Eye-tracking Film Music

Miguel Mera and Simone Stumpf

ABSTRACT

Film music scholars, composers, directors and audiences have always implicitly believed that music can help determine the focus of an audience's visual attention, but researchers have not as yet been able to prove this empirically. Eye-tracking research—the process of measuring either the point of gaze or the motion of the eyes—has grown exponentially in recent years. This paper reports on a foundational, empirical eye-tracking study that examined the effects of contextual musical attributes on visual attention, emotion and user experience during exploration tasks in moving images. Our results show that music is able to direct how we see by switching attention to target foci more quickly as well as lengthening fixations, and that music can also encourage greater exploration of visual scenes outside targets. Our work contributes the first step in understanding how music shapes visual attention using eye-tracking techniques. We encourage wider adoption of this approach which has the potential to enhance understanding of the complex processes of audiovisual perception in action.

INTRODUCTION

How does music help focus visual attention in the film experience? Music is often seen as influencing emotion but relatively little research has been done on the impact of music on visual attention. The Congruence-Associationist Framework, first proposed by Marshall and Cohen (1988) and subsequently extended and developed by Cohen, aimed to explain the role that music plays in film.[1] Congruence refers to the structural properties of music that overlap with structures in the visual domain. Association refers to the meanings that music can bring to mind. In essence, Cohen proposed that when audiences experience film, they generate a conscious working narrative. The working narrative results from top-down inferences that best match outputs of a bottom-up analysis of various audiovisual channels. Within the Framework, Cohen put forward the idea that music can sometimes determine the focus of visual attention: "When the auditory information and visual information are structurally congruent, the visually congruent information becomes the figure, the visual focus of attention".[2] To a certain extent, this theorization simply repeats what film music scholars, composers, directors and audiences always implicitly believed to be true. However, the purpose of empirical research is to make explicit what has previously been implicit; to develop experiments that prove assumptions are either correct or need further refinement. Filmmakers and composers employ deliberate strategies of narrative focus through the use of music, but do they actually work? After more than twenty-five years of empirical research in this area, we are only beginning to comprehend some of the ways in which music and moving images interact to create meaning, and there is still much more to be done before we can understand the complexity of these interrelationships. A major problem has been that the vast majority of research has, as described by Lipscomb, employed “highly reductionist stimuli” rather than “more ecologically valid film and animations.”[3] Lipscomb referred to numerous experiments exploring, for example, relationships between single-tone sequences or simple single-object animations, where neither the visual or aural stimuli, or their combination, reflected real-world experiences. The challenge for experimental design is that real-world experiences are extraordinarily intricate making it very difficult to unpick meaningful, discrete, and reliable metrics. Happily, exciting developments and innovative approaches are emerging in this field.[4] Yet, given that the emphasis is on listening and looking, astonishingly little work has been done on the ways in which music might affect visual attention using eye-tracking methodologies.[5] Research in this area is underdeveloped in relation to both music and cinematic moving images, and is entirely absent when the two media are combined.

Principally located within the fields of psychology and human computer interaction, eye-tracking research has grown exponentially in recent years. It is the process of measuring either the point of gaze or the motion of the eyes. Humans do not look at a scene in steady or rigid ways. Our eyes dart from place to place, locating interesting parts of the scene, bringing them into high resolution so that we can see in greater detail the central direction of the gaze. When scanning a scene the eyes make fast jumpy movements, stop several times, and move very quickly between each gaze point which helps us to develop a three-dimensional, mental ‘map’. The jerky movements are known as saccades and are the fastest movements produced by the human body, normally taking thirty to eighty milliseconds to complete. When the eye lingers on a particular object it is known as a fixation and can last anywhere from eighty milliseconds up to several seconds. Fixations and saccades are the most frequently measured and reported events in eye-tracking studies and are further nuanced by detailed examinations of position, movement, numerosity and latency. Effective summaries of the historical precedents, technical approaches, and the current state of research in the field are provided by Duchowski, Land and Tatler, Liversedge et al., and Holmqvist et al.[6]

Eye-tracking research exploring how users interact with *moving images* is still in its infancy. Marchant et al. argue that this is because “the study of visual perception of dynamic scenes is inherently more difficult as the objects that are being fixated upon are in constant motion.” It is, therefore, more challenging to relate the position of the fixation to any “particular object or compare multiple viewing experiences without a painstaking visual analysis of the scene.”[7] Previous studies have shown that spectators typically focus their visual attention onto a particular common area.[8] This seems to be influenced by both low level sensory triggers such as contrast and motion, and ‘top-down' conscious control mechanisms such as attraction to the human face. Music may be able to focus attention in multisensory moving images sequences, as Cohen suggested, because the gestural patterns and/or emotional associations generated by music attach themselves automatically to the visual focus of attention, but researchers have not as yet been able to prove this empirically.

Music-related eye-tracking studies have focused on sight-reading of western musical notation by performing musicians,[9] the arousing or distracting effects of music tempo on multi-attribute decision-making,[10] and the effects of background music on reading.[11] Recent research employing eye-tracking data suggests that 'pleasant music' can increase visual attention in stroke patients when exploring *static images*.[12] This work assumes that music has an effect on mood, and mood in turn has an effect on the cognitive and perceptual system. Within the field of cognitive psychology there has been a great deal of empirically oriented research on music and emotion illustrating how and why music may evoke emotion in a listener, and the effects of emotional arousal and valence on visual perception.[13] However, in relation to film music it has been demonstrated that arousal identified within self-reported indices of emotion is not always matched in the attendant physiology, and that the effect of music on emotional responses to viewing films is especially complex.[14]

Some research has been conducted studying the effect of music on visual attention in *dynamic real-life visual situations*, although it has not generally used eye-tracking methodologies. Much of the research in this vein has been conducted in the automobile industry and has examined the effect of 'background' music on driving performance.[15] This research has focused on the comparison of musical parameters, such as tempo and loudness, and their effect on driving rather than the direct impact on visual attention. Furthermore, some of this research, which supposedly isolates musical parameters, is often fundamentally flawed from a musicological perspective. In Brodsky's experiment, for example: "Only neutral sounding instrumental pieces were considered. All prospective stimuli characterized a fusion music incorporating pop, rock, jazz, blues, funk, and country genres."[16] The idea of the fusion of these multifarious genres is already problematic, especially when the pieces of music that were actually used in the study are examined. It is clearly ludicrous to suggest that smooth jazz tracks by George Benson and electronic dance music created by DJ Jurgen are equally 'neutral' except for differences in tempo. Aside from the enormous variance in the means of production, structure, contour, instrumentation, spatialization, harmonic rhythm, timbre, texture and intensity, the tracks by DJ Jurgen were specifically designed for the purpose of dancing, to encourage reactions in the body. But because it is a relatively easy parameter to measure, this study claimed that "the *tempo* [our italics] of background music consistently affected both simulated driving speed and perceived speed estimate."[17] We are quite sure that the different music in this study affected driving speed, but it is far too simplistic to suggest that this was caused only by differences in tempo, or that this was the only factor that affected the perception of speed. It would not be difficult, for example, to think of pieces in the same genre with exactly the same measured metronome mark where the perception of speed is completely different. This brief digression is relevant to our study because it highlights both the dangers of reductive attitudes and the need for rigorous interdisciplinary collaboration in experimental design.

Our work is concerned with music's effects on visual attention in moving image sequences. Whereas previous research has frequently attempted to deconstruct music into discrete parameters, we take a different approach: *music* rarely isolates any component elements, and once these are removed from their context only dismembered *sounds* remain. Music is multidimensional, "like a complex chemical catalyst to our brain."[18] We use the term *musical attributes* to encompass a range of interdependent musical features that include rhythm, harmony, timbre, intensity, and gesture. These musical attributes cannot be divorced from the macro-musical structure or characteristics of a piece; they work symbiotically to create music. Likewise, we used an actual narrative film sequence, rather than an animation of abstract geometric shapes. Though this made data collection much more difficult, the encounter—complete with a complex set of inherent grammatical constructs relating to editing, lighting, *mise en scène,* character relationships, etc.—was closer to the real world experience of film. It is our contention that further refinement of this work should continue in the same direction, seeking to reconcile the obvious challenges it presents.

