eye-tracker.

2. CONTEXT OF RESEARCH

The exact nature of the processes involved in music-reading are still a matter of debate. One of the more persuasive models, proposed by Kinsler and Carpenter comprises a tripartite process "encoder, processor, executive" (1995: 1455). The pace at which visual symbols are encoded and processed is a key issue bearing the effectiveness of a screen score. This model proposes that the sight-reading process begins with encoding through preliminary transformation of the patterns into neural activity by retinal and central mechanisms, which is processed through the interpretation of musical symbols (traditional or otherwise), and that information about pitch and duration is finally transformed into appropriate patterns of commands to the executant muscles.

Eye-tracking studies focus at the surface of the encoding phase of this process by observing where, in what order and for what duration the eye is focusing, and moving. Performers visually acquire music notation with a combination of fixation upon graphical features and rapid repositioning of the eye (saccades). Sight-reading studies of traditional music are in agreement that fixation durations, extracting information from the score, fall within the range of 200-400 ms (Goolsby 1994a; 1994b, Truitt et al. 1997, Waters et al. 1997, Waters and Underwood 1998, and Gilman and Underwood 2003). The durations of saccades between fixations while sight-reading fall within in the tens of milliseconds (Gilman and Underwood 203:221).

Each fixation takes in a region termed the gaze frame or perceptual span which has been demonstrated to extend for "approximately 1 measure right of the fixation point" (Truitt et al. 1997) (equivalent to 3-5cm). Contrary to popular belief, the extent to which performers read ahead of their execution, termed the Eye-hand Span is relatively small, being between approximately 2 and 4 beats (Rayner & Pollatsek, 1997). Even in skilled readers visual processing of notation is not very far ahead of the hands and the actual position in the score (Gunter et al 2003:742).

The extremely short durations of some eye-movements (in the range of 25-30 milliseconds (Rayner, 1978)) meant that early systems using film (Weaver 1943) were unable to capture subtle eye movements (Goolsby, 1994a). Since the mid 1970's, notably the release and availability of the home computer, technology has allowed the revival of eye-tracking research. Rayner (1998: 372) identified this as the 'third era' of research that enabled the collection of more accurate data and larger amounts of data. Goolsby (1994a)'s study is said to have established modern musical eye-tracking practice because it was the first to utilise a dedicated eye-tracking device which measured the horizontal and vertical eye position every millisecond (Goolsby, 1994a; Madell & Hébert, 2008). Contemporary systems including the one used in these studies generally use the reflection of infrared light off the cornea to determine the angle of the eyes.

The applicability of the significant literature exploring the mechanism of sight-reading of traditional musical notation is limited in a number of ways. The tasks in many sight-reading studies involve quite simple musical examples (especially in comparison to the works of many composers) and because of the ease of collecting accurate data from MIDI keyboards, sight-reading studies have typically focused upon keyboard players, and have therefore not taken into account variation in the performance of instrumentalists who must place musical notation at a significant distance: a keyboard player might typically read from a score at a distance of 50 cm, whereas a percussionist may need to place a score 150 cm away in order to allow for a large instrument or set up.

Very few studies specifically address the issue of reading music from screens. In 1997 Picking compared a number of presentation styles for musical notation including traditional paper-based medium, as well as its screen-based counterpart however the study used bit-mapped notation presented in hypercard stacks, now many technological generations of out of date, and was undertaken in an era when smart phone and tablet literacy was not ubiquitous.

Importantly there are currently no studies of reading of nonstandard musical notation. This is perhaps due to its degree of specialisation involved, as well as the idiosyncratic nature and variety of graphic notations. Similarly the effect of performer restraint via a click-track had also not been studied.

3. INDICTATIVE RESULTS

Both experiments were designed as pilot studies to identify issues for further investigation. The reasons that neither study was able to attain fully valid results are discussed in section 4. For this reason only the raw data and its potential implications are discussed.

3.1 Vickery: Screening the Score

The graphs in figures 6-11 show a multicoloured band representing the average result for each participant for each case and then a black bar representing the average result for that case.

Average fixation durations were compared between *notation* types and found to increase in the following order: traditional notation 647ms, sonographic representation 681ms, non-semantic graphical notation 803ms, text 965ms and semantic graphical notation 2604ms (Fig. 6.).

