Eye Movements in Time
Auditory Influences on Oculomotor Timing

Jonathan Batten

Department of Psychological Sciences
Birkbeck, University of London

A thesis submitted for the degree of
Doctor of Philosophy
February 2018

Originality Statement

‘I, Jonathan Batten, declare that the work in this submitted thesis is my own.’

Signed:

February 21st 2018

Abstract

The dominant models of eye movement timing consider only visual factors as modulators of when gaze orients (e.g. EZ-Reader, SWIFT, CRISP, LATEST). Yet real-world perception is multimodal, and temporal information from audition can both aid the predictive orienting of gaze (to relevant audiovisual onsets in time), and inform visual orientation decisions known to modulate saccade timing, e.g. where to orient. The aim of this thesis was to further the current understanding of eye movement timing to incorporate auditory information; specifically investigating the implicit and explicit capacity for musical beats to influence (and entrain) eye movements, and to quantify the capacity and limitations of direct control when volitionally matching eye movements to auditory onsets. To achieve this, a highly-simplified gaze-contingent visual search paradigm was refined that minimised visual and task factors in order to measure auditory influence. The findings of this thesis present evidence that self-paced eye movements are impervious to implicit auditory influences. The explicit control of eye movements, as small corrections in time to align with similarly timed music, was very limited. In contrast, when visual transitions were externally timed, audiovisual correspondence systematically delayed fixation durations.

The thesis also measured the extent of direct control that can be exerted on eye movements, including the role of auditory feedback, as well as modulating visual complexity to further increase inhibition and temporal precision. These studies show a predictive relationship between the level of direct volitional control that an individual can affect and how synchronised they are. Additionally, these studies quantify a large subpopulation of quick eye movements that are impervious to direct control. These findings are discussed as provocation for revised oculomotor models, future work that

considers the temporal relationship between shifts of attention and gaze, and implications for wider psychological research that employs timed eye movement measures.

Table of Contents

Acknowledgements

If someone had told me ten years ago that I would care about eye movements, I would not have taken them seriously. That I do care now is because of Dr Tim Smith, my PhD supervisor. When I was a final year undergraduate frustrated by project choices, Rebecca stepped in and introduced me to Tim. Unsurprisingly, Tim said yes to a highly ambitious eye-tracking project (that required days of dynamic region coding) and sparked my interest in active vision research. Since that day, Tim has actively involved me in his lab and even encouraged me to present that initial study at ECEM in Sweden. Since then and throughout my PhD, Tim has continued to support my somewhat ambitious ideas. He has been a fount of knowledge on all things vision and shown a friendly face when "it doesn't work…". I am incredibly grateful to Tim for his scientific rigour, his relaxed mentoring style, and most notably his support when my son was born. I am indebted.

I would also like to thank my second supervisor, Professor Fred Dick, for his advice, encouragement and engagement with my research. I am grateful to Leslie Tucker for relocating me from the darkness (MERLiN basement) and bringing me into the light, and to Harish Patel for aiding me in solving a myriad of problems. I'm convinced that Harish is the glue that holds the Psychology Department together. I am also grateful to the Economic and Social Research Council for funding this PhD.

Throughout this PhD, I have had the privilege of meeting and working alongside some wonderful people, especially Rebecca, Nick, Daniel, Rosy, Jen, Sinead, Jenny, Ana, and Celeste. Rebecca, thank you for sorting everything – you're priceless. Nick, you are a friend and an invaluable support. Daniel, Rosy, and Sinead, I have enjoyed our discussions on life, research and everything. I could not have asked for better people to share this experience with; you are all most excellent scientists.

I would like to thank my family for the support you have provided to Louise, Luke and I. This process has not always been easy and your love and support has meant a lot. Mum and Dad, thank for encouraging me to persevere in life. Daniel and Annette, without your help with Luke this process would not have been possible. Lastly, Lesley, your encouragement and support have helped me immensely.

Finally, I would like to thank Louise. You are truly the most wonderful wife. Your love, support, and encouragement have kept me going. The hours have been long and moments have been stressful, but having you in my life has kept me sane, fed, and put academic drama into perspective. I love you. This thesis is dedicated to you and my beautiful son, Luke.

List of Figures

List of Tables

Chapter 1

General Introduction

Real-world perception is a multimodal experience, yet much of psychological research has investigated perception and attention of each modality in isolation (Fodor, 1983). This thesis investigates the interaction between vision and audition modalities specifically considering the influence of musical beats on eye movement timing. When listening to music with a strongly accented beat it can be difficult to resist the urge to move in time. With relative ease, humans can entrain physical movements to correspond with musical beats, even involuntarily moving in synchrony with it. The ability to entrain movements to music is (mostly) uniquely human, and universal across cultures (Brown & Jordania, 2013; Repp, 2005). In addition to facilitating creative expression, for example, dance and musical performance, the ability to entrain movement is evidence of a wider capacity to predict and align action to correspond with external events in time (Jones & Boltz, 1989).

Saccadic eye movements are one of the body's most frequent movements and are vital for orienting the limited visual perspective to produce a vivid and coherent percept of the world (Yarbus, 1967). Vision has a field of view that is limited to around 210º (Traquair & Dott, 1938), and detailed (high acuity) vision is further limited at around 4º (fovea), and degrades with eccentricity from the fovea (Polyak, 1941). To produce a smooth and broad visual percept of the world, where and when gaze orients varies subject to the spatial and temporal demands in the world (Findlay & Gilchrist, 2003). The current understanding of eye movement timing is limited, considering only visual and physiological factors, yet real-world perception is multimodal. For example, sensory information from auditory and visual modalities interact, redefining both what requires visual attention (i.e. *what* is relevant) in the world and when it is required (i.e. *when* it is relevant).

An aim of this thesis is to consider how predictable auditory information, specifically musical beats, influence the timing of eye movements. The bulk of research investigating audio-visual integration has measured behaviours (e.g. reaction time) and neural responses (event-related potentials from electroencephalography recordings) during passive vision tasks, such as when a centrally presented visual stimuli is combined with variations of auditory or visual spatiotemporal correspondence (Schröger & Widmann, 1998; Spence, 2011). These studies provide compelling evidence of early bimodal perceptual integration of aligned audio-visual onsets, but omit the dynamics of eye movements as mediators or facilitators of this integration. Similarly, within vision research there has been a notable emphasis on the spatial characteristics of visual information, i.e. where saccades orient to (Itti, Koch, & Niebur, 1998; Loftus & Mackworth, 1978). More recently, studies have considered the dynamics of eye movements (fixation durations) as a representation of the demands of visual information or task (Henderson, 2003; Henderson, Weeks, & Hollingworth, 1999; Nuthmann, 2016). Much less is known about whether auditory information informs or influences the dynamics of eye movements.

As there is little research considering the auditory influences on eye movement timing, there are many open questions. This paucity of research has motivated this thesis. Key questions addressed in this thesis are, first, do musical beats (as predictable auditory referents) implicitly influence the timing of eye movements, either altering them in duration or when they orient in relative phase? Secondly, considering the relative ease of synchronising other body movements to external auditory beats, can eye movements explicitly synchronise self-produced (gaze-contingent) visual onsets to correspond with simple musical beats? Finally, what limitations are there in the direct control of eye movements, both in how temporally precise they can be, and in the proportion of all ocular movements that can be controlled in time?

In order to address these questions, this thesis will review prominent eye movement timing models, which propose different mechanisms that modulate ocular timing. These mechanisms, for example a saccadic timer (Engbert, Nuthmann, Richter, & Kliegl, 2005; Nuthmann, Smith, Engbert, & Henderson, 2010) and direct control processes, will be proposed as candidates for auditory influence or entrainment. Additionally a review of sensorimotor synchronisation literature will detail the characteristics and mechanisms underlying synchronised behaviour to contrast motor with oculomotor movements.

1.2 Eye Movement Timing

Throughout our waking lives, human gaze reorients spatially with a ballistic movement (i.e. saccade) around three times per second (Findlay & Gilchrist, 2003; Rayner, 1998). Both cognitive and visual features can alter the duration of fixations, for example visually degraded images increase fixation durations (Henderson, Olejarczyk, Luke, & Schmidt, 2014; Wooding, Ruddock, & Mannan, 1995), and conducting a visual search generates shorter fixations than a memorisation task of the same scene (Henderson et al., 1999; Luke, Smith, Schmidt, & Henderson, 2014). Auditory information can influence the spatial orientation of eye movements (McDonald, Teder-Sälejärvi, & Hillyard, 2000), and enhance the salience of visual onsets when corresponding in time at fixation (Escoffier, Sheng, & Schirmer, 2010; Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010). Importantly, both the orientation decision and the relative salience of objects in the world influence the timing of eye movements (Tatler, Brockmole, & Carpenter, 2017). Despite the viability of audiovisual influence, the current understanding of ocular motor timing is limited to visual factors. In the ensuing overview, models of eye movement timing will be discussed with a focus on how they account for the visual, cognitive and physiological aspects of eye movement timing. The overview of these models is multipurpose; firstly, to consolidate the specific visual and task influences that can characterise ocular timing, with the intention to design a visual search paradigm that limits these visual factors. Secondly, to utilise this control of visual factors to identify potential mechanisms of eye movement timing that are susceptible to auditory entrainment or influence. Finally, to consider the limitations of direct control of eye movement timing as conceptualised across the different models, to motivate investigations that quantify this capacity.

1.2.1 Top-down control: direct, indirect and mixed control.

Saccadic movements are essential for aligning the limited high-acuity vision of the fovea with visual features in the world that are to be processed with the most detail (Findlay & Gilchrist, 2003). There is a physiological necessity to move, as in the absence of movement the retinal image fades disrupting visual perception (Riggs, Ratliff, Cornsweet, & Cornsweet, 1953). The sensory demands of the visual world and limitations of high acuity vision motivate saccadic movements, but the need to reorient is mediated by physiological constraints. Much of visual information is perceived during periods of relative stable gaze (fixations), and supressed during saccades (Bridgeman, Hendry, & Stark, 1975). For fixation durations to provide a smooth visual percept, the duration of the current fixation must remain stable long enough to sufficiently process the current and next saccadic target locations. This requires time to account for neural delays between the retina and the visual cortex (Poghosyan & Ioannides, 2007) and for target selection of the next saccade (Fuchs, et al., 1985). Beyond these low-level (bottom-up) constraints, the duration of fixations varies subject to perceptual and cognitive demands (top-down influences).

The top-down influences on eye movement timing have historically been conceptualised differently across three frameworks: direct, indirect, and mixed control. Firstly, direct-control models proposed that the duration of a fixation is the product of visual processing at the current location. For example, Rayner and McConkie, (1976), found that the processing difficulty of words was predictive of individual fixation durations when reading; specifically less frequent and more complex words had longer fixation durations. As such, the emphasis of direct-control is that there is a very close relationship between cognitive processing demands and the execution of saccades.

The second framework, indirect control, proposes that there is no relationship between the currently fixated visual information and the duration of fixations. Rather, fixation durations are modulated by the visual task and the physiological demands, for example the ‘eye-brain’ lag and neural delay due to saccade programming. Visual input from the retina requires around 50ms to reach the visual cortex (Poghosyan & Ioannides, 2007), and saccade programming requires between 100 and 150ms (Schall & Thompson, 1999), although there is some contention regarding the existence of a saccade program (e.g. Tatler, Brockmole, & Carpenter, 2017; this point is discussed in more detail in 1.2.7). These physiological constraints are considered notable contributors to the fixation duration distribution of eye movements, which averages between 200 and 300ms (Henderson, 2003; Rayner, 1998). From the indirect control perspective, a saccade is produced by an automatic process that is both task sensitive and a product of physiological constraint. Global demands of a task influence the distribution of fixation durations from which each fixation is automatically sampled irrespective of the currently fixated scene content, such as shortening all fixations in visual search or slowing them down during memorisation tasks. For example Hooge and Erkelens (1996) found when orienting around a Landolt-C array, between 20 to 35% of saccades were not in the correct direction (not the product of foveal processing or direct control), and that speeding the presentation time of visual information did not alter fixation durations. The indirect-control theory argues that fixation durations are controlled by a mechanism that predicts foveal demands based on the demands of previously processed stimuli, and importantly is not the product of currently fixated visual processing.

The third perspective is that of mixed control, where both the cognitive processing of currently fixated information and indirect control mechanisms influence the timing of eye movements. Mixed control is more prominent in recent models, for example SWIFT (Engbert et al., 2005), CRISP (Nuthmann et al., 2010) and ICAT, (Trukenbrod & Engbert, 2014). A key paradigm that provides evidence of mixed-control is the scene onset delay (SOD: Henderson & Pierce, 2008; Henderson & Smith, 2009), a modification of the classic stimuli onset delay in Morrison (1984). In the SOD participants view photographs of scenes under varying task conditions (memorisation or search), and during a saccade to a pre-specified crucial fixation the scene is replaced by a mask, which delays its onset. The duration of the mask is variant, and produces two distinct populations of fixation; the first sub-population varies as a function of the delay size and the second remains relatively constant irrespective of delay. If direct-control was acting on each crucial fixation, the duration of these fixations would be expected to all vary as a product of the delay as if they end before the scene returns they are essentially saccading to nothing. As only some fixations were subject to the visual delay, there is some limitation on the direct control of fixations, considered the product of an automatic process that is not subject to visual demands but is the random execution of saccades, or a ‘stochastic timer’ (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). The presence of two subpopulations within distributions of fixations, those fixations that are sensitive to the currently fixated information and those that are not, is strong evidence for mixed control.

To further expand on the conceptual differences in the mechanisms underpinning eye movement timing, this thesis will review six prominent models of oculomotor timing (sections 1.2.2 to 1.2.7). There are notable models omitted in this review (e.g. Adeli, Vitu, & Zelinsky, 2017; Yang & McConkie, 2001). The omission of these models is a function of brevity and not a reflection of their intellectual contribution. The models included each present different mechanisms that modulate timing (summarised in 1.2.8), which inform the predictions of the thesis.

1.2.2 Findlay and Walker model.

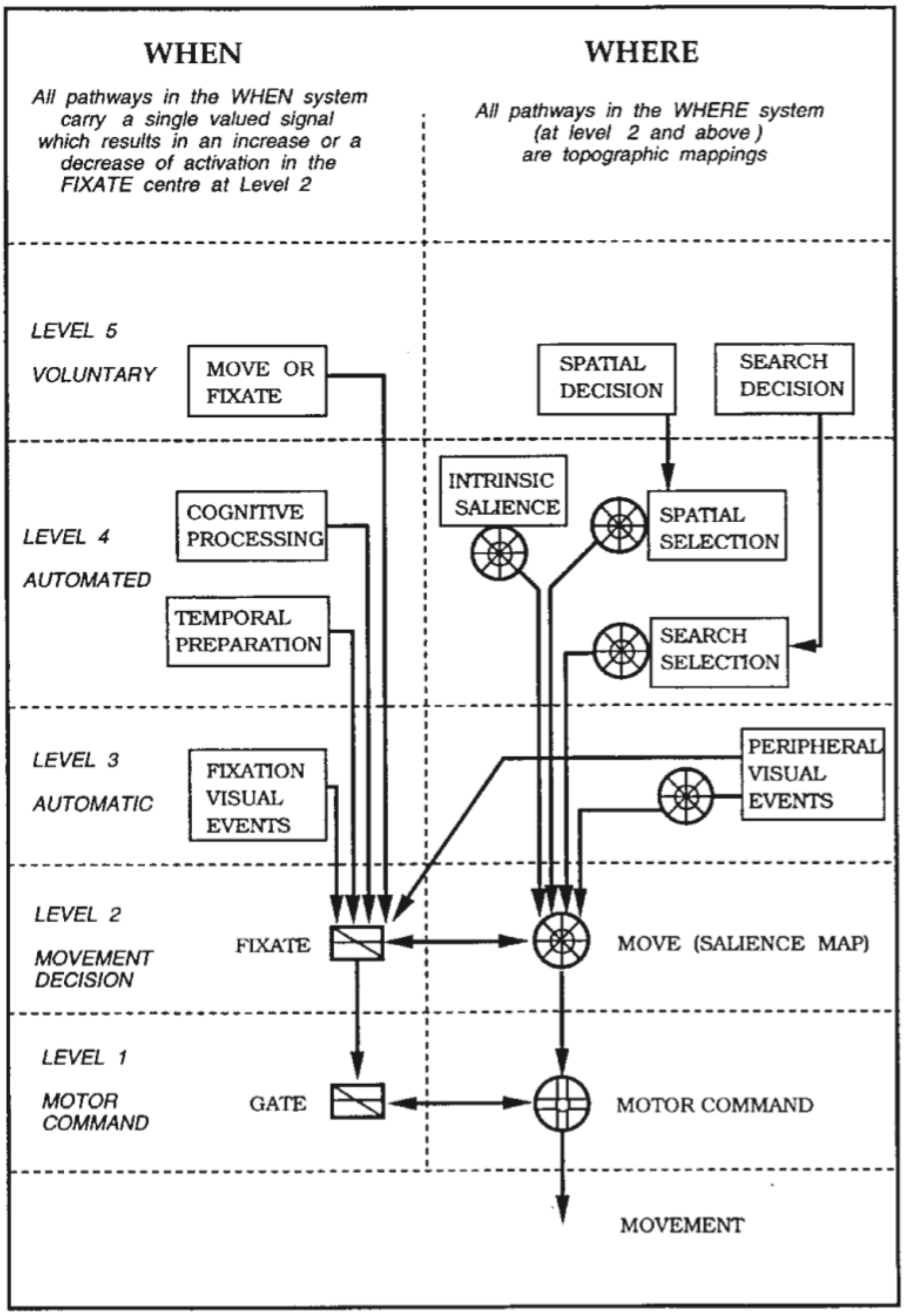


Figure 1-1 Findlay and Walker Framework of the information flow leading to saccade execution, from Findlay & Walker, (1999), p662

The Findlay and Walker, (1999) model of saccade generation conceptualised two parallel pathways for the When and the Where decisions in visual processing, based on physiological and behavioural evidence (see Figure 1-1). The interaction between the two pathways is conceptualised at the two final stages of processing. Firstly (at *Level 2)* during the movement decision, which is a competitor (push-pull) relationship between the salience of peripheral targets and the demands of the currently fixated information, whereby a saccade is executed when activity at the fixate centre drops to threshold. Secondly at the motor command (*Level 1*), which is the triggering of a saccade (conceptualised as a gate release), which has a target location informed by the 'Where' stream. Potential features of relevance to this thesis within the framework are the two higher-order influences that act on the 'When' pathway (*Level 4*). The first is a dynamic modulation of sequences of eye movements to ensure that the duration of fixations is sufficient to promote a smooth visual percept (sufficient time to allow for cognitive processing). Secondly, the authors argue that predictable targets allow for some temporal preparation, which promotes disengagement in the fixate centre. This could be interpreted as temporal prediction of visual targets modulating the timing profile of eye movements. The temporal prediction of visual targets is a potential mechanism by which predictable auditory information could modulate eye movement timing as the temporal processing of rhythmic visual information engages auditory processing (Repp & Penel, 2002). A final component of the model is voluntary control (level 5), which accommodates the volitional intent to saccade. The authors describe the voluntary overriding of low-level processes to trigger specifically timed saccades as unusual, a likely reason why the addition of volition to eye movement timing is unique to this framework and not accounted for in later conceptualisations (see below).

1.2.3 E-Z Reader model.

Much of the early quantification of eye movement timing stem from visual behaviours when reading. This is not surprising when you consider that reading presents a simple and commonly learned active vision task, that has quantifiable complexities (e.g. word frequency and length) and simple measures for perception (e.g. comprehension; Findlay & Gilchrist, 2003). A prominent model of eye movement timing when reading is the E-Z Reader model (Reichle, Pollastek, & Rayner, 2004; Rayner, Reichle, & Pollatsek, 1998; Reichle, Pollatsek, & Rayner, 2012; Reichle, Rayner, & Pollatsek, 1999). This evolving model is classified as a serial model, (*SAS*: sequential attention shifts / serial attention shifts), where cognitive processing is strictly sequential (one word at a time). In line with this serial shift, parafoveal processing (covert shifts of attention) of the next word can facilitate word jumps (fixations that skip a word in sequence), and allow shorter fixations on familiar words that have been previewed by parafoveal vision. Serial processing is a point of contention with other eye movement models, for example the SWIFT model (Engbert et al., 2005), which argues for the simultaneous processing of multiple words when reading. A notable component of the E-Z Reader model is that after a period that is just sufficient for lexical processing, the initiation of a saccade program begins. The saccade program itself is divided into two distinct stages, based on the findings from the double-step paradigm (Becker & Jürgens, 1979), and an earlier model of eye movement timing (Morrison, 1984). The two stages are a relatively short early stage when a saccade can be cancelled (*labile stage*), and a second later stage when cancellation can no-longer occur (*non-labile stage*). It is important to note the absence of a saccade timer underlying routines or the timing of eye movements. The EZ-Reader model proposes that the timing of eye movements is the product of direct-control, and represent an individual's lexical processing ability within ocular physiological constraints. The distribution of eye movement timings formed during reading are the product of moment-to-moment lexical decisions, the individual's experience with the reading material and their ability. EZ-Reader does not accommodate indirect or mixed-control of eye movement timing.

1.2.4 SWIFT model.

A second influential model of eye movement timing is the (Autonomous) **S**accade-Generation **W**ith **I**nhibition by **F**oveal **T**argets (SWIFT) model (Engbert et al., 2005). As previously mentioned, the SWIFT model does not conform to the serial processing of the E-Z Reader model, rather it incorporates a spatially distributed processing of the visual activation field, commonly considered a gradient model (*GAG*, guidance by attentional gradient). The visual activation field is the functionally useful field that surrounds fixation (detailed processing degrades with eccentricity from the foveal region). It is also called perceptual span (Rayner, Slattery, & Bélanger, 2010), and accounts for around 8° of the visual field from the centre of fixation. In practise the GAG facilitates the processing of multiple words within the visual field with a reduction in processing with increased eccentricity from foveal vision.

As in the Findlay and Walker (1999) framework, the SWIFT model also conforms to separate pathways for saccade timing and target selection. A novel component of the model is the addition of a random saccade generator, or timing mechanism, which approaches threshold using a random-walk model (a stepped rise in activation that can delay at any step prior to threshold) for word recognition. The authors note that the random-walk reaches threshold prior to full word comprehension. The distribution from which the timer operates is representative of that individual's predefined mean (varies by individual not by task). The random-walk can be inhibited (slowed) by foveal processing demands (e.g. word complexity, or a broad distribution of fixations) following a delay. The stochastic timer and inhibition of the random-walk to threshold are the main contributors to the production of random fixation durations. A commonality with the E-Z Reader model is the two-stage programming of saccades, the *labile* and *non-labile* stage. Interestingly, performing model simulations, the authors found that 90% of saccade programs were not cancelled and only 1.1% were cancelled twice, which suggest the main source of direct control variance in timing is inhibition (rather than cancellation). In summary, the two mechanisms of direct control are the inhibition of the random-walk and the cancellation of saccades during the labile stage, both of which can only result in a longer fixation duration. This is a variation from the Findlay & Walker (1999) model, where in addition to cognitive demand, temporal predictabilities could modulate the activation threshold at the fixation centre, promoting the "Move" signal to increase the speed of saccades when visual demands require (e.g. warning or predictable events). The absence of any account of temporal predictabilities in a reading model are unsurprising as reading does not normally involve transient visual information and the pace of visual information presentation is reader-dictated.

1.2.5 CRISP model.

The timer **C**ontrolled **R**andom-walk with **I**nhibition for **S**accade **P**lanning (CRISP) model (Nuthmann et al., 2010) is an extension of the mechanisms of the SWIFT model to account for eye movement timing during scene viewing. In contrast to reading, scene viewing presents a highly complex challenge for the modelling of visual behaviour, most notably because of the ambiguities in visual processing, for example defining salience at the pixel or object level (Borji & Itti, 2013; Malcolm & Shomstein, 2015), as well as the highly variant demands on visual processing (e.g. variance in colour, contrast, luminance, edges, object recognition, occlusions, and that saccade target candidates are from 360° of the visual field), when compared to reading (Henderson, 2003). The CRISP model avoids these complexities by addressing only the "When" stream, in line with the parallel processing conceptualisation of Findlay & Walker (1999). The core principles of CRISP are similar to SWIFT. There is a stochastic timer (although the timer may have some rhythmic connectivity with wider motor activations from the central nervous system (McAuley, Rothwell, & Marsden, 1999), that begins each random-walk to threshold, during which cognitive processing demands can inhibit (slow) throughout the random-walk. In contrast with SWIFT this engagement of inhibition in the random-walk is not subject to any delay, and allows continuous cross-talk between cognitive processing and saccadic pre-planning. A saccade program begins when the random-walk reaches threshold, and has two stages a *labile* stage (when cancellation of a program can occur), and a *non-labile* stage, following which the execution of a saccade is imminent. These two stages somewhat parallel the final two levels in Findlay & Walker (1999) model, the Movement Decision (level 2) parallels the *labile* stage, as the demands of the currently fixated location can cancel saccades, and the *non-labile* stage is the final execution of the Motor Command (level 1). A notable feature of the CRISP model is that random-walks can co-occur in time, as observed in the temporal scheme, Figure 1-2. The execution of a saccade occurs when the random-walk has reached threshold and the ensuing saccade program reaches its *non-labile* phase. As with the SWIFT model, the influence of direct control is either inhibition of the stepped transitions in the random-walk or the cancelation of the saccade, both of which are limited by the persistence of the autonomous timer (i.e. the random-walk cannot be delayed indefinitely).

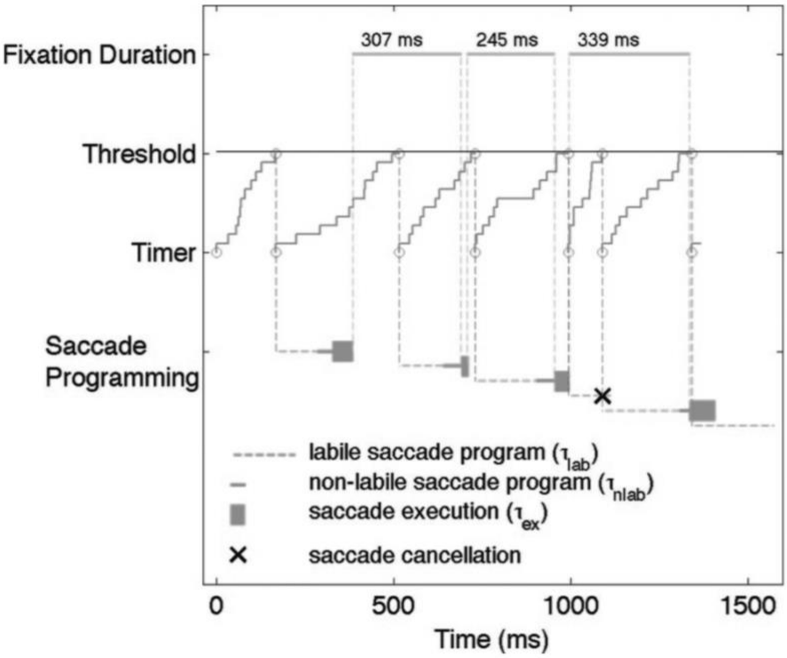


Figure 1-2 The temporal scheme of the saccade timing and programming of the CRISP model from Nuthmann et al., (2010), p386

1.2.6 ICAT model.

A further framework building on SWIFT and CRISP is the **I**nhibitory **C**ontrol with **A**daptive **T**imer (ICAT) model by Trukenbrod & Engbert (2014). For the most part, the ICAT model conforms to the mixed control with stochastic timer components of SWIFT and CRISP, but differs in how the timing distribution for the random-walk operates. Both SWIFT and CRISP describe the distribution underlying the random-walk as individually variant and a product of physiological constraint (e.g. eye brain lag and saccade programming) that deviates around an optimal mean duration that has sufficient time for these constraints and the demands of cognitive processing (inhibition of the random-walk when required) of current visual information. Deviation in timing within SWIFT and CRISP is the product of inhibition in the random-walk due to the current visual processing demands. The ICAT model proposes the addition of a global modulator of this timing distribution, which accumulates evidence of processing difficulty in previous fixations to modulate the timing distribution subject to anticipated demands. The evidence for this is based on research by Hooge & Erkelens (1996, 1998), who in addition to identifying mixed-control behaviour in saccades (only some eye movements were controlled), found that eye movement timings did not immediately adapt to changes in visual presentation time, which they argue is evidence of previous timing estimations determining saccade timing. This is conceptualised in ICAT as a global (lagged) estimator of future demands that shifts the distribution of the saccade timer, which is adding further bias towards indirect control in the distribution of movements. The addition of this global estimator is difficult to tease from the existing inhibition mechanisms.

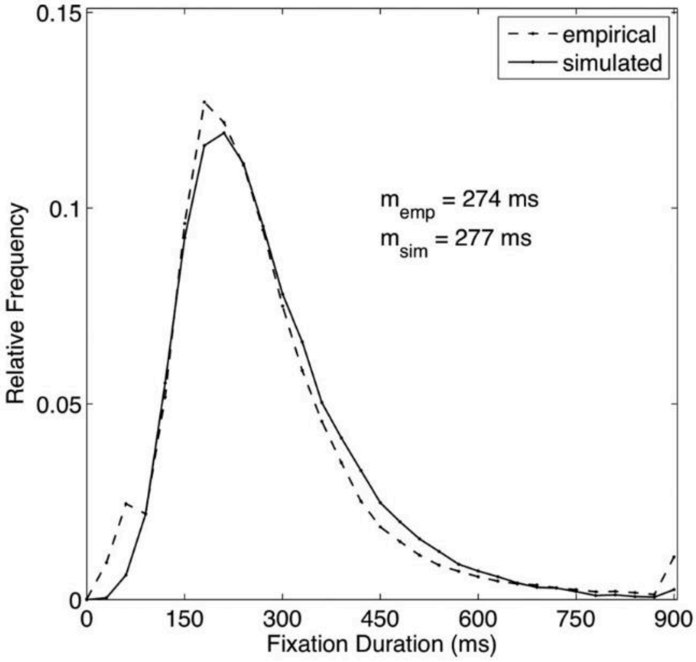


Figure 1-3 An example of both common empirical and CRISP model simulated fixation duration distributions, demonstrating the ex-Gaussian distribution with both a normal distribution component and an exponential tail, sourced from Nuthmann et al. (2010), p388

Furthermore, irrespective of visual demands eye movement timings consistently display an ex-Gaussian type distribution (e.g. Figure 1-3), with fixations that are both too short and notably too long for the current or prior processing demands (Findlay & Gilchrist, 2003; Henderson, 2003; Rayner et al., 1998). A key function of this global timer would generally be inhibitory, due to the lower threshold physiological constraints on timing. In the context of an ex-Gaussian distribution, this global component would shift in the mean of the normal distribution (μ, the initial peak), rather than the exponential component (τ, the positive skew), which has previously been described as exhibiting inhibition in the random-walk. (de Urabain, Nuthmann, Johnson, & Smith, 2017).

1.2.7 LATEST model.

A recent model of eye movements is the **L**inear **A**pproach to **T**hreshold **E**xplaining **S**pace and **T**ime (LATEST) model by Tatler, Brockmole, and Carpenter (2017), a modification of the LATER model (Carpenter, Reddi, & Anderson, 2009; Noorani & Carpenter, 2016). The novel application of this scene viewing model is that it promotes a continuous interactive relationship between the where and when pathways, conceptualised as decisions. The inclusion of the two when / where pathways were considered parallel streams in the Findlay and Walker (1999) framework, with minimal interaction. This separation has some consensus and much of the current research into visual behaviour has conceptualised spatial evaluations separately from movement decisions (Engbert et al., 2005; Itti et al., 1998; Itti & Koch, 2001; Nuthmann et al., 2010). The LATEST model proposes a much more integrated decision stream that simultaneously emulates the when and where as a combined evaluation. The four fundamental components of the LATEST model are: first, that the decision to saccade is the product of peripheral evaluations, or evidence, to support a movement away from the currently fixated location. Secondly, movement decisions are the product of multiple 'stay-go' evaluations of peripheral targets that simultaneously weigh the need to remain in the currently fixated location against the evidence for saccade orientation to an evaluated target. The strength of a go decision in the periphery is represented in the model by the steepness of the slope in the rate of rise for the decision to go (at threshold). Thirdly, the duration of each fixation is equivalent to the decision time for the winning 'stay-go' evaluation. A saccade to a specific location is the product of the winning peripheral go decision (the steepest rate of rise). Importantly, and in contrast with the CRISP, SWIFT and ICAT models (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014), the duration of fixations are not subject to either two-stage saccade programming or an overarching saccade timer. The winning peripheral location (saccade target) is considered the product of the visual evaluation with the steepest 'rate of rise' (decision time), resulting in a saccade that is immediately executed in a winner takes all fashion. All task or visual processing variation in saccade execution, both global or local, is accounted for in the rate of rise. The fourth component accounts for what the authors describe as a subpopulation (~10%) of fixation durations, which are less than 100ms in duration. These short fixations are deemed incapable of representing 'stay-go' decisions, simply because they are too short to allow for sufficient evaluation of the current and peripheral information. These short fixations are classified as the product of a maverick saccade generator, which is in competition with the main decision process, and serves to promote randomness in visual orientation. This randomness is argued as an adaptive biologically advantageous neural process, as in Carpenter (1999). Therefore, the variation in eye movement timing is both the product of a maverick generator which occasionally wins against the main stay-go evaluations (around 10% of the time), and mostly the product of competition across stay-go evaluations over multiple peripheral targets. The apparent randomness in the timing of saccades in the main component is directly related to current visual processing demands, indicative of direct control.