To our knowledge our work is the first that looks in detail at the effects of contextual musical attributes on visual attention, emotion and user experience during challenging exploration tasks in moving images. To investigate these aspects, we conducted an empirical study with eighteen participants which collected eye-tracking, emotion and user experience data. We employed a between-subject design where six participants each watched the same moving images clip but experienced different versions of accompaniment. We had three main research questions:

- RQ1: How do musical attributes affect users in terms of emotional response while watching moving images?
- RQ2: How do musical attributes impact users in terms of perceived experience?
- RQ3: How do musical attributes influence users' visual attention?

Our work contributes the first step in a fundamental and detailed understanding of how music influences visual attention and could have important lessons for the design of music that supports visual exploration tasks in complex dynamic scenes. The findings in our 'pilot study' point to the potentially significant impact of music in encouraging participants to fixate on visual foci, and suggest that further expansion of this research is necessary. As a result of our initial results, therefore, we make two calls to arms. The first is for wider use of eye-tracking as a methodology in sound and multimedia studies. As one of a combination of empirical methodologies it has the potential to enhance our understanding of the complex processes of perception in action. The second, echoing Tan et al., is a call for genuine interdisciplinary research so that exciting, applicable work can be generated and both scientific and musical rigor can be maintained.[19]

EXPERIMENT SET-UP

We conducted an empirical between-participant study which investigated how users respond to music and moving images using eye-tracking data and questionnaires. Our team included a human-computer interaction researcher, and a musicologist-composer to provide their respective expertise in this inter-disciplinary research project.

We recruited eighteen participants to take part in our study via university email lists, comprising seven males and eleven females with an average age of 26.9. A background questionnaire revealed that the majority of our participants watched videos several times a week, went to the cinema on a monthly basis and listened to music more or less every day. Ten participants were undergraduate or Masters students, four research or academic staff, and four gave other occupations. Six participants had some background in making videos/films and twelve had some background in music.

We used a 2:38-minute clip from the motion picture *The Artist* (2011). This clip, starting at thirty-five minutes and twenty-six seconds into the movie, depicts the chance encounter of the two main characters, the established movie star George and the rising star Peppy, on some busy studio stairs. It consists of thirty-one shots, the majority of which are over-the-shoulder mid-shots or medium close-ups with a turn-taking focus on one of the two main characters (Figure 1). Two scenes break this pattern, and were of particular interest to us because they allowed us to investigate the effect of music on visual attention more clearly. Scene *stairs_both* (Figure 2), at 37.14 seconds into the clip lasting 8.8. seconds, is a wide-shot of the two main characters with other people walking past them. This scene was interesting because it introduced competing visual foci in the form of a good deal of passing traffic on the stairs. We hypothesized that this would cause a reduced focus on the main characters.

Figure 1. A mid-shot scene showing George in a conversation with Peppy (over-the-shoulder shot).

Figure 2. Scene *stairs_both*. A wide-shot scene of George and Peppy on the stairs; note passing traffic which provide competing visual foci.

Figure 3. A mid-shot of Peppy immediately followed by Figure 4

Figure 4. Scene *stairs_george*. Extreme wide shot of George on the stairs dramatically shifting the focus of attention.

Scene *stairs_george* (Figure 4), at 137.19 seconds lasting till the end of the clip, shows an extreme wide-shot of George on the stairs after Peppy has departed, then slowly walking down the stairs while passing traffic enters. This scene presents a particularly challenging visual exploration task in a moving images setting. In the midshot of Peppy (Figure 3) which appears just before the cut to *stairs_george* (Figure 4) we would normally expect the eyes to focus on the face at the center-top of the screen. When Peppy exits she quickly travels towards the right of the screen and we would normally expect the eyes to both anticipate and follow this trajectory in smooth pursuit. To locate George on the next cut the eyes then need to refocus on the center of the screen where he is standing motionless. Yet at the same time there is considerable movement around the staircase which draws attention away from George because the eyes are automatically attracted to motion. We hypothesized that participants would find it especially difficult to focus their gaze on George at the start of the scene and would also be distracted by the competing visual foci.

We chose this movie since it is a modern *silent* movie. Hence we did not need to remove any sound effects or dialogue and could manipulate the music in isolation. Using the same moving images clip, we constructed three versions:

- A *silent* version without any music soundtrack (a control version)
- A *'distracting'* music version
- A *'focusing'* music version

Since this was a pilot study, we designed only two music versions which allowed us to examine different combinations of focusing or distracting musical attributes. Our aim was to understand general tendencies by using contrasting yet stylistically similar pieces of music that could be altered to highlight or disregard particular on-screen moments. A larger study would be able to employ a much more complex multi-factorial experimental design.

The 'distracting' version is an adaptation of American composer Leroy Anderson's orchestral piece *Fiddle Faddle*. The bright and exciting orchestral music is characterized by rapid sixteenth-note movement in 2/2 time and a high-energy performance. An extremely fast half-note pulse of 170 BPM is sustained in the recording we employed.[20] Furthermore, the audio was edited to maintain the lively sixteenth-note movement throughout by removing any moments of structural, textural or rhythmic repose (such as those provided by the central pizzicato section in measures 70–85), and to ensure that there were several repetitions of the principal thematic material contributing to an insistent character.[21] It was our expectation that the relentless, almost manic character of the edited piece would encourage participants' visual attention to shift frequently from one area to another over the whole sequence, but especially on the two stairs shots where there are competing visual foci in different parts of the screen. The perpetual motion of the music with a lack of textural, timbral or rhythmic difference essentially 'ignored' the two stairs shots. The constant level of intensity in the music was designed to run parallel to the moving images, discouraging congruent audiovisual connections with particular characters or locations within the scene.

Conversely, the 'focusing' version aimed to draw attention to the two stairs scenes through specific shifts in musical attributes and directive gestural devices. We used music composed by Ludovic Bource for the film *The Artist* itself, but not the score that was originally created for the scene as it did not attempt specifically to highlight the wide shots. We adapted music from tracks six and seven of the soundtrack album. Starting at fifty-two seconds into 'Fantaisie D'amour' and leading directly into the first fifteen seconds of 'Waltz for Peppy', the music begins in 2/2 with a half note pulse of 154 BPM. This music is also orchestral and has a playful character but is much more relaxed than the 'distracting' version. The 'focusing' version emphasizes the interplay between the central characters with melodic, textural, and orchestrational materials that change with greater fluidly to match the narrative dynamics of the scene.

Noticeable shifts in the music for the focusing version were synchronized with the two wide-shots on the stairs. On the cut to the *stairs_both* shot the music changed from a swung to a straight performance. Although swing remains a complex and elusive term within musicology, it is most simply understood as a rhythmic performance style, primarily associated with jazz, resulting from the conflict between a fixed pulse and the wide variety of accent and expressive nuance that a performer plays against it. On the wide-shot the music shifted from the relaxed freedom of the swung performance to the rigidity of a straight, metronomically precise performance. Furthermore, a solo violin playing long, high harmonics ascending the first, third and fifth of the dominant chord was added to the music at this point.[22] This addition aimed to provide a noticeable textural difference without also seeming like an alien stylistic presence. We hoped that the changes in the music here would signal subtle narrative connection and encourage participants to seek out the two main characters.

On the cut to the *stairs_george* scene the music was synchronized so that a quick gestural reduction in texture would occur. The shift from the large orchestra to a single high pedal note played by the violins is a classic 'focusing' film music device, akin to a camera zoom. It is also significant that at this point the piano marks out the change from duple to triple meter as the 'Waltz For Peppy' is heard under the high pedal note. The shift at this moment, therefore, results in a clear change in meter, pace, intensity, texture and timbre.

Following the Congruence-Associationist Framework, it was our expectation that the musical focusing devices present in the two scenes would encourage participants to concentrate their visual attention on the central characters and would, to some extent, reduce the participants' tendency to search for focal points elsewhere in the scene. The visual shot/reverse-shot dialogue between the two main characters had been closely matched by similar musical interplay. It seemed reasonable, therefore, to hypothesize that the two congruent musical shifts in the focusing version would direct participants towards the characters. Furthermore, we expected the *stairs_george* scene music to be much more directive because the musical shift at this point was bolder, because high long notes associated with a cut to a wide shot had already been partially established in *stairs_both*, and because this scene appeared at the end of the sequence so that music and image had more time to develop narrative connections.[23]

Procedure

In order to avoid learning effects yet still provide the ability of within-participant comparisons, we assigned participants to three different session set-ups. In session set-up A, participants were exposed to the silent version first, followed by the distracting version; during B, they watched the distracting version first and then the focusing version; in session set-up C, the focusing version was first, and the distracting version second. The participants were randomly assigned to one of three session set-ups, however we ensured there were six participants in each set-up.