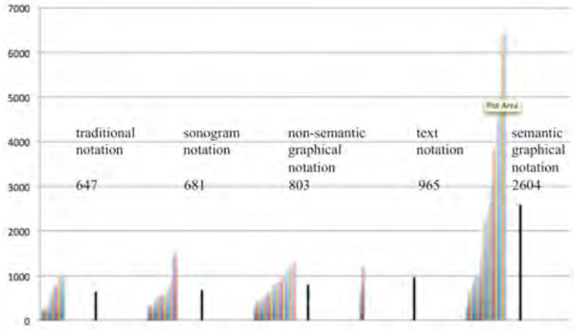

Figure 6. Average fixation length (ms) versus notation type.

These results may reflect that notation that requires the performer to continuously follow the evolution of graphical shapes require greater fixation durations and that graphical shapes with specific semantic meanings require greater fixation durations than graphical shapes that allow greater interpretation by the performer. Average fixation durations were compared between score *presentation* types and found to increase in the following manner: traditional score 566ms, 2D scrolling 819ms, rhizomatic 823ms, scatter display 893ms and scrolling score 1560ms (Fig. 7.).

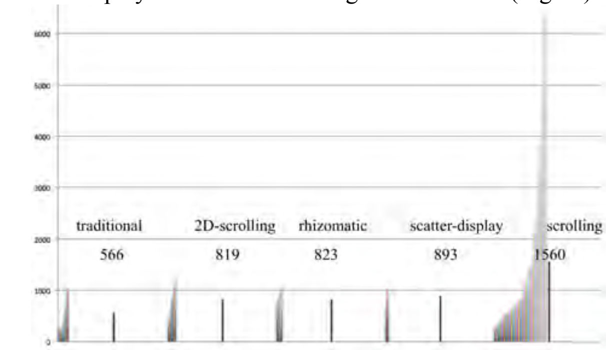

Figure 7. Average fixation length (ms) versus presentation type.

However, the range between the lowest and highest of fixation lengths in each category (staff 785ms, 2D scrolling 804ms, rhizomatic 279ms, scatter display 400ms, scrolling 6147ms) suggest that notation type may play a more important role than presentation type in determining fixation length.

Saccades, the movements of the eye between fixations, were also examined. Average saccade durations were compared between *notation* types and found to increase in the following manner: semantic graphical notation 86.5ms, text 99ms, staff notation 108.25ms, sonographic representation 120.8ms and non-semantic graphical notation 122.1ms (Fig. 8.). When average saccade durations were compared between score *presentation* types it was found that average fixation durations increased in the following manner: rhizomatic 92ms, scatter display 96ms, 2D scrolling 97ms, traditional score 108ms, and scrolling score 113ms (Fig. 9.). Again the data suggests that notation type has a more important effect upon saccade duration than presentation type.

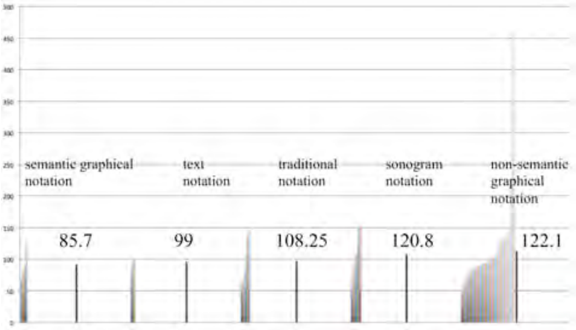

Figure 8. Average saccade length (ms) versus notation type.

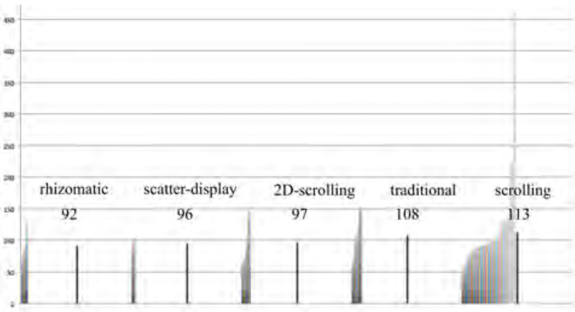

Figure 9. Average saccade length (ms) versus presentation type.