1.2.8 Model summary.

Table 1-1 Feature comparison of the Findlay and Walker (1999), EZ-Reader (Reichle et al., 2012), SWIFT (Engbert et al., 2005), CRISP (Nuthmann et al., 2010), ICAT (Trukenbrod & Engbert, 2014) and LATEST (Tatler et al., 2017) Eye Movement Timing Models. The Source of Randomisation factor details what mechanism within the model accounts for apparent randomness in the duration of fixations. The Saccade Program feature contrasts the inclusion of a two-stage (*labile* and *non-labile*) program, which is a short period prior to a saccade that facilitates either the cancelation or execution of the saccade. The Top-Down Influence factor contrasts how visual and task features impact on fixation durations. The Volitional Intent factor contrasts the presence of voluntary control mechanism (operationally defined here as a decision irrespective of visual demands) as an independent factor within the model

	Source of Randomisation	**Saccade Program**	**Top-Down Influence**	**Volitional Intent**
Findlay and Walker	None	Yes	Direct	Yes
EZ-Reader	None	Yes	Direct	No
SWIFT	Stochastic timer	Yes	Mixed (with delay in direct control)	No
CRISP	Stochastic timer	Yes	Mixed	No
ICAT	Stochastic timer	Yes	Mixed with Indirect Global Estimation	No
LATEST	Sub-component: maverick	No	Direct (for main component)	No

The purpose of reviewing these eye movement timing models was firstly to consider visual and task influences on ocular timing. Secondly, to identify mechanisms that are candidates for influence by auditory information. Finally, to consider existing accounts of volitional control (specifically the control of each eye movement at the local level). Across each of the eye movement timing models there is agreement that both visual and cognitive influences alter the timing of eye movements, either at the point of fixation (direct control), at a more global level (indirect control), or both (mixed control). The generation of a visual task that intends to limit variance in eye movements would need to reduce visual information at fixation and in the periphery (Clarke & Tatler, 2014; Tatler et al., 2017), avoid ambiguity in saccade targets (Tatler & Vincent, 2008), limit return to the same spatial location (Klein, 2000), produce highly predictable movement sequences (Hooge & Erkelens, 1998), and have minimal task demands (Nuthmann et al., 2010; Henderson & Pierce, 2008), as each of these factors have been shown to account for some top-down delay in eye movement timing.

Auditory information can alter visual perceptions, especially if the auditory information is rhythmic (Repp & Penel, 2002b), or corresponds in time (McGurk & MacDonald, 1976; Quigley, Onat, Harding, & Cooke, 2008; Thurlow & Jack, 1973). As summarised in Table 1-1, there are notable conceptualisation differences in the proposed mechanisms underlying eye-movement timing across the models. The more cognitive models (Reichle et al., 2012; Tatler et al., 2017) ascribe to direct-control, with emphasis on visual processing during fixation as the main modulator or inhibitor of saccadic movement. The mixed-control models argue for some limitation on direct control, i.e. that not all fixations represent processing demands of the current fixation. But under both direct and mixed control theories, eye movements respond to increases in visual or task complexity with inhibition, increasing fixation durations. It is plausible that auditory information could both increase and reduce moment to moment visual processing demands at the point of fixation (through modulating fixated or peripheral visual salience, or perceptual ambiguity), which would both increase or reduce inhibition respectively. The existence of a saccadic timer and saccade program is contentious (Reichle et al., 2012; Tatler et al., 2017), though there is some agreement that a random process (e.g. a stochastic timer or a maverick competitor) mediates the systematic delay of physiological and visual processing demands. Although, the distribution of the timer may shift subject to a global estimator of cognitive demand (Trukenbrod & Engbert, 2014).

A potential limitation in the attribution of randomness to the timer, is that each of these timing frameworks are modelled on viewing behaviour to static stimuli, yet saccadic movements are highly sensitive to dynamic visual information (Fischer & Boch, 1983). Both the temporal dynamics of attention (Jones & Boltz, 1989; Large & Jones, 1999) and dynamic audio-visual information (Mital, Smith, Hill, & Henderson, 2010; Smith & Mital, 2013) are all potential contributors to eye movement timing, and may account for some of the variation in the proposed random mechanisms. Fixation durations and orienting decisions have been shown to be influenced by auditory information within dynamic scenes (Coutrot & Guyader, 2013; Coutrot, Guyader, Ionescu, & Caplier, 2014; Smith, 2014). Whilst outside of the remit of this thesis, other research conducted during this PhD investigated auditory influences (film sound-design) on eye movements during film viewing (included within appendix 1), and was published recently (Batten & Smith, 2018). These studies identify significant influence from auditory information on the spatial orientation of eye movements to dynamic scenes. It is plausible therefore, that the saccade timer is susceptible to the entraining influence of musical beats, especially if it corresponds with dynamic visual onsets.

Only the Findlay and Walker model accounts for voluntary eye movements as a *uncommon* factor in when the eye moves (Findlay & Walker, 1999). There is an abundance of evidence in psychological tasks that the volitional control of eye movements is possible, for example the anti-saccade paradigm (Munoz & Everling, 2004). Participants are tasked to orient their eye movements either toward or away from a visual onset, i.e. a voluntary eye movements away from a visually salient target. Additionally, human gaze shifts can be a social signal (Friesen & Kingstone, 1998; Langton, Watt, & Bruce, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Senju & Johnson, 2009), which inherently requires some voluntary control over ocular timing irrespective of the currently fixated visual information. The capacity for volitional intent as either direct control or a separate entity within eye movements is not currently quantified within the more recent eye movement timing models. Unlike other direct control mechanisms, volition can both promote or inhibit movements irrespective of the currently fixated visual information. A motivation of this thesis is to measure the limitations of volitional eye movement control (controlling every eye movement in sequence). In the light of the differences in top-down control as either mixed-control (that only some eye movements are subject to top-down influence) or direct-control (that top-down factors influence each movement), evidence for these differences would be quantifiable in the distributions. For example, as in LATEST (Tatler et al., 2017), where the main component is subject to direct control, voluntary inhibition would shift most of the distribution as a function of the inhibition period (a unimodal shift in duration by promoting the “stay” decision). Under mixed control (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014), two distributions would emerge. As the period of inhibition increases, the movements that are subject to direct control would form a second distribution, as seen in the SOD paradigm (Henderson & Smith, 2009), where the proportion of fixations in the second distribution would inform the capacity for direct control. Through the use of an auditory onset referent the temporal precision and predictive nature of voluntary eye movement control could also be quantified. The nature of such an auditory onset and how it has been shown to influence the timing of other sensory-motor processes will now be reviewed.

1.3 Sensory Motor Synchronisation

There is a widely-researched ability in humans to correspond the timing of movement with an external rhythm, commonly classified as sensorimotor synchronisation (SMS). SMS underlies the human capacity for musical performance, co-ordinated dance, and within scientific research is generally observed in the synchronisation of finger-taps with a metronomic sequence (Repp, 2011; Repp & Su, 2013). There is general consensus that most humans can produce synchronised movements (McNeill, 1995; Repp, 2005), and the evidence for SMS emerges relatively early in human development (Phillips-Silver & Trainor, 2007; 2008). The prevalence of SMS is also notable in the rarity of its deficit, ‘beat deafness’, which when identified in an individual has motivated a case study, for example, Bégel et al. (2017). For the purpose of brevity, this review of SMS research will focus on features of SMS that directly relate to the later eye movements studies. Specifically, the differences between auditory and visual rhythms, error correction processes, predictive responses, musical expertise, music perception, implicit entrainment, and the capacity for explicit and implicit SMS of saccadic movements. For a more detailed overview of SMS literature, there are (at least) two comprehensive reviews, the first published in 2005 (Repp, 2005) and the second in 2013 (Repp & Su, 2013).

1.3.1 Auditory and visual rhythms.

Synchronisation performance is enhanced when tapping in time to an auditory rather than a visual sequence. Tapping performance to an auditory sequence can reproduce faster intervals than visual (non-moving) sequences (Repp, 2003). The variability of synchronised responses is greater to visual stimuli (Patel, Iversen, Chen, & Repp, 2005), and becomes more pronounced when the timed interval exceeds 3Hz (Chen, Repp, & Patel, 2002). Employing a moving visual sequence does reduce some of the modality difference in variability, for example, tapping to referent finger movements (Hove, Spivey, & Krumhansl, 2010), or to a bouncing ball (Hove, Iversen, Zhang, & Repp, 2012). The difference in temporal capacity in vision is somewhat unsurprising considering the differences in temporal processing demands between visual and audition modalities is directly a product of the two sensory systems. Sound is perceived through changes in air pressure over time, whereas visual information is largely a product of the difference in photon absorption across the retina. The perception of auditory information requires the binding of a continuous stream of auditory features through time by temporally grouping sound based on phonic similarities (Bregman, 1990). In contrast, a core function of visual perception is processing changes in brightness projected spatially across the retina, in order to identify edges and to bind these spatial components together to form objects (Marr, 1982). Therefore, whilst there are both spatial and temporal components to both vision and audition, there is a clear dominance in audition for temporal processing and a spatial dominance in vision. Audition and vision interact to produce coherent precept of the world though the binding of spatio-temporal information. Temporal ambiguities (onset differences between vision and audition) are generally aligned to the auditory percept and spatial ambiguities (sound source location or position) are aligned to the visual (Thurlow & Jack, 1973; Watanabe & Shimojo, 2001), although as observed in the widely known ‘McGurk effect’, a novel percept as a fusion of the two can occur (McGurk & MacDonald, 1976). There is evidence that the temporal perception of visual rhythms engages auditory processing. Repp and Penel (2002) found that when tapping to a visual rhythm the addition of auditory distractor sequence notably impaired performance, yet the inverse, a visual distractor sequence with an auditory referent, did not. The error correction of tap responses to visual sequences is also influenced by auditory distractors, with the same pattern of unidirectional impairment (Kato & Konishi, 2006). Beyond simply resolving temporal ambiguity between the modalities, the entrainment of rhythmic movement (prediction of rhythmic action) and the employment of error correction is reliant on auditory processing irrespective of the referent modality.

1.3.2 Error correction.

The error correction processes of SMS are essential for the timed control of rhythmic movement. In their absence, small misalignments would compound and eliminate any relation between the movement and external rhythm (Vorberg & Wing, 1996). The general consensus is that there are two independent mechanisms that facilitate alignment of movement with an external interval, *period correction* and *phase correction* (Repp & Su, 2013). Period correction is considered an intentional shift in the interval between sequential movements to align them with the phase duration. Secondly, phase correction, which is considered an automatic process, utilises sensory feedback to align movements to cluster responses in phase, for example, landing the tap of a finger at the point of tone onset (Repp, 2005). There are various models that account for the variability and error correction in responses. A seminal model by Wing and Kristofferson (Wing & Kristofferson, 1973) conceptualised two sources of variance. Firstly, the motor implementation is subject to some processing delay in the execution and perception of movement (a source of negative mean asynchrony, which will be discussed in section 1.3.3.). Secondly, a timekeeper which is conceptualised as a clock and varies as a product of motor movements (in self-paced tapping) and external rhythms (combined with a feedback process during SMS). The timekeeper is the main source of variance in

SMS responses (the motor implementation is considered a constant delay), and both period and phase error corrections modulate the timekeeper. These ideas were extended by Mates (1994), to incorporate a dual-process error correction model, where period correction alters the timekeeper to align in duration, and phase correction leaves the internal timekeeper unchanged. These dual processes are the product of two perceptual mismatches, phase correction is the percept of asynchrony and involves some automatic alignment, and period correction is the product of mismatch between the external referent in the internal timer and is resolved through direct control (intention). A final alternate model is the Dynamic Attending Theory (DAT) by Jones and Boltz (1989), which rather than utilising an internal timekeeper, considers attention as energy that is entrained by the external sequence. Within the DAT model, attention ebbs and flows based on the temporal demands of the external stimuli. The entrainment of attention is therefore the allocation (a burst of energy) of attention at predicted moments and correction in phase and period is the modulation of temporal expectation based on perceptual mismatch (altering the ebb and flow). Potential evidence for DAT is observed in the connectivity and bursts of activation in neural motor areas when perceiving external rhythms (in the absence of production), reflecting both the sensitivity of attention to external temporal dynamics, and the time-sensitive allocation of attention as readiness for future action (Alluri et al., 2012; De Freitas, Liverence, & Scholl, 2014). In summary, the production of SMS responses relies on corrections in both period and phase, which are the product of either perceived asynchronies with an internal timekeeper, or the modulation of a supra-modal attentional resource that fluctuates energy allocation subject to temporal demands.

1.3.3 Negative mean asynchrony.

A notably feature of SMS is that participants' response times tend to cluster prior to the onset of stimuli in phase, a behaviour coined negative mean asynchrony (NMA). This is considered evidence that when synchronising participants anticipate rather than react to beat onsets (Mates, Müller, Radil, & Pöppel, 1994; Repp, 2005). Although as previously noted by Repp and others, any response that occurs less than 150ms after the beat is also anticipatory due to the inherent delays in movement planning (Mates et al., 1994; Repp & Su, 2013). The delay in NMA tap responses tends to increase linearly as the inter-onset interval (IOI) increases (Repp, London, & Keller, 2008), and the addition of subdivision referent tones reduces the NMA by reducing the required interval of estimation (Loehr, Large, & Palmer, 2011). When the tap interval is held constant and the referent IOI duration is slowed, the NMA increases (Zendel, Ross, & Fujioka, 2011). The NMA suggests perceptual underestimation which is particularly notable during a pseudo-SMS task. During this task at a predefined point a IOI referent is replaced by a tone representation of the subjects tap onset. Participants, during the self-paced tap responses accelerate away from the previously entrained IOI, consistently underestimating themselves (Flach, 2005). As such, the NMA is likely a product of perceptual underestimation, with temporal prediction errors increasing linearly as a function of the estimated interval (as described by Weber's law). The function of this underestimation is argued to be an accommodation for delay in the integration of sensory information (Drewing, 2012), although it may also be a product of variability in the neural response to sound (Tierney & Kraus, 2013b).

1.3.4 Musical expertise.

Training and expertise influence SMS performance; there is an abundance of studies that have compared musicians to non-musicians. Adult musicians are generally more synchronised (less variable) than adult non-musicians (Repp, 2005). When tapping to a very slow beat (> 2 seconds), non-musicians switch from anticipating the beat (a NMA) to reacting to it, while musicians do not (they are continue predicting) and are more accurate (Repp & Doggett, 2007). The type of musical expertise matters. Krause, Pollok, and Schnitzler (2010), compared percussionists to other musicians and non-musicians and found that percussionists have both a smaller NMA (closer to the IOI) and enhanced perceptual sensitivity to the beat interval than both pianists, singers and non-musicians. When musical training begins in the life-span also matters; Bailey & Penhune (2010) compared tapping ability of those who had musical training prior to 7-years old to those who trained later (matching for musical training and experience), and found that early-trained musicians had a less variable SMS (which they argue is evidence of critical periods in development). Recent auditory brain-stem response studies (ABR) have investigated the neural correlates of SMS (Kraus, 2011; Tierney & Kraus, 2013a; 2016). A key finding from these studies is that variability in SMS performance is related to early neural responses to sound. Participants with a closer neural phase coherence are both faster and more able to synchronise motor responses. Musical training with a specific focus on temporal synchronisation is argued to increase this ABR phase coherence, which may directly transfer to counter other linguistic impairments related to temporal dynamics of sound (Tierney & Kraus, 2013a). In summary, those with musical training (especially expertise in rhythmic actions), are less variable, closer to the IOI, and show a closer neural representation of the temporal dynamics of sound when synchronising movement to external rhythms.

1.3.5 Music perception and groove

Sensorimotor synchronisation to a musical beat differs from corresponding movements to a simple metronomic auditory sequence. Music is a spectotemporally complex stimulus and perception of it requires segregating and monitoring differences in multiple sound attributes through time (e.g. timbre, tone, tonality and metrical structures). This complexity is integral to the enjoyment of music, with temporal expectation and violations of expectation engaging emotion and arousal systems in the brain (Alluri et al., 2012; Koelsch, Fritz, & Schlaug, 2008). The perception of a musical beat often termed the 'pulse' of music (as opposed to the metrical structure or rhythm of metronomic stimuli), is an internal representation of a period structure in the music commonly represented in motor action, for example, tapping a finger or foot (Toiviainen, Luck, & Thompson, 2009). To perceive a music beat requires converting temporal information from an external timekeeper (e.g. regularities in a drum pattern) into an internal temporal representation. This internal representation often corresponds with accents in the music (especially when isochronous), although the volitional determination of *what* and *when* forms this pulse can alter the percept of the musical stimulus by altering attention through time to musical attributes (Patel & Iversen, 2014). The previously described error-correction processes for SMS are also utilised in musical beat perception, but the product of synchronisation is a phenomena unique to music: the percept of groove as a positive affect that occurs when SMS is aligned to music (Janata, Tomic, & Haberman, 2012). The feeling of groove differs across genre and songs and is most strongly associated with a compulsion to move and a positive feelings. Importantly the phenomena of groove is commonplace in both musicians and non-musicians (Janata et al., 2012). In summary, musical perception engages a wide network of neural responses, including those in auditory, motor, emotion and arousal regions. The product of this activation is the phenomena of groove, a pleasant affect that occurs when moving to an underlying pulse or beat internalised from the music.

1.3.6 Implicit SMS.

When previously discussing error correction, period correction was termed an intentional behaviour that explicitly aligns perceived production asynchronies to an external rhythm (the phase error correction is automatic), as detailed in Repp (2005). The requirement of intention in synchronisation is contentious, as in the absence of instruction implicit entrainment of movements can occur. During a line-drawing task with irrelevant background music, Kuribayashi & Nittono, (2014) found that the pace of line-drawing increased as the tempi of the music increased. Similarly, implicit physiological responses have shown sensitivity to musical tempi; both heart-rate and respiration intervals have increased when the tempi of music increases (Khalfa, Roy, Rainville, Dalla Bella, & Peretz, 2008). Finally, during a walking experiment that manipulated the adaptive nature of music to participants gait speed, Moens et al. (2014) showed evidence of implicit synchrony in gait; both correcting the period and phase (error correction behaviours) adapted participants' gait timing to align with the music. This walking behaviour is evidence of implicit error correction altering gait timing to synchronise to musical beats. Importantly, implicit entrainment did require a starting interval in the music that was closely related to the natural gait speed. The facilitation of implicit entrainment is likely the product of interplay between auditory and motor cortices (Alluri et al., 2012). This integration with the motor cortex is a likely cause for why the percept of a beat (tactus) is one of movement and not simply an auditory one (Stupacher, Hove, Novembre, Schütz-Bosbach, & Keller, 2013). A recent study by Leow, Waclawik, & Grahn, (2017) argued that intention and attention (to the music) are required for the synchronisation of gait. This study compared gait entrainment with and without instruction to synchronise. In both conditions there was some synchronisation, although performance when explicitly instructed was much more prevalent. Furthermore, attention to the beat of the music was also required to synchronise gait during this task; synchronisation was much worse when distracted. There is therefore mixed evidence for implicit entrainment, as musical beats can entrain movement but the effect may be limited in the absence of attention or intention to synchronise to the beat.

The dynamic attending theory (Jones & Boltz, 1989) argues that regular temporal cues are used to predict future events. This predictive ability facilitates or entrains the dynamic modulation of cognitive resource to enhance predicted ‘salient’ moments in time. This modulation allocates ‘energy’ or attention in time. As the locus of visual attention is highly correlated with where gaze is allocated (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995) and eye movement is subject to some voluntary control, the entrainment of visual attention to a musical beat should coincide with saccadic movements that serve to optimise the perceptual processing of visual information, ultimately regulating the timing of eye movements to synchronise with a predictable audio-visual beat. Evidence for implicit entrainment of visual attention has been observed in studies that have found an on-beat perceptual advantage for visual onsets that are aligned to a predictable external auditory beat. A study by Van der Burg, Olivers, Bronkhorst, and Theeuwes, (2008) demonstrated an increase in the perceived visual salience of a target within a search array when the presentation synchronised with an auditory tone (a ‘pip). The enhanced visual salience (a ‘pop out’) was evident as quicker detection of an ambiguous target. A second visual discrimination study by Escoffier, Sheng, and Schirmer, (2010) required speeded identification of whether an image of a face or house was upright or inverted. The task was conducted with background music or in silence and when presented with music the presentation of the images was either in or out of synchrony with the beat. The authors observed that reaction times to images presented on the musical beat were faster than those presented offbeat or in silence, which they argue is the entrainment of visual attention to the musical beat, altering salience through time. The dynamic influence of entrained visual attention on eye movements would be most apparent when gaze shifts are required to access the timed presentation of sequential visual information. Therefore, gaze-contingent sequential presentation could test whether eye movement timing is actively modulated by the entrained dynamics of visual attention (which is sensitive to musical beat), and if this is employed to enhance visual perception at predictable moments in time.

1.3.7 Saccadic SMS.

The previously discussed SMS research has almost entirely been compiled from finger-tap responses to various types of external rhythm. The body movements of interest for this thesis are eye movements (saccades). There are notable differences between finger movements in SMS and saccades. Finger movements involve flexion and extension phases followed by a dwell period. These distinct phases allow for different strategies for synchronised movement, for example, speeding the downward movement (emphasis on the flexion phase), or a strategy observed in percussionists, in speeding both the flexion and extension phases (Krause et al., 2010). Saccades in contrast are considered ballistic movements (Fuchs et al., 1985), which can only be directly controlled during the ‘dwell’ phase (i.e. during fixation). For example, once a saccade has been executed it can no longer be cancelled (Becker & Jürgens, 1979). A series of studies have measured the timing of horizontal saccades between two visual targets. Shelhamer and Joiner (2003) manipulated the frequency of alternating horizontal visual targets (15° apart), and measured the relationship between target orienting saccades and a visual onset. When the interval required a spatial shift faster than 500ms the saccadic timing was reactionary, with no evidence of temporal (corrective) relationship between the saccades. The saccades to intervals greater than 500ms (the authors argue for a switching point) were notably predictive and evidenced corrective behaviour (period and phase shifts). A second study by Joiner, Lee, Lasker, and Shelhamer (2007a) extended this research with auditory stimuli as the referent signal. The authors found that saccades between fixed horizontal visual targets could be entrained to an auditory referent with intervals between 500 and 1000ms. A final study by Hornof and Vessey (2011) contrasted audible representations of eye movements (tone onsets at first fixation to spatial locations) as they oriented between fixed horizontal visual targets, with the intention of identifying the most representative temporal representation of the eye movement. The authors found that a fixation based auditory referent was most accurate (tones at fixation start times), and that the eyes could orient between targets at 500ms, but notably dropped in performance with 250ms intervals. Taken together, both visual and auditory intervals can inform saccadic movements between two horizontal targets. A notable omission in these studies is the profile of eye movements during the inter-saccade interval. Unlike finger-taps the eyes must move (Martinez-Conde, Otero-Millan, & Macknik, 2013), and a notable unanswered question is whether the auditory and or visual interval modulated or entrained the timing of all of the eye movements, or is limited to the volitionally controlled shift between targets. A further consideration is the horizontal task, as saccades can certainly shift between two targets as previously described. Although, some of the variance in saccadic behaviour may be attributed to inhibition of return, or a reflexive shift of visual attention away from previously attended targets (Klein, 2000). Performance on these sub-500ms intervals may be enhanced were the visual task orienting around a sequential array of multiple targets, rather than alternating between two targets.

There is some evidence of implicit effects of music on the timing of eye movements in less constrained visual environments. Schäfer and Fachner (2014) found that both when viewing a dynamic video (driving along a country road) and a static scene (an external view of a house and garden), the presence of music reduced fixation durations. The authors account for this as an arousal response, i.e. music increased arousal which speeded eye movements. Wallengren and Strukelj (2015) conducted a pilot study with variations in film score compared across film clips and found a reduction in fixation durations to faster paced (more intense) musical scores. A final study by Day, Lin, Huang, and Chuang (2009), measured eye movement timing during a decision making task with irrelevant music. The authors identified a shortening of fixation durations as the tempo of the music increased, suggestive of some implicit period correction (although the fixation durations did not match the tempi). There is therefore some evidence that musical beats can influence the timing of eye movements both explicitly and, if visual controls were implemented to remove visual sources of variance, implicitly.

1.3.8 Summary of SMS.

The intention of this review of SMS literature is to motivate studies that measure the implicit and explicit entrainment of eye movements to simple musical beats. Much of the literature has considered the production of synchronised finger-taps to external rhythms (Repp & Su, 2013). These studies have identified an auditory dominance over visual rhythms in the production, error correction, and rate-limits of synchronised movements; with convincing evidence that the synchronisation to a visual rhythm relies on auditory processing (Repp & Penel, 2002). Finger-taps are subject to much greater control during their movement trajectory than are saccades (which are ballistic movements). A feature of error correction models in SMS, for example Mates (1994) as well as Wing and Kristofferson, (1973), is that there is a systematic (constant) motor delay and that variance in SMS production is the product of the timekeeper. This assumption does not reconcile with the current understanding of saccadic movements, which in addition to systematic physiological delays may also be influenced by a random timer or maverick

(Engbert et al., 2005; Nuthmann et al., 2010; Tatler, et al., 2017), and have notable limitations in the number of movements in a sequence that can be directly or voluntarily controlled (Henderson et al., 2009; Morrison, 1984, Munoz & Everling, 2004). Despite these modality limitations which would be expected to impair the synchronisation of saccadic movements, eye movements to horizontal targets can produce a synchronised response at relatively slow intervals (greater than 500ms) displaying evidence of anticipation or prediction (NMA). Spatial shifts of gaze can align with an external auditory rhythm, although these studies (in contrast to the finger tap studies) have erroneous movements which are ignored within the analysis and do not involve musical stimuli which could enhance the compulsion to move. A more representative timing measure of eye movements during these tasks would be to include the interval between saccades rather than the interval between spatial targets, as the former directly represents the capacity for direct (voluntary) control and the later does not.

1.4 Conclusion and Research Aims

This thesis aims to further the current understanding of eye movement timing to incorporate auditory information; specifically investigating the implicit and explicit capacity for musical beats to influence (and entrain) eye movements, and to quantify the capacity and limitations of direct control when volitionally matching each eye movement to auditory onsets.

A motivation of the review of eye movement literature was to identify visual and task factors that influence eye movement timing with the aim of producing a visual paradigm that limits the visual variance to enhance any auditory influence. These factors included: i) the level of information at fixation and in the periphery, which motivate a simple and constant visual discrimination, ii) controlling for the timing of visual onsets by employing a gaze-contingent display, iii) removing ambiguity in the saccadic decisions regarding the size and location of the next saccade by keeping them constant in and predictably sequential in location, and iv) limiting the need to return to the previous saccade location by using multiple sequential targets. By limiting these visual and task factors, the profile of the saccade timer (or the slope of stay-go decisions) will be relatively enhanced (unimpeded) within the distribution, as the visual and task reasons for direct (or indirect) control at any fixation will be minimised.

Within chapters 1, 2 and 3, the main visual search paradigm developed during this thesis project will be refined and the implicit influence of musical beats on the timing of eye movements will be quantified, measuring both period and phase corrective behaviours and any presence effect of music. Chapters 3 and 4 consider the capacity for error correction in eye movement timing, when implicitly and explicitly synchronising eye movements with music that begins at a similar period to the saccade timer (chapter three also includes finger-tap timing for contrast). Eye movement timing will be contrasted between self-paced and when responding to externally timed visual onsets. Chapter 5 extends the visual search paradigm to consider the capacity for volitional intent as a form of direct control, by statistically modelling eye movement distributions when volitionally synchronising sequential eye movements to incrementally slower beat intervals. This capacity is contrasted across variations of the task with auditory representations of eye movements (Chapter 5), and with variation in difficulty to contrast the influence of cognitive processing demands and volitional control (Chapter 6). These effects and the implications for the current conceptualisation of eye movement timing will be discussed, including the relationship between attention and eye movements and broader implications for psychological research (Chapter 7).

Chapter 2

Implicit Entrainment of Eye Movements to Music

2.1 Chapter Overview

The question addressed within this chapter is whether musical beats implicitly influence the timing of eye movements. The first experiment introduces a highly controlled visual search paradigm that promotes a sequence of eye movements with minimal influences from task, visual or programming decisions. Within Experiment 1 the visual search task was accompanied by different instrumental music genres at commonly found musical tempi (80 and 120 beats per minute). These music manipulations did not implicitly affect eye movements. Experiment 2 employed simplified musical beats which corresponded to the natural period (mean ±1 SD) produced by eye movements when orienting in the task. Again, there was no clear evidence of implicit influence (some trend indicative of slowing was observed). Visual factors in the paradigm are evaluated as potential confounding influences.

2.2 Experiment One: Implicit Entrainment at 500 and 750ms

The temporal correspondence of visual events and musical beats can increase the salience of these moments in time (Escoffier et al., 2010; Escoffier, Herrmann, & Schirmer, 2015), evidenced with faster responses to on-beat stimuli (as opposed to off-beat). The mechanism of this on-beat enhancement is well framed within the dynamic attending theory (Jones & Boltz, 1989). The allocation of visual attention (a limited resource in time) predicts the onset of relevant audiovisual events to enhance attention at key moments in time. As eye movements are deemed the locus of visual attention (Deubel & Schneider, 1996; Kowler et al., 1995), it is plausible that the entrainment of visual attention to musical beats will also entrain the timing of eye movements. This behaviour would be pronounced in a task where eye movements could be aligned in period and phase to produce predictably timed spatial shifts, to facilitate the enhanced perception of predictable temporally aligned audiovisual events.

Music can spontaneously induce entrained movements, from tapping a finger to dancing (Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2014). Yet studies that reliably observe implicit entrainment of movements to music are relatively rare, for example the previously discussed line-drawing task (Kuribayashi & Nittono, 2014), or when walking on a treadmill (Moens et al., 2014). There are two eye-tracking studies have identified timing changes in eye movement behaviour subject to background music. Firstly Day et al. (2009) found that when musical tempi increased it shortened fixation durations during a decision making task. Secondly, Schäfer and Fachner (2014) found that simply the presence of music shortened fixation durations during free viewing of scenes (neither provides clear evidence for entrainment, which would be evidenced by period shifts of fixation durations that match the musical tempo, or some phase alignment between the two). Evidence from both a metronome finger tap study (Repp & Keller, 2004) and recent music and gait research (Leow et al., 2017; Mendonça, Oliveira, Fontes, & Santos, 2014) found that without the intention to synchronise, participants did not significantly alter their timing or synchronise motor movements with auditory beats. Therefore, there is mixed evidence for implicit synchronisation. The presence of music can implicitly influence the timing of motor movements, such as gait and hand movements while drawing (Kuribayashi & Nittono, 2014; Moens et al., 2014), as well as eye movements (Day et al., 2009; Schäfer & Fachner, 2014); although the magnitude of implicit influence may be limited in the absence of the explicit intention to synchronise (Leow et al., 2017).

When spontaneously producing a tapping rhythm (Fraisse, 1982) or walking (Murray, Drought, & Kory, 1964; Styns, van Noorden, Moelants, & Leman, 2007), multiple studies have identified a ‘preferred tempo’ around 500ms or 120 beats per minute (bpm). This preference is generally considered to stem from an embodied resonance with gait durations (Fraisse, 1982). As a beat duration around 500ms is commonly identified in the production of spontaneous motor responses, the first study will compare eye movement timings of a silent control to the 500ms IOI ‘preferred tempo’ music, as well as a slower 750ms (80bpm) IOI, which Escoffier et al. (2010) employed to entrain visual attention in their face/house detection task. In addition to eye movements the ability to synchronise a more common movement (finger-taps on a track-pad) will be recorded for each IOI level as a measure of explicit beat synchronisation ability. There are competing predictions regarding the spontaneous synchronisation of eye movements to music. As the timing of visual attention is enhanced on irrelevant musical beats, the timing of eye movements is predicted to correspond (synchronise) with the musical beat, as this would serve to utilise the enhanced attention when on-beat. It is predicted that this synchronisation will produce faster movements in the 500ms condition than 750ms, although both musical conditions are predicted to have shorter durations than silent (as in study by Schäfer & Fachner, 2014). The predicted synchronisation of eye movements to the music will be measured not only with latency variance but also a circular measure of synchronisation (mean resultant length), as circular statistics are considered most appropriate when accuracy is more variable (Repp & Su, 2013). The alternative null-hypothesis is plausible, as intention may be required to reliably synchronise movements; in the absence of explicit instruction, eye movements will not synchronise. As the null hypothesis is also plausible, this thesis will complement the use of frequentist statistics with Bayesian evidence for the null hypothesis where appropriate.

2.2.1 Method.

Participants.