While each of these participants saw two versions of the same film clip, our analysis of emotional responses and eye-tracking used data only from the first version they saw. We decided that participants could not 'unsee' the first film clip and therefore using the emotional and gaze data of the second version they saw in analysis would have introduced a learning effect. However, contrasting two versions was very useful when investigating the user experience as it clarified certain qualitative differences.

A facilitator ran each session with individual participants, lasting thirty minutes. The session started with the participant filling in a background questionnaire and the calibration of the eye-tracking equipment and headphone sound levels. As per their condition assignment, the participants watched the first version, followed by a post-task questionnaire. They then watched the second version and afterwards filled out a post-task questionnaire and a post-session questionnaire.

Data Collection and Analysis

We employed a Tobii X60 eye tracker with Tobii Studio 3.2 software. Participants viewed the clips on a 19-inch display monitor with a resolution of 1280x1024, with the clips displaying in 640x480 actual size. The eye tracker measures the raw gaze points every 16 ms and groups them together into fixations. There are a variety of fixation filters to derive fixations from the grouping of the raw gaze data; a discussion of these details are beyond the scope of this paper.[24] We chose the Tobii default fixation filter for this project.

A common, intuitive approach to analyzing eye-tracking data is through visualizations, such as bee swarms, gaze plots, heat maps or clusters.[25] Unfortunately, none of these visualizations provide a detailed understanding of the focus of visual attention, especially in dynamic scenes. Hence, we needed to resort to eye-tracking metrics. These metrics are calculated by the software, based on fixations and areas-of-interest (AOIs). In our project, we defined three AOIs: the entire visual scene in the whole clip, the two main characters standing on the stairs in scene *stairs_both*, and George in *stairs_george*. For the latter two, we created dynamic AOIs covering the main characters' entire visible bodies. For practical reasons, we used simple rectangles tightly overlaid with the characters' movements as AOIs, otherwise each frame would have required drawing a new multi-point shape outline without gaining much more precision.

In our study we aimed to understand in detail *where* and *how* participants attended in the visual scene. Therefore, we analyzed how often they dwelled on the main characters, how quickly music was able to direct visual attention to the main target characters, how long they looked when they dwelled on a point as well as in total on either of the main characters, and the size of the visual scene to which participants attended. Most of these can be obtained from eye-tracking metrics provided by the software:

- Fixation count: number of all fixations in an AOI (N)
- Time to first fixation: how long it takes for a user to fixate on a AOI for the first time (seconds)
- First fixation duration: how long the user looks at the AOI during their first fixation (seconds)
- Fixation duration: the duration of each individual fixation in an AOI, also called fixation length (seconds)
- Total fixation duration: sum of the duration of all fixations in an AOI (seconds)

For the two scenes on the stairs, we were also able to obtain metrics for fixations *not* on the AOIs, i.e. when participants did not look at the main characters and looked elsewhere on the screen. To measure the size of the area to which participants attended, we introduced *fixation spread* as a novel measure; this metric is not provided by the Tobii Studio software. Fixation spread measures the average distance between all fixations and gives an indication of how tightly fixations are clustered; the greater the distance, the more fixations are spread out from each other. In effect, the fixation spread also indicates how much of the visual scene has been explored through fixations. We calculated the distances by first exporting the fixation data with their point coordinates, then constructing a matrix of distances between each fixation, and finally calculating the average over all these distances. Due to the data-intensive nature of these calculations, we only determined the fixation spreads for the *stairs_george* scene where we were particularly interested in the effects of music on an especially difficult visual exploration task.

We also collected subjective data from participants in the form of questionnaires. In order to investigate possible links between emotion and visual attention, we collected aspects of the participants' emotional experience in a post-task questionnaire, asking about how they felt watching the film clips. We elicited these responses using a two-scale version of the Self-Assessment Manikin (SAM, Figure 5), capturing the valence (positive/happy versus negative/unhappy) and arousal (high/excited versus low/relaxed) on a nine-point, non-textual scale.[26] The Self-Assessment Manikin is a well-established technique that gathers and measures affective responses to stimuli through a pictorial representation.[27] It thus avoids the problem of translating an essentially non-verbal response into semantically-equivalent words.

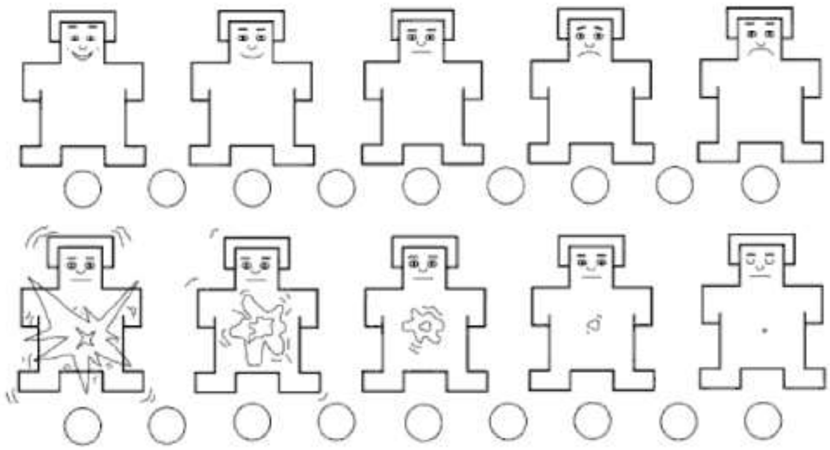

Figure 5. Self-Assessment Manikins for valence (top) and arousal (bottom).

We also asked participants to describe their experience overall, using a list of twenty-two suggested adjectives derived from the Microsoft Desirability Toolkit,[28] with the opportunity for participants to provide their own terms. We then probed their perceived experience of attention in the clip, whether the music helped focus their attention, and whether the music was suited to the clip, both on five-point scales, as well as open comments. In the post-session questionnaire, we simply asked participants to contrast their experiences of the first and second versions in their own words.

We analyzed our data both qualitatively and quantitatively. It should be noted that the sample of the population was relatively low and we have mainly aimed to understand trends in our data, which a study with a larger sample size could substantiate. Statistical analysis of eye-tracking data is also notoriously difficult because extensive manipulation of raw gaze data is required. To reiterate, when we analyzed the eye-tracking metrics, we only took participants' data after their first encounter with the clip with respect to a specific version, before they had learned what to expect which may have influenced their visual attention.

RESULTS

Emotion (RQ1)

It has previously been argued that visual attention is primarily affected by emotion and that emotion is affected by music.[29] In this chain of reasoning, music does not directly influence visual attention as it is mediated by a listener's emotional response to the music. Hence, our first research question addressed the connection between music and emotion. In our study we asked the participants to report their emotion after watching the versions through SAM, which captures valence and arousal.

Participants rated the emotional response to the different sequences fairly similarly (Figure 6). In terms of valence, participants ratings tended towards being positive (happy), however participants perceived the silent version as slightly more neutral than the music versions. In terms of arousal, the ratings were consistently in the neutral range for all versions. A Kruskal-Wallis test for both valence ($H(2,N=18)=1.80$, $p=0.407$) and arousal ($H(2,N=18)=0.11$, $p=0.945$) did not show any significant statistical differences in ratings between versions.[30] Surprisingly, according to the SAM responses in our study, music, as well as the absence of music, did not have an important influence on the reported emotions of participants. Since emotion has been assumed to be the main driver for visual attention we wondered whether music had any more direct effects on felt experience or actual visual attention.