The ratio (%) of the total number of fixations versus the total number of saccades was also examined. Amongst *notation* types the ratio increased in the following manner: traditional 72.2%, non-semantic 72.7%, spectrogram 78.9%, semantic 83.9%, text 85.7% (Fig. 10.). Amongst *presentation* types the ratios found were: traditional 66%, 2D scrolling non-semantic 76%, rhizomatic 78%, scrolling 79% and scatter display 82%, (Fig. 11.). The results showed more fixation activity in non-traditional scores and the highest activity in scatter displays in which the notation was changing in multiple sectors of the screen.

Significant variation was found in the width of participants scan patterns: the broadest gaze point width was for a Sonogram-style notation work, with a standard deviation of 272 px (X axis) and 224 px (Y axis); the narrowest was for a the semantic graphical notation work, with a standard deviation of 100 px (X axis) and 56 px (Y axis) (Figure 13.).

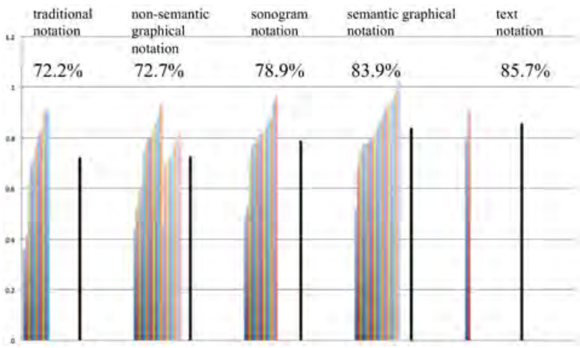

Figure 10. Fixation number to saccade number ratio (%) versus notation type.

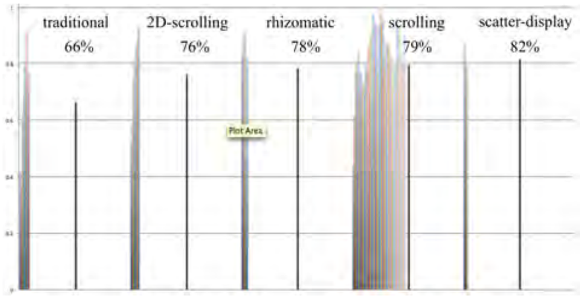

Figure 11. Fixation number to saccade number ratio versus presentation type.

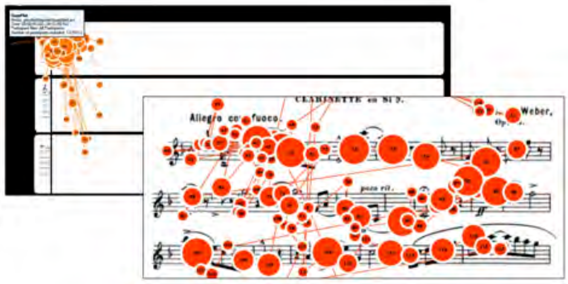

Figure 12. Comparison of Scrolling score gaze-plot in which the eye is restricted to a small area (above) and traditional notation gaze plot (below) in which the eye scans across staves.

3.2 Goh: Sight-Reading Polyphonic Musical Textures

The most significant effect was seen in reading with a click-track, which decreased the number of fixations per second and eye-hand span, and increased the average fixations duration. This was attributed to a higher cognitive load involved in performing with a click-track. Musical texture also interacted with eye movements. Polyphonic excerpts containing more voices correlated with a higher number of fixations per second and lower eye-hand span and fixation durations. Unusual musical features accounted for abnormally large numbers of fixations and time spent looking at areas of the score.

The principle findings of the results were that eye movement patterns were affected by the number of voices in the excerpt and the click-track condition, but not the number of encounters. Furthermore, the fixation positions were influenced by features of the music and notation.

The click-track condition performances were markedly slower and more difficult for the participants than the comfortable condition. This was reflected in the average fixation duration, number of fixations per second and eye-hand spans of the excerpts. The average number of fixations and fixation duration were notably higher when participants performed with a click-track. However, when accounting for the time it took to perform the excerpts the number of fixations per second of performance was actually lower in the click-track condition. In addition, the eye-hand span was markedly smaller when playing with the click-track (1.36 beats compared to 1.86 when played comfortably.)

Excerpts containing more voices had slightly more average fixations, more progressive fixations and shorter average fixation durations than excerpts with less voices.

The number of fixations per second was positively correlated with the number of voices in each score (Fig. 13).