A total of 20[1] participants completed the study; two were replaced due to eye-tracking complications (N = 1) and data loss (N = 1). This left 18 participants (10 Female), with an age range from 18 to 64 (M = 33.61, SD = 11.46), who completed the study for payment. The musical training factor from the Goldsmiths Musical Sophistication Index (Gold-MSI) provided responses on a scale from 7 indicating no training, to 49 indicating a level of training greater than 99% of the population (Müllensiefen, Gingras, Musil, & Stewart, 2014). The training scores indicated a spread of musical training from 7 to 47 (M = 19.56, SD = 12.6). Those with higher levels of musical training did not produce a significantly more synchronised tapping response, $R^2 = 0.025$ $p = 0.266$. Of the 18 participants, 2 were not able to consistently tap in time to both tempi; initial analysis will include all participants, though further analysis that excludes them is discussed.

Apparatus.

Monocular eye movements of the participant's dominant eye were recorded using an EyeLink 1000 eye-tracker at 1000Hz, with a chinrest to reduce head movements and keep a constant 60cm distance from the screen. The stimuli were controlled using Experiment Builder (SR Research) and presented on a 21" Ilyama monitor with the resolution set to 800x600 (85Hz refresh rate). The subtended visual angles were 37.3° horizontal and 28.5° vertical. The music was played at a pre-set comfortable volume

---

[1] No directly comparable demonstration of the expected effects exists, so the sample size cannot be powered. This sample size is larger than the previously discussed saccadic SMS studies which employed around 10 participants (Hornof & Vessey, 2011; Joiner, Lee, Lasker, & Shelhamer, 2007b)

through Sennheiser HD01 headphones which were worn for the entirety of the experiment. Tap responses were recorded from space-bar presses on a keyboard.

Stimuli.

The visual stimulus was a circular array of 12 small black circles (each circle was 1.2° in diameter and the distance between each circle was 3.5°), presented centrally on the screen (see Figure 2-1). A small white letter (0.35° vertical), was displayed in the centre of each black circle, either T or L. Only one letter was displayed at any time.

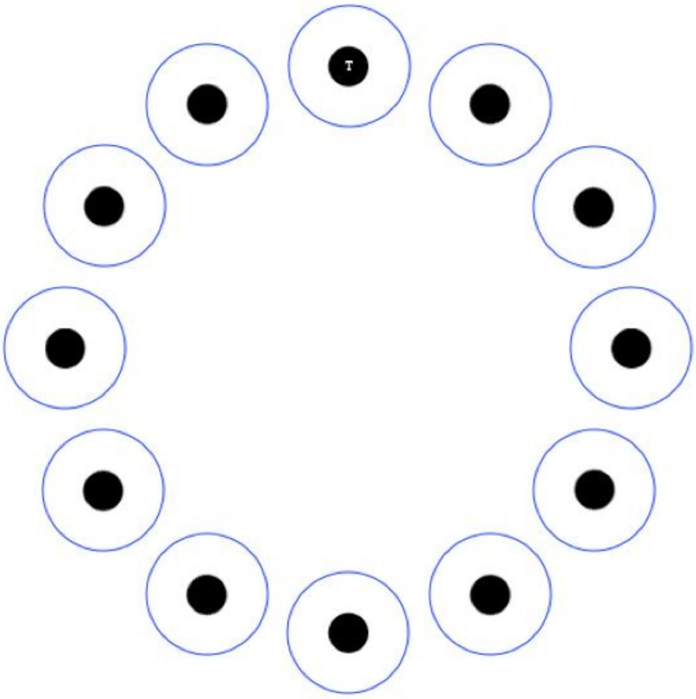


Figure 2-1 A sample image of the visual stimuli, only one black circle displayed a white letter (either T or L) at any time. The blue circles represent invisible boundaries which were used to trigger gaze-contingent changes (i.e. they were not seen during the experiment)

The presentation of a letter within each circle was triggered by a fixation within the region around the circle (gaze-contingent). When a participant's gaze was held within the invisible bounds surrounding a circle for 15ms, and the current circle fixated was the next circle in a clockwise sequence, a letter would be displayed. The 15ms requirement served to enforce a fixation at each circle location. Without this rule, a saccade through multiple boundaries could present a sequence of letters in error (off-task behaviour). The clockwise requirement kept the saccade amplitude and planning decisions constant, i.e. the next saccade will be ~ 4.5° to the next circle in the sequence clockwise. The decision rule for which letter was displayed from circle to circle (T or L) was based on a dynamic counter that updated the number of times the current letter had been displayed in sequence and compared that number to the current value held from a list of 'change letter after' values. The length of each sequence of a single letter was either from between 25-34 circle transitions (2/3 of the time), or from a 4-11 circles range (1/3 of the time, omitted from analysis), for example the letter T would be displayed 28 times in a row, followed by the letter L for eight circles, then T for eleven. Multiple letter transitions occurred within a trial, and the trial duration was two minutes. The addition of the faster (4-11 circles) sequence range was to avoid boredom during the task and eliminate predictability.

The musical stimuli were two stylistically different instrumental musical tracks generated for the experiment using Apple's Logic Pro© software, both tracks had a 750ms and 500ms IOI version. The first track was formed to have an aesthetic similar to 'house' music[2], with synthesised sounds and a clear driving drum rhythm that accented

---

[2] An example of the 500ms house music can be heard here: https://goo.gl/FyZDqS

each beat. The second musical track was similar to 'funk'[3] music with a high sense of groove conveyed through the bass guitar riff and Hammond style organ. This was overlaid with a clear and punctate drum rhythm that again accented each beat. Both musical tracks were formed of a repeating loop and all composite sounds for both tracks were generated from midi files with sound durations that varied subject to the tempo. This facilitated the manipulation of the same track to form different tempo levels without pitch warping.

Procedure.

Following informed consent participants put on the headphones and were made comfortable on the chinrest to minimise head movements. A 9-point gaze calibration confirmed accuracy to within 1°. This calibration was repeated if their head moved from the chinrest or if the inter-trial drift correction exceeded 1°. The instructions for the task were to look from circle to circle clockwise (starting each trial at the top '12 o'clock' circle), identifying any change in the letter being displayed from the previous circle. When the letter changed, the participants were tasked to quickly press the space-bar on the keyboard and continue looking for the next change. The first trial was a practice trial in silence to aid familiarity with the gaze-contingent display and the requirement for a sequential clockwise viewing pattern. Following this, participants were given information that some trials would be in silence and others would have background music, throughout the task of identifying letter changes remained the same. The first four trials were in silence and the remaining eight trials were accompanied by a random order of each of the musical tracks (house and funk styles repeated twice). Breaks were

---

[3] An example of the 750ms funk music can be heard here: https://goo.gl/j2mubf

provided when required or after six trials. Following completion of the gaze-contingent task the participants were provided new instructions to tap in time to the beat of the music on the space bar. Here there was no visual task (eye movements were not recorded and they were not on the chinrest). Each tapping trial was 20 seconds and each musical track was repeated twice. The final task was a short 7-question musical training factor questionnaire from the Gold-MSI. The total duration of the experiment was 50 minutes.

Data analysis plan.

All eye movement fixation and saccade classifications were identified using Eyelink's default eye-movement algorithm. Some data cleaning steps were taken with the eye movement data. Firstly, fixation classifications that were preceded by, prior to, or contained a blink were removed. Secondly, the fixations that were not following the clockwise sequence (off target) were removed. The previous saccade was required to begin one circle prior and end in a fixation on the next circle clockwise. This removed all re-fixations within the current interest area. Finally, very short fixation durations (less than 100ms, insufficient time to process the new visual information and plan the next movement (Findlay & Gilchrist, 2003), and those greater than 1500ms (indicative of off-task behaviour) were removed. These cleaning steps removed 32.6% of fixations, leaving an average of 548 fixations per condition for each participant for analysis. Analysis of average run lengths around the circles ranged between 2.4 and 8.4 circles (M = 4.11, SD = 1.60). These values are particularly low as the sequence was broken by blinks. Following these cleaning steps, each fixation was summed with the preceding saccade duration to generate the eye movement *latency* measure which is representative of the time interval of movements between circles.

A goal of this study was to identify a measure of entrained response in eye movement timing to two musical tempi. The relationship of interest is the timed responses as they align with the period and the phase of the auditory beat. Therefore the eye-movements were analysed using both measures of period (latency) and phased responses using circular methods, specifically the *mean resultant length* (MRL) measure (Fisher, 1995) with the 'circular' package for R (Agostinelli & Lund, 2017). Circular methods are particularly informative when synchronisation performance is poor (Repp & Su, 2013).

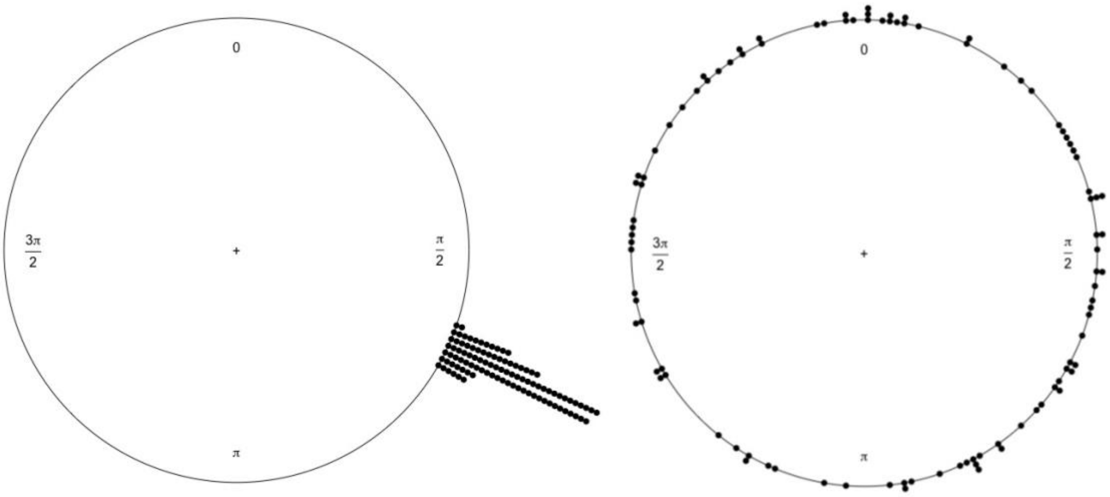


Figure 2-2 Two circular plots of simulated data as examples of synchronised responses with an MRL of .99 (left), where dots are individual responses to an isochronous rhythm, and the right are random or uniform responses with a very low MRL of .09. Rayleigh's test (Wilkie, 1983) is significant for the left but not the right

The MRL is a deviation measure of the circular mean, indicative of how tightly clustered response vectors are to the circular mean based on a pre-set phase (the musical IOI). The MRL has a range from 1 (all the vectors are equal to the circular mean), to 0 (a random distribution whereby the vectors bear no relationship to the circular mean), visualised in Figure 2-2. A statistical test of the MRL values, the Rayleigh test (Wilkie, 1983) confirms if the clustering of the vectors (MRL) within a condition are significantly representative of the circular mean (significantly non-uniformly distributed). This test is the minimum requirement for a circular representation of vectors, a previous employed threshold for finger tapping synchronisation is MRL > .73 (Konvalinka, Vuust, Roepstorff, & Frith, 2010; Tognoli, Lagarde, DeGuzman, & Kelso, 2007). The eye movement representation that was coded for the timed response (each vector) was the fixation start time (the end of a saccade). This aspect of an eye movement best parallels an entrained motor movement, i.e. the landing of a finger (or tap) which is the time in the movement aligned with the beat (Repp & Su, 2013). A MRL value was generated and tested for significance (Rayleigh test) for each participant at both the Genre and musical IOI levels.

2.2.2 Results.

Latency.

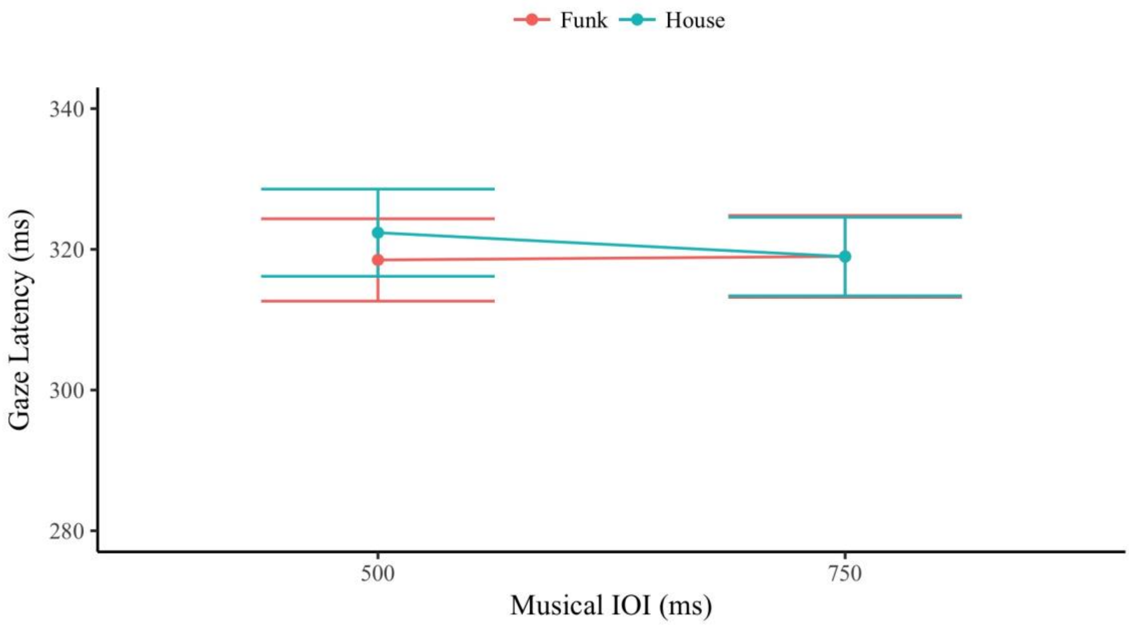


Figure 2-3 Mean Latency (ms) ±1SE, by Genre and Musical IOI

Two participants were unable to tap in time to both musical IOI (a MRL when tapping below 0.5). The subsequent analysis below includes their eye movement data, as omitting them did not influence the results or the strength of the Bayesian evidence. Whilst there are no predictions regarding genre as an effector of eye movement timing, the design of the study motivated a repeated measures ANOVA contrasting musical IOI and Genre. As is apparent in Figure 2-3, there was no significant main effect of Genre, $F(1,17) = 1.170$, $p = .295$, $\eta^2 = .064$, $BF_{01} = 2.55$. An initial prediction was that the 500ms IOI would produce shorter latencies than 750ms, however there was no significant main effect of musical IOI, $F(1,17) = .562$, $p = .464$, $\eta^2 = .032$, $BF_{01} = 3.32$, rather there was positive evidence for invariance between the IOI levels (Kass & Raftery, 1995). The interaction between genre and musical IOI was also not significant, $F(1,17) = 1.613$, $p = .221$, $\eta^2 = .087$, $BF_{01} = 7.911$.

A prediction of the study was that the presence of music would reduce the latency of eye movements when compared to silent. Collapsing for genre (which did not significantly influence latency) and employing Bonferroni correction, neither the 500ms IOI (M = 320.32ms, SD = 24.95) music, $t(17) = .513$, $p = .594$, $BF_{01} = 3.45$, or the 750ms IOI (M = 318.86ms, SD = 23.68), $t(17) = .543$, $p = .614$, $BF_{01} = 4.02$, significantly differed from the silent condition (M = 319.4, SD = 25.85ms). Rather there was positive evidence for the null hypothesis, that music did not vary the timing of eye movements during the task. This invariance is particularly apparent in the density Figure 2-4.

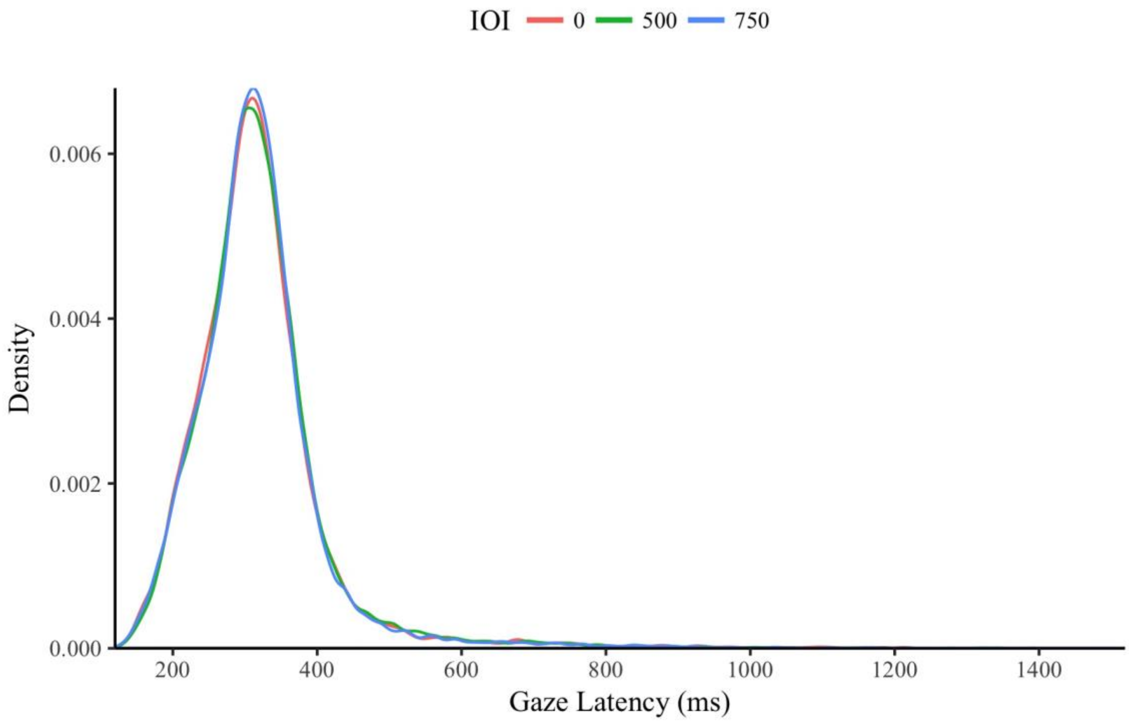


Figure 2-4 A density plot of all gaze latency values across the three musical IOI levels (silent, 500 and 750ms). The density estimation utilised a bandwidth smoothing that is .9 times the minimum standard deviation

The distributions are also positively skewed. A $\log_{10}$ transformation of the data confirmed the same pattern of invariant mean latency results ($F < 1.3$, $p > .27$) with no difference in the effect size or variance in the strength of Bayesian evidence.

Mean resultant length.

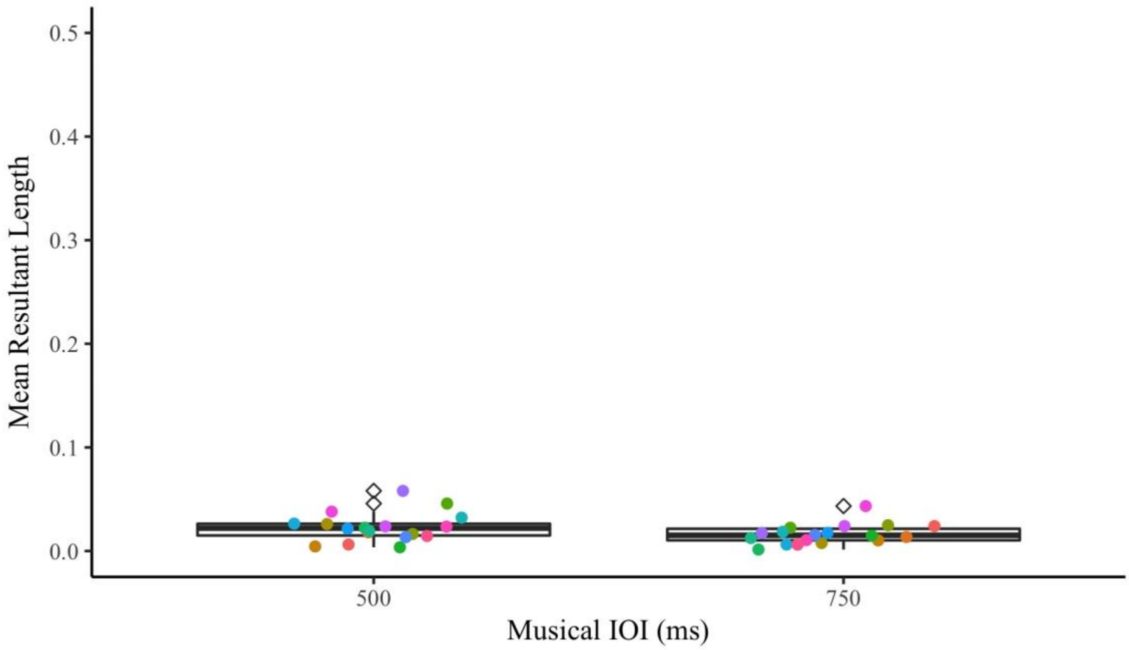


Figure 2-5 A Box-Plot of Mean Resultant Length values by Musical IOI (colours represent participants, the MRL were collapsed across Genre), the diamonds are slightly less random potential outliers

The final prediction was that the eye movements would entrain in phase with the musical beat. An initial test of whether the eye movement behaviour presented a phased response to the IOI, the Rayleigh's test, found that no participant's eye movements were significantly clustered in phase at either 500ms or 750ms IOI (all *p* > .3 across both Genre or Musical IOI). Further confirmation of this asynchronous response is observed in the MRL values themselves (see Figure 2-5). At both 500ms (M = .020, SE = .002), and at 750ms (M = .011, SE = .002), the timing of fixations was random in relation to the musical beat.

2.2.3 Discussion of Experiment One

The absence of any musical tempi influence on the timing of eye movements was surprising but not completely unexpected considering the previous studies that have found an intention (or instruction) to synchronise was essential in producing synchronised tapping or gait behaviour (Leow et al., 2017). The absence of any effect of music (e.g. a presence effect) is puzzling considering the previous research which described the entrainment of visual attention to music (Escoffier et al., 2010), and changes in eye movement timing subject to variance in musical tempi or the presence of irrelevant music (Day et al., 2009; Schäfer & Fachner, 2014; Wallengren & Strukelj, 2015). This study found positive Bayesian evidence for invariance across condition. When completing this highly-controlled gaze-contingent visual search task the addition of music with different tempi did not affect eye movements or systematically vary eye movement timing. The initial evidence is that eye movement timing is not influenced by irrelevant music irrespective of changes in its tempo. Alternatively, it is plausible that the phase in the music was too slow to resonate with eye movement timing.

The rationale for the use of the 'preferred tempo' was that the 500ms periodicity resonates with the natural pace of motor movement, especially the pace of walking (Burger et al., 2014; Fraisse, 1982; Moelants, 2002). Our eyes reorient in space with a ballistic movement (saccade) around three times a second (Henderson, 2003; Rayner, 1998), which is a much faster movement period than the IOI used in E1. The absence of eye movement entrainment could be because the size of the period shift in timing required to produce a match with the musical IOI and form a synchronised relationship was too large for the pace of eye movements (to achieve a 500ms mean latency required 7.2 standard deviations from the mean, at 750ms 18.2 standard deviations above the mean were required). Walking may spontaneously entrain with music at 120bpm (500ms IOI) because the natural pace of walking is around 500ms (Styns, van Noorden, Moelants, & Leman, 2007), and spontaneous phase alignment would require minimal corrections in phase. Eye movement timing may display entrainment behaviour if the musical beats were closer aligned in period to the natural eye movement timing distribution.

2.3 Experiment Two: Implicit Entrainment at the Natural Pace of Eye Movements

The first study found no evidence of implicit musical influence on the timing of eye movements. A factor that may have contributed to this invariance is the size of the period difference between the music and eye movement timings, which is commonly around three times per second (Rayner, 1998). This study will utilise three musical IOI levels that re-represent the natural pace of eye movements when orienting around this gaze-contingent task: the mean as well as one standard deviation above and below the mean. We predict that, when only a small phase correction is required to produce on-beat eye movements, eye movements will synchronise and the latencies will vary linearly with the changes in musical IOI. As with E1, evidence for the null hypothesises will also be tested, as it is plausible that eye movements are not implicitly influenced by background music.

2.3.1 Method.

Design Modifications.

The characteristics of the previously used background music (tempi or genre) could have mediated the observed invariance in behaviour. This experiment will therefore alter the musical stimuli, firstly to align the musical beats with the natural pace of eye movement for the task, and secondly to simplify the musical score to use just one highly simplified and clearly accented beat. A further consideration is the 2-minute trial duration used in experiment 1. A shorter duration (1 minute) could enhance the effect of music as a presence effect may occur at the initial exposure to the new tempi and would be more represented in shorter trials.

The visual stimuli, screen, headphones and eye-tracking setup remained the same as in experiment 1. To generate the new music stimuli five naïve participants completed a paid 5-trial pilot version of the experiment in silence with the trial duration shortened to 1-minute. The mean gaze latency of the pilot subjects was 312.44ms, with a standard deviation of 24.31. These values were used to form three musical tempi, the mean latency (312.44ms 192 BPM), the mean plus one standard deviation 336.75ms (178 BPM), and minus one standard deviation 288.13ms (208 BPM). These three IOI levels (337, 312 and 288ms) represent a natural range of eye movement latencies when completing the task. As the pace of these IOI levels is much faster than the previous IOI used (500 and 750ms) the musical stimuli had to be simplified to clearly accent the beat.

The new musical tracks were formed from a single simple drum rhythm, with an electronic style kick-drum and snare-drum at equal intervals accenting the beat, and a quieter high-hat half-note to provide an intensity contrast (so that there are weak and strong beats). The music was percussive and had no melody. The aesthetic was similar to minimalist electronic music or a basic arcade game soundtrack[4].

Participants.

Twenty-two naïve participants were recruited for a paid study. One was unable to complete the experiment due to eye-tracking issues, which left twenty-one participants (12 Female; Mean age = 29.9, SD = 6.9, range 20-40 years). As detailed in experiment 1, the absence of a previous effect size to inform this experiment negates power analysis,

---

[4] An example of the new music at 312ms IOI can be heard here: https://goo.gl/YR4oG2

therefore a sample size of 21 was selected as a starting point. All participants self-reported normal hearing, and normal or corrected to normal vision. The Gold-MSI musical training factor scores (M = 19.4, SD = 11.4) ranged from no training (7) to 36 (out of 49). When tasked to tap a finger in time with the three musical tempi, all but three participants were significantly able to synchronise to all three tempi. The synchronised tapping was positively correlated with musical training ($R^2 = .19$, $p = .028$, one-tailed). The initial analysis will include all 21 participants, the impact of the three poor tapping performers will be discussed within the results.

Procedure.

The procedure and instructions remained the same as E1. Each participant completed a silent practise trial, two trials in silence, a random order of the three musical IOI levels (337, 312 and 288ms) repeated four times. The gaze-contingent component ended with two further silent trials to account for any fatigue effects. This generated four gaze-contingent trials per IOI for each participant. The total duration for the experiment was 50 minutes.

Data Cleaning.

The data was treated to the same rules as E1, which removed 33% of the gaze data. This left an average of 533 latency values/vectors (fixation + previous saccade duration) per condition per participant.

2.3.2 Results

Latency.

The three musical tempi IOI were generated to represent the natural timing range of eye movements when completing the task. The prediction was that implicit entrainment will occur when only small timing changes were required to align with the beat. Three Bonferroni corrected pairwise t-tests compared the eye movement latencies in silence to the three musical IOI levels. As observed in the mean latencies in Figure 2-6 and in the Density plot Figure 2-7, no IOI level significantly differed from silent (all $t(20) < 1.8$, $p > .3$). The Bayesian evidence BF_{01} ranged between 1.26 and 2.8, and was weak evidence for the null hypothesis. Each of the IOI levels had a slightly slower mean latency than silent (ns).

A repeated-measures ANOVA of the eye movement latencies at the three IOI levels (337, 312 and 288ms) found a marginally significant main effect of IOI, $F(2,40) = 2.558$, $p = .090$, $\eta^2 = .113$, $BF_{01} = 1.251$ (weak evidence for the null). The mean latencies in the 337ms condition were slower than the two faster IOI levels, although there were no significant differences between the three IOI conditions (Bonferroni corrected post-hoc comparisons, all $p > .2$). As with E1 the $\log_{10}$ corrected latency values were analysed and produced a near identical non-significant main effect of IOI, $F(2,40) = 2.529$, $p = .092$, $\eta^2 = .112$, $BF_{01} = 1.248$. As the pattern of results in the $\log_{10}$ corrected analysis are consistently near-identical to those without correction, the future analysis in this thesis will not report log corrected results for brevity and simplicity of interpretation.

Excluding the three participants who did not tap in time again produced a similar marginally significant main effect of IOI, $F(2,34) = 2.673$, $p = .083$, $\eta^2 = .14$, $BF_{01} = 1.109$ (weak evidence for the null). The trend in the data from the previous analysis remained the same with a (non-significant) slowing only in the 337ms condition (Bonferroni corrected pairwise *p* = .236).

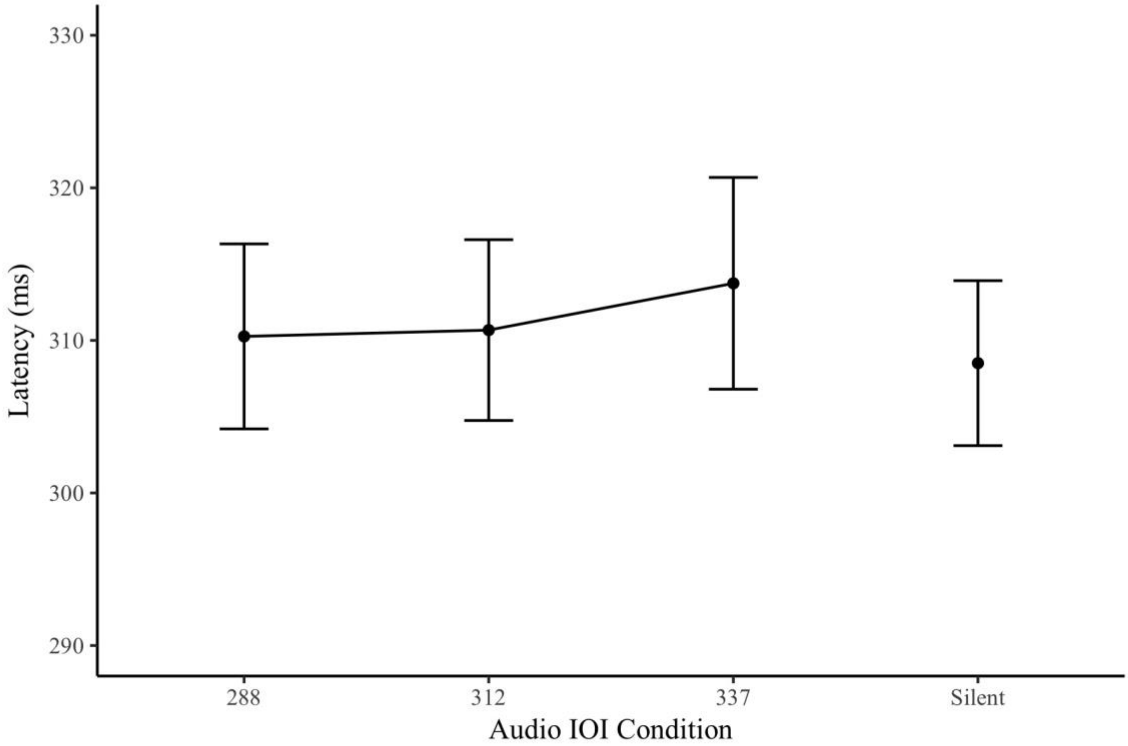


Figure 2-6 Mean Gaze Latency (ms ±1 SE) by Musical IOI and Silent Conditions

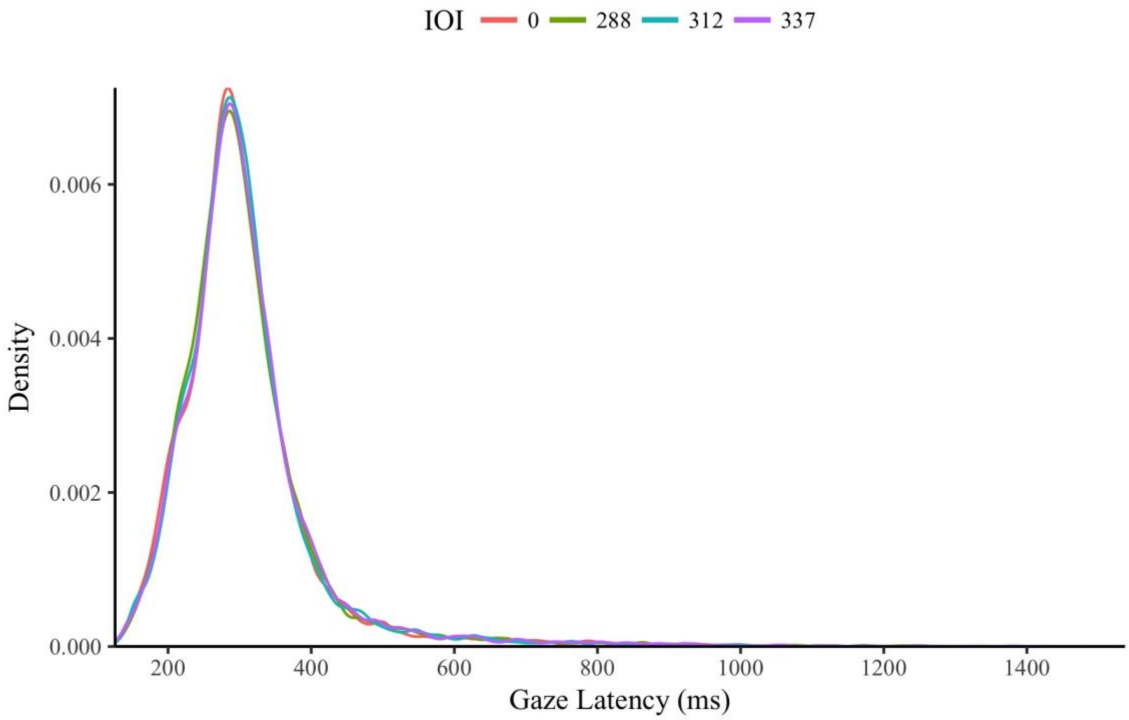


Figure 2-7 Gaze Density by Musical IOI Conditions and Silent Control (0), The density estimation utilised a bandwidth smoothing that is .9 times the minimum standard deviation

Mean Resultant Length.