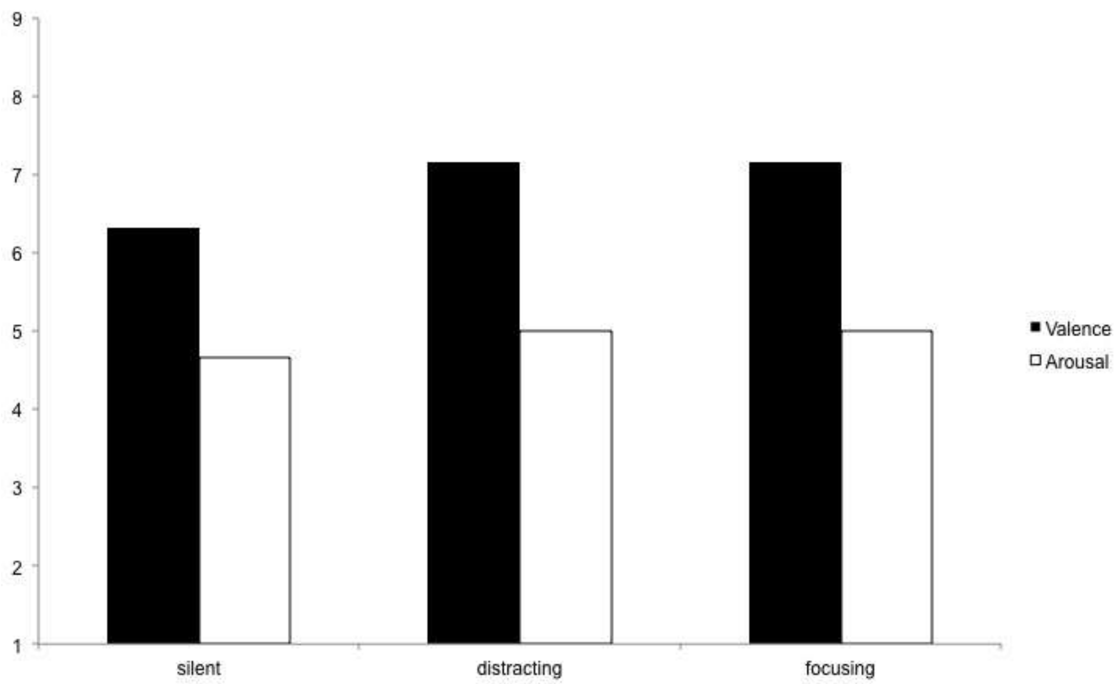

Figure 6. Average valence and arousal ratings for the silent, distracting and focusing versions.

User Experience (RQ2)

Our second research question concerned the perceived experience given different music. We have already seen that the self-reported emotion of the participants did not differ significantly between versions but music could still have an impact on how users engaged with the moving images. In order to investigate this we looked at participants' responses to the post-task and post-session questionnaires, in which we asked them to choose words or provide open comments that best described their experience of watching the film clips.

In the post-task questionnaires, participants were asked to select from a list of adjectives that best described their experience and they could also provide their own words. As the number of responses was unequally distributed and low, with some words only provided once, statistical analysis was inappropriate. However, these results still provide a qualitative understanding about user experience. Figures 7, 8 and 9 show word clouds of terms that participants used to describe the experience of watching the moving images. In these word clouds, the size of the word indicates their relative frequency of use by participants applied to that version. As can be seen, the most frequent terms used to describe the silent version were "understandable" and "clear", which are fairly neutral. In contrast, in both music versions, the terms "entertaining" and "interesting" featured most, which suggests an increased level of immersion and flow when moving images were accompanied by music.

Figure 7. Word cloud for silent version.

Figure 8. Word cloud for distracting version.

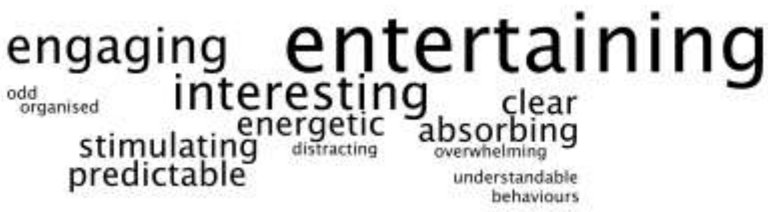

Figure 9. Word cloud for focusing version.

It also appears that our music design had at least some of the intended effects. Within the top most used terms to describe the versions, participants used "energetic" more frequently in the distracting version than in the focusing version. This aligns well with our intended design of the distracting version featuring a perpetual insistent motion and character.

Overall, we did not find any notable differences in ratings of participants between versions relating to their perceived experience of attention or the suitability of the music to the moving images. Responses to various 5-point Likert-scale questions such as "I think the music suited the video clip" or "The music helped me understand what was going on in the story" revealed only small statistical differences. However, analyzing the post-task questionnaire comments, participants seemed to find the focusing version more helpful in conveying the overall mood of the story. Seven participants specifically commented on the positive aspects of the focusing version in supporting the narrative, for example:

> *"The music matched with the different shots and actions, for example the sudden change when the woman whistles and dances." (P9)*
>
> *"The music told a story through tempo and different motifs from different instruments of the orchestra." (P11)*

When asked specifically about how effectively the distracting music suited the moving images, however, participants thought that it did not match the flow of the story. Eight participants shared similar sentiments to these after they experienced the distracting version:

> *"It was far too fast-paced and would've been more suitable for a scene with more action." (P13)*
>
> *"It seemed a bit disconnected, without enough changes to accompany the flow of the story." (P14)*
>
> *"No I disagree, the music was way too fast and started making me feel highly anxious." (P18)*

Interestingly, two participants mentioned the effect of the music on their visual attention explicitly in their post-session summary comments:

"I believe the 1st clip [the distracting version] *with more dynamic track instigated me to keep a faster track of 'foreground' and 'background' action." (P8*)

"It was definitely more difficult to know what to focus on the first time around [the distracting version]*, particularly since the music was overwhelming and slightly obnoxious. I felt like I had to focus on one image to keep my bearings. The second time around, I felt more comfortable to focus where the music 'guided' me." (P12)*

As these results indicate, it was very difficult for participants to externalize complex internal processes that are tacit and pre-cognitive. Both the distracting and focusing versions attracted positive and negative comments, however, the distracting version appeared to overwhelm the participants more than the focusing version while the focusing version appeared to help participants make more sense of the story. We wondered whether this would also be evident in the eye-tracking metrics, which objectively measured visual attention.

Visual Attention (RQ3)

Overall Effects of Music on Visual Attention

We analyzed the eye-tracking metrics over the whole of the clip to investigate whether the music we designed had any major effects overall on participants' visual attention. We found some striking and significant differences in eye-tracking measures between the versions (Figure 10). First, we analyzed the fixation count i.e. how many times the participants dwelled on a point in the visual scene. Overall, there was a highly significant difference in the number of fixations between versions ($\chi^2(2,N=4257)=22.25$, $p<0.001$), with fixation count highest in the distracting version. Each participant on average fixated 240.83 times in the silent version, 259.83 times in the distracting version yet only 209.33 times in the focusing version. The average fixation duration between versions was also different (ANOVA, $F(2,4257)=13.30$, p=<0.001),[31] with participants in the silent version spending on average 0.57 seconds per fixation, 0.56 seconds in the distracting version but 0.69 seconds in the focusing version.

The total fixation duration is the sum of all individual fixation's durations. There was no statistical significance between participants' length of total fixations (Kruskal-Wallis, $H(2,N=18)$, $p=0.236$) although we can note that, on average, the total fixation duration in the silent version was the shortest (131.77 seconds), the distracting version was in the middle (146.35 seconds) while the total fixation duration in the focusing version was the longest (150.34 seconds).[32] In effect, the longer average length of fixations in the focusing version cancelled out the much fewer number of fixations. It should be noted that the total fixation duration is shorter than the length of the clip; the difference is accounted for by the length of saccades—when the gaze shifts—and when the gaze does not dwell long enough to be classed as a fixation. In this sense, participants spent most time lingering on the visual scene in the focusing version.

These results have implications for the design of music and the exploration of moving images. First, while the overall time participants spent exploring the moving images was very similar, the versions that they experienced were able to direct *how* they explored the visual scenes, in terms of number of fixations and the average duration of those fixations. We designed the distracting version with an unrelenting dynamic character and this kind of music was able to directly affect the number of times that gaze shifted, indicated by increased fixation counts, something that was also intuitively perceived by some of our participants and reflected in their comments. The focusing music had the ability to increase how long participants gazed on certain points in a visual scene and in total, so our design of this music had a 'settling' effect. Taken together, this implies that music can influence fixation count and duration independently. By carefully designing the music we may be able to direct how users explore a visual scene, either encouraging them to look around more (or less) or to linger on a point of focus for longer (or shorter).