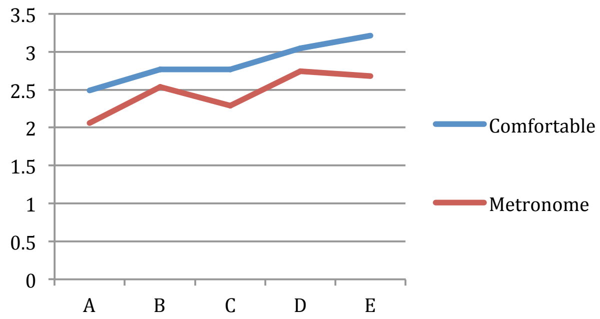

Figure 13. Number of Fixations per Second.

Fixation duration was slightly negatively correlated with the number of voices of each excerpt (Fig. 14.) The eye-hand span was negatively correlated to the number of voices in each excerpt in the comfortable condition, but not in the click-track condition.

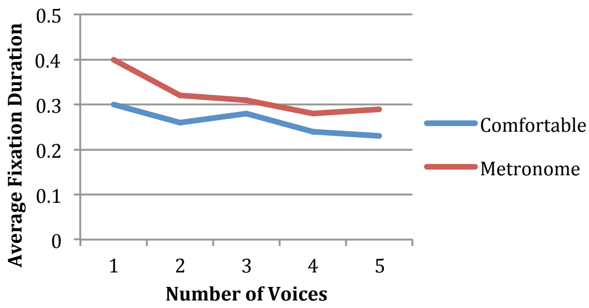

Figure 14. Average Fixation Duration Across Excerpts.

In general, fixations were made within a stave but not on note-heads themselves. Fixations tended to occur between staves in the two-handed excerpts, and underneath the bass stave in the one-handed excerpt. Subjects tended to read in a horizontal fashion, a trend reported in Weaver (1943) for contrapuntal music. Overall, the fixation positions suggest that musicians rely heavily upon perceptual span and peripheral vision. This could be an element that is important in chunking behaviour.

In Excerpt A, which only involved the left hand, fixations tended to occur slightly below the note heads themselves (Fig. 15.). Excerpt E shows a pattern of focus between the treble and bass staves rather than on any one line itself (Fig. 16.). Although this could be due to a calibration error, a more likely explanation is that peripheral vision was important in reading these excerpts. Peripheral vision and the perceptual span have an important role in chunking groups of notes, rather than focusing on one note at a time. Other studies, such as Gilman and Underwood (2003) and Truitt et al. (1997), have also commented on the proportion of blank space that was fixated upon. Although these studies did not specify where the 'blank space' was, the current study has also found a focus on the empty space between staves and underneath the staff in the one-handed excerpt.

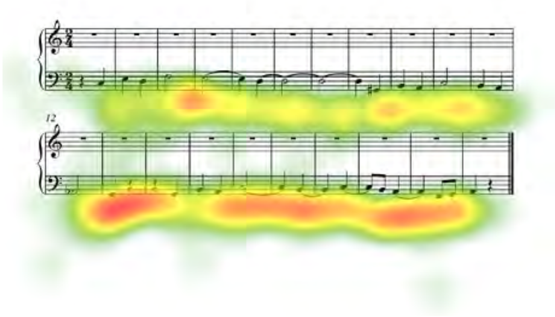

Figure 15. Heatmap: Excerpt A, Number of Fixations.

Another striking feature of the fixation positions was the patterns created by sequential fixations. Although the fixations were rarely on the notes themselves, especially those in the treble clef, the eyes tended to move according to the main features of each bar of music.

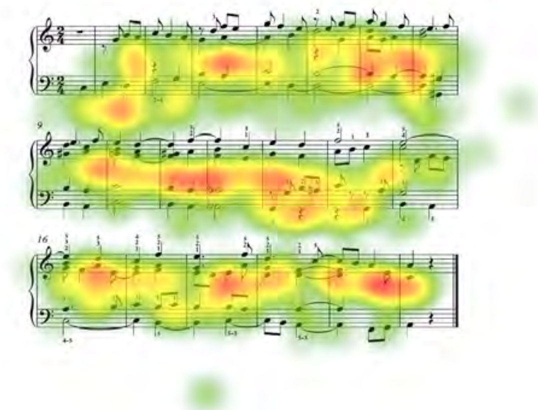

Figure 16. Heatmap: Excerpt E, Number of Fixations.