As observed in the boxplot Figure 2-8, all of the MRL values were less than 0.2 (on a scale from 0 to 1, where 0 is random or uniform in phase), and no participant achieved anywhere near the .73 threshold.

The Rayleigh’s test of the fixation start times (as a phased response to the IOI) by participant at each IOI condition identified just three of the twenty-one participants who had a significantly non-uniform response at 288ms, ten (48%) at 312ms, and four (19%) at 337ms IOI. Just one participant had significantly non-random responses to all IOI levels (the MRL values were still lower than .2). More participants were able to produce a significantly non-uniform phase at the 312ms IOI than the outer (±1 SD) IOI values. This is likely due to the design of the experiment as we have produced a referent IOI which itself is very close in period to the eye movement latencies. It is likely that the slight increase in MRL values in the 312ms condition is a phase alignment by chance rather than any implicit musical modulation. In summary, despite producing a referent IOI that was very close in phase to the natural pace of eye movements, no participant displayed a convincing pattern of synchronised eye movement behaviour.

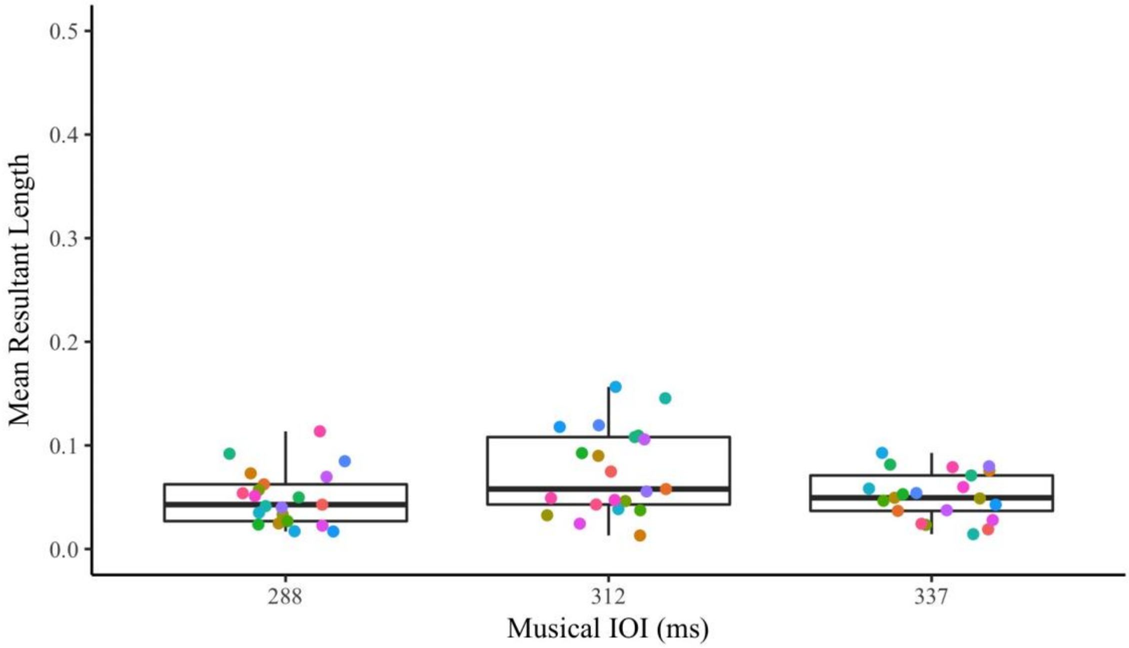


Figure 2-8 A Boxplot of Mean Resultant Length (MRL) values by Musical IOI Condition, the Colours Represent Participants

Post Hoc Analysis: Saccade Direction.

An intended feature of the circular gaze-contingent design was to facilitate a sequence of eye movements that could be equal in size, in angle of orientation, and in visual features. This aim was to produce a series of eye movements that were subject to the same visual information at each point and to promote sequences of uninterrupted movements that could be entrained. To confirm these assumptions the saccade and fixation behaviour at each circle location was analysed separately. The circular design required saccades that oriented in different directions. To analyse the effect this had on the duration of saccades (only those saccades that correctly oriented between the circles), the direction the saccades oriented in were grouped into four orientation 90º directions, Up (between 303.75º and 33.75º), Down (between 123.75 and 213.75º), Left (between 213.75º and 303.75º), and Right (between 33.75º and 123.75º). There was a significant main effect of saccade direction on the duration of saccades, $F(3,60) = 11.81, p < .001, \eta^2 = .371$, seen in Figure 2-9. Bonferroni corrected pairwise comparisons confirmed that saccades orienting downward were significantly slower (a longer duration) when compared to all other directions (all $p < .007$, no other directions significantly varied). Additionally, Bayesian repeated measures confirmed very strong evidence for the main effect of direction, $BF_{10} > 1000$. Considering this saccade directional effect, the prior IOI analysis was extended to observe whether the direction of the saccade influenced the synchronisation (MRL) of eye movements. A repeated measures ANOVA of the four saccade directions (down, up, left and right) found no significant main effect of saccade direction on synchronisation, $F(3,60) = .048, p = .986, \eta^2 = .0002$. The slower downward saccades affected the latency of eye movements but did not increase synchronised behaviour (which was generally very low).

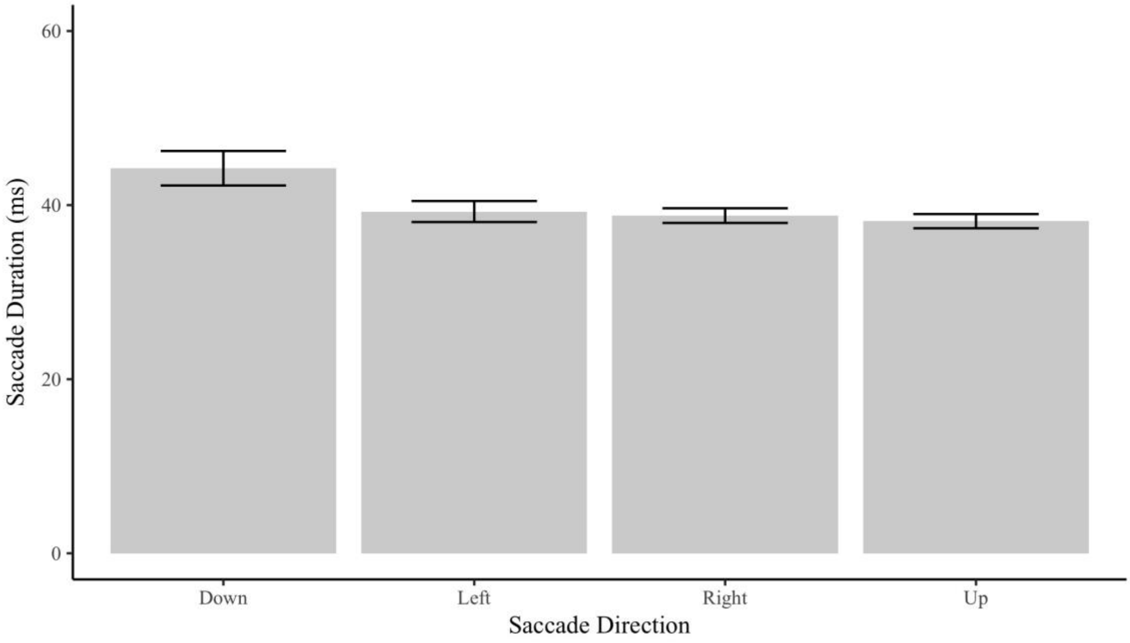


Figure 2-9 Mean Saccade Duration (ms) by Saccade Direction ±1SE

2.4 Discussion of Experiment 1 and 2

The question motivating these two studies was: are eye movements implicitly entrained by music? In the first study, which used commonly occurring musical IOI levels (500ms and 750ms) we found evidence for the absence of variance in eye movement timing subject to the presence of irrelevant music, or changes in the musical IOI. The timing of eye movements did not present any indication of entrained behaviour to music at these tempi. This is contrary to the prior evidence of presence effects and fixation duration differences subject to tempi change (Day et al., 2009; Schäfer & Fachner, 2014). The evidence in these previous studies had notably unconstrained visual environments, where oculomotor profiles could not be standardised across audio conditions. As shown within this task, even small changes within a highly constrained visual task (the direction of a saccade) can have a large effect on eye movement timing irrespective of auditory influence.

In the second experiment, we changed the duration of trials (from 2 to 1 minute) and employed three much faster musical IOI (288, 312, 337ms) that were aligned with the normal period (latency) of eye movements during the task. As with the first experiment the addition of music did not significantly alter the timing of eye movements. There was a trend to slow eye movement latencies at the slowest IOI, with weak evidence of variance between the IOI levels. The slowest IOI (337ms) resulted in numerically slower eye movement latencies (not significantly), but this slowing did not result in a higher level of synchronisation. The weak Bayesian evidence for the main effect indicates that the second study (N = 21) was likely underpowered and/or that any implicit effect of music is small. Future studies will employ a larger more appropriately powered sample size. Notably, in contrast to the previous studies Day et al. (2009) and Schäfer and Fachner (2014) which reported eye movement differences in much less constrained visual environments, we fail to find any clear evidence of irrelevant music beats significantly influencing eye movement timing.

In both studies (E1 and E2), there was a subset of participants who were not able to tap in time to the music at each IOI. An assumption of participant recruitment was that tapping in time to a musical beat was a commonplace human ability (Burger et al., 2014; Large, 2000). Whilst the exclusion of these participants did not have a noticeable effect on the (absence of) effects reported, future recruitment will actively recruit and pre-test individuals to ensure good beat production ability.

A final consideration of these studies was the paradigm employed. The task was generated to promote optimal conditions for eye movements to be executed by limiting the visual factors that could inhibit or cancel saccades. Visual decisions regarding where to saccade, the visual information at fixation, and the task demands were minimised to promote the underlying saccade timer and observe the effect of music. To this end the experiments appear quite successful, as the distributions of eye movements have a much smaller exponential component in the distribution than commonly found in more normal viewing tasks (Henderson, 2003; Rayner, 1998). One consideration not accounted for was saccadic direction. As previously discussed the downward saccades were significantly slower than all other directions. This difference between downward saccade latencies and other saccade directions has been identified previously in single saccade studies, and in scene research (Honda & Findlay, 1992; Ludwig & Gilchrist, 2003; Tatler & Vincent, 2008). What is clear for this paradigm is that future implementations should limit the requirement for vertical saccades and alter the task so that it is maximally horizontal, a preferred saccade orientation (Foulsham & Kingstone, 2010). Furthermore, considering the absence of implicit effects, contrasting the synchronisation eye movements with finger-taps, and explicitly tasking synchronisation will allow confirmation that direct control of eye movements can align with a musical beat as either a predictive synchronised response to exogenous visual onsets (tap-generated or timed) or when explicitly synchronising gaze-contingent onsets.

Chapter 3

Eye Can’t Dance: Entraining Motor and Oculomotor Movements to Music

3.1 Chapter Overview

The experiment within this chapter extends the previous two studies by refining the visual search paradigm to account for extraneous visual and beat production influences. It contrasted not only the implicit effect of music on eye movements, but also modality differences between eye movements and finger-tap timing. The paradigm was also extended further to measure the capacity for explicit synchronisation of finger and eye movements. The study provides clear evidence that eye movement timing is not implicitly influenced by either the presence of, or changes in the timing of music. Furthermore, the explicit synchronisation of eye movements, manifested as small corrections in period and phase, was notably poor when self-paced and slightly improved when a response to tap-timed visual referents.

3.2: Experiment Three: Implicit and Explicit Eye and Finger-Tap Movements to Music

Finger-taps are a common measure of beat production ability and most human subjects are able to produce a synchronised response that adapts to changes in tempo and can continue after a beat referent is no longer present (Repp & Keller, 2004; Repp & Su, 2013). By comparison, the volitional control of eye movements is imprecise (Munoz & Everling, 2004), but there is some evidence that eye movements can be controlled to explicitly match an auditory rhythm (Hornof & Vessey, 2011; Joiner, Lee, Lasker, & Shelhamer, 2007a). The purpose of this experiment is three-fold. Firstly, to overcome the potential confounding issues of the previous studies and optimise the paradigm to create the best opportunity to find any implicit entrainment, no matter how small. The second purpose is to contrast the latency behaviour of eye movements with the more commonly tested finger-tap, both for implicit entrainment and explicitly synchronised movements in time. Thirdly, to contrast eye movements that are an explicitly synchronised gaze-contingent response to those that are responding to a tap-contingent response. In light of the previous differences in tapping ability across participants, that suggest a broader individual difference in ability than the SMS literature imply (Repp, 2005), this and future studies will positively recruit participants with rhythmic ability.

The three factors manipulated are: Task at two levels (*Implicit, Explicit*), Mode at two levels (*Gaze* and *Tap* Contingent), and Musical IOI at three levels (264, 293, 323ms). The three dependent measures (DV) of interest are, first, latency which as before is a millisecond measure of duration of movement between the circles. When gaze-contingent, this is formed by combining each fixation duration with the previous saccade duration. When tap-contingent, this is the duration between two taps (key press times) in sequence. The second DV is the absolute Error between the Latency and the target IOI (ms) as a measure of period approximation of the target IOI. The final DV is the measure of synchronisation: the mean resultant length (MRL). The tap time and the start of fixation durations (the landing of an eye movements) are the time-points employed as the referents in time.

The prediction for implicit entrainment of eye movements remains motivated by the previous research which identified implicit influences of music on eye movements (Day et al., 2009; Schäfer & Fachner, 2014). Whilst the previous studies (E1 and E2) showed evidence of invariance, the previous paradigms employed may have obscured the music effect. Additionally, there was a trend at the slowest IOI within E2 that indicated a IOI related slowing of eye movement latencies. As previous research has identified implicit motor entrainment both when drawing lines (Kuribayashi & Nittono, 2014) and walking (Moens et al., 2014), in addition to the large body of SMS literature that employs finger-taps (Repp & Su, 2013), we predict that implicit entrainment would be more pronounced in finger-tap movements than eye movements.

The lack of any effects in E1 and E2, and the underlying consistency in the timing distribution of eye movements during the task, question whether the timing shifts predicted to occur implicitly are even possible under explicit conditions. The saccadic SMS studies reviewed within chapter 1 showed that explicitly controlled large horizontal saccadic movements can synchronise with an external rhythm with some rate limitations and without any visual task demands (Joiner et al., 2007a; Shelhamer & Joiner, 2003). Considering the limitations of direct control of eye movements demonstrated with the SOD paradigm (Henderson & Pierce, 2008; Henderson & Smith, 2009), and the broad SMS literature using finger-taps (Repp, 2005; Repp & Su, 2013), the explicit synchronisation and temporal control of finger-taps are predicted to be closer in latency, have less error, and show a higher MRL than the explicitly (directly) controlled eye movements. That said, both modalities are predicted to synchronise.

A final purpose of the study is exploratory, specifically to investigate the temporal precision in the direct control of eye movements, both when producing and synchronising gaze-contingent visual changes and when responding to an motor-produced visual onset in the form of a tap-generated visually timed referent. When completing this task with a gaze-contingent response, the visual onset is a representation of the timing of gaze shifts. In contrast, when following an motor-timed visual onset, there are visual transients that determine the temporal demands on visual timing. A question addressed within this exploratory analysis is: Can participants align gaze-contingent onsets equally easily as tap-contingent visual transients?

3.3 Method

3.3.1 Participants.

A power analysis using G^*Power (Faul, Erdfelder, Lang, & Buchner, 2007) of the prior study (E2) measured the implicit effect of music ($\eta^2 = .15$) on eye movement latencies. This effect size was used to identify an appropriate sample size (N = 30) for this study powered to $d = .8$ when $\alpha = .05$. The study tested 39 participants. Of these, 9 were unable to either tap in time (5/9) or had lost eye tracking data in at least one condition (4/9). This left 30 participants (22 Female), with an age range from 18 to 32 (M = 22.5, SD = 4.26), who contributed data for analysis. The musical training scores (on a scale from 7: no training to 49: more training than 95% of the population) indicated a range of training from 11 to 43 (M = 24.8, SD = 9.3); none of the participants were professional musicians. There was a positive (although non-significant) relationship between musical training score and MRL ($p = .18$). Those with more training did tap slightly more accurately, although lack of a significant correlation is likely limited by the participation criteria which required (and positively recruited) those with a good tapping / beat production ability.

3.3.2 Apparatus.

Monocular eye movements of the participant's dominant eye were captured using an EyeLink 1000, recording at 1000Hz. The monitor and image resolution were updated to an ASUS VG248QE 24-inch screen, with the resolution set to 1024 x 768 pixels at 144Hz (with black vertical boxing). The subtended visual angles were 39.17º horizontal and 29.95º vertical. The headphones were also updated to over-ear Beyerdynamic DT-770 PRO 80 OHMS.

3.3.3 Design modifications.

The visual search paradigm was refined to limit the influence of saccade direction on the latency of movements between each fixation target. As downward saccade durations were found to be significantly slower in the previous study (E2), the visual trajectory around the target circles was altered to be maximally horizontal, whilst still enforcing a pattern of sequential movement. To achieve this, the circular shape of the previous 12 circles was converted to an elliptical shape of 8 circles, visualised in Figure 3-1. Each circle in the elliptical array was 1.2° in diameter, and the distance between each circle was 7.8°. The longest letter edge of the T and L within each circle was 0.3°.

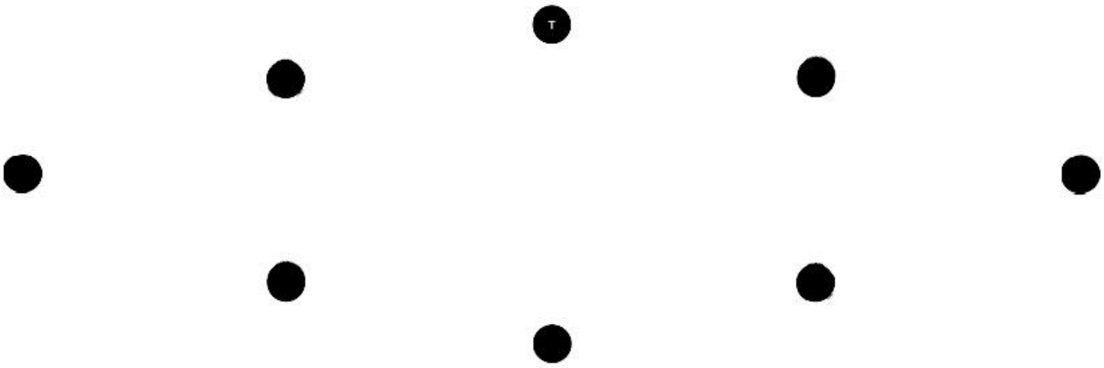


Figure 3-1 The Refined Elliptical Visual Search Paradigm with Eight Circles

A pilot study was conducted on a further five naïve participants, orienting around the elliptical shape in silence (1 practice trial and 5 trials for analysis), identifying infrequent changes in letter (between 28 and 31 circle movements before a change, from 5 to 7 changes per trial). This pilot served two purposes. Firstly, the duration of saccades between each circle around the ellipse was analysed. There were no significant differences between the saccade durations at any location around the ellipse, $F(7, 28) = .700$, $p = .672$. The mean saccade duration between each circle location was consistently within 2ms of the mean (45ms). This confirmed that the changes in the design removed the directional confound (downward saccades) of the prior studies. Secondly, the mean latency whilst completing the task was used to inform three new musical tempi based on the revised visual demands. The mean latency was 293.32ms, with a standard deviation of 29.51ms. Therefore, the three new musical IOI levels based on the 8-circle elliptical design are 264, 293 and 323ms. The musical sequence of kick and snare sounds was further simplified to just single a kick and snare sound[5].

Final modifications were to the duration of trials and the number of sequence changes in a trial. The end of a trial was altered from the previous duration based (two or one minute) cut-off to one that is sequentially contingent. A trial end occurs when 180 circles have been sequentially fixated (a gaze contingent letter has been displayed around 180 circles), or subject to a prolonged timeout (three minutes). This change provided the maximum possible opportunity for the data collected (number of fixations) across participants to be similar in size. The number of sequence changes was altered from the previous experiments to a random change after 28 to 31 circles. This increased the capacity for entrained movement by removing the cognitive demands of more frequent changes.

3.3.4 Procedure.

The consent and calibration procedure did not change from the previous experiments. Participants were recruited to a “Visual Search with Music” study, with the overarching task of identifying infrequent changes in letter sequences. Following consent each participant completed a short-version of the Edinburgh Handedness inventory (Veale,

---

[5] An example of the revised musical stimuli at 293ms can be heard here: https://goo.gl/AnRpnT

2013), to identify which hand was dominant and to be employed for key-press responses. All keyboard and tap responses were made with the index finger on their dominant hand.

The orientation *Mode* around the ellipse was either gaze-contingent, looking at the next circle (clockwise) displayed the next letter; or tap-contingent, pressing the correct key for the currently displayed letter moved the letter presentation to the next circle clockwise (*V* for **L**, and *B* for **T**). When the *Mode* was gaze-contingent the task was to press the identifying letter on the keyboard (the key that corresponded to the changed letter, e.g. pressing V if the new letter sequence was L) as soon as the change occurred. When tap-contingent the task required a correct key-response (using the index finger on their dominant hand) to progress (the number of incorrect key presses were shown at the end of each trial, with the encouragement to have none).

Both modality types began with the *Implicit Task*, which like in E1 and E2 was completing the search task with irrelevant music. Following these implicit trials, participants received *Explicit Task* instructions, specifically to move in time with the music, matching every beat in sequence. The orientation mode was counter-balanced across participants. Those in condition A completed each of the four trials (each IOI and a silent randomised trial) in Implicit Gaze, followed by Implicit Tap, then Explicit Gaze, and ending with Explicit Tap. Those in condition B began with Implicit Tap, then Implicit Gaze, Explicit Tap, and ended with Explicit Gaze. Whenever there was a change in task there was a practice trial to ensure compliance (note that there was a randomly inserted silent trial in both the implicit and explicit blocks for each modality). The experiment ended with the task to tap in time to each IOI level 60 times per trial on an Apple track-pad and to complete the previous GOLD-MSI musical training questionnaire.

3.3.5 Data cleaning.

As with the previous experiments, fixation durations that were less than 100ms and greater than 1500ms were removed from analysis, as were fixations that contained or followed a blink and movements that were out of sequence (the saccade did not begin at the previous circle and land in the next correct circle in sequence), or fixations that contained a change response (key press) during the gaze-contingent trials. Incorrect key-presses were removed as were those with latencies less than 100ms and greater than 1500ms (indicating off-task behaviour). Finally, the key-presses at each sequence change were removed from analysis. A final control was to remove participants based on their performance in the final tapping task (simply tapping on a track-pad to music). The MRL of the tap responses had to average above 0.5, a more generous threshold than used by Konvalinka et al. (2010). This easier synchronisation level was in light of the fast-paced tempi used, although five participants did not meet the .5 MRL threshold.

3.4 Results.

Latency.

The musical IOI was generated from the latency distribution of five subjects completing a pilot study. To confirm that the Musical IOI matched the natural pace of the participants' eye movements during this experiment we employed a Bayesian one-sampled t-test to compare the gaze latencies from the silent trials (M = 291.10, SD = 27.77) to the pilot mean latency (292.32). The was positive evidence for the null, BF_{01} = 4.70 ($t(29) = .444$, $p = .66$), and confirmation that the musical tempi employed does represent the natural latency of eye movements for the task.

An initial test of whether the music implicitly influenced the timing of tap and eye movements is to contrast the latency of the three music IOI levels to the silent trials.

As observed in Figure 3-2, Bonferroni corrected t-tests showed no significant differences between the silent means and those with irrelevant music (all $p > .3$). The Bayes factors for the Tap latencies indicated positive evidence (Kass & Raftery, 1995) for the null hypothesis ($BF_{01} > 4$), the Bayes factors for the three gaze-contingent conditions were less strong, but again provided evidence of invariance (264 IOI: $BF_{01} = 4.3$; 293: $BF_{01} = 2.1$; 323: $BF_{01} = 1.5$). The presence of music did not significantly alter the latency of either modality, with minimal difference compared to the silent condition.

A three-way within subject ANOVA compared latency across the two modalities (gaze and tap), to the two tasks (implicit and explicit), and the three IOI levels (264, 293 and 323ms), visualised in Figure 3-2. There was a significant main effect of modality, $F(1,29) = 65.213$, $p < .001$, $\eta^2 = .692$, in that tap-contingent latencies (M = 314.5ms, SD = 36.8) were significantly slower than gaze-contingent (M = 286.0ms, SD = 28.5). The main effect of task was significant, $F(1,29) = 18.133$, $p < .001$, $\eta^2 = .385$, as explicit latencies (M = 289.1ms, SD = 27.9) were significantly faster than implicit (M = 308.2ms, SD = 40.2). The main effect of musical IOI was also significant, $F(2,58) = 10.428$, $p < .001$, $\eta^2 = .264$, as the latencies varied linearly ($p < .001$), increasing as the musical IOI increased.

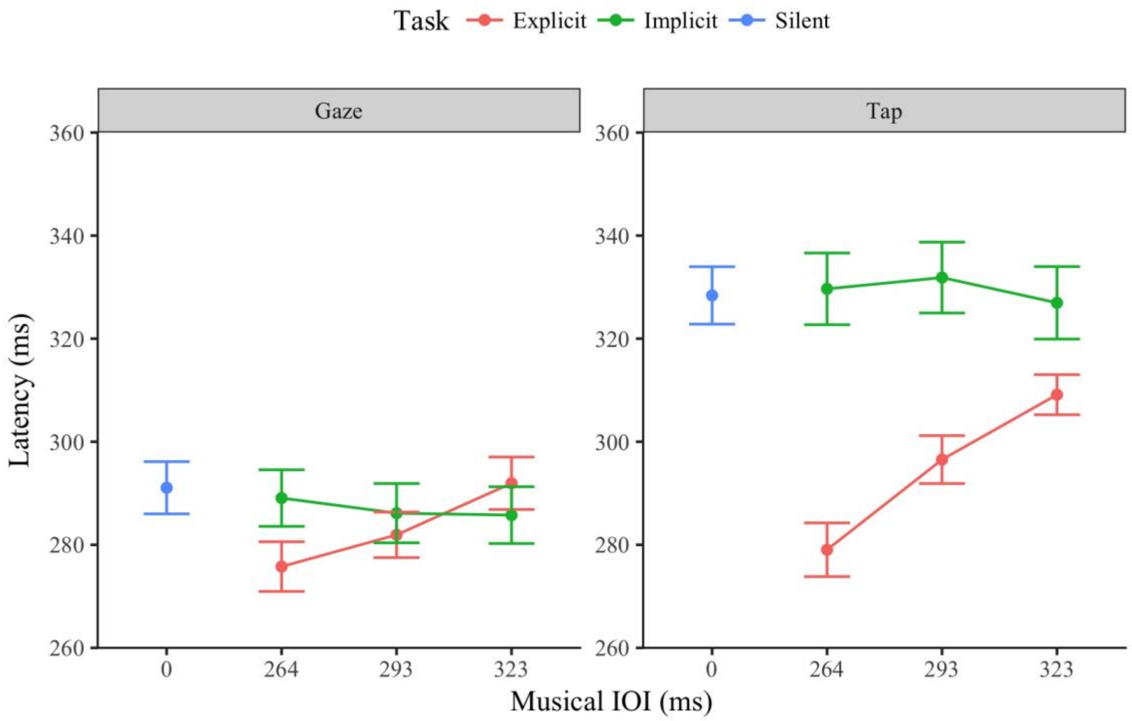


Figure 3-2 Mean Latency (ms) ± 1 SE by Modality, Task and Musical IOI

The interaction between task and modality was significant, *F*(1,29) = 25.300, $p < .001$, $\eta^2 = .466$. The tap-contingent latencies in the explicit task (M = 294.9ms, SD = 27.9) were significantly faster than in the implicit condition (M = 329.5ms, SD = 37.7), *t*(29) = 5.14, $p < .001$. This is unsurprising as the musical IOI levels represent the self-paced timing of eye movements and not finger-tap movements which are slower. In contrast the gaze-contingent latencies in the explicit task (M = 283.2ms, SD = 26.8) did not significantly differ from implicit (M = 287.0ms, SD = 30.3) latencies, *t*(29) = 1.01, $p = .646$.

The interaction between task and musical IOI was significant, *F*(2,58) = 27.351, $p < .001$, $\eta^2 = .485$. The latencies in the explicit Task varied significantly across the musical IOI levels and the implicit task latencies did not (collapsed across modality). The main effect of IOI in the explicit task on latencies was highly significant *F*(2,58) =

44.847, $p < .001$, $\eta^2 = .607$, as the latencies increased linearly as the musical IOI increased ($p < .001$). In contrast, for the implicit trials there was no significant main effect of musical IOI, $F(2,58) = .535$, $p < .588$, $\eta^2 = .018$, rather there is positive Bayesian evidence for the null hypothesis ($BF_{01} = 6.755$), that there was no evidence for variance across musical IOI levels.

The interaction between modality and musical IOI was marginally significant, $F(2,58) = 2.952$, $p = .060$, $\eta^2 = .092$. This is likely due to the size of the variance or steepness in the slope between the explicit tap and gaze-contingent conditions, observed in Figure 3-2. There was a significant main effect of musical IOI on latency for the gaze-contingent condition $F(2,58) = 4.108$, $p < .021$, $\eta^2 = .124$, although the tap-contingent condition had a larger effect size $F(2,58) = 8.927$, $p < .001$, $\eta^2 = .235$. Both modalities varied in the same direction, the tap-contingent latencies varied more across the musical IOI levels.

Finally, the three-way interaction between modality, task and musical IOI was not significant $F(2,58) = 1.642$, $p = .202$, $\eta^2 = .054$. In summary, both the gaze and tap modalities varied linearly with the musical IOI during the explicit task, evidence that both modalities can vary in timing subject to a predictable auditory referent. During the implicit task, the presence of the irrelevant music for both modalities showed evidence of invariance in latency across the IOI levels. Therefore, the musical tempi changes did not implicitly alter the timing of eye or tap movements. There was an overarching difference in latency responses between the two modalities, especially in the implicit tap-contingent condition, which was much slower than the gaze and explicit latencies. It is plausible that the musical IOI used were too fast to promote implicit entrainment in tap responses, although the tempi were achievable, as-per the explicit task.

Mean Latency Error from the Musical IOI

Neither modality significantly varied their latency over the IOI levels in the implicit condition, whereas in the explicit condition the tap responses varied significantly more in latency than the gaze latencies. An appropriate measure of performance is the proximity of responses in relation to the target IOI (as an absolute value), specifically the difference (or error) between the mean latency of each modality and the target IOI in the explicit synchrony condition.