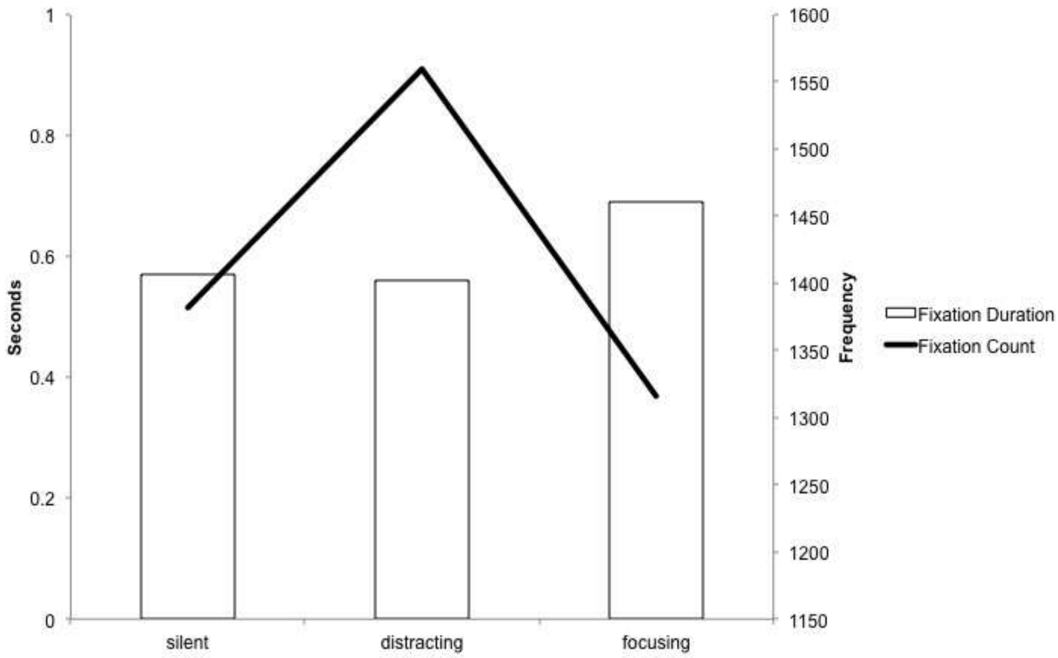

Figure 10. Average fixation duration for the three versions (white bar chart, y-axis left) compared to total fixation count (black line, y-axis right).

Music's Influence on Competing Targets for Visual Attention

We carefully analyzed the *stairs_both* scene because it introduced strong competing targets for visual attention—the passing traffic on the stairs. We hypothesized that the focusing music might help participants to focus more on the main characters in the visual scene, largely ignoring competing visual foci, while the distracting version would encourage participants to explore more i.e. focus more on the passing traffic. Both music versions were expected to do better than the silent version.

First, we examined the fixation count and fixation duration for both the target, George and Peppy, as well as *not* on target i.e. anywhere else in the visual scene (Table 1). For this scene, we found that combined average fixation counts for on target and not on target were higher in *both* music versions during this scene of the clip than the silent version (this only happened for the distracting version over the whole clip). However, the ratio of whether participants looked on the target versus not on the target was skewed in favor of the distracting version: in the focusing version a smaller number of fixations overall were on the target (0.69) as compared to the distracting version (1.16). Hence, surprisingly, participants who had experienced the distracting version looked at George and Peppy slightly more frequently instead of on other, competing foci for their visual attention. Music also had a very small effect on the time participants spent fixating on the target versus everywhere else. On average, the focusing version encouraged participants to fixate the longest per fixation on the target but average fixation duration when visual attention rested anywhere else in the visual scene was very similar to participants in the distracting version. However, there was no statistically significant effect of version on either fixation counts or duration. We noted larger differences in the total fixation duration, however these were also not statistically significant. While both music versions had longer total fixation duration on George and Peppy, participants in both music versions also fixated for longer everywhere else.

These results, especially when contrasted to the initial findings across the whole clip, present a much more complex picture of music's effect on visual attention. It appears that *any* music encouraged more and lengthier exploration of the visual scene when competing visual foci were present, including exploration of 'distractions'. Intriguingly, it points to the possibility that the distracting music we designed may have been equally good for helping participants to resist competing foci for visual attention in this scene.

	Silent	Distracting	Focusing
Participants on target	6	6	6
Average Fixation Count on target	5.50	9.33	5.67
Average Fixation Count not on target	2.50	8.00	8.17
Average Fixation Duration on target	0.65	0.54	0.86
Average Fixation Duration not on target	0.24	0.43	0.46
Average Total Fixation Duration on target	3.57	5.04	4.87
Average Total Fixation Duration not on target	0.60	3.44	3.74

Table 1. Eye-tracking measures for the *stairs_both* scene for the silent, distracting and focusing versions.

Music's Effect on Directing Focus

In the final scene of the clip, *stairs_george,* we were particularly interested in investigating whether music was able to direct visual attention to a particular character more quickly and consistently, especially given the particular challenge of acquiring the target while competing visual foci were present. We designed the focusing version specifically to include musical attributes to help focus on the character George.

First, we measured the fixation spread for the *stairs_george* scene. This gave us an indication of how closely fixations were clustered within the entire visual scene presented to participants. We found that the average distance between all fixations was highest in the silent version (Figure 11). In the silent version the fixation spread was on average 195.16 pixels, whereas in the distracting version it was 156.09 pixels and in the focusing version 152.74 pixels. Although the focusing version had the lowest fixation spread, it was not very different from the distracting version. The gap between the silent version fixation spread and music versions was thus at least 0.45 inches, given the screen resolution and screen size used in this study (86.27 PPI). Since the fixation spread measures how closely fixations are clustered together, this means that participants explored slightly more of the visual scene when there was no music, whereas they stayed in a smaller cluster for the music versions. It appears that music, whatever its nature, had the effect of concentrating participants' gaze in the visual scene.

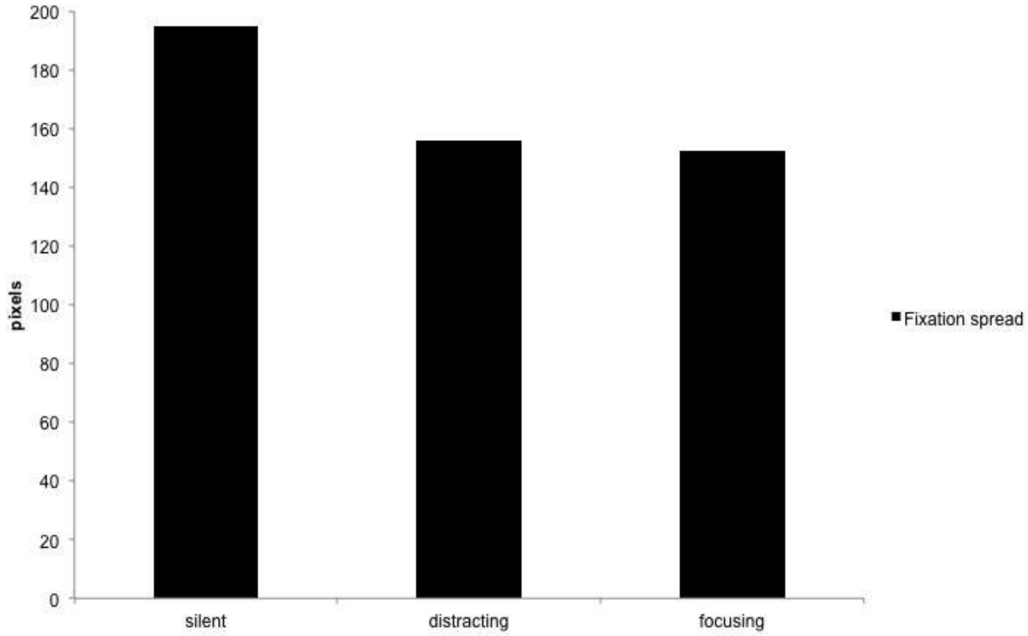

Figure 11. The average distance between fixations in the silent, distracting and focusing versions; silent had a larger spread than music versions.

Participants may have attended to a smaller area in the visual scene in the music versions, yet they were not oblivious to what else was going on in the scene. We also investigated when participants *did not* gaze at the main characters (Table 2, shaded rows). We found significant differences between versions in the total fixation duration and fixation counts that were not on target, where the silent version was significantly lower than the music versions for the number of fixations (Kruskal-Wallis, $H(2,N=18)$, $p=0.007$) and total fixation duration (Kruskal-Wallis, $H(2,N=18)$, $p=0.006$). This means that although participants in the music versions concentrated their gaze in a smaller area overall they also looked more frequently and longer outside of the main target than when no music was present. It appears, therefore, that music, whatever its nature, not only had the effect of concentrating participants' gaze on the central character but also of encouraging greater exploration of the overall scene.