4. LIMITATIONS OF THE EXPERIMENTS

Vickery's study was deliberately broad in order to identify issues for further investigation. While it was arguably successful in this regard there were too great a number of independent variables in the study to make conclusive reliable findings. Sight-reading studies suggest that the baseline for the duration of fixations on musical symbols is between 200-400 ms per fixation (Goolsby 1994a; 1994b, Truitt et al. 1997, Waters et al.

1997, Waters and Underwood 1998, and Gilman and Underwood 2003). The average fixation length of the participants in this study for traditional notation was 556ms with a range between 272 and 1057ms. Average fixation durations amongst participants in this study ranged from 450ms (player 1) to 2089ms (player 8). It is not clear whether this accurately reflects the reading style of the participants or the difficulties of standardising data acquisition across such a wide range of instruments. For example Infrared scanning may have been affected by the distance between the performer and the apparatus and/or the reflectivity of some instruments.

Goolsby (1994) has suggested that notational complexity is a strong determinant in predicting eye-movement in music) and this effect was not compensated for. Future work should include the development of means of evaluating notation complexity, in order to allow accurate comparison between notation types. Although Vickery has speculated on the role of scroll speed in scrolling scores (Vickery 2014), this issue was also not compensated for.

Susan George claims that "in the most general sense the score is comprised of units. Sometimes these units are primitive elements themselves, and sometimes they are composite so that the primitive elements must be extracted from the units themselves" (2004: 157). In website analysis these are referred to as "semantically meaningful units", and used to measure the interaction of a reader with the screened page. Such an approach (Fig. 17), might usefully be adopted in the study of performers' interaction with screenscores, as a means of measuring the number of semantically meaningful units that a performer is able to capture in a single fixation and the rate at which they can be captured.

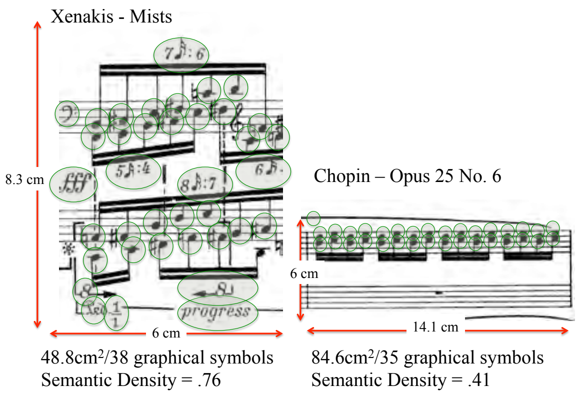

Figure 17. Calculating Semantic Density by measuring graphical symbols per cm^2.

In Goh's study the small number of participants restricts the validity of the results. The most unusual finding – that the subjects did not generally improve playing performance across encounters - is unusual, as it is contrary to the gradual improvement in perceptual span across performances reported by Burman and Booth (2009) and the improvement in musical performance reported by Goolsby (1994a). A possible explanation is that the excerpts were too simple for the participants for any improvement to be recorded, or that the three performances analysed were insufficient to record improvements in performance. Future studies should focus on a more longitudinal approach and a greater variety of musical textures to observe the effects of rehearsal on music reading performance.

5. DISCUSSION

Given the limitations upon the validity of the findings suggested above, the results suggest that:

1. a score's Notation type has a greater influence on eye-movement in music reading than Presentation type.

2. participants fixation length was shortest when reading traditional notation, increased through more interpretive forms of non-semantic graphical notation to a peak when reading semantic graphical notation. This may be the result of the fact traditional notation consists of a symbolic language that is read in discrete chunks, whereas graphical notation implies more continuous monitoring of notational changes and that semantic graphical notation, the most precise form of graphical notation demands the most continuous attention and therefore has the longest fixation lengths.

3. Semantic graphical notation had the shortest average saccade time and non-semantic graphical notation the longest. Similarly, the ratio of "unclassified gaze event types" was shortest for semantic and longest of non-semantic forms of notation and the scan pattern width was also smallest for semantic notation. These findings support the notion that semantic notation requires the eye to remain relatively fixed in a confined area, whereas non-semantic graphical notation encourages a more open, interrogative approach to reading.

4. performance with a click-track, decreased the number of fixations per second and the eye-hand span. This was thought to be an effect of the high cognitive load involved in playing with a click-track. There are many other factors that can influence musical sight-reading like psychomotor speed and general cognitive skills (Kopiez & In Lee, 2008), and so there may be other explanations for this phenomenon.