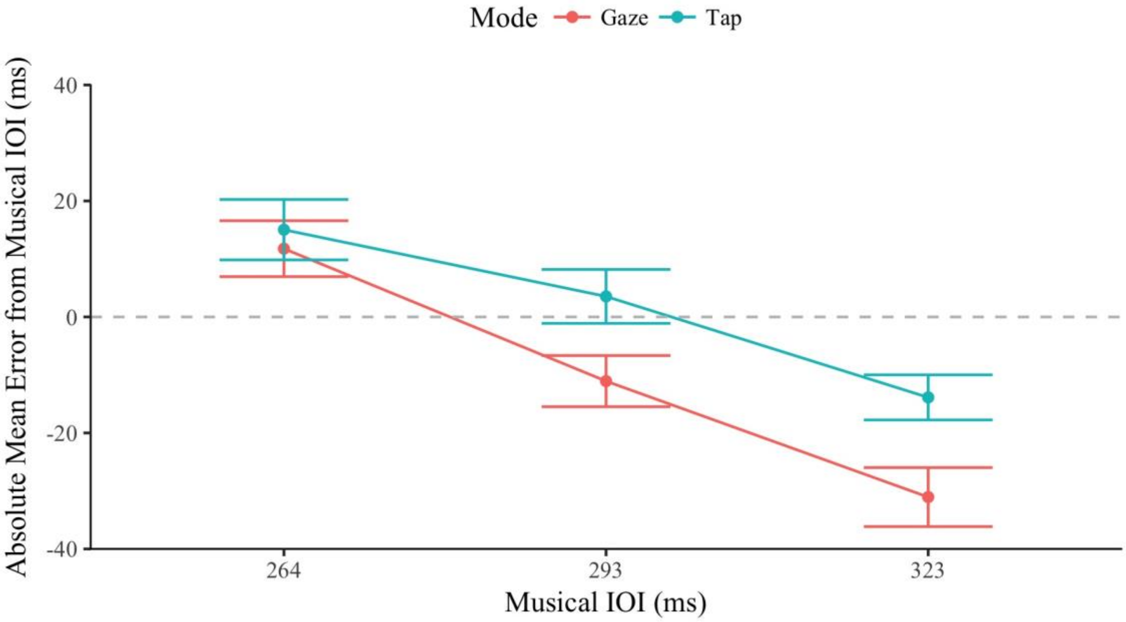


Figure 3-3 Absolute Mean Latency Error (±1SE) by Target Musical IOI for Each Modality During the Explicit Task

As observed in Figure 3-3, there was a significant main effect of modality, $F(1,29) = 11.449$, $p = .002$, $\eta^2 = .283$, as the tap Error (M = 1.57ms, SD = 14.6) was significantly smaller than the gaze error (M = -10.12ms, SD = 21.4). There was a significant main effect of IOI, $F(2,58) = 107.921$, $p < .001$, $\eta^2 = .788$, and Bonferroni corrected pairwise comparisons confirmed significant differences in error between all

IOI levels. The error was a positive value (faster mean latencies than the target IOI) at 264m*s* (M = 13.4ms, SD = 2.3), and decreased linearly to a negative error (slower mean latencies than the target IOI) at 323ms (M = -22.5ms, SD = 12.2). The 293ms condition had the smallest absolute error (M = -3.8ms, SD = 10.3). There was also a significant interaction between modality and musical IOI, $F(2,58) = 5.728$, $p = .005$, $\eta^2 = .165$. To further unpack the interaction Bonferroni corrected t-tests (correction for three tests, $\alpha = .0167$) compared the error values of the modality levels at each level of IOI. At 264ms there was no significant difference in error between the two modalities, $t(29) = .744$, $p = .463$. At 293ms there was a significant difference in error; the gaze error was greater (-11.7ms, SD = 27.2) than the finger-tap error (M = 3.5ms, 25.5), $t(29) = 3.290$, $p = .003$. Finally the error at 323ms was much greater in the gaze condition (M = -31.1ms, SD = 27.9) than the tap condition (M = -13.9ms, SD = 21.3), $t(29) = 4.302$, $p < .001$. The interaction between the musical IOI and modality factors is the result of similarity in error at the fastest 264ms IOI between the two modalities. As the musical IOI slowed the gaze error significantly worsened; the gaze latencies were increasingly too fast when the IOI target slowed.

In summary, the tap-contingent responses were a closer approximation of the musical IOI than the gaze-contingent responses. This is confirmation that the differences in latency between the two modalities across the musical IOI levels (marginally significant interaction in the prior ANOVA) suggest some limitation in the direct control of eye movements that limits volitional control over when the eye moves. This is most notable when the referent interval moves away from the natural pace of movement.

Gaze behaviour differences by orientation modality.

The difference between gaze-contingent and tap-contingent visual onsets provide an interesting timing difference for the comparison of eye movements. Are gaze-contingent movements more or less temporally precise than those produced by tap-contingent responses? As previously discussed, the gaze-contingent responses had a greater mean error from the target IOI than the tap-contingent latencies (in the explicit condition). As the tap timings were more representative of the IOI levels than the gaze latencies, the eye movements produced by each orientation modality may have different timing and provide insight into the differences between gaze contingent and tap-contingent visual timing.

In the following analysis, the eye movement latencies will be compared by Orientation Modality (Gaze or Tap-contingent) and Musical IOI. Significant differences in latency across Musical IOI levels was only observed in the Explicit task (there was positive Bayesian evidence for invariance across musical IOI in the implicit condition, $BF_{01} = 6.755$), therefore this analysis will focus only on explicit synchronisation.

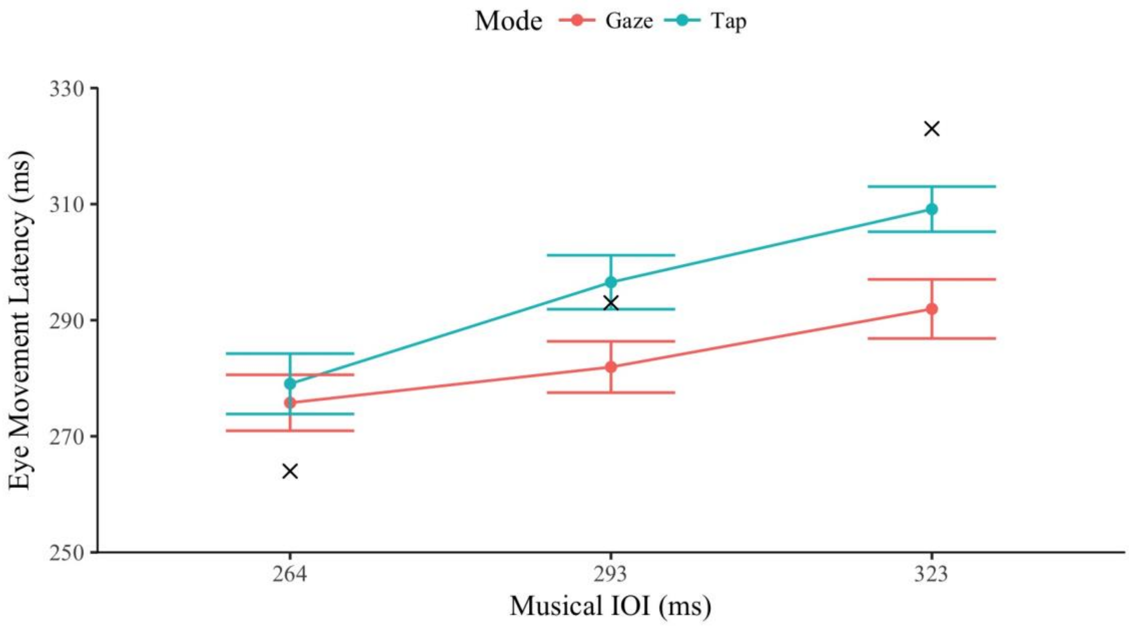


Figure 3-4 Eye Movement Latencies (ms) ±1SE from the Explicit Task by Orientation Modality and Musical IOI (the X represents the actual musical IOI)

A two-way repeated measures ANVOVA compared eye movement latencies by orientation modality and musical IOI. There was a significant main effect of orientation modality on eye movement latencies, $F(1,29) = 11.449$, $p = .002$, $\eta^2 = .283$, the tap-contingent eye movement latencies were significantly slower (M = 294.9ms, SD = 27.9) than the gaze-contingent latencies (M = 283.2ms, SD = 26.8). The main effect of musical IOI was also significant, $F(2,58) = 44.847$, $p < .001$, $\eta^2 = .283$, the eye-movement latencies increased significantly as the IOI increased; Bonferroni corrected pairwise comparisons confirmed significant differences between each level of IOI ($p < .001$). Finally, there was a significant interaction between orientation modality and musical IOI, $F(2,58) = 5.728$, $p = .005$, $\eta^2 = .165$. To investigate the interaction three Bonferroni-corrected t-tests ($\alpha = .0167$) compared the two orientation mode eye-movement latencies at each level of Musical IOI. There were no significant differences between the two orientation modality latencies at 264ms, *t*(29) = .744, *p* = .463. At both the 293ms and 323ms musical IOI the tap-contingent eye-movements were significantly slower than gaze-contingent eye-movements (293ms: *t*(29) = 3.290, *p* = .003; 323ms: *t*(29) = 4.302, *p* < .001). The eye movements in the tap-contingent condition differed significantly by IOI and were a closer approximation of the target IOI (see Figure 3-4) than the gaze-contingent (or directly controlled) eye-movements.

3.4.1 Summary of latency results.

There was no evidence of variation in latency in either the tap or gaze modalities during the implicit task, rather there was convincing evidence for the null hypothesis; the presence of irrelevant music or changes in the IOI of background music did not significantly alter the timing of either modality. The explicit synchrony task did influence the latency of movements for both modalities. The magnitude of variance in latency across the three IOI levels was greater in the tap condition, which was a closer approximation of the target IOI. The gaze-contingent explicit latencies did vary in the predicted direction, but were more conservative (less accurate) shifts in relation to the target IOI. At the fastest and slowest IOI they were too slow and too fast respectively.

An interesting behaviour of note during the experiment was the ability of participants to volitionally quicken their eye movement latencies and move around the circles at a faster speed during the 264ms explicit trials. Eye movement timing models, for example, CRISP (Nuthmann et al., 2010), consider direct control as an inhibitory-only process; the engagement of direct control (cognitive demand) inhibits the rate of rise, random walk to threshold, or cancels saccades. The influence of volitional or direct control as a promoter of eye movements is absent and seemingly contrary to these models. These data suggest it is possible to volitionally speed the timing of eye movements, without a visual reason to do so (unlike express saccades, Carpenter, 2001).

A plausible explanation is that top-down influence could deemphasise the moment to moment cognitive demands of a task (e.g. by processing less deeply or responding less accurately), which could facilitate a reduction in inhibition and promote faster movements.

Finally, the eye movement latencies during the tap-contingent explicit trials varied significantly by IOI, and were generally more aligned with the referent IOI than when directly controlled in the gaze-contingent condition. This novel finding is evidence that firstly, the volitional direct control of eye-movements is less temporally precise than the automatic response to motor-produced visual timing demands, and secondly, that eye-movement timing is highly sensitive to, and modulated by the timing of visual information produced by motor movements.

Mean resultant length.

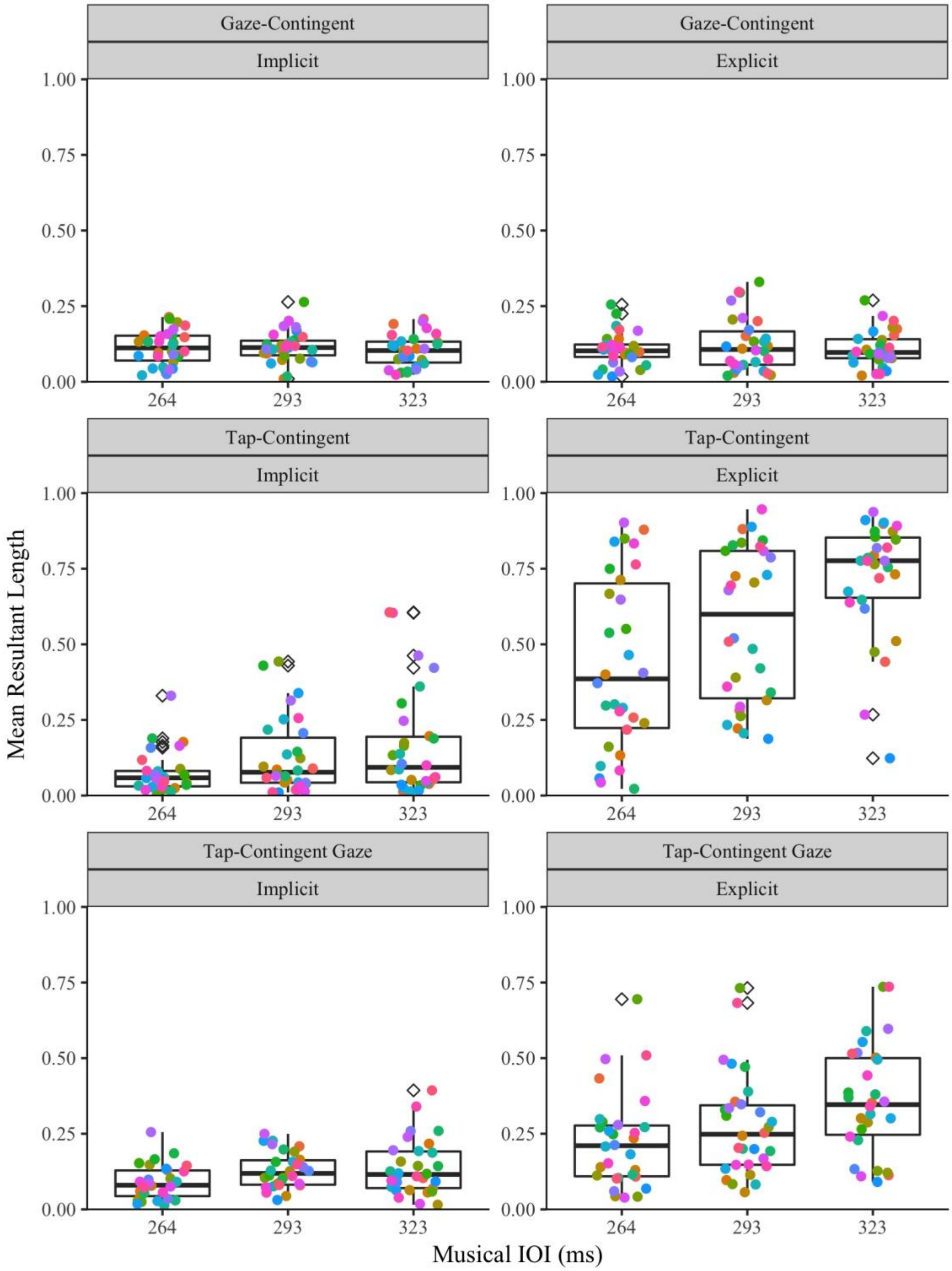


Figure 3-5 Mean Resultant Length Boxplots by Modality Measured, Task and IOI (colours represent individual participants), the diamonds represent potential outliers who performed notably better or worse than the group

The mean resultant length provides a scaled measure of synchronisation performance, from 0 (random) to 1 (perfectly synchronised). As observed in the boxplots of Figure 3-5, no participant in the gaze-contingent condition achieved an obviously synchronised MRL score (e.g. above .5 threshold). The minimum test of circularity in vectors is the Rayleigh test (Wilkie, 1983), which confirms if the vectors are significantly not uniform in phase. As observed in the descriptive Table 3-1, both gaze-contingent conditions had a small proportion of participants at the threshold of phasic behaviour. In contrast, the proportion of phasic tap-contingent responses in the implicit condition increased notably with a slowing in musical IOI (to 43% of participants).

Table 3-1 The Proportion of Participants with Significantly Circular Data by Modality, Task and IOI (significantly non-uniform with the Rayleigh test)

	Gaze		**Tap**	
	Implicit	Explicit	Implicit	Explicit
264 IOI	.10	.03	.17	.83
293 IOI	.10	.17	.33	1
323 IOI	.03	.13	.43	.97

The following ANOVA should be interpreted with some caution, as it is comparing gradations of randomness for some participants (especially in the gaze-contingent conditions). A three-way repeated measures ANOVA compared MRL values (synchronisation) by modality (gaze and tap-contingent), task (implicit and explicit) and musical IOI (264, 293 and 323ms). There was a significant main effect of modality on MRL values, $F(1,29) = 130.165, p < .001, \eta^2 = .818$, as the tap-contingent responses (M = .348, SD = .312) were significantly higher than gaze-contingent (M = .111, SD =

.062). The main effect of task on MRL values was significant, $F(1,29) = 113.369$, $p < .001$, $\eta^2 = .796$, explicit responses (M = .344 , SD = .305) were significantly higher than implicit (M = .116, SD = .103). The main effect of Musical IOI was also significant, $F(2,58) = 23.162$, $p < .001$, $\eta^2 = .444$; the MRL values increased linearly (p < .001) as the Musical IOI level increased, the synchronisation improved when the IOI slowed.

The interaction between modality and task was significant, $F(1,29) = 110.329$, $p < .001$, $\eta^2 = .792$. Bonferroni corrected t-tests of the two task levels for each modality confirmed significant differences between the explicit (M = .575, SD = .218) and implicit tap MRL scores (M = .121, SD = .089), $t(29) = 10.75$, $p < .001$. There was no evidence for differences between the explicit (M = .113, SD = .046) and implicit gaze MRL (M = .110, SD = .035), $t(29) = .404$, $p = .689$, $BF_{01} = 4.769$.

The interaction between modality and IOI was significant, $F(2,58) = 27.598$ $p < .001$, $\eta^2 = .488$. MRL performance in the tap contingent condition varied significantly by musical IOI, $F(2,58) = 30.390$, $p < .001$, $\eta^2 = .512$, as the musical tempi slowed synchronisation values increased ($p < .001$). In contrast, gaze-contingent MRL performance showed no evidence of variance over the tempi, $F(2,58) = .847$, $p = .434$, $\eta^2 = .028$, $BF_{01} = 5.014$.

The interaction between task and musical IOI was significant, $F(2,58) = 9.305$, $p < .001$, $\eta^2 = .243$. The MRL performance during the explicit task varied significantly by musical IOI, $F(2,58) = 25.340$, $p < .001$, $\eta^2 = .466$, the synchronisation performance improved linearly as the musical IOI slowed ($p < .001$). The implicit MRL values also varied by musical IOI, $F(2,58) = 3.212$, $p < .048$, $\eta^2 = .100$, as with the explicit condition the MRL values increased linearly as the musical IOI slowed ($p = .017$). The interaction is explained by the difference in IOI effect size between the two tasks; the increase in performance as the IOI slowed was larger (steeper) during the explicit task.

The three-way interaction between modality, task and musical IOI was also significant, $F(2,58) = 6.888$, $p = .002$, $\eta^2 = .192$, observed below in Figure 3-6. To unpack the three-way interaction further analysis will separate the data by modality.

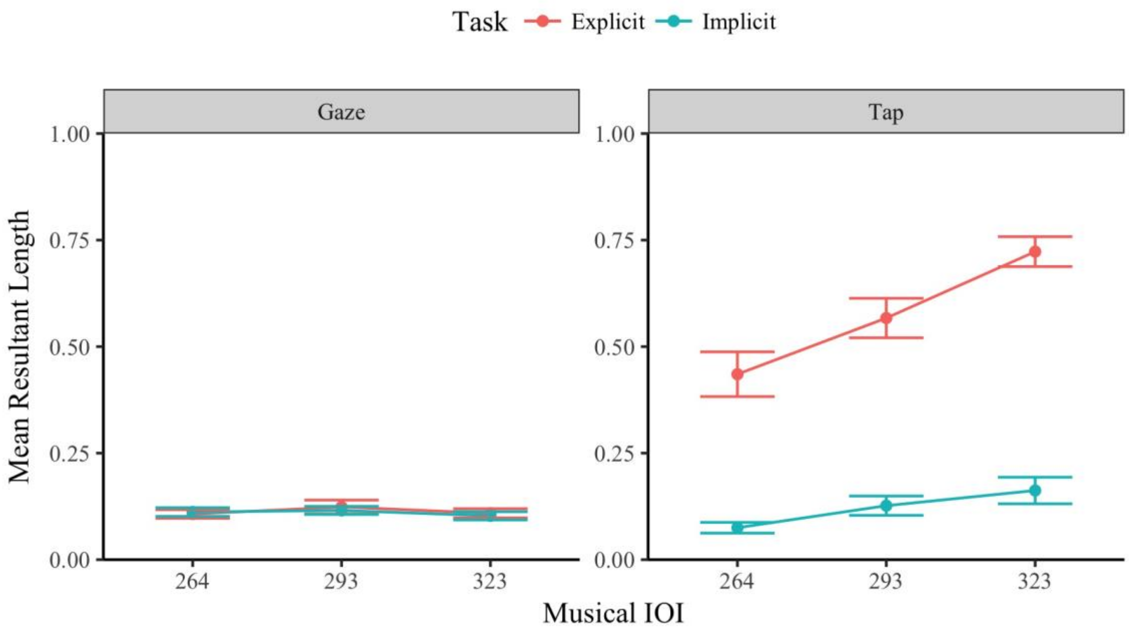


Figure 3-6 Mean Resultant Length Values (± 1 SE) by Modality, Task and Musical IOI

A two-way repeated measures ANOVA compared gaze-contingent MRL values by task (implicit and explicit) and musical IOI (264, 293 and 323ms). As observed in Figure 3-6, there was no evidence for variance between the two tasks ($BF_{01} = 5.944$), or across the three musical IOI levels ($BF_{01} = 7.384$), or as an interaction between task and musical IOI ($BF_{01} = 43.404$); all frequentist tests were not significant ($F < 1$, $p > .4$, $\eta^2 < .03$). The synchronisation performance during the gaze-contingent trials was uniformly poor (random) across task and musical IOI level; the explicit control of eye movements may have altered the latency of gaze timing toward the phase duration of the IOI, but did not generate a synchronised response.

A two-way repeated measures ANOVA compared the tap-contingent MRL values by task and musical IOI. There was a significant main effect of task, $F(1,29) = 115.516, p < .001, \eta^2 = .799$. Synchronisation performance was better when explicitly tapping (M = .575, SD = .272), than in the implicit condition (M = .121, SD = .132). The main effect of musical IOI was significant, $F(2,58) = 30.392, p < .001, \eta^2 = .512$, as synchronisation performance increased linearly ($p < .001$) as the musical IOI slowed. The interaction between task and musical IOI was also significant, $F(2,58) = 9.434, p < .001, \eta^2 = .245$, as the effect size of the linear increase was higher ($\eta^2 = .507$) than the effect size of the implicit increase ($\eta^2 = .136$). Both tasks had a significantly linear increase in MRL as the musical IOI slowed (Explicit $p < .001$, Implicit $p = .014$). The only indication of wide-spread synchronisation during the task occurred in the explicit tap-contingent condition. The performance on the two gaze-contingent tasks was invariantly low. A comparison between the two modalities during the implicit task was conducted to observe whether the implicit tap responses were more synchronised than gaze. The implicit tap-contingent MRL (M = .121, SD = .089) was not significantly different from the implicit gaze-contingent MRL (M = .110, SD = .035), $t(29) = .615, p = .543, BF_{01} = 4.320$. The prior analysis displayed a significant linear increase in implicit tap-contingent responses across IOI levels. Despite this increase, the tap-contingent performance was not generally more synchronised than the gaze-contingent condition.

Eye movement synchronisation by orientation modality.

A final comparison of interest is the difference in eye-movement synchronisation between the two orientation modes (gaze or tap-contingent) in the explicit condition. As the tap timing was much more synchronised than the gaze-contingent timing, analysis of the eye movement MRL values between the two conditions would confirm whether the eyes can move in time to a self-generated motor-produced (somewhat isochronous) visual presentation.

The Rayleigh's Test of circularity in the tap-contingent eye movements showed that a larger proportion of participants were able produce a phasic eye movement response when following an exogenous visual rhythm (264 IOI: 50% of participants; 293 IOI: 63%, and 323 IOI: 73%). The Figure 3-7 displays the relationship between synchronised responses (MRL) and where in phase the fixation began (mean direction; values to the right of 0, e.g. 1 to 90 indicate fixations beginning late, values to the left of 0, e.g. 270-359, indicate early fixations). Those participants who had a more-synchronised response (a longer line) appear to start fixations prior to the IOI onset (before 0) in phase, increasingly in the 323ms condition. This could be some initial evidence that the timing mechanisms underlying eye movements were predicting the visual onsets produced by the motor-movements through time (as opposed to reacting to a peripheral visual target).

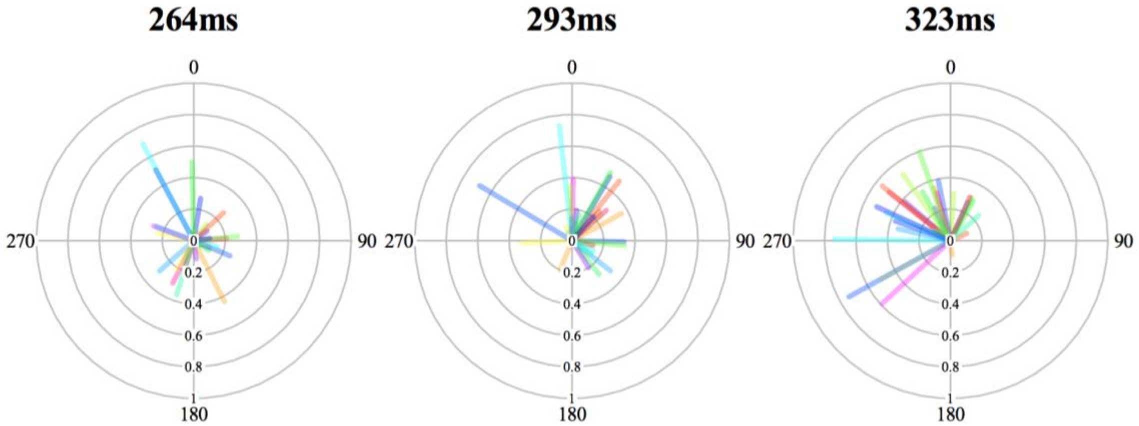


Figure 3-7 Mean Direction (phase location, 0 is the visual or audiovisual onset) and Mean Resultant Length (line length where 1 is the circle radius) by IOI (colours represent participants) of the Explicit Tap-Contingent Gaze Responses

The tap-contingent eye movements (M = .287, SD = .146) were significantly more synchronised than the gaze-contingent eye movements (M = .113, SD = .046), *t*(29) = 6.994, *p* < .001, as observed in the boxplot Figure 3-5. The eye movements during the tap-contingent task were significantly less synchronised than the tap-contingent responses themselves, *t*(29) = 9.461, *p* < .001.

3.4.2 Summary of mean resultant length results.

Explicit tap responses were significantly more synchronised than eye movements, both when completing the tap-contingent task or when directly controlling visual timing in the gaze-contingent condition. There was minimal evidence for implicit synchronisation in either modality, although the tap condition did show some increase in the number of phasic responses and the level of MRL increased notably as the musical IOI increased. There was no evidence of difference in synchronisation between the implicit and explicit gaze-contingent conditions, as both were generally random in phase. This particularly interesting as the 293ms musical IOI was very close in phase duration to the natural pace of movement in silence, requiring only minimal change in phase location to produce synchronised movement. These small phase changes were either exceeding the capacity of direct control to modulate eye movements, or were too small for an appropriate temporal shift.

The participants were attempting to match the speed of the target IOI, which did not automatically facilitate or evidence a phasic response. An interesting finding was that the eye-movement synchronisation to the imperfectly timed motor-produced visual sequence (the tap times were not consistent in time) was significantly more synchronous than when directly controlling the timing of eye movements. This suggests that the timing of eye movements can respond to temporal regularities generated by the tap-contingent display albeit less precisely than the taps themselves. Furthermore, an motor-produced visual referent in time appears to be perceived and predicted differently to a gaze-contingent visual onset. The temporal control of visual information at fixation appears to incur some delay not present when following or predicting a motor generated visual referent.

3.5 Discussion

The initial aim of this experiment was to measure the implicit and explicit musical entrainment of eye movements, and contrast it with finger-tap measures. The predictions for implicit entrainment were mixed, as the previous experiments E1 and E2 did not provide compelling evidence of implicit music or tempo effects on eye movements. Prior research had identified implicit effects of musical tempi on motor and eye movements (Day et al., 2009; Kuribayashi & Nittono, 2014), although there is recent evidence that intention is required for synchronisation of motor movements (Leow et al., 2017). Neither eye movement or finger taps displayed evidence of variance in latency between the silent control or the three musical IOI levels. Few participants displayed even a minimally phasic response during the implicit task (although there were more participants with significantly circular implicit tap responses than in gaze). The synchronisation (MRL) was generally random for both modalities during the implicit task.

The absence of synchronisation in eye movements is more than simply the absence of intent or instruction. When contrasting the mean resultant length performance between the implicit and explicit task, we find no evidence of any increase in synchronisation. The duration between eye movements varied during the explicit task as a function of the musical IOI, what could be loosely classified as correction in phase duration (approximating the interval duration), although these shifts were much more conservative than those observed for tap responses. What was notably absent was the smaller corrections (inhibition) to the timing of gaze-contingent movements within phase to match the beat (small corrections in time to correspond to the same location in phase, e.g. moving on the tone in time). A likely explanation is that randomness is inherent to the timing of saccades, as incorporated in the SWIFT, CRISP and ICAT models (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014), and serves a wider adaptive process (Carpenter, 1999) that protects the timing of visual processing and in this instance produced uniformly circular responses. An alternative explanation is that the demands of the task, which require remembering which key to press when the letter sequence changed, could have impaired the temporal control of eye movements. The explicit control of eye movements may improve if the demands were reduced.

An interesting outcome from the exploratory analysis of eye movements between the orientation modalities (gaze and tap-contingent visual timing) was the notable increase in performance for both the latency (durations varied more across the IOI levels as a closer approximation of the IOI) and MRL values when the eye movements corresponded to the tap-contingent visual transitions, compared to the gaze-contingent timings. The tap-contingent visual timings, whilst not perfectly synchronised themselves, did significantly increase the synchronised performance of eye movements. There was clearly a difference in how the visual system processed the motor-prodcued visual onsets as opposed to the fixation-triggered onsets that facilitated predictive movements in the former but not the latter. One implication could be that synchronised attention synchrony between visual onsets and musical beats described in Escoffier et al. (2010) is only present when the visual onsets are not gaze-contingent. The perceptual impairment in processing the timing of gaze-contingent onsets is likely due to the lack of self-awareness of fixations (Clarke, Mahon, Irvine, & Hunt, 2016), and the blurring of time across saccades (Yarrow, Haggard, & Rothwell, 2008; Yarrow, Haggard, Heal, Brown, & Rothwell, 2001). In contrast, when the visual onsets are not the product of a gaze-contingent display, the visual system appears to predict future visual onsets and modulate its timing to coincide in time (both speeding and slowing of the phase duration), even when the timing of the taps in this experiment were not exactly synchronised. The addition of musical beats could be providing an additive temporal cue that aids the predictive timing of eye movements to corresponding visual onsets. A perfectly isochronous sequence would facilitate analysis of the capacity for synchronised movements to a perfectly predictable sequence in addition to the additive influence of corresponding musical beats, which may inform the predictive timing of eye movements.

Chapter 4

Following the Lead: Endogenous and Exogenous Eye Movement Synchronisation

4.1 Chapter Overview

Chapter three identified differences in eye movement timing and synchronisation behaviour when responding to motor-generated visual transitions, compared to self-paced gaze-contingent visual transitions. This chapter replicates the gaze-contingent findings of the previous chapter, and then extends the motor-produced enhancement to test whether music aids the predictive timing of eye movements when corresponding with (or competing with) exogenously controlled visual onsets. The addition of music did not enhance the synchronisation of eye movements, but did systematically slow them by 3ms when corresponding in time.

4.2: Experiment Four: Endogenous and Exogenous Eye Movement Synchronisation

Exploratory analysis in Experiment three (E3) compared eye movement timing and synchronisation performance across two orientation types: gaze and tap-contingent visual onsets. The visual onsets (letters presented within small black circles) that were produced by the finger-taps had corresponding eye movement latencies that were closer to the referent IOI (better period correction), and were more synchronised than those produced by gaze-contingent (endogenous) timing. The timing and synchronisation of eye movements during the tap-contingent trials of E3 indicated some predictive response in the timing of gaze-shifts. The temporal control of volitionally timed eye movements around the gaze-contingent display was notably poor by comparison.

The previously discussed, research of Escoffier et al. (2010; 2015) found that on-beat audiovisual presentations (visual onsets corresponding with musical beats) had faster reaction-time responses than off-beat ones. They argue that this is evidence of musical beats entraining visual attention. The prior experiments of this thesis (E1 to E3) extended this research to consider whether eye movements, which are deemed the locus of attention (Deubel & Schneider, 1996; Kowler et al., 1995), also implicitly entrained to musical beats. These gaze-contingent experiments provide evidence that the timing of eye movements are impervious to the implicit entraining effect of musical beats, likely due to inherent randomness of saccade timing (Nuthmann et al., 2010), the limitations in self-awareness of fixation behaviour (Clarke et al., 2016; Clarke & Tatlet, 2014), and the absence of intention to synchronise (Leow et al., 2017). The Escoffier et al. (2010)'s task did not require spatial shifts in gaze to process the on-beat audiovisual onsets as all visual stimuli were presented centrally. In the prior experiments in this thesis there was an expectation that participants would synchronise their eye movements with the musical beats to optimise attention as shown in Escoffier et al. (2010) overt task. The participants could not do this, therefore this study will consider eye movements to exogenous rhythms (the interval of interest is set by an external source), to observe whether the addition of musical beats that match these visual rhythms will entrain eye movements and aid their predictive timing. To achieve this requires an active vision task (a visual task involving spatial shifts that reorient gaze to discriminate visual differences). Therefore, this experiment will extend the existing visual search paradigm to measure eye movements to sequences of pre-set isochronous (equal time-points between the onsets) visual timing intervals (T and L letter movements around circles), which are either matched by a corresponding musical beat, visually isochronous without music, or randomly visually timed (with and without isochronous musical beats). In addition to the eye movement measures as employed in Escoffier et al. (2010) the reaction times to changes in sequences of letters will be contrasted as a behavioural measure of entrained attention.