Astonishingly, three of the six participants experiencing the silent version and two of the six in the distracting version *never* fixated on the target (George) at all during this scene, whereas all participants in the focusing version attended to the target. There are clearly narrative implications for the construction of film sequences when the attributes of a particular piece of music could mean that significant sections of an audience might not even notice the central character at an important dramatic moment. Table 2 also shows that it took participants on average only 2.66 seconds to first fixate on George in the focusing version, considerably faster than the distracting version, and almost half the time as in the silent version. Unfortunately, a statistical test did not show this result as significant due to the low sample size of the number of participants that fixated on George in the first place. We had already noticed a similar effect on time to first fixation in the *stairs_both* scene but here it was very much more pronounced, perhaps because the music was more gesturally bold and synchronous. The focusing music also had a slight effect on the length of time participants fixated on the target; their first fixation duration was the longest as was their fixation duration and their total fixation duration. There were no noticeable differences between fixation counts on George for the three versions. While all of these differences cannot be classed as statistically significant, we can nonetheless note that the focusing version appeared to help participants in the manner which we intended in the design by attracting visual attention to the main character more quickly.

	Silent	**Distracting**	**Focusing**
Participants on target	3	4	6
Average Time to first fixation, from start of scene	5.02	3.76	2.66
Average First fixation duration	0.35	0.37	0.41
Average Fixation Duration	0.51	0.53	0.59
Average Fixation Duration not on target	0.39	0.42	0.50
Average Total Fixation duration on target	2.91	2.92	3.24
Average Total Fixation duration not on target *	2.35	17.47	17.38
Average Fixation Count on target	5.67	5.50	5.50
Average Fixation Count not on target *	6.00	41.67	35.00

Table 2. Eye-tracking measures for the *stairs_george* scene for silent, distracting and focusing versions; statistically significant results marked with *.

DISCUSSION

A detailed understanding of how music affects visual exploration of moving images could help in the analysis and design of complex dynamic scenes. Our study, using an inter-disciplinary team, aimed to explore how musical attributes can shape visual attention. The results of this study provide some important lessons in designing music for visual attention and future work.

We found no effect in our study of music on participants' self-reported emotion. This may be due to a number of reasons. First, the two pieces of music and the film excerpt itself appear to be broadly similar in emotional terms, as our results demonstrated. The Self-Assessment Manikin may not be sufficiently finely graded when there is a lack of differentiation between emotions that share the valence-arousal space. Measuring emotional responses, especially for extended dynamic audiovisual scenes, is difficult. Ellis and Simons used only six-second clips for their study and also attempted to cover the entire spectrum of valence-arousal dimensions.[33] Our excerpt was relatively lengthy and therefore the self-assessment tool we used may not be accurate for capturing emotional responses in complex multi-sensory sequences, where there will likely be a subtle range of emotional shifts over time. This simply highlights the potential usefulness of eye-tracking methodology as a means of filtering user experience data. Future studies should consider using a combination of more nuanced feedback tools. The Geneva Emotional Music Scale (GEMS) model developed by Zentner, Grandjean and Scherer might provide a more useful method for measuring mixed-emotion, or emotions of potentially differing valence and arousal felt simultaneously.[34] Biometric measures such as electroencephalography (EEG), galvanic skin response (GSR), electromyographic (EMG) activity, and/or heart rate, might also provide useful real-time data to identify emotional response.[35] Some recent research even suggests that it may be possible to measure emotion using eye-tracking methodologies themselves, although this remains highly speculative.[36] Biometric tools have their own problematic aspects but may help in better understanding the link between emotion, music and visual attention. Last, existing theory posits that music affects emotion and emotion in turn affects visual attention. Our results suggest that this theory may need further refinement as it appears that music has a much more direct influence on visual attention than previously suggested.

Responses to Likert-scale questions revealed only small statistical differences between music versions, suggesting that users may not be consciously aware how music affects their cognitive processes. Participants in our study also reported both positive and negative comments in terms of suitability and narrative support of the focusing and distracting music, yet they seemed to find the focusing version more helpful in conveying the overall narrative mood. While we obtained qualitative data of participants' perceptions that pointed to increased immersion and flow in the focusing version, this is less valuable for the detailed study of visual attention. Hence, capturing eye-tracking metrics in addition to perceived user experience data is essential in the study of complex visual exploration tasks.

The results based on eye-tracking measures showed that the influence of music on visual attention and exploration is nuanced and complex. Our findings suggest that certain kinds of contextual musical attributes and specific gestures have the ability to make users fixate more frequently and for longer on points in the visual scene. Perhaps the most significant finding is that visually congruent, bold musical focusing gestures appear to affect time to first fixation, drawing attention to particular points more quickly in complex visual exploration tasks. Returning to the

Congruence-Associationist Framework we, therefore, suggest that structural audio-visual congruence can indeed help focus attention on the visual figure, although it is the speed of visual focus that is most significantly affected rather than the total fixation duration. This idea finds some resonance with Lipscomb's research on the alignment of cross-modal accent structures. In line with Lipscomb, our results may be seen to expand the congruence component of the Congruence-Associationist Framework.[37]

Both pieces of music in our experiment had the effect of concentrating participants' gaze in the overall visual scene. One possible explanation is that music, whatever its nature, has a direct impact on cognitive load and this in turn has a direct impact on the amount of visual exploration. Capacity for cognitive processing is limited when we are engaged in challenging tasks. Further study focused on the graded intensity of musical attributes in combination with moving images might help understand the potential demands on cognitive load more fully.

Clearly, more work is needed to tease apart the effects of musical attributes. In our experiment we used pre-existing music and film, albeit with some significant editing of the music. In future, however, we suggest that more detailed and nuanced results could be achieved if both original high-quality moving images and music/sound are created specifically for experimental purposes. Stimuli that remain ecologically valid yet also modify specific measurable parameters will yield results that can be related to real-world experiences. Of course, this cannot be achieved without the participation of filmmakers or composers. We can already point to the value of genuine interdisciplinary research in our study. Without the different skillsets brought to experimental design the research questions themselves could not have been formulated. Furthermore, we have found balancing different perspectives and approaches useful and productive. Most typically this has involved questioning the disciplinary biases that have shaped the intuitive frameworks in which we think.

We are encouraged that we are moving in the right direction. The music we designed was able to direct participants' attention more quickly to the desired targets of focus and to shrink the area in which visual exploration took place. However, it did not reduce exploration outside of the target, meaning that it did not focus attention *too much*. This implies that the right kind of music could in fact help users in complex visual exploration tasks to acquire targets more easily while not ignoring contextual surrounding information, which seems entirely appropriate as one of the functional roles of music in narrative film sequences.

CONCLUSION

This paper reported on the results of a foundational, empirical eye-tracking study that investigated the effects of contextual musical attributes on visual attention, emotion and user experience during challenging exploration tasks in moving images. We found that:

- Music did not have a noticeable effect on emotions in our study. However, fine-grained biometric feedback or a mixed-emotion model could prove helpful in exploring possible links between music and emotion.
- Participants had mixed responses to the music in our study, designed to have either distracting or focusing musical attributes. The distracting version was perceived as overwhelming but the focusing version seemed to help with immersion and engagement.
- Music had an effect on visual attention; the distracting version increased fixations while the focusing version increased fixation duration.
- Music in fact seemed to encourage participants to fixate on competing foci for visual attention, however it also appears to have helped to attract visual attention to a target more quickly and to shrink the overall focus of attention. These effects taken together seem especially beneficial for a complex visual exploration task in moving images.

There is limited understanding of how music shapes visual attention in audiovisual interactions and our study is the first fundamental step in designing music that directs what we see. Eye-tracking as a methodology is necessary because it provides invaluable empirical data for the study of visual attention that cannot be secured by other means. Finally, we can report that there have been many occasions when the assumptions inherent in our respective disciplines of music and human computer interaction have been fruitfully challenged by our research process. It is only through this inter-disciplinary approach that we have come closer to understanding the full implications of our research questions and have moved towards work that is both scientifically and musically relevant.

ACKNOWLEDGMENTS

We thank the participants in our study and especially Milena Markova for running the eye-tracking experiment.