5. A correlation was observed between musical texture and eye movements as polyphonic excerpts with higher numbers of voices tended to produce a smaller eye-hand span and more fixations per second of performance, and these fixations lasted a shorter duration. As no other studies have focussed on polyphonic textures, this investigation is the first to report this effect.

6. CONCLUSION

The (slightly miraculous) process by which performers encode, process, execute notated musical structures in an extremely precise temporal frame may never be fully understood. These studies aimed to shed light on some aspects of critical importance to composers seeking to coordinate performers through computer control. They examined the effects upon eye-movements of a range of notation and presentation types in score-reading and the sight-reading of polyphonic musical textures with a click-track beat through several encounters. A number of interesting results were recorded and require further isolated study in order to develop reliable and valid data. Additional studies will need to be conducted to see if these findings apply to a wider population and to make valid generalisations from the results.

Although the findings of Goh's study agree with the current sight-reading literature, future studies should incorporate multiple tempo conditions to fully examine the effect that click-track playing has on the cognitive load.

As eye-tracking technology becomes more advanced and accessible, music sight-reading studies will reach beyond the basic phenomena in this fascinating field. This study has explored the effects forms of notation and score presentation, polyphonic textures, multiple encounters, cognitive load and musical features on sight-reading. Future studies should take all aspects of the sight-reading context into consideration in order to explain eye movement data. It is hoped that such studies will provide solid and beneficial data for the increasing number of composers working with coordination of performers through computer via notation on screen and/or click-track.

7. REFERENCES

Burman, D. D., & Booth, J. R. 2009. Music rehearsal increases the perceptual span for notation.

George, S. E. 2004. Visual Perception of Music Notation: On-line and Off-line Recognition. IRM Press: Hersey, PA.

Gilman, E., & Underwood, G. 2003. Restricting the field of view to investigate the perceptual spans of pianists. Visual Cognition, 10(2), 201-232.

Goolsby, T. W. 1994a. Eye movement in music reading: Effects of reading ability, notational complexity, and encounters. Music Perception, 77-96.

Goolsby, T. W. 1994b. Profiles of processing: Eye movements during sightreading. Music Perception, 97-123.

Gunter, T. C., B.-H. Schmidt, et al. 2003. Let's face the music: A behavioral and electrophysiological exploration of score reading. Psychophysiology 40: 742–751.

Kinsler, V., & Carpenter, R. 1995. Saccadic eye movements while reading music. Vision Research, 35(10), 1447-1458.

Kopiez, R., & In Lee, J. 2008. Towards a general model of skills involved in sight reading music. Music Education Research, 10(1), 41-62.

Madell, J., & Hébert, S. 2008. Eye Movements and Music Reading: Where Do We Look Next? Music Perception: An Interdisciplinary Journal, 26(2), 157-170. doi: 10.1525/mp.2008.26.2.157

Picking, R. 1997. Reading Music From Screens Vs Paper, Behaviour & Information Technology, 162:72-78

Rayner, K. 1978. Eye movements in reading and information processing. Psychological Bulletin, 85(3), 618.

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. 2009. Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8), 1457-1506. doi: 10.1080/17470210902816461

Rayner, K., & Pollatsek, A. 1997. Eye Movements, the Eye-Hand Span, and the Perceptual Span During Sight-Reading of Music. Current Directions in Psychological Science, 6(2), 49-53. doi: 10.1111/1467-8721.ep11512647

Tobii X2-30 Eye Tracker User's Manual. 2014 Retrieved 1 November, 2014, from http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_X2-30_EyeTrackerUserManual_WEB.pdf

Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The Perceptual Span and the Eye-Hand Span in Sight Reading Music. Visual Cognition, 4(2), 143-161. doi: 10.1080/713756756

Vickery, L. 2014. The Limitations of Representing Sound and Notation on Screen. Organised Sound 19(3). Cambridge University Press.

Waters, A. J., & Underwood, G. 1998. Eye Movements in a Simple Music Reading Task: A Study of Expert and Novice Musicians. Psychology of Music, 26(1), 46-60. doi: 10.1177/0305735698261005

Waters, A. J., Underwood, G., & Findlay, J. M. 1997. Studying expertise in music reading: Use of a pattern-matching paradigm. Perception & psychophysics, 59(4), 477-488.

Weaver, H. E. 1943. Studies of ocular behavior in music reading. Psychological Monographs, 55(1), i-50. doi: 10.1037/h0093537