An initial subsection of this experiment (4A) will replicate the previous implicit and explicit gaze-contingent element of E3 with a simplification of the task. A plausible limitation in the previous experimental paradigm was that the cognitive effort required to recall and press the correct key impaired synchronisation performance. To simplify this decision, this modified experiment will require a simple click of a right trigger (on a game-controller) when letter changes occur irrespective of the letter. This decision simplification should further reduce the difficulty of the task and facilitate more cognitive resource allocated to the synchronisation of movement. The prediction for the 4A subsection is that the reduction in task difficulty will facilitate greater explicit control, evidenced by the latency, error and synchronisation performance differences between the explicit and implicit conditions. There are no predictions for the implicit condition due to the strength of the prior evidence for invariance in eye movement timing across IOI levels.

The second subsection (4B), will measure eye movement timing and synchronisation as well as reaction times to target-letter changes across four condition types: Isochronous Music (the visual sequence and musical IOI match), Isochronous Silent (a consistent visual sequence with no music), Random Music (the musical IOI is constant and the visual sequence is randomly timed), and Random Silent (a randomly timed visual sequence in silence). The core prediction is that the Isochronous Music condition would have the most salient and predictable moments in time (as the onsets match across modalities), facilitating the temporal allocation of attention through time (Jones & Boltz, 1989). Therefore, the Isochronous Music condition will have enhanced eye movement synchronisation and latencies that vary by IOI as well as speeded reaction times when compared to the random and two silent conditions. The reaction time performance will be the slowest in the Random Music condition as the timing information from the two modalities is competing and mismatched, limiting or dividing the predictive allocation of attention through time.

4.3 Method.

4.3.1 Participants.

Power analysis using G*Power (Faul et al., 2007) of the prior study (E3) utilised the effect size of the interaction between explicit and implicit gaze-contingent latencies (η^2 = .225) to identify a sample size for this study (initial indications suggested N=16, powered to d = .95 with α = .05). As the implicit latency effect was very small (η^2 = .020) and the explicit effect much larger (η^2 = .291), we wanted to ensure the study was sufficiently powered to evidence some implicit effect (if present), so pre-set the sample size to a compromised value of 24 (the interaction was powered to d = .997), sufficient to power a medium effect size within the design. The study tested 34 naïve participants and of these 10 were either unable to tap in time (2/10), had some lost eye tracking data in at least one condition (4/10), or had drop in response accuracy below 60% during 4B. (4/10). This left 24 participants (17 Female), with an age range from 20 to 43 (M = 24.67, SD = 5.52), who contributed data for analysis. The Gold-MSI (Müllensiefen et al., 2014) musical training scores (on a scale from 7: no training to 49: more training than 95% of the population) indicated a range of training from 8 to 37 (M = 22.625, SD = 9.28); none of the participants were professional musicians.

4.3.2 Design modifications.

There were minimal changes to the experimental paradigm for the first subsection of the experiment (4A). The only modification was how the response to changes in visual sequences were collected. Rather than a keyboard key-press that corresponded to the correct letter on screen, the new experiment required simple click of the right trigger on a Microsoft game-controller. This reduced the difficulty of the task by reducing the decision at a change. This change also facilitated a consistent instruction with the latter half of the experiment (4B), which required a fast (minimal equipment latency) response-type that only the controller could provide. The three musical IOI levels (264, 293 and 323ms) all remained the same as E3, as did the visual stimuli, screen type, eye-tracker, eye-tracking setup, and data cleaning steps.

4.3.3 Procedure.

This experiment was divided into two subsections (4A and 4B). Firstly, each participant completed a replication of the implicit gaze-contingent trials of E3 (4A). This involved navigating clockwise from circle to circle around the elliptical gaze-contingent display, clicking the right trigger each time the letter sequence changed between T and L (or L and T). The first trial was a practice trial in silence to familiarise participants with the task, and was followed by four implicit trials (the three IOI levels and a silent trial in random order). The task order was constant (as in E3): the Implicit task was followed by the Explicit synchrony task (1 practice at 293ms, then 4 trials, 3 with music and a silent trial). The frequency of letter sequence changes for both tasks was relatively infrequent (between 5-7 per trial) as in E3. Each trial ended after 180 circles, or timed-out after three minutes.

Experiment 4B manipulated the timing of visual presentation and the presence of a corresponding (competing or absent) musical beat. The instructions were simple; follow the visual sequence of letters closely as they move around the circles, and click the right trigger quickly when the sequence of letters changes (between T and L). The changes in letter sequence occurred much more frequently than the previous trials (after 7 to 13 circles) to increase the number of reaction times collected in a trial and to promote close following of the letter sequence. The timing of the visual sequence (Sequence) condition was either Isochronous (letter transitions were at a constant interval, one of the three IOI levels) or Random (the time between each letter transition was randomly one of the three IOI levels). There were two Audio conditions: Firstly, Music which either matched the visual sequence (Isochronous Music), or added a competing Musical IOI when Random (Random Music). Secondly, a Silent condition which formed Isochronous Silent and Random Silent conditions. These four conditions were contrasted across the three IOI levels.

The IOI factor varied in which modality presented the interval (either visual, auditory, i.e. music, or both in the case of Isochronous Music), but was always one of the three IOI levels (264, 293 and 323ms) as used in the previous study. As detailed in Table 4-1, when the visual sequence was isochronous the IOI was either visual (when silent) or audiovisual (with music). When the Sequence was random the IOI was the musical beat or a randomised pre-assigned pseudo IOI level when silent (to balance the design), these different IOI representations are visualised in Figure 4-1. The mean resultant length value was calculated from the time difference (as a circular vector) between the fixation start time and the IOI. For the pseudo-allocated IOI in the random silent condition, the IOI period was a constant (at one of the three intervals) and had no relation to the stimuli presentation.

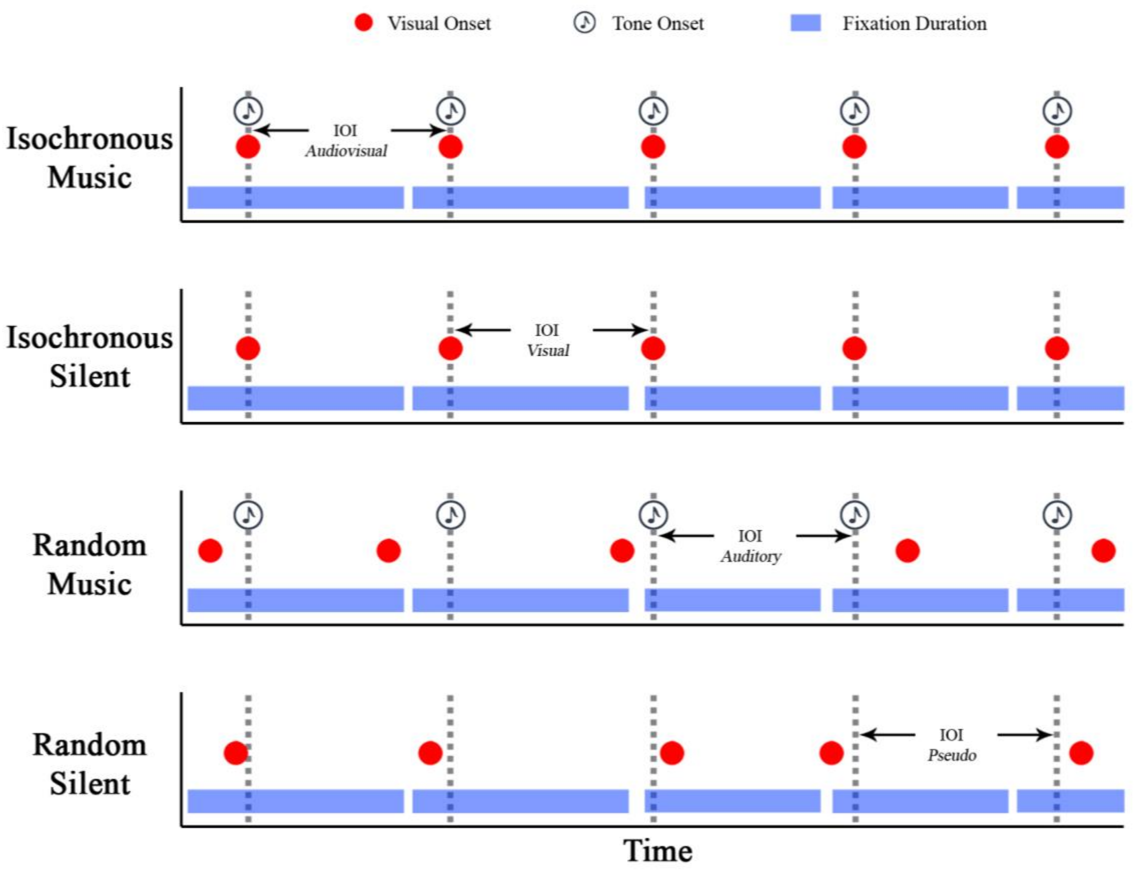


Figure 4-1 A Simulated Example of Fixation Behaviour Over Time in Relation to the IOI (dashed grey line) as an Audiovisual, Visual, Auditory or Pseudo Allocated Constant Through Time Across the Four Conditions

Table 4-1 Experiment 4B Conditions, with IOI Representation

Sequence	**Audio**	**IOI Representation**
Isochronous	Music	Audiovisual
	Silent	Visual
Random	Music	Auditory
	Silent	Pseudo Allocation

Each of the twelve trial types, Sequence (2 levels) x Audio (2 levels) x IOI (3 levels), were repeated three times. The trials were randomised within each block of twelve. Breaks were provided after twelve trials or if the participant moved from the chinrest. Each trial ended after 10 letter sequence changes (at the 11th change to allow a reaction time to all 10 changes to be collected). Following each trial (including the practice) participants were presented their accuracy in detecting the letter changes (accurate responses (hit) minus false alarms as a percentage of all responses). If this detection accuracy was above 60%, the experiment continued with the written encouragement to aim for 100% on every trial. If accuracy dropped below 60% the experiment was held for the experimenter, who terminated it and paid the participant for their time. The task was simple and off-task behaviour, evidenced by low accuracy, indicated a loose following of the visual transitions, which impacted on the amount of usable data.

The experiments ended with six tapping in time trials on an Apple Magic Trackpad©, and each IOI was repeated randomly twice. As per the previous experiments each trial collected 60 taps. The final task was to complete the musical training questionnaire and demographic questions.

4.4 Results

4.4.1 Experiment 4A: Implicit and explicit entrainment of gaze-contingent responses.

Latency.

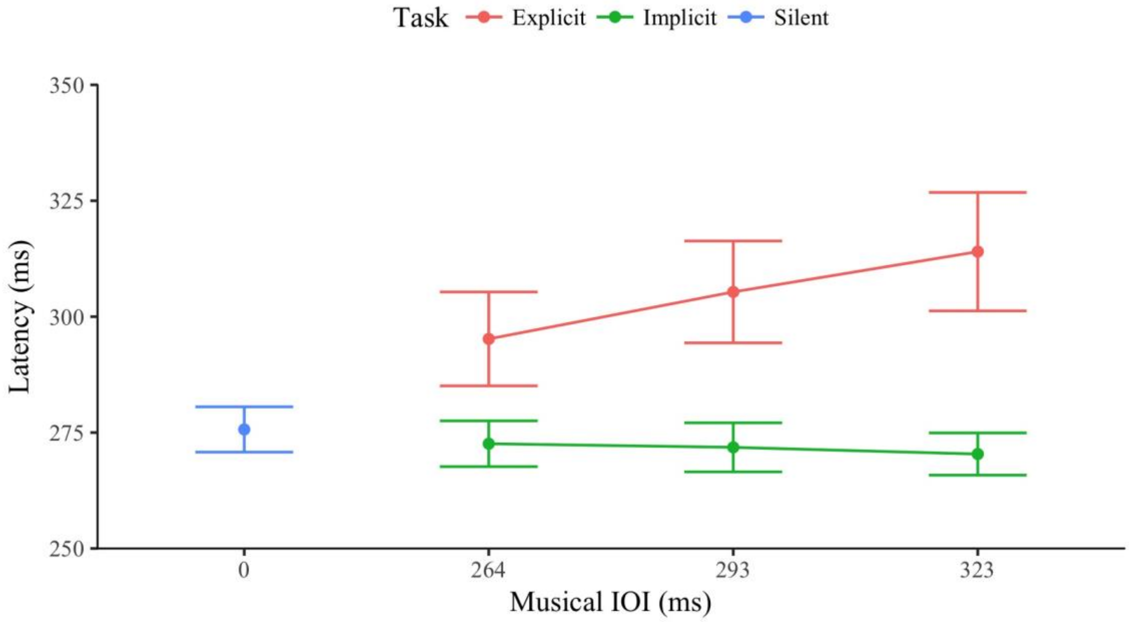


Figure 4-2 Mean Eye Movement Latency (±1SE) by Task and Musical IOI

A two-way repeated measures ANOVA compared eye movement latencies by task (implicit and explicit) and musical IOI (264, 293 and 323ms). As observed in Figure 4-2, there was a significant main effect of task, $F(1,23) = 12.660$, $p = .002$, $\eta^2 = .350$, the Explicit latencies (M = 304.9ms, SD = 55.4) were significantly slower than the Implicit latencies (M = 271.6ms, SD = 23.9). The implicit latencies in this experiment are notably faster than the E3 implicit eye movement latencies (M = 287.0ms, SD = 30.3), the reduction in task difficulty appears to have quickened eye movements. There was also a significant main effect of musical IOI, $F(2,46) = 5.007$, $p = .011$, $\eta^2 = .179$, Bonferroni corrected pairwise confirmed the linear trend ($p = .037$), eye movements slowed as the IOI slowed.

There was a significant interaction between task and musical IOI, $F(2,46) = 8.845$, $p = .001$, $\eta^2 = .278$. The latencies in the explicit condition varied significantly by musical IOI, $F(2,46) = 7.973$, $p = .001$, $\eta^2 = .257$; the latency slowed linearly as the musical IOI slowed ($p < .01$). The latencies were invariant across musical IOI levels in the implicit task, $BF_{01} = 5.32$ ($F < .7$, $p > .5$).

Mean absolute error from musical IOI.

Error responses in the explicit task were compared across musical IOI levels with a repeated measures ANOVA. There was a significant main effect of musical IOI, $F(2,46) = 36.282$, $p < .001$, $\eta^2 = .612$. The fastest 264ms musical IOI (M = 31.2ms, SD = 49.6) had a positive error (latencies were much slower than the target IOI). The 264ms IOI was significantly further from the referent IOI than both the 293ms (M = 12.3ms, SD = 53.8) and the 323ms IOI (M = -9.0ms, SD = 62.6). As in E3, the mean latencies in the 323ms condition were too fast to match the phase of the target IOI (all Bonferroni corrected, $p < .001$). These error values confirm a conservative pattern of gaze that did vary in the direction of the musical IOI, but notably failed to be either fast or slow enough as the IOI deviated from the natural pace of movement (293ms condition). The mean error values were generally smaller than those in E3 (264: 13.4ms, 293: -11.7ms, and 323ms: -31.1ms), with the exception of the 264ms condition. The reduction in task difficulty appears to have facilitated some further control over the duration of eye movements in the slower conditions.

Mean resultant length.

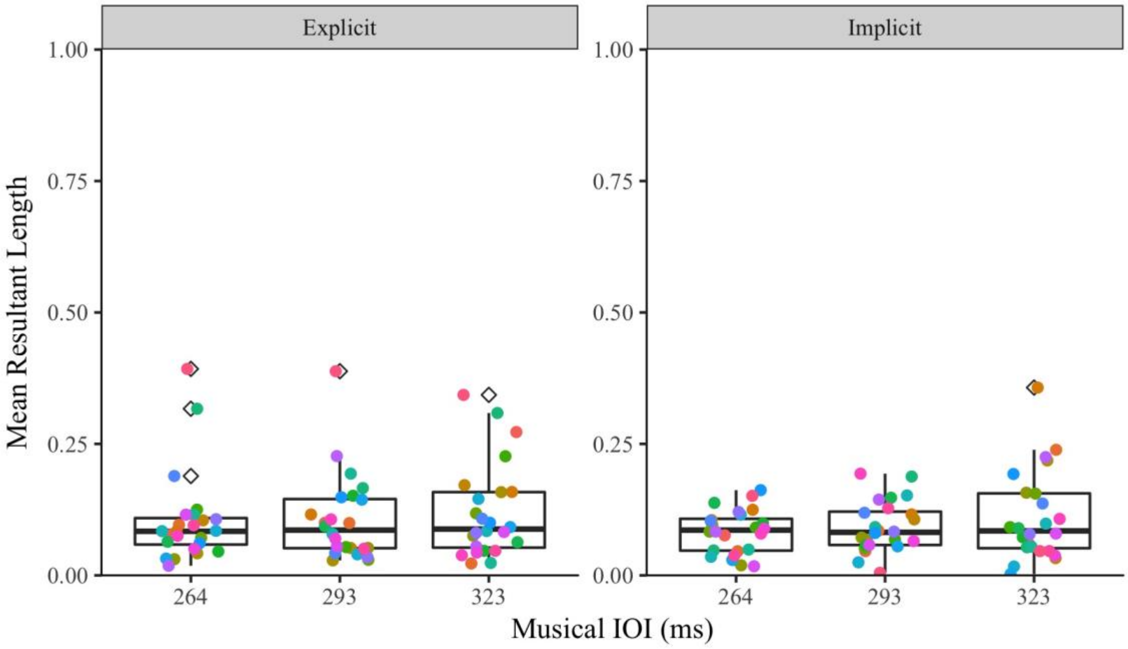


Figure 4-3 Mean Resultant Length values by Task and Musical IOI (colours represent individual participants, diamonds represent potential outliers)

The Rayleigh's test of circularity (a circular measure that tests the uniformity of circular data) found that only 12.5% of participants in the 264ms IOI, 29% in the 293ms and 33% in the 323ms were significantly non-uniform in the explicit condition. This generally randomly phased performance is observed in the boxplot Figure 4-3. No participant achieved a clearly synchronised MRL value, although one participant did achieve MRL values around .4.

A two-way repeated measures ANOVA compared the MRL values across the task and musical IOI levels. The main effect of task had weak evidence for the null hypothesis ($BF_{01} = 2.82$). The main effect of musical IOI had positive evidence for the null ($BF_{01} = 4.85$). There was also evidence for no interaction between task and musical IOI ($BF_{01} = 13.75$). The frequentist measures were all not significant ($F < 1.8$, $p > .15$).

Not only was MRL synchronisation performance generally poor, it did not improve when explicitly attempting to synchronise; the change in task difficulty did not aid performance.

4.4.2 Summary of replication (4A) results.

The results of this replication generally mirror those of experiment three (E3). There is a robustly invariant pattern of implicit latencies across the IOI levels and an imprecise (too conservative) variance in gaze latency in the explicit condition (error). The revised instructions and response method did alter eye movement latencies, and the error was smaller for the slower IOI (the latencies were closer to the target IOI). The implicit latencies were generally faster than in E3, with evidence that the prior task demands slowed the duration of eye movements. The faster implicit latency observed in this replication resulted in explicit latencies that were slower than implicit at each IOI (even though the target IOI was faster than the mean implicit latency in the 264ms condition). This may be evidence that the volitional speeding of eye movements (observed in E3) is the modulation in the depth of cognitive processing. As the task was highly simplified in this version, it likely limited the level of down-regulation of task demands that was possible. The prediction of the replication study was that the reduction in task difficulty would produce differences in synchronisation between the implicit and explicit tasks, but this was not supported. The synchronisation attempts ultimately had aphasic responses despite the IOI range being within the natural pace of eye movements for this task, and the closer phase duration shifts toward the target IOI during the explicit condition.

4.4.3 Experiment 4B: Isochronous and random visual sequences with music and silence.

The purpose of this subsection is to measure the influence of musical beats on the timing of eye movements (and reaction times) to exogenous visual sequences. We predicted that the corresponding audiovisual condition (isochronous music) will have higher levels of synchronisation, greater shifts in latency of eye movements, and speeded reactions when compared to the other conditions where the IOI is either, only visual (isochronous silent), musical (random music) or absent (random silent).

Latency.

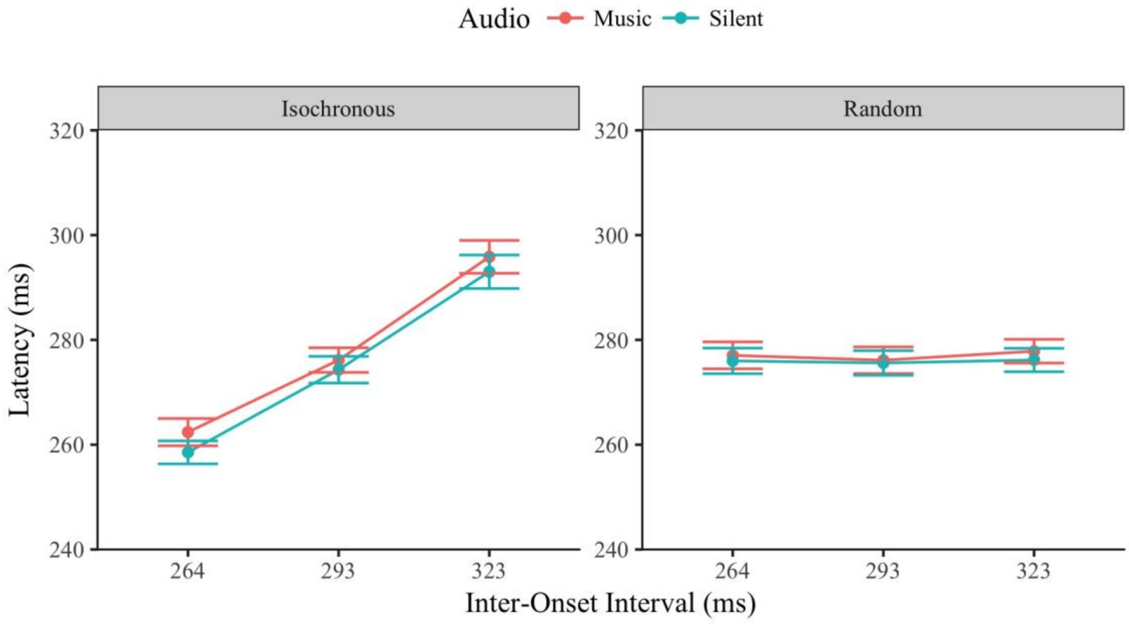


Figure 4-4 Mean Latency of Eye Movements (±1SE) by Sequence, Audio and IOI

A three-way repeated measures ANOVA compared the latency of eye movements across the sequence (isochronous and random), audio (music and silent) and IOI (264, 293, 323ms) conditions. There was no main effect of sequence, $F(1,23) = .259$, $p = .616$, $\eta^2 = .011$. The main effect of audio was significant, $F(1,23) = 17.630$, $p < .001$, $\eta^2 = .434$; the music trials (M = 277.6ms, SD = 15.9) had a small but significant slowing of latencies compared to silent trials (M = 275.6, SD = 15.7). There was also a main effect of IOI, $F(2,46) = 176.212, p < .001, \eta^2 = .885$, as the IOI slowed the eye movement latencies slowed linearly ($p < .001$).

The interaction between sequence and audio was significant, $F(1,23) = 5.391, p = .029, \eta 2 = .190$. Analysis of the effect of audio between the two sequence types confirmed the trend in Figure 4-4, that the isochronous sequence had slower eye movement latencies with matching musical beats (M = 278.1ms, SD = 11.99) than when silent (M = 275.3ms, SD = 11.78), $t(23) = 4.947, p < .001$. The random sequence did not significantly differ between the two audio conditions, $t(23) = 1.729, p = .10$. Note that only the audiovisual correspondences were slowed (3ms), the effect of music was not simply a presence of music effect as it did not influence the eye movements during the random sequence.

The interaction between sequence type and IOI was also significant, $F(2,46) = 123.859, p < .001, \eta^2 = .843$. The isochronous sequence varied significantly by IOI, $F(2,46) = 179.645, p < .001, \eta^2 = .887$, as eye movement latencies slowed linearly as the IOI slowed ($p < .001$). The random sequence had positive evidence for invariance across the IOI levels, $BF_{01} = 4.355$ (F < 1, p > .4), which is unsurprising considering the use of pseudo IOI allocation in the random silent condition making prediction of the timing onset impossible and the competing random visual sequence (to the musical IOI) in the random music condition.

Mean resultant length.

An initial test of circularity, the Rayleigh test was performed on the eye movements in each of the Sequence, Audio and IOI conditions. As observed in Table 4-2, the majority of participants showed significantly non-uniform phasic patterns of eye movements when the visual sequence was isochronous (all participants when the IOI was 323ms). In contrast with E3, the use of a perfectly isochronous sequence (as opposed to a tap-generated one) has notably improved the synchronisation of eye movements. The relatively high number of significant participants in the random 293ms condition is likely a product of the random timing having a mean interval of 293ms (each of the three IOI values were randomised to form the random sequence). This could be evidence that eye movement timing varies as a product of the average visual timing demands (producing a phase relationship at the 293ms interval).

Table 4-2 The Proportion of Participants with Significantly Circular Data by Modality, Task and IOI (significantly non-uniform with the Rayleigh test)

	Isochronous		**Random**	
	Music	Silent	Music	Silent
264 IOI	.83	.83	.21	.21
293 IOI	.88	.92	.38	.42
323 IOI	1	1	.17	.08

A three-way repeated measures ANOVA compared the MRL values by sequence, audio and IOI. There was a significant main effect of sequence, $F(1,23) = 97.679$, $p < .001$, $\eta 2 = .809$, as the isochronous MRL values (M = .329, SD = .135) were significantly higher than the random sequences (M = .081, SD = .028). The main effect of audio was not significant, $F(1,23) = .564$, $p = .460$, $\eta 2 = .024$ ($BF_{01} = 4.60$), music did not increase the synchronisation of eye movements. The main effect of IOI was significant, $F(2,46) = 10.510$, $p < .001$, $\eta 2 = .314$, with the MRL values increasing linearly ($p < .001$) as the IOI slowed.

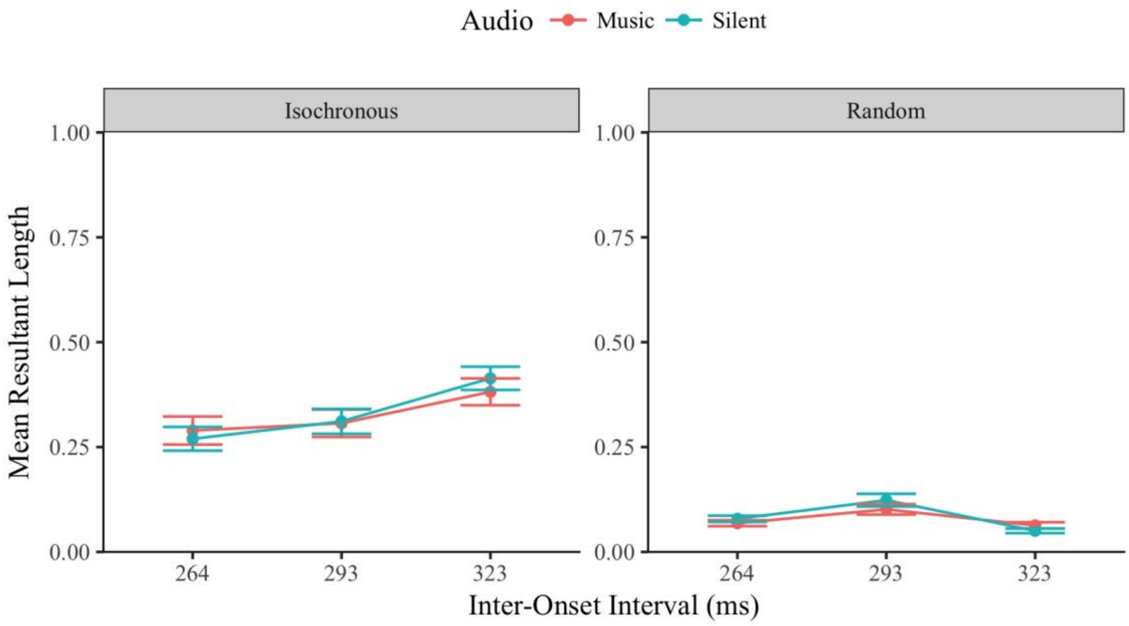


Figure 4-5 Mean Resultant Length (±1 SE) by Sequence, Audio and IOI conditions

The interaction between sequence and audio was not significant, *F*(1,23) = .003, *p* = .953, *η2* < .001. As observed in the difference between the two plots in Figure 4-5, the interaction between sequence and IOI was significant, *F*(2,46) = 32.771, *p* < .001, *η2* = .588, as the isochronous MRL values varied significantly, *F*(2,46) = 22.5, *p* < .001, *η2* = .495, increasing linearly as the IOI level slowed (*p* < .001). The random MRL values also varied by IOI level, *F*(2,46) = 13.980, *p* < .001, *η2* = .378, and Bonferroni corrected pairwise comparisons showed significant differences between the three levels (*p* < .033). The 293ms condition had the highest MRL values followed by the 264ms and the 323ms conditions respectively. The interaction between Audio and IOI was not significant, *F*(2,46) = .701, *p* = .501, *η2* = .030.

The three-way interaction between sequence, audio and IOI was significant, *F*(2,46) = 3.376, *p* = .043, *η2* = .128. For the isochronous sequence, there was no main effect of audio, *F*(1,23) = .166, *p* = .687, *η2* = .007; the addition of music did not increase the synchronisation of eye movements as predicted (BF_{01} = 5.259). As previously detailed, the MRL values did significantly increase as the IOI slowed, there was no significant interaction between audio and IOI ($F < 2$, $p > 1.5$). For the random sequence, there was no significant main effect of audio, $F(1,23) = 1.034$, $p = .320$, $\eta 2 = .043$ ($BF_{01} = 4.275$). The interaction between audio and IOI was also not significant, $F(2,46) = 2.158$, $p > .127$, $\eta 2 = .086$.

Exploratory mean direction analysis.

As the isochronous trials produced significantly non-uniform mean resultant length values, further analysis of the mean phase direction is informative for understanding when in relation to the visual phase the eye movements begin fixations. A common finding in finger-tap studies is the negative mean asynchrony (see section 1.3.3), observed as a clustering of responses prior to the beat onset.