[1] Cohen initially explored film and music before employing the term multimedia to encompass a broad range of audiovisual media, although all her experiments focused on film material. As the research has developed she has explored the wider soundtrack including sound effects and dialogue as well as music. Her papers on this topic include: “Associationism and Musical Soundtrack Phenomena,” *Contemporary Music Review* 9, (1993): 163–178; “Film Music:

Perspectives from Cognitive Psychology," in *Music and Cinema*, ed. James Buhler, Caryl Flinn, and David Neumeyer (Hanover, NH: University Press of New England, 2000), 360–377; "Music Cognition and the Cognitive Psychology of Film Structure," *Canadian Psychology* 43, 4 (2002): 215–232; "How Music Influences the Interpretation of Film and Video: Approaches from Experimental Psychology," in *Perspectives in Systematic Musicology*, eds. Roger Kendall and Roger Savage (Los Angeles: University of California, 2005), 15–36; "Music as a Source of Emotion in Film," in *Handbook of Music and Emotion: Theory, Research, and Applications*, ed. Patrik Juslin and John Sloboda (Oxford: Oxford University Press, 2010), 879–908; "Film Music and the Unfolding Narrative," in *Language, Music and the Brain (Strüngmann Forum Reports)*, ed. Michael A. Arbib (Cambridge, MA: MIT Press, 2013), 173–201; "Film Music From the Perspective of Cognitive Science," in *The Oxford Handbook of Film Music Studies*, ed. David Neumeyer (Oxford: Oxford University Press, 2013), 96–130.

[2] Cohen, "Film Music; Perspectives from Cognitive Psychology," 371; also "Film Music From the Perspective of Cognitive Science," 122.

[3] Scott Lipscomb, "Cross-Modal Alignment of Accent Structures in Multimedia," in *The Psychology of Music in Multimedia*, eds. Siu-Lan Tan, Annabel Cohen, Scott Lipscomb, and Roger Kendall (Oxford: Oxford University Press, 2013), 209.

[4] See for example: Siu-Lan Tan et al., *The Psychology of Music in Multimedia* (Oxford: Oxford University Press, 2013); Arthur Shimamura, *Psychocinematics: Exploring Cognition at the Movies* (Oxford: Oxford University Press, 2013).

[5] Surprisingly, in the first and currently the only book dedicated to empirical research focusing on music's interaction with moving images, Tan et al.'s *The Psychology of Multimedia*, not one single experiment outlined in the volume employs eye-tracking, the methodology is not mentioned in the final chapter, Future Research Directions for Music and Sound in Multimedia, and the word eye-tracking does not even appear in the index.

[6] Andrew Duchowski, *Eye Tracking Methodology: Theory and Practice* (London: Springer, 2007); Michael Land and Benjamin Tatler, *Looking and Acting: Vision and Eye Movements in Natural Behaviour* (Oxford: Oxford University Press, 2009); Simon Liversedge, Iain Gilchrist and Stefan Everling, eds., *The Oxford Handbook of Eye Movements* (Oxford: Oxford University Press, 2011); Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van der Weijer, eds., *Eye Tracking: A Comprehensive Guide to Methods and Measures* (Oxford: Oxford University Press, 2011).

[7] Paul Marchant, David Raybould, Tony Renshaw and Richard Stevens "Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes," *Digital Creativity* 20, 3 (2009): 154.

[8] Tore Vesterby, Jonas Voss, John Paulin Hansen, Arne John Glenstrup, Dan Witzner Hansen, and Mark Rudolph "Gaze-guided viewing of interactive movies," *Digital Creativity* 16, 4 (2005): 193–204; Robert Goldstein, Russell Woods, and Eli Peli "Where people look when watching movies: do all viewers look at the same place?" *Computers in Biology and Medicine* 37, 7 (2007): 957–964; Tim Smith "Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory," in *Psychocinematics: Exploring Cognition at the Movies*, ed. Arthur Shimamura (Oxford: Oxford University Press, 2013), 165–189. We should also mention research by the Diem Project [http://thediemproject.wordpress.com/] which resulted in the creation of CARPE (Computational and Algorithmic Representation and Processing of Eye-movements), an open-source software tool that allowed researchers to represent eye-movement data in dynamic heat-maps which identify where people look during scene viewing.

[9] Jamie Madell and Sylvie Hébert "Eye Movements and Music Reading: Where Do We Look Next?" *Music Perception* 26, 2 (2008): 157–170.

[10] Rong-Fuh Day, Chien-Huang Lin, Wen-Hung Huang, and Sheng-Hsiung Chuang "Effects of music tempo and task difficulty on multi-attribute decision-making: An eye-tracking approach," *Computers in Human Behavior 25*, 1 (2009): 130–143.

[11] Fabrice Cauchard, James Cane, and Ulrich Weger "Influence of background speech and music in interrupted reading: An eye-tracking study," *Applied Cognitive Psychology* 26, 3 (2012): 381–390.

[12] Mei-Ching Chen, Pei-Luen Tsai, Yu-Tung Huan, and Keh-Chung Lin "Pleasant music improves visual attention in patients with unilateral neglect after stroke," *Brain Injury* 27, 1 (2013): 75–82.

[13] For an effective overview see Patrik Juslin and John Sloboda, *Handbook of Music and Emotion: Theory, Research, and Applications* (Oxford: Oxford University Press, 2010).

[14] Robert Ellis and Robert Simons "The Impact of Music on Subjective and Physiological Indices of Emotion While Viewing Films," *Psychomusicology* 19, (2005): 15–40.

[15] Warren Brodsky "The effects of music tempo on simulated driving performance and vehicular control," *Transportation Research Part F: Traffic Psychology and Behaviour* 4, 4 (2002): 219–241; Ayça Berfu Ünal, Dick de Waard, Kai Epstude, and Linda Steg "Driving with Music: Effects on Arousal and Performance," *Transportation Research* 21, (2013): 52–65.

[16] Brodsky, "Effects of music tempo on simulated driving performance," 225.

[17] Brodsky, "Effects of music tempo on simulated driving performance," 238.

[18] Day, "Effects of music tempo and task difficulty on multi-attribute decision-making," 132.

[19] Tan et al. *The Psychology of Music in Multimedia*, 204.

[20] Anderson, L. Recording of Leroy Anderson's 'Fiddle Faddle' https://itunes.apple.com/gb/album/the-

typewriter/id353970686.

[21] The edited version of the piece was constructed sequentially as follows: measures 40–70, 90–166 (including repeats), 9–42, and 166–202.

[22] The violin harmonics were the only new addition to the edited track. They were rendered using a virtual orchestral software and sample library (the Vienna Symphonic Library - Solo Strings). The quality of the samples and the flexibility with which they can be manipulated within the software resulted in an addition that was largely indistinguishable from a real instrumental recording. The harmonics in this 'focusing' version were also carefully mixed and balanced so that they blended seamlessly into the existing track.

[23] It could be argued that any sort of auditory cue could be used to evoke a learned association with a perspective change. This is not to trivialize our approach, rather to suggest that behind these congruencies/associations may lie learned, audio-visual pattern matching.

[24] See Duchowski, *Eye Tracking Methodology: Theory and Practice*; Oleg Špakov, "Comparison of eye movement filters used in HCI," *Proceedings of the Symposium on Eye Tracking Research and Applications* (New York: ACM, 2012): 281–284.

[25] Bee swarms display the gaze points of numerous participants simultaneously over time; gaze plots summarize gaze data from one or multiple recordings and display gaze points, fixations, and scan paths; heat maps show an agglomerated analysis of the visual exploration patterns in a group of users based on the summary of gaze data from multiple recordings. The 'hot' zones with higher density designate where the users focused their gazes with higher frequency; clusters visualize the true areas of interest, polygons display the areas with the highest concentration of gaze points recorded during the test.

[26] Daniel Västfjäll, Penny Bergman, Anders Sköld, Ana Tajadura, and Pontus Larsson "Emotional Responses to Information and Warning Sounds," *Journal of Ergonomics* 2, 3 (2012).

[27] Margaret M. Bradley and Peter J. Lang "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994): 49-59.

[28] Joey Benedek and Trish Miner "Measuring Desirability: New methods for evaluating desirability in a usability lab setting," in *Proc. UPA 2002*, (2002): 8–12.

[29] Lisa Jefferies, Daniel Smilek, Erik Eich and James Enns "Emotional Valence and Arousal Interact in Attentional Control," *Psychological Science* 19, 3 (2008): 290–295; Brett Ford, Maya Tamir, Tad Brunyé, William Shirer, Caroline Mahoney, and Holly Taylor "Keeping Your Eyes on the Prize: Anger and Visual Attention to Threats and Rewards," *Psychological Science* 21, 8 (2010): 1098–1105.

[30] Named after William Kruskal and W. Allen Wallis, the Kruskal-Wallis analysis of variance test is a method for identifying whether samples originate from the same distribution. It is used for comparing more than two samples that are independent, or not related.

[31] Analysis of variance (ANOVA) is a statistical method for making simultaneous comparisons between two or more means that yields values which can be tested to determine whether a significant relation exists between variables.

[32] The data is too sparse and violates major normality assumptions therefore making it unsuitable for a parametric test.

[33] Ellis and Simons, "The Impact of Music on Subjective and Physiological Indices of Emotion."

[34] Marcel Zentner, Didier Grandjean, and Klaus Scherer "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion* 8, (2008): 494–521.

[35] Electroencephalography (EEG) is the recording of electrical brain activity usually by placing a series of electrodes on the scalp of the subject; Galvanic Skin Response (GSR) measures the electrical conductance of the skin. Sweat glands are controlled by the sympathetic nervous system so GSR can be used as an indication of physiological arousal; Electromyograpy or (EMG) is a method for recording the electrical impulses produced by the facial muscles when they contract and has been shown to be useful in measuring emotional reactions. All of these methods are potentially useful because they provide real-time data and cannot be 'faked'.

[36] Margaret Bradley, Laura Miccoli, Miguel Escrig, and Peter Lang "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology* 45, 4 (2008): 602–607; Vidas Raudonis, Gintaras Dervinis, and Andrius Vilkauskas "Evaluation of Human Emotion from Eye Motions," *International Journal of Advanced Computer Science and Applications* 4, 8 (2013), 79–84.

[37] Lipscomb, "Cross-Modal Alignment of Accent Structures in Multimedia."

BIBLIOGRAPHY

Benedek, Joey and Trish Miner. "Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting." *Proceedings from the Usability Professionals Association 2002*, (2002): 8–12.

Bradley, Margaret M. and Peter J. Lang. "Measuring Emotion: the Self-assessment Manikin and the Semantic Differential." *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994): 49-59.

Bradley, Margaret, Laura Miccoli, Miguel Escrig, and Peter Lang. "The Pupil as a Measure of Emotional Arousal and Autonomic Activation." *Psychophysiology* 45, 4 (2008): 602–607.

Brodsky, Warren. "The Effects of Music Tempo on Simulated Driving Performance and Vehicular Control." *Transportation Research Part F: Traffic Psychology and Behaviour* 4, 4 (2002): 219–241.

Cauchard, Fabrice, James Cane, and Ulrich Weger. "Influence of Background Speech and Music in Interrupted Reading: An Eye-tracking Study." *Applied Cognitive Psychology* 26, 3 (2012): 381–390.

Chen, Mei-Ching , Pei-Luen Tsai, Yu-Tung Huan, and Keh-Chung Lin. "Pleasant Music Improves Visual Attention in Patients with Unilateral Neglect After Stroke." *Brain Injury* 27, 1 (2013): 75–82.

Cohen, Annabel. "Associationism and Musical Soundtrack Phenomena." *Contemporary Music Review* 9, (1993): 163–178.

__________. "Film Music: Perspectives from Cognitive Psychology." In *Music and Cinema*, ed. James Buhler, Caryl Flinn, and David Neumeyer, 360–377. Hanover, NH: University Press of New England, 2000.

__________. "Music Cognition and the Cognitive Psychology of Film Structure." *Canadian Psychology* 43, 4 (2002): 215–232.

__________. "How Music Influences the Interpretation of Film and Video: Approaches from Experimental Psychology." In *Perspectives in Systematic Musicology*, eds. Roger Kendall and Roger Savage, 15–36. Los Angeles: University of California, 2005.

__________. "Music as a Source of Emotion in Film." In *Handbook of Music and Emotion: Theory, Research, and Applications*, ed. Patrik Juslin and John Sloboda, 879–908. Oxford: Oxford University Press, 2010.

__________. "Film Music and the Unfolding Narrative." In *Language, Music and the Brain (Strüngmann Forum Reports)*, ed. Michael A. Arbib, 173–201. Cambridge, MA: MIT Press, 2013.

__________. "Film Music From the Perspective of Cognitive Science." In *The Oxford Handbook of Film Music Studies*, ed. David Neumeyer, 96–130. Oxford: Oxford University Press, 2013.

Day, Rong-Fuh, Chien-Huang Lin, Wen-Hung Huang, and Sheng-Hsiung Chuang. "Effects of Music Tempo and Task Difficulty on Multi-attribute Decision-making: An eye-tracking Approach." *Computers in Human Behavior 25*, 1 (2009): 130–143.

Duchowski, Andrew. *Eye Tracking Methodology: Theory and Practice.* London: Springer, 2007.

Ellis, Robert and Robert Simons. "The Impact of Music on Subjective and Physiological Indices of Emotion While Viewing Films." *Psychomusicology* 19, (2005): 15–40.

Ford, Brett, Maya Tamir, Tad Brunyé, William Shirer, Caroline Mahoney, and Holly Taylor "Keeping Your Eyes on the Prize: Anger and Visual Attention to Threats and Rewards." *Psychological Science* 21, 8 (2010): 1098–1105.

Goldstein, Robert, Russell Woods, and Eli Peli. "Where people look when watching movies:

do all viewers look at the same place?" *Computers in Biology and Medicine* 37, 7 (2007): 957–964.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van der Weijer, eds. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press, 2011.

Jefferies, Lisa, Daniel Smilek, Erik Eich and James Enns. "Emotional Valence and Arousal Interact in Attentional Control." *Psychological Science* 19, 3 (2008): 290–295.

Juslin, Patrik and John Sloboda, eds. *Handbook of Music and Emotion: Theory, Research, and Applications*. Oxford: Oxford University Press, 2010.

Land, Michael and Benjamin Tatler. *Looking and Acting: Vision and Eye Movements in Natural Behaviour*. Oxford: Oxford University Press, 2009.

Lipscomb, Scott. "Cross-Modal Alignment of Accent Structures in Multimedia." In *The Psychology of Music in Multimedia*, eds. Siu-Lan Tan, Annabel Cohen, Scott Lipscomb, and Roger Kendall, 192–213. Oxford: Oxford University Press, 2013.

Liversedge, Simon, Iain Gilchrist and Stefan Everling, eds. *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press, 2011.

Madell, Jamie and Sylvie Hébert. "Eye Movements and Music Reading: Where Do We Look Next?" *Music Perception* 26, 2 (2008): 157–170.

Marchant, Paul, David Raybould, Tony Renshaw and Richard Stevens. "Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes." *Digital Creativity* 20, 3 (2009): 153–163.

Marshall, Sandra and Cohen, Annabel. "Effects of Musical Soundtracks on Attitudes to Geometric Figures." *Music Perception* 6, (1998): 95–112.

Raudonis, Vidas, Gintaras Dervinis, and Andrius Vilkauskas. "Evaluation of Human Emotion from Eye Motions," *International Journal of Advanced Computer Science and Applications* 4, 8 (2013), 79–84.

Shimamura, Arthur, ed. *Psychocinematics: Exploring Cognition at the Movies*. Oxford: Oxford University Press, 2013.

Smith, Tim. "Watching You Watch Movies: Using Eye Tracking to Inform Cognitive Film Theory." In *Psychocinematics: Exploring Cognition at the Movies*, ed. Arthur Shimamura, 165–189. Oxford: Oxford University Press, 2013.

Špakov, Oleg "Comparison of Eye Movement Filters Used in HCI." *Proceedings of the Symposium on Eye Tracking Research and Applications.* New York: ACM, (2012): 281–284.

Tan, Siu-Lan, Annabel Cohen, Scott Lipscomb and Roger Kendall, eds. *The Psychology of Music in Multimedia*. Oxford: Oxford University Press, 2013.

Ünal, Ayça Berfu, Dick de Waard, Kai Epstude, and Linda Steg. "Driving with Music: Effects on Arousal and Performance." *Transportation Research Part F: Traffic Psychology and Behaviour* 21, (2013): 52–65.

Västfjäll, Daniel, Penny Bergman, Anders Sköld, Ana Tajadura, and Pontus Larsson. "Emotional Responses to Information and Warning Sounds." *Journal of Ergonomics* 2, 3 (2012). doi: 10.4172/2165-7556.1000106.

Vesterby, Tore, Jonas Voss, John Paulin Hansen, Arne John Glenstrup, Dan Witzner Hansen, and Mark Rudolph. "Gaze-guided Viewing of Interactive Movies." *Digital Creativity* 16, 4 (2005): 193–204.

Zentner, Marcel, Didier Grandjean, and Klaus Scherer. "Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement." *Emotion* 8, (2008): 494–521.