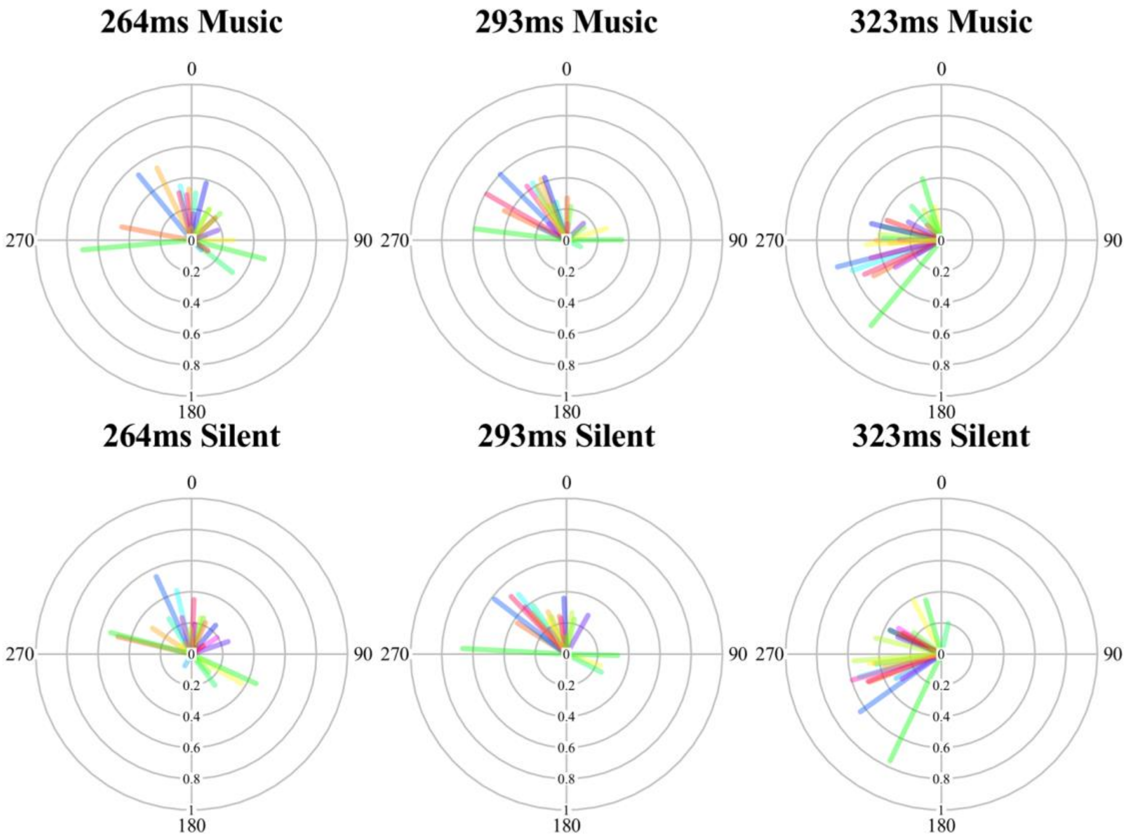


Figure 4-6 Mean Direction (phase location, 0 is the visual or audiovisual onset) and Mean Resultant Length (line length where 1 is the circle radius) by Audio Condition and IOI (colours represent participants)

The Figure 4-6 displays the relationship between synchronised responses (MRL) and where in phase the fixation began (mean direction; values to the right of 0, e.g. 1 to 90 indicate fixations beginning late, values to the left of 0, e.g. 270-359, indicate early fixations). Those participants who had a more-synchronised response (a longer line) appear to start fixations prior to the IOI onset (before 0) in phase, increasingly in the 293ms and 323ms conditions. This could be initial evidence that the timing mechanisms underlying eye movements were predicting the visual onset in time, as opposed to reacting to a peripheral visual target. As there is no known within-subject ANOVA equivalent for circular data, the circular mean (M) was converted to a linear measure of absolute distance from the IOI. This was achieved by converting the (radian) circular mean value (which ranged from +π to -π, where 0 is the musical beat) to a degree value, and then re-representing this degree value as a proportion of the IOI (represented as an equation below).

$$Mean\ Distance = \left(\frac{\text{M} \times 180}{\pi}\right) \times \left(\frac{\text{IOI}}{360}\right)$$

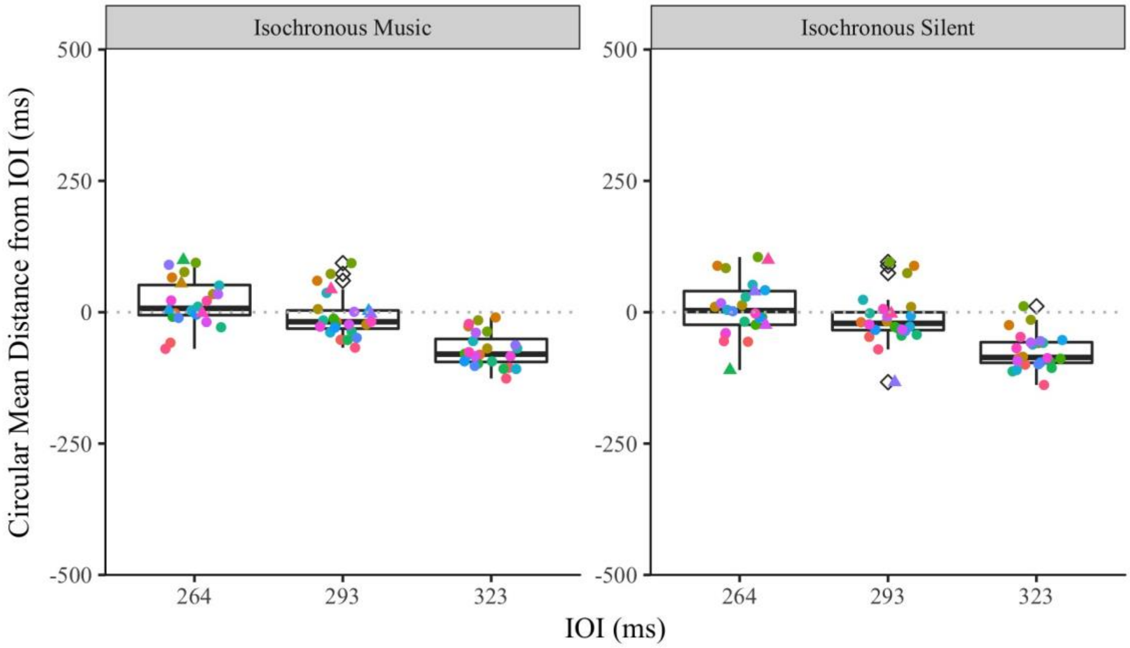


Figure 4-7 A boxplot of Circular Mean Distance (ms) from the IOI, colours represent participants, (triangles represent non-significant Rayleigh test or circular uniformity), the diamonds are potential outliers, kept within the later analysis

An intention of the study was to measure whether the addition of corresponding musical beats to the visual sequence would aid the predictive timing of eye movements. This further analysis of the circular mean distance from the target onset facilitates a comparison between the music and silent Isochronous sequence conditions and the three IOI levels. A two-way repeated measures ANOVA compared the circular mean distance across audio and IOI isochronous conditions. There was no significant main effect of music, $F(1,23) = 2.137$, $p = .157$, $\eta 2 = .085$ ($BF_{01} = 4.165$); the presence of musical beats did not influence where in time fixations started in relation to the IOI phase (see Figure 4-7). There was a significant main effect of IOI, $F(2,46) = 183.194$, $p < .001$, $\eta 2 = .888$, as fixation start times were increasingly prior to the visual onset as the IOI levels slowed. The 264ms IOI (M = 13.722ms, SD = 49.11) had slightly late response (after the letter onset); the 293ms IOI (M = -10.12ms, SD = 44.95) and 323ms IOI (M = -73.09ms, SD = 32.90) had increasingly early responses in phase (Bonferroni corrected pairwise all significant, $p < .001$). The interaction between audio and IOI was not significant, $F(2,46) = .352$, $p = .705$, $\eta 2 = .015$ ($BF_{01} > 1000$). The predictive behaviour in eye movements appears to co-vary with performance as synchronisation performance also increased as the IOI slowed.

To further confirm this pattern within individuals, a simple linear regression was performed to predict the synchronisation performance (MRL) based on the circular mean distance from the IOI in the 323ms Isochronous Music condition (this condition had 100% significantly non-uniform phasic responses).

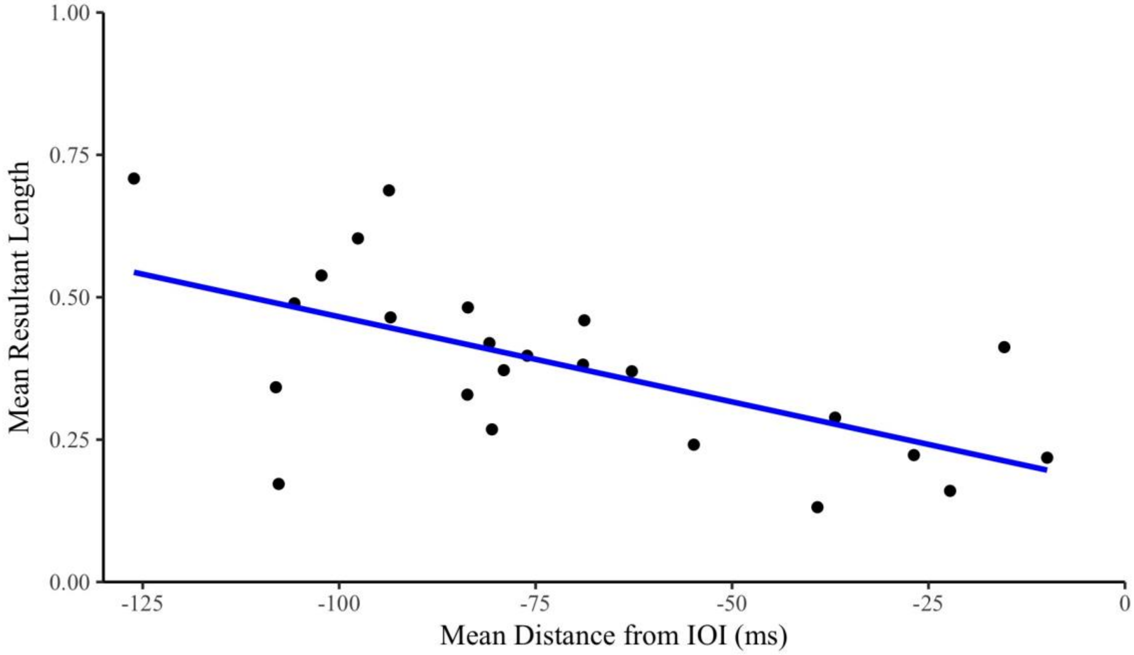


Figure 4-8 A Scatterplot and Regression Line (blue) of the Relationship between the Mean Distance from the Target IOI (ms) and the Mean Resultant Length values in the Isochronous Music 323ms Condition

As observed in Figure 4-8, a significant negative relationship was identified between the mean distance from the target IOI and synchronisation performance, $F(1,22) = 13.39$, $p = .001$, with an R^2 of .350. An increasingly early response from the IOI onset (fixating early in phase) predicted a higher level of synchronisation. This predictive behaviour may reflect some inbuilt accommodation in the timing of eye movements for physiological constraints in ocular motor timing, for example eye-brain lag. Although this account does not accommodate the increase in NMA when the IOI slowed; the delay would be constant if this physiological constraint was the only contributor to the delay.

Reaction time.

The prediction for the reaction time performance was that the isochronous music condition would have faster reaction times than all other conditions, as evidence of audiovisual entrained attention. A three-way repeated measures ANOVA compared the reaction times across sequence, audio and IOI conditions. The main effect of sequence was significant, $F(1,23) = 11.984$, $p = .002$, $\eta^2 = .343$, as the isochronous trials were faster (M = 543.7ms, SD = 72.6) than the random trials (M = 555.9ms, SD = 69.8); that is, predictable visual onsets were responded to faster than random ones. The main effect of audio was not significant, $F(1,23) = 2.474$, $p = 129$, $\eta^2 = .097$. The main effect of IOI was significant, $F(2,46) = 6.148$, $p = .004$, $\eta^2 = .211$, as the IOI slowed the reaction time decreased linearly ($p = .006$).

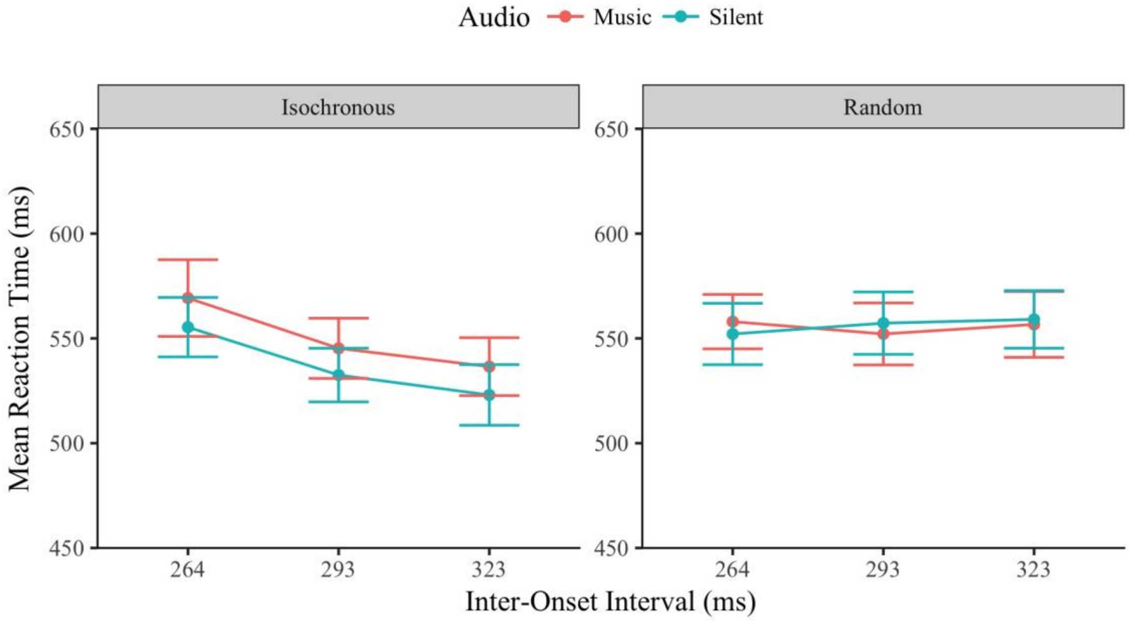


Figure 4-9 Mean Reaction Time (ms ±1SE) by Sequence, Audio and IOI Conditions

There was a marginally significant interaction between sequence and audio $F(1,23) = 3.638$, $p = .069$, $\eta^2 = .137$. Analysis of the effect of audio between the two sequence types confirmed the trend observed between the two plots in Figure 4-9. The isochronous sequence had faster reaction times when silent (M = 537.0ms, SD = 63.7) than with music (M = 550.3ms, SD = 72.1), $t(23) = 2.121$, $p < .045$. This is in the opposite direction predicted, as the synchronised audiovisual presentation was predicted to speed RT performance. The random sequence did not differ between the audio conditions, $t(23) = 0.123$, $p = .903$ ($BF_{01} = 4.626$), which again is contrary to the prediction that competing audiovisual rhythms would impair RT performance.

The interaction between sequence and IOI was significant, $F(2,46) = 7.811$, $p = .001$, $\eta^2 = .254$. The reaction times to isochronous sequences varied significantly by IOI, $F(2,46) = 14.936$, $p < .001$, $\eta^2 = .394$, as the RT's decreased linearly as the IOI slowed ($p < .01$). RTs for the randomly timed visual sequences did not vary across the IOI levels, $BF_{01} = 7.964$ ($F < 1$, $p > .5$), which is unsurprising considering these are pseudo-IOIs and each level was actually made up of a random selection of all three IOI levels. Finally, neither the interaction between the audio and IOI conditions or the three-way interaction between sequence, audio, and IOI were significant ($F < .3$, $p > .7$).

4.5 Discussion

The intention of these two studies was to both replicate the previous experiments with a further refinement to the paradigm, and to measure the influence of complementary and competing musical beats (with exogenously timed visual stimuli) on the timing and synchronisation of eye movements. In the gaze-contingent visual search task (E4a), the reduced response complexity notably quickened the latency of eye movements in silence and with irrelevant music. There was also a reduction in the error difference between the mean latency and the target IOI compared to the previous experiment; the reduction in difficulty may have made it easier to volitionally control the latency of eye movements. This initial 4A study directly replicated the previous findings: there was again evidence of invariance in eye movement timing subject to changes in the tempo of irrelevant musical beats indicating that we again find no evidence of eye movements implicitly entraining to music. The volitional control of eye movements in the explicit task produced mean duration shifts that were in the direction of the target musical IOI but too conservative, resulting in phase durations that did not match the target and random responses in phase. The explicit synchronisation of volitionally controlled eye movements (explicit gaze-contingent conditions in E4a) was both poorly controlled and no better than when not trying. This suggests that the explicit control involved period shifts which somewhat approximated the IOI but there was no evidence of phase correction aligning the eye movements in time (e.g. on the beat). There was some slight indication that at a slower beat interval the volitional control may improve.

The results of this (and the previous) study provide compelling evidence that there is minimal temporal precision in the volitional control of eye movements to allow for small corrections (as either a period correction that quickens or inhibits, or a phase correction that aligns with on the beat) when the eye movements align with a point in phase and produce a clearly synchronised response. This raises more questions around the limitations of volitional intent as a direct control of eye movements than it answers. With the exception of the Findlay and Walker (1999) model, no other eye movement model considers volitional intent influences on timing, beyond having a wide remit for cognitive influences or direct-control. The results in this experiment evidence clear limitations to the volitional speeding of eye movements, but there are notable unknowns regarding any capacity to volitionally cancel or inhibit saccades in order to maintain sequences of prolonged fixation during an active vision task.

The second experiment (4B), which manipulated the timing of visual sequences and the presence of musical beats found that the sequence type significantly influenced the timing of eye movements. When isochronous, the timing varied as a function of the IOI level, yet when random the latencies were constant, equivalent to the isochronous responses to 293ms, irrespective of musical IOI (beats). This suggests that the timing of eye movements is an active process that predicts temporal visual characteristics in the environment and produces latencies that encompass those durations (both quickening, slowing and averaging around an appropriate pace). A notable finding of interest is the small but constant (3ms) slowing in eye movements during the Isochronous Music condition compared to Isochronous Silent. This increase in duration is suggestive of difference in eye movement timing that is only to exogenous audiovisual temporal correspondences. This effect was not present with the addition of music to the random sequence or in the gaze-contingent data with and without music. This effect may be evidence that exogenous audiovisual temporal correspondences require more time during fixation, a delay incurred when binding an audiovisual representation. An alternate account is that concurrent audiovisual presentation increases cognitive load demands, and that the load effect had already delayed eye movements in the random condition. The gaze-contingent onsets (E4a) were not influenced by the presence of audio in the same way, which provides further evidence that there is notable difference in the processing of temporal information between exogenous visual onsets and gaze-contingent visual onsets.

The volitional (explicit) control of when the eyes move is notably imprecise, which could be due to direct control limitations and perceptual smoothing of time in visual perception. As identified in the SOD paradigm, not all eye movements are subject to direct control (Henderson & Smith, 2009). The visual system presents a temporally smooth visual percept, both of visual events in the world (Yarrow et al., 2001) and by reducing awareness of saccades that would interrupt the smoothed perceptual stream of visual information (Clarke et al., 2016). This perceptual smoothing appears to limit the visual feedback on when fixations begin, which during a gaze-contingent task could impair the precision of corrective movements, especially at such fast intervals. The eye movement response to the externally timed visual sequences (E4b) was more predictive and had latencies that were closer aligned to the IOI, notably orienting with a faster mean latency to the 264ms condition than was achieved in the gaze-contingent experiment

(E4a). This suggests a capacity of the visual system to respond and predict exogenous temporal information that was diminished when volitionally controlled. This may be due to the absence of visual transients.

The second part of the study (4B) predicted an enhanced reaction time performance when musical beats corresponded with visual onsets, due to the attentional entrainment effect described by Escoffier et al. (2010). Rather than any additive influence of music, it was the predictability in the visual sequence and the pace of the interval that influenced reaction times. When the visual sequence was predictable, the reaction times were faster, i.e. the visual sequence entrained visual attention. Interestingly, in addition to slowing eye movements, the addition of corresponding musical beats slowed reaction times uniformly across the IOI levels. The correspondence of audiovisual onsets, rather than entraining the allocation of attention in time (Jones & Boltz, 1989), appear to have increased the cognitive demands of visual processing at the visually entrained moments in time.

The synchronisation of eye movements to the isochronous visual sequence was notably more clustered than the tap-contingent eye movements in E3. All participants had a significantly non-uniform response at 323ms and the MRL values, whilst not at the .73 threshold (Konvalinka et al., 2010), were notably higher than all previous experiments. The addition of musical beats to the isochronous sequence did not increase the synchronisation performance (as predicted). Eye movement timing is clearly sensitive to predictable exogenous visual timings (much higher MRL values when isochronous than when directly controlled), but not aided by corresponding auditory information. The exploratory analysis of mean direction and mean distance from the IOI found a clear relationship between predictive responses (early in phase) and better synchronisation, which contrasts with the findings from SMS. The NMA tends to decrease (approach the tone-onset) with better performance (Krause, Pollok, and

Schnitzler (2010). These increasingly early NMA values within the more synchronised eye movements suggests some eye movement specific accommodation.

The findings of this study provide evidence that eye movement timing is not only sensitive to and predictive of exogenous visual onsets, but is systematically slowed by audiovisual correspondences. We also provide evidence that eye movements can synchronise with a regular interval when visual processing demands require it. Finally, we confirm that the volitional control of eye movements is limited and much less precise than eye movements to exogenous visual onsets. None of these effects are currently accounted for within eye movement timing models.

Chapter 5

Temporal Precision of Directly Controlled Eye Movements

5.1 Chapter Overview

This chapter addresses the capacity to volitionally control self-paced eye movements at increasingly slower intervals during a simple visual search paradigm. Note that the operational definition of volitional control within this thesis is the intent to control the timing of every eye movement in sequence. Experiment 5 statistically modelled fixation durations to account for the emergence of bi-modal distributions, similar to those observed in SOD paradigms (Henderson & Pierce, 2008; Morrison, 1984). The volitional control of eye movements formed two distributions, one that had quick and relatively invariant fixations and a second that increased linearly as a function of the referent interval. These fixations were somewhat synchronised to the musical interval. Experiment 6 extended this approach to consider auditory feedback on eye movements, as either a tone after each saccade or a tone following a spatial shift between targets. The study directly replicated the previous results but found that the addition of an auditory feedback tone did not increase the amount of direct control or the synchronisation of eye movements.

5.2: Experiment Five: Inhibiting Eye Movements to Match Musical Beats

Much like Monty Python’s *The Meaning of Life*, there are two perspectives on eye movements: those who believe every fixation is sacred, and those who don’t. The former (direct control) theories argue that the moment to moment demands of visual processing are reflected in the duration of fixations, for example, the EZ-Reader model by Reichle et al. (2012). Theories for the later (mixed control) argue that only a subpopulation of fixations represent the moment-to-moment demands of the currently fixated stimuli, for example, the CRISP model (Nuthmann et al., 2010).

The weight of evidence in favour of mixed control is strong. As previously reviewed in Chapter 1, Hooge and Erkelens (1996, 1998) identified a sub-proportion of eye movements oriented incorrectly around a Landolt-C array, seemingly insensitive to the currently fixated directional information. Onset-delay paradigms employed in both reading (Morrison, 1984; Yang & McConkie, 2001) and in static scene studies (Henderson & Pierce, 2008; Henderson & Smith, 2009), describe the emergence of two populations within fixation duration distributions in response to the increasing delays of visual information presentation. One population varies as a product of the onset delay (with a mode that approaches the delay duration), which Henderson and Smith (2009) describe as an immediate effect of the currently fixated information (not a global change altering the following fixations). The second population is not sensitive to visual demands and is an optimally fast distribution, considered the product of a stochastic saccade timer (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). These studies agree that direct control is limited to some fixations, yet with the exception of the Findlay and Walker (1999) model, which describes volitional control as uncommon, there are currently no quantifications of the capacity for voluntary control as a direct control influence on eye movement timing.

The voluntary control of eye movements differs from other top-down factors. As previously discussed within chapter 1, direct and mixed control theories describe the influence of visual and task factors on the duration of fixations, commonly reporting how the visual or task demands increased the inhibition of eye movements, represented in average fixation duration increases (Nuthmann, 2016). Volitional control of eye movements, for example the anti-saccade task, demand a different response that is not the product of the visual demands or global task properties, i.e. to quickly orient away from a visually salient target (Munoz & Everling, 2004). The capacity to volitionally orient the eyes irrespective of visual demands may not be as unusual as Findlay and Walker describe. There is a growing body of evidence that eye movements are important active social signals, providing others with timely cues of emotion and guiding their attention (Langton et al., 2000; Puce et al., 1998). Furthermore, eye movement decisions during real-world tasks orient towards task relevant rather than low-level visual factors, directly representing volitional influences on *when* and *where* the eyes move (Henderson, 2017; Tatler, Hayhoe, Land, & Ballard, 2011). Rather than unusual or uncommon, the volitional decision to orient the eye is likely an essential and currently underestimated capacity that modulates the timing profile of eye movements.

The temporal precision of voluntary saccadic movements has been investigated within saccadic SMS research. As described in chapter 1 (*1.3.6 Saccadic SMS*), saccades to horizontal visual onsets that exceeded 500ms were predictive of (arriving prior to) the visual onset (Shelhamer & Joiner, 2003). Joiner, Lee, Lasker, and Shelhamer (2007a) found a similar 500ms threshold in synchronised saccades to static horizontal visual targets using an auditory metronome. This study argued that the temporal precision of saccades is similar to that of finger-taps, with some rate limitations. These studies have two notable limitations: first, they measure only the eye movements that orient between the visual targets, omitting all other saccades that re-fixate on a target. This is a missed opportunity to measure the proportion of eye movements that can be directly controlled as a factor underlying synchronised saccades. Secondly, they involve no visual task beyond a voluntary orientating decision. It is plausible that visual demands modulate the capacity for volitional control. This study will account for these factors.

The intention of this study is to quantify the direct volitional control of eye movements when synchronising sequential eye movements to auditory IOI. The predictions of this experiment are: first, as the auditory IOI referent slows, the requirement for saccade inhibition and cancelation will increase. As only some eye movements are subject to direct control (Henderson & Pierce, 2008; Henderson & Smith, 2009), we predict a second mode will emerge and vary as a response to the referent IOI. Secondly, the proportion of eye movements subject to delay will increase as the IOI slows, both as a product of the shifted distribution (the slower distribution will be more distinct when further from the first) and due to the temporal demands of the referent IOI (holding fixation for 800ms requires more inhibition than 500ms so may involve two fixations that are delayed rather than one for the entire period). The third prediction is that synchronisation of eye movements will increase as the IOI slows. This is because the volitional timing will be subject to less interference as the fixation is inhibited beyond the first mode (the range of the timer). Eye movements to slower IOI levels are predicted to be increasingly predictive with a circular mean that is prior to the IOI in phase as described in Joiner and Shelhamer (2006) and Joiner, Lee, Lasker, and Shelhamer (2007b).

5.2.1 Methods.

Participants.

A power analysis was calculated using G*Power3 (Faul et al., 2007) based on the main effect of latency ($\eta^2 = .257$) in the explicit synchrony of gaze-contingent movement in experiment 4A. At a 95% chance of finding an effect at $d = .8$ the study required a sample size of 17. A total of 20 participants completed the study, two were replaced as they were unable to produce an in-time tapping response (to all IOI beats). This left 18 participants (10 Female), with an age range from 21 to 48 (M = 27.39, SD = 5.9), who completed the study for payment. The musical training scores (from Gold-MSI) indicated a range of training from no training (7) to 40, more training than 75% of the population (M = 21, SD = 11.92); none of the participants were professional musicians.

Design modifications.

The equipment, including screen, eye-tracking setup with chinrest, headphones, game-controller and Apple Trackpad remained the same as that employed in E4. The musical stimuli were formed of the same kick-drum and snare drum midi-sounds used in the previous experiments. Eight new musical tracks were generated to represent IOI levels at 100ms increments starting at 300ms (from 300 to 1000ms). The visual stimuli were altered from previous experiments; the two notable changes were the number of circles in the ellipse (12 rather than 8) and the gaze-contingent visual onset that was the colour of the circle (and not a T or L letter).

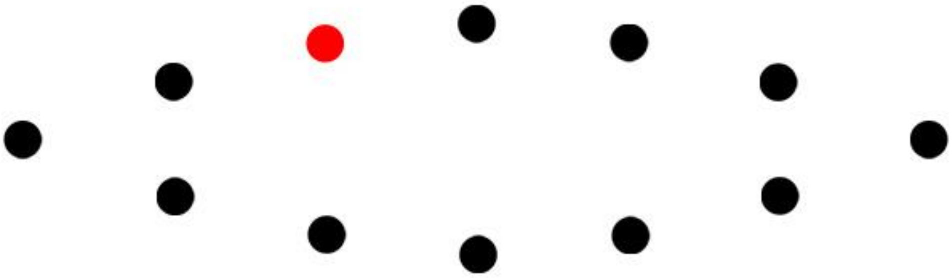


Figure 5-1 An example of the gaze-contingent twelve-circle elliptical visual stimuli. One circle was in colour, either blue or red, at any point

As seen in Figure 5-1, the colour referent (either blue or red) encompassed the entire circle (1.2º); this change from previous experiments (E1:E4) was to further simplify the visual task demands. The requirement for precise saccade targeting when using small (0.3º) centralised letters may have inadvertently increased the prevalence of short corrective (express) saccades within the distribution. To avoid these corrective movements, which could interrupt the direct control of each eye movement, the entire circle was in colour reducing the requirement for precision. The distance between each circle target was also reduced (4.7º between circles as opposed to 7.8º in E3 and E4).

This change altered the number of circles to 12 circles within the elliptical shape rather than 8. To further reduce the visual demands within the task, the new visual task had a reduced number of visual changes (colour changes) within a trial sequence, from 28-31 in E3 and E4 to between 36-45 circles. Each trial ended after 180 sequential circles, therefore each trial required the identification of between 4 and 5 colour changes.

Procedure.

The participants were recruited to a 'Music and Eye Movement Study', the sign-up information detailed that the study was measuring on-beat movements, both finger taps and eye movements. Participants with rhythmic ability were encouraged to participate. The eye-tracking calibration procedure remained the same as E4. The first trial was a familiarisation practice of gaze-contingent navigation without music. The participants were instructed to look from circle to circle clockwise, clicking a game-controller trigger when the colour sequence changed. Each trial ended after a sequence of 180 circles, or timed out after 3-minutes. The second trial was in silence with the same instructions. The third trial was a further practice trial that introduced an additional instruction to synchronise when the eyes moved with the music (set to 500ms). During this practice, participants were instructed to hold fixation on each circle and to control each movement to match the musical beat. The participants then completed 18 trials, and each IOI level was randomised and repeated twice, with a 19th trial which acted as a final silent trial (breaks were provided when needed). When completing each trial, the instructions were to match the circle transitions with the musical beat and to click the trigger when the colour sequence changed. Following completion of these eye movement trials, the trials tasked tapping in time on the Apple trackpad to each IOI (each trial recorded 60 taps, there was no visual task) and to fill in a short musical training questionnaire.

Data analysis plan.

To investigate the influence of different IOI targets on the dynamics of eye movement timing, both the automatic and directly controlled elements of the movement distribution must be defined. As identified in the scene-onset-delay (SOD) paradigm (Henderson & Pierce, 2008; Henderson & Smith, 2009) and statistically modelled within recent eye movement timing accounts (e.g. CRISP, Nuthmann et al., 2010) there appears to be two timing mechanisms that act on the distribution of fixation durations. The first is an early and relatively invariant distribution that is indicative of the automatic execution of saccades (visually invariant and optimally fast). The second distribution is the product of some delay (saccade cancelation or inhibition), and are the eye movements subject to direct control (cognitive or explicit inhibition of saccades to prolong the current fixation).

The statistical modelling of the multi-modal distribution generated during this task presents a unique opportunity to describe both the capacity and the precision of direct control through the analysis of the relative size of each distribution (*lambda*) as well as the mean and deviation of each of the two distributions (*mu* and *sigma*), see example fit in Figure 5-3. The preliminary visualisation of the data, for example Figure 5-2, shows the emergence of bimodality in the fixation durations is relatively consistent across participants. The modelling of multimodal parameters is achieved using the 'mixtools' package built for R statistical software (Benaglia, Chauveau, Hunter, & Young, 2009). This method of normal mixture modelling identifies the number of clusters (*k*) across multiple iterations to achieve the least worst (or best) fit for the data. Parameters for the number of clusters, potential starting values for the distributions, and iteration limitations can all be set independently. As empirical estimations of these parameters were unknown for this paradigm, these values were set with a conservative starting point. Specifically, the maximum iterations were set to 400, and no starting values for number of clusters (distributions), lambda, mu or sigma estimates were set.

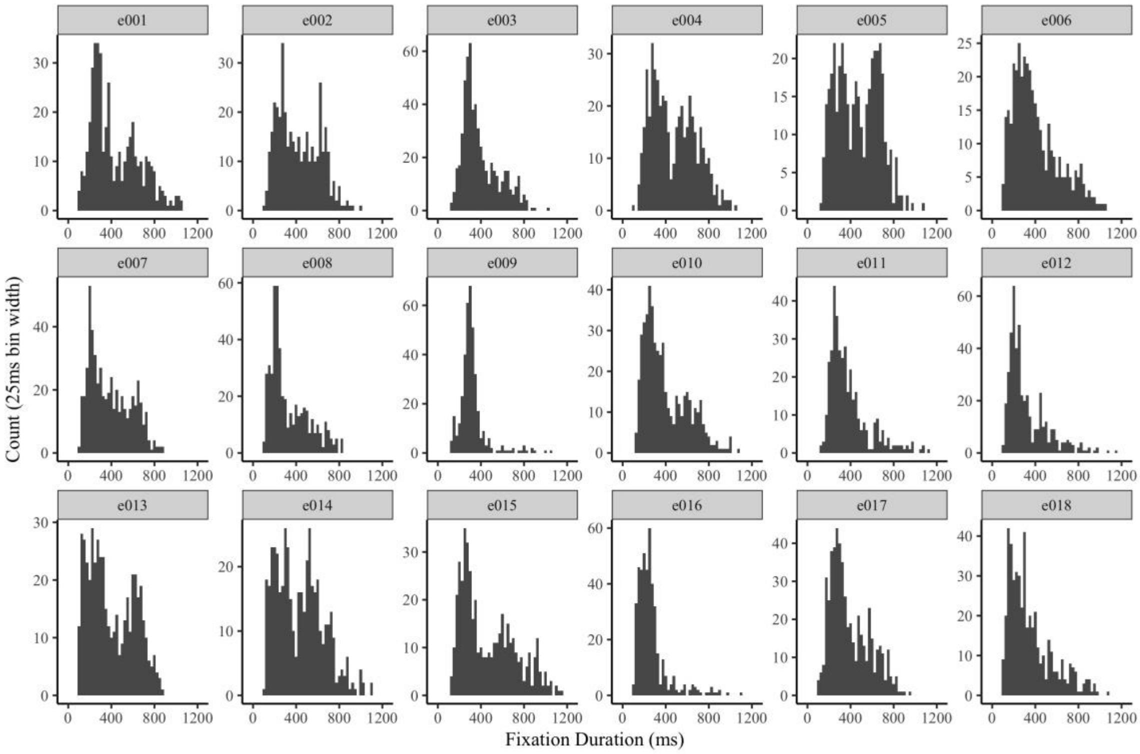


Figure 5-2 A histogram for each participant's fixation durations (ms) in the 700ms IOI condition (bin width set to 25ms)

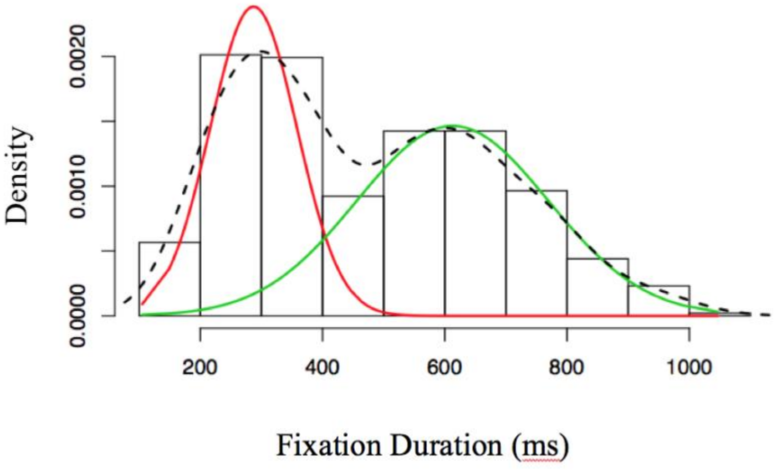


Figure 5-3 An example model fit to a single participant's data from the 700ms IOI. The red line is Distribution 1 and the green line is Distribution 2. The dotted line is the density curve fit of the entire multimodal distribution

The 'mixtools' statistical modelling converged on a bi-modal fit for all participants at each musical IOI level. Visual comparisons were made between these bi-modal fits and a simple unimodal fit. In the 300 and 400ms IOI levels the second mode was diminished yet present within the data, as observed in the 300ms IOI displayed in the Figure 5-4. At the slower levels (500-1000ms) the second distribution is clearly identifiable and distinct from the first. Importantly, even at the 300ms IOI, the bimodal parameters visually fit the data better than a normal distribution.

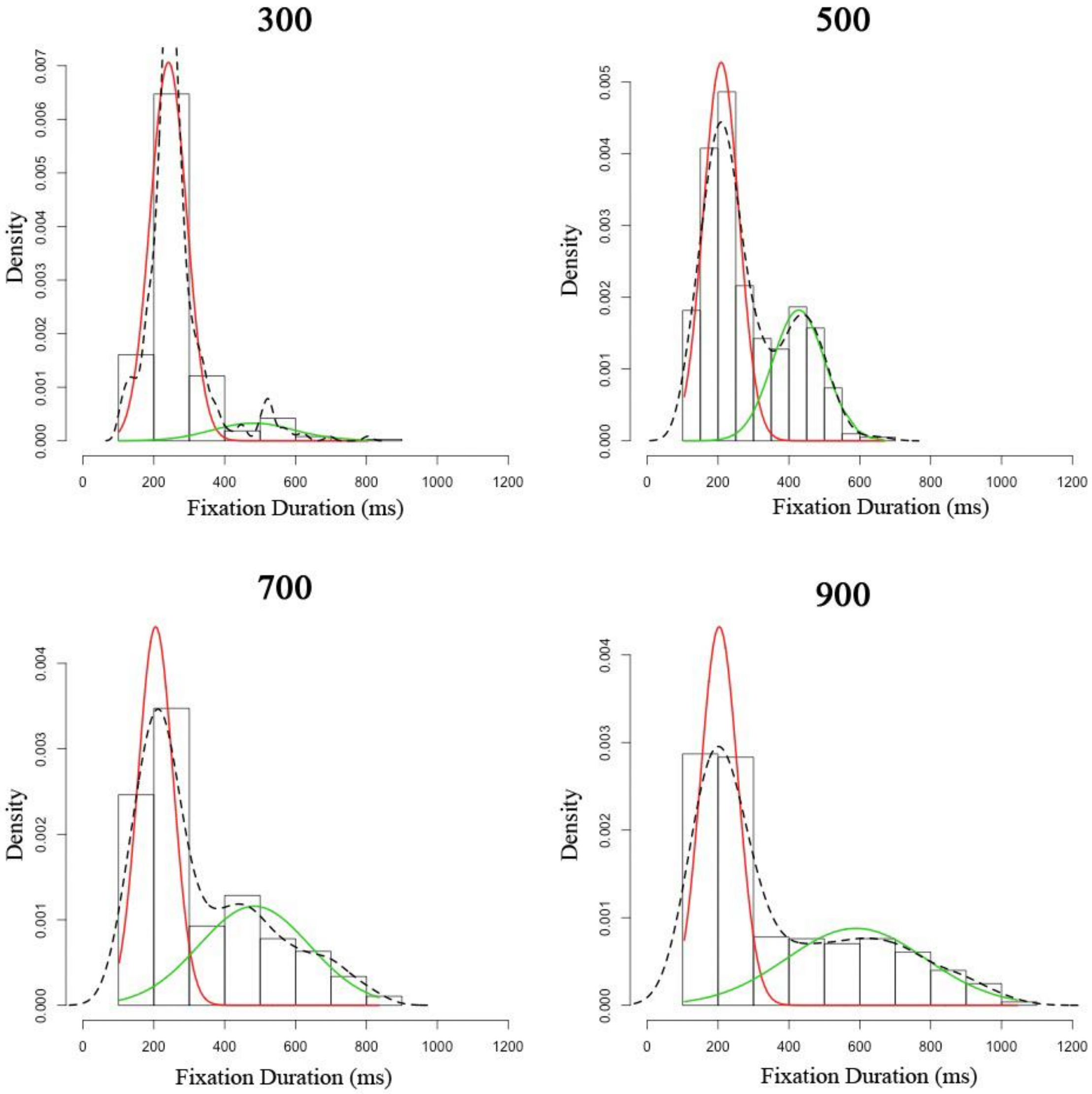


Figure 5-4 Bimodal Histograms of a Single Participant's Fixation Durations at 300, 500, 700 and 900ms. The red line is the first distribution estimation, the green line is the second distribution, and the dotted line is a density curve fit to both distributions

The *lambda* value of all distributions (*k* clusters) sums to 1 and is a proportional measure for each distribution. The mean (*mu*) and deviation measure (*sigma*) for each distribution provide an opportunity to observe the influence of volitional control on the timing characteristics of both the automatic (*mu 1*) and directly controlled eye movements (*mu 2*). The visualised examples in the two figures (5-3 and 5-4) conform well to the two-distribution fits. A point of contrast with the previous experiments in this thesis is the use of fixation duration as the measure of interest rather than latency (a combined measure of fixation duration and saccade duration). Measuring fixation duration for these studies allows for easier comparison with previous models of fixation behaviour (Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Additionally, considering the previously identified limitations in the direct control of eye movements (E3 and E4), there is no clear benefit to adding a further source of variance (i.e. saccade duration), as the duration between points on the ellipse are expected to vary from the referent interval. Prior to modelling the fixation duration data, the same data cleaning steps were taken to remove fixations that were surrounded by blinks, less than 100ms and longer than 1500ms, as in previous studies. This left an average of 430 fixations at each IOI level per participant. Some fixations that were removed in prior experiments were not in this study, specifically when a participant re-fixated within the currently coloured circle. These fixations are likely representative of fixations that are not subject to direct control and inform the predictions. Fixations that indicate off-task behaviour (those outside the predefined areas of interest) were removed; see chapter 2 for an example of the area of interest region size.

The modelled parameters will be employed alongside the circular measures, the circular mean and the mean resultant length (MRL), which will be tested for uniformity with the Rayleigh test (Wilkie, 1983). These circular measures are both formed from fixations start times following a shift from circle to circle (as in previous experiments). This differs from the fixation durations modelled within the mixture modelling, which included re-fixations at each circle. The task was to synchronise eye movements between the circles to match the musical IOI, therefore these intentional movements were isolated as a measure of synchronisation ability (as in the previous experiments). By omitting the involuntary saccadic movements, the circular measures will best represent the controlled shifts of gaze between the circle locations (keeping these fixations within the mixture modelling facilitates analysis of the relationship between temporal precision and direct control).

An intention of this study is to measure whether gaze is predictive of the external rhythm. As the MRL values of gaze-contingent shifts are notably less tightly clustered representations of the circular mean phase than other movements (e.g. finger-taps, observed in E3:E4), the circular mean of significantly non-uniform (Rayleigh's test) eye movements will be converted to a linear measure of absolute distance (ms) from the target IOI. This re-represents the circular mean as a linear measure and reduces the noise that would be introduced with standard measures of negative mean asynchrony (which measures the distance between all circular vectors and the external rhythm). A core assumption underlying this analysis decision is that those fixations that are most clustered in phase (the circular mean) are most representative of direct control. This analysis within the study is intentionally omitting reporting the mean fixation duration of eye movements to the musical IOI levels, as we know that the distributions are bimodal (violating the assumptions of means testing). Therefore, the dependent variables for analysis are the lambda, mu, and sigma values for each distribution as well as the MRL and absolute distance from target IOI (ms) synchronisation measures.

5.2.2 Results

Distribution parameters: Lambda.

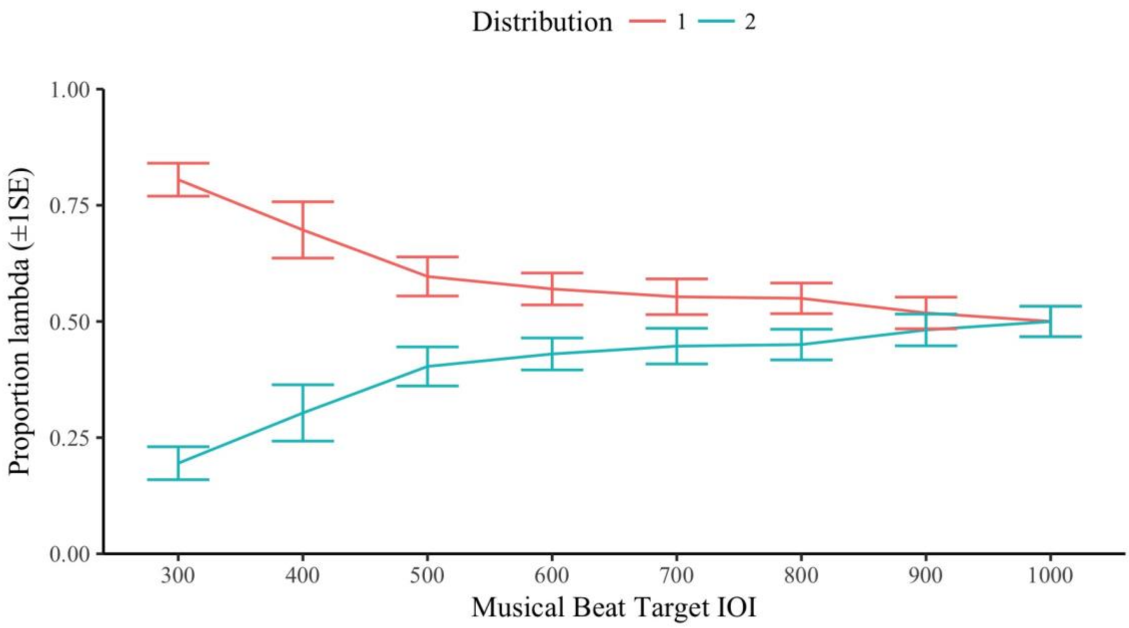


Figure 5-5 The *lambda* of each distribution (both sum to 1), ±1SE. A higher proportion in the second distribution indicates more fixations are subject to inhibition

A repeated measures ANOVA of the proportion of the second distribution, *lambda* (there is no reason to analyse both distributions as they sum to 1), by musical IOI (found a significant main effect of IOI, $F(7,119) = 11.510$, $p < .001$, $\eta^2 = .404$. As observed in Figure 5-5, there was a significant linear trend ($p < .001$); the proportion of fixations within the second distribution increased as the musical IOI slowed. The quadratic effect was also significant ($p < .001$) as there is a notable asymptote at 500ms, where the linear proportion increase slows.

Distribution parameters: Mu.

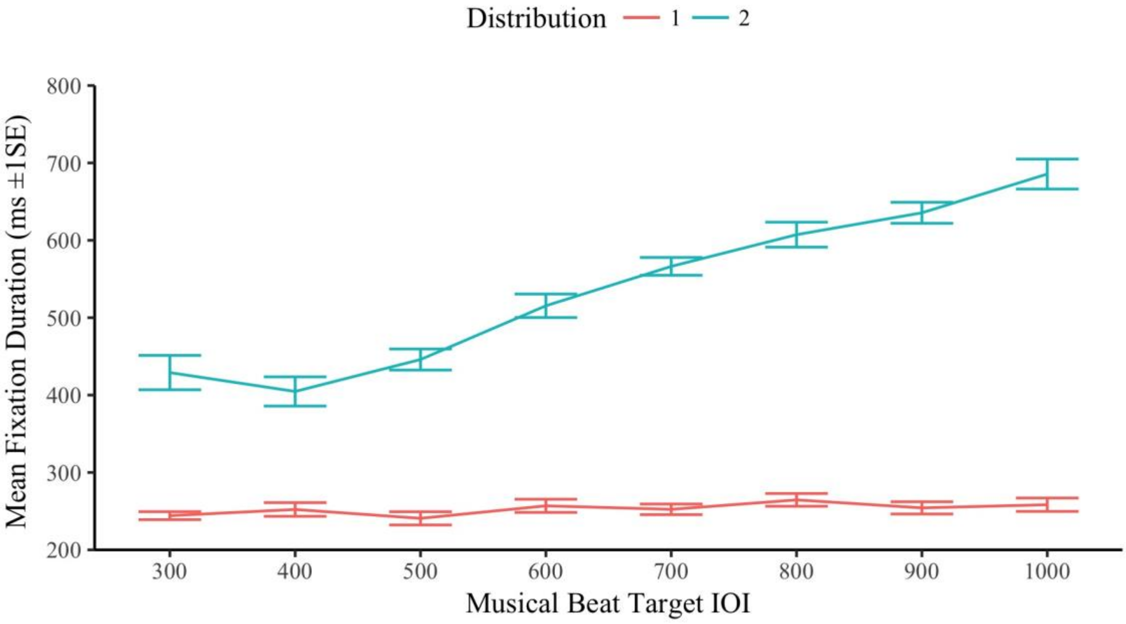


Figure 5-6 The mean (*mu*) of each fixation duration distribution (ms) ±1 SE by Musical IOI

A repeated measures ANOVA of the mean fixation durations of the two distributions and musical IOI levels found a significant main effect of distribution, $F(1,17) = 1665.79$, $p < .001$, $\eta^2 = .990$. The first distribution (M = 252.95ms, SD = 33.59) had significantly shorter fixation durations than the second distribution (M = 536.23ms, SD = 119.38). The main effect of musical IOI was significant, $F(7,119) = 41.91$, $p < .001$, $\eta^2 = .711$. The interaction between distribution and IOI was also significant, $F(7,119) = 47.91$, $p < .001$, $\eta^2 = .738$, as the fixation durations in the second distribution lengthened significantly as the IOI slowed, $F(7,119) = 48.50$, $p < .001$, $\eta^2 = .740$. The fixation durations in the first distribution also significantly lengthened as the IOI slowed, $F(7,119) = 2.68$, $p = .045$, $\eta^2 = .136$. As observed in Figure 5-6, the linear slowing of fixation durations in the first distribution ($p = .019$) was notably more conservative than the slope in second distribution ($p < .001$).

Distribution parameters: Sigma.

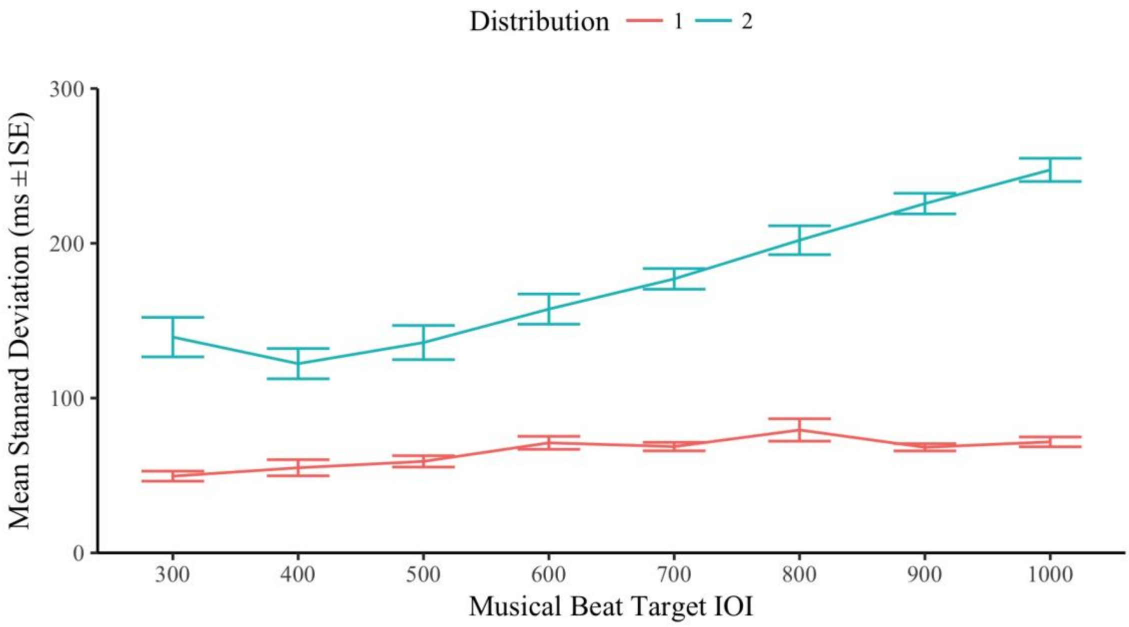


Figure 5-7 The deviation (*sigma*) of each distribution (ms) ±1 SE by Musical IOI

A repeated measures ANOVA of the deviation measures (*sigma*) by distribution and musical IOI showed a significant main effect of distribution, $F(1,17) = 225.63$, $p < .001$, $\eta^2 = .930$. The first distribution (M = 65.34ms, SD = 19.90) had significantly less deviation than the second distribution (M = 175.89ms, SD = 57.7). The main effect of IOI was significant, $F(7,119) = 47.68$, $p < .001$, $\eta^2 = .737$. As observed in slope differences of Figure 5-7, the interaction between distribution and IOI was also significant, $F(7,119) = 25.51$, $p < .001$, $\eta^2 = .597$. The deviation of fixation durations significantly increased as the IOI slowed in both the first distribution, $F(7,119) = 7.44$, $p < .001$, $\eta^2 = .304$, and the second distribution, $F(7,119) = 44.52$, $p = .001$, $\eta^2 = .724$. The interaction is due to the slope differences between the two linear effects (both $p < .001$); the second distribution had a much larger increase in deviation as the IOI slowed than the first distribution.

Mean resultant length.

Prior to reporting the MRL values, the circular uniformity of the data was tested with the Rayleigh's test (Wilkie, 1983). As previously mentioned it tests the significance of the representative nature of the MRL and circular means based on the spread of the distribution. The proportion of participants who had significantly non-uniform responses per musical IOI condition are detailed in Table 5-1.

Table 5-1 Proportion of Participant's Rayleigh Tests outcomes that had significantly non-uniform responses by Musical IOI level (higher values indicate more participants had a phasic response to the IOI)

Musical IOI	**300**	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
Proportion	.39	.67	.83	.83	.83	.78	.83	.94

One participant did not significantly synchronise their eye movements to any of IOI levels (even though the tapping check showed that they could accurately perceive the beat and produce it manually), and another achieved significance at only one IOI (1/8). These two participants were removed from the following circular analysis (remaining N = 16); only 5 participants achieved significance to all IOI, mostly because of apparent difficulties in synchronising at 300 and 400ms. These participants were kept within the prior distribution analysis as they complied with the task and were able to tap in time (the prerequisites for inclusion).

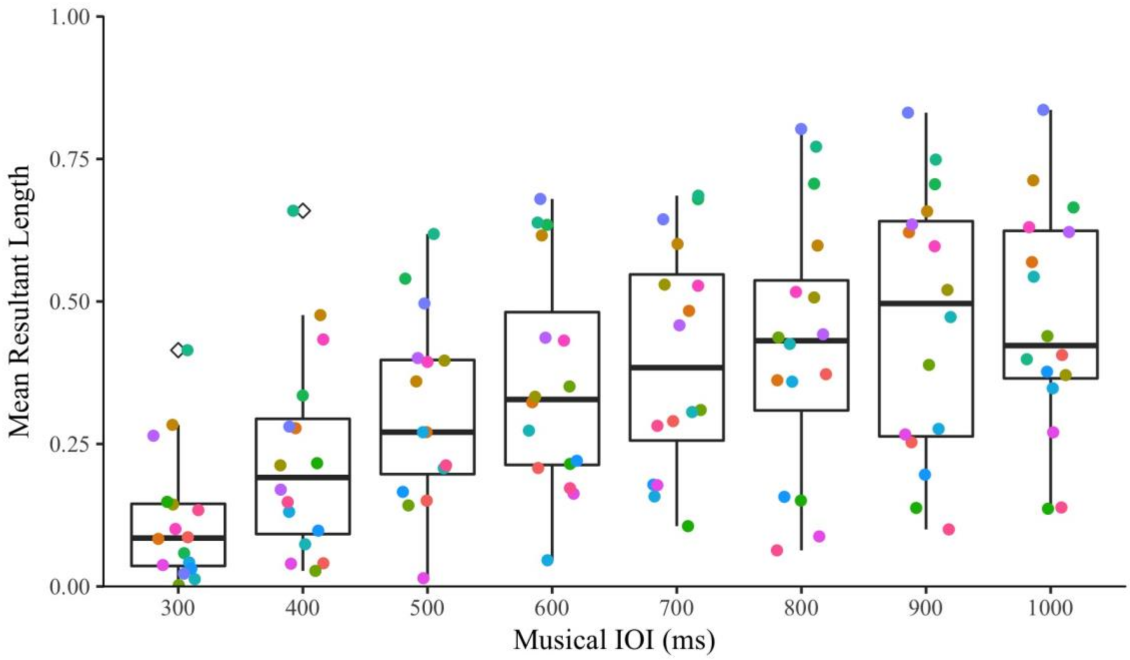


Figure 5-8 A box-plot of Mean Resultant Length values by Musical IOI, colours represent participants, (diamonds represent potential outliers with better performance at 300 and 400ms)

As observed in the boxplot Figure 5-8, the MRL values in the 300ms condition were very low. Additionally, the MRL values for many participants irrespective of IOI very rarely exceeded .73, a previous threshold for synchronised tapping responses in Konvalinka et al. (2010). These MRL values were notably higher than those achieved in the prior explicit synchrony conditions to fast paced IOI within E3 and E4. Two responses in the 300 and 400ms conditions were notably good (marked as outliers). They were from a single participant, and were kept within the following analysis as the responses across all conditions were all consistently highly synchronised.

A repeated measures ANOVA of the MRL values by musical IOI showed a significant main effect of IOI, $F(7,105) = 7.062$, $p = .001$, $\eta^2 = .320$. The MRL values increased linearly ($p < .001$) as the musical IOI slowed (the quadratic effect was not significant, $p = .26$).

Circular mean.

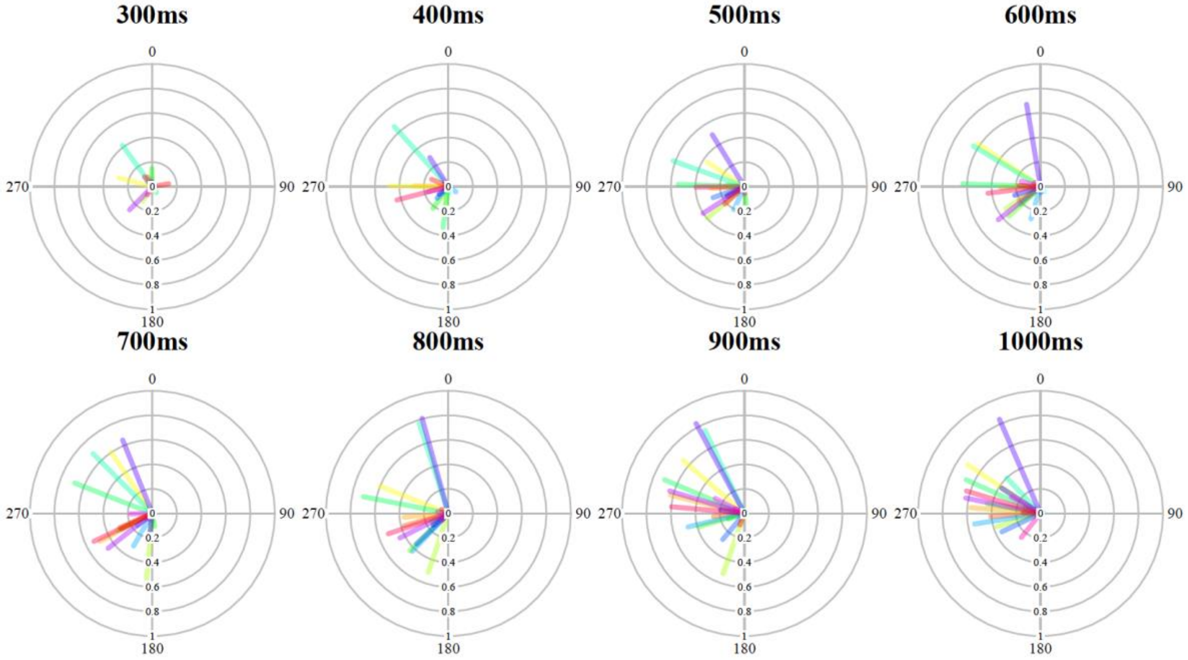


Figure 5-9 The circular mean (vector orientation) and mean resultant length (vector length, the radius is MRL =1) by musical IOI level. Colours represent participants

As observed in Figure 5-9, early responses (after 180 and prior to 0), indicate a circular mean that is before the musical beat. Responses between 0 and 180º are after the musical beat. Note that the longer lines (more synchronised responses) tend to cluster prior to 0, indicating a predictive rather than reactive response. As there is no known within-subject ANOVA equivalent for circular data, the circular mean (M) was converted to a linear measure of absolute distance from the IOI, as detailed in chapter 4.

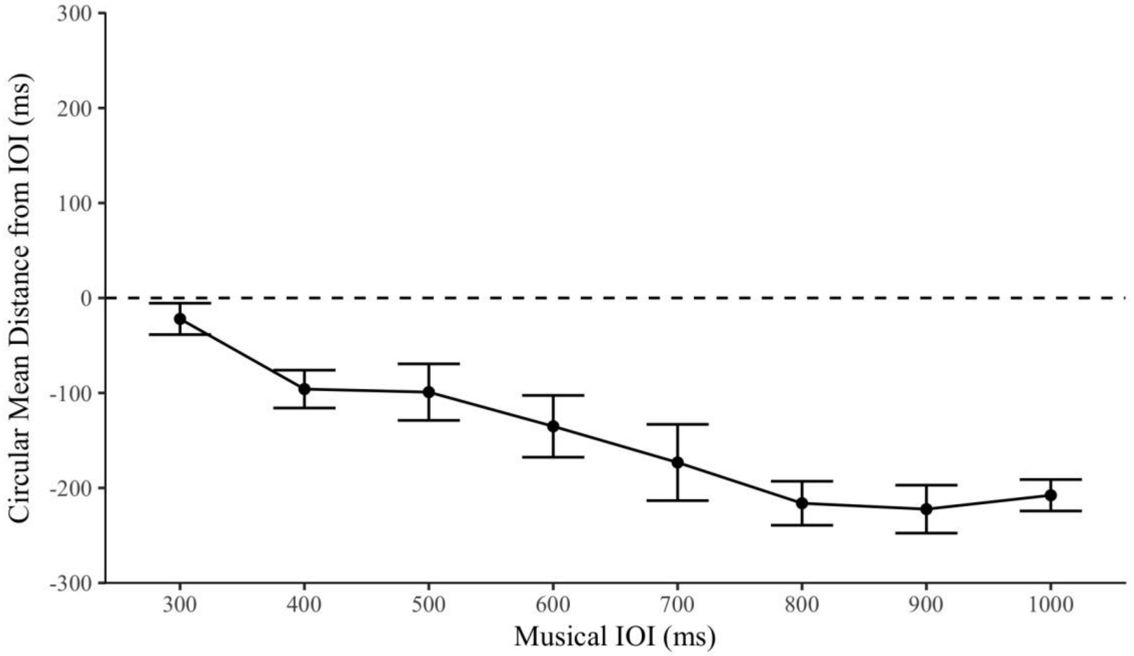


Figure 5-10 Mean Distance (ms) ± 1 SE of the Circular Mean from the Musical IOI

A repeated measures ANOVA of the mean distance across the musical IOI levels showed a significant main effect of IOI, $F(7,105) = 9.131$, $p = .001$, $\eta^2 = .378$, visualised in Figure 5-10. The early (predictive) response of the circular mean increased linearly as the IOI slowed ($p < .001$). The quadratic effect was not significant ($p = .069$), although considering the asymptote at 800ms (the mean distance plateaus at around -210ms), it would be expected to emerge if the IOI level continued past 1000ms. Visualisation of the circular measures across the conditions show that with the exception of the 300ms condition (which had minimal phase relation), the following IOI levels have a relatively constant early response in phase between 70 and 100º prior to the onset.

Exploratory analysis: Correlations between direct control and synchronisation. A linear regression analysis was calculated to predict MRL based on the proportion of lambda in distribution two (the number of fixations subject to direct control). As observed in Figure 5-11, a significant association was found between Lambda and MRL,

$F(1,14) = 5.608$, $p = .033$, with $R^2 = .235$. The participants predicted MRL was -.17 + 1.18*lambda. In other words, the participants' MRL increased by 1.18 when the lambda increased by 1.

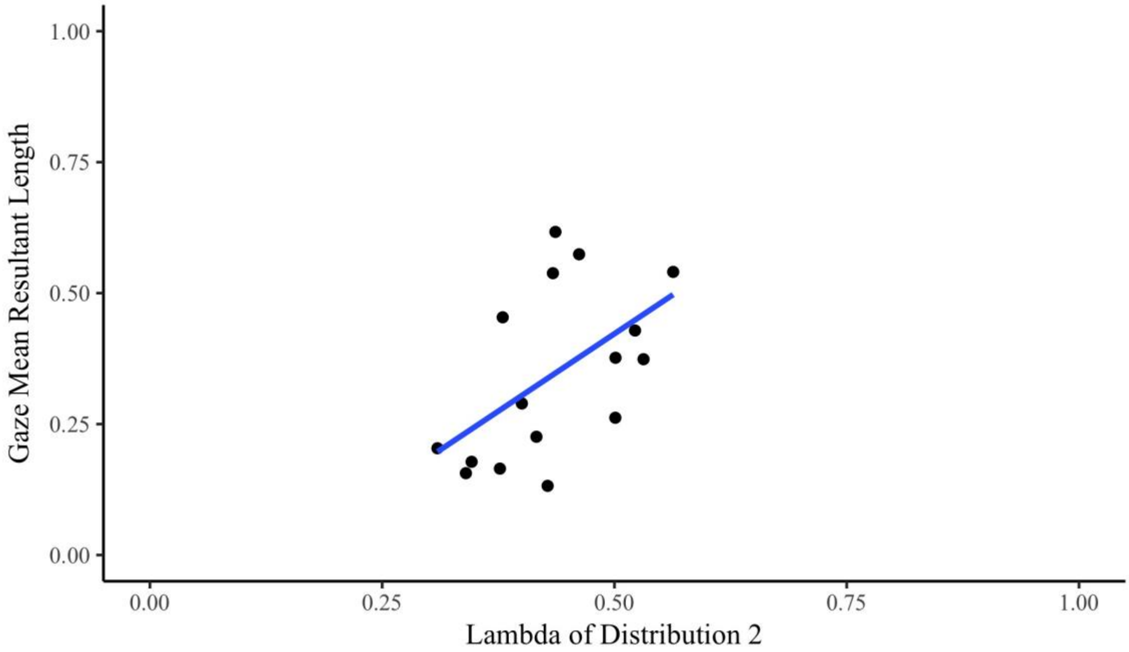


Figure 5-11 A scatterplot with linear regression line of the relationship between the proportion of lambda in distribution two and the synchronisation measure mean resultant length of eye movements

A further linear regression was calculated to predict MRL of eye movements with the MRL of the tap performance at the end of the study. A near but not significant association was found, $F(1,14) = 2.26$, $p = .155$, $R^2 = .078$. As observed in Figure 5-12, there was a slight (non-significant) positive relationship between the two MRL values; participants gaze MRL increased by .58 when their tap MRL increased by 1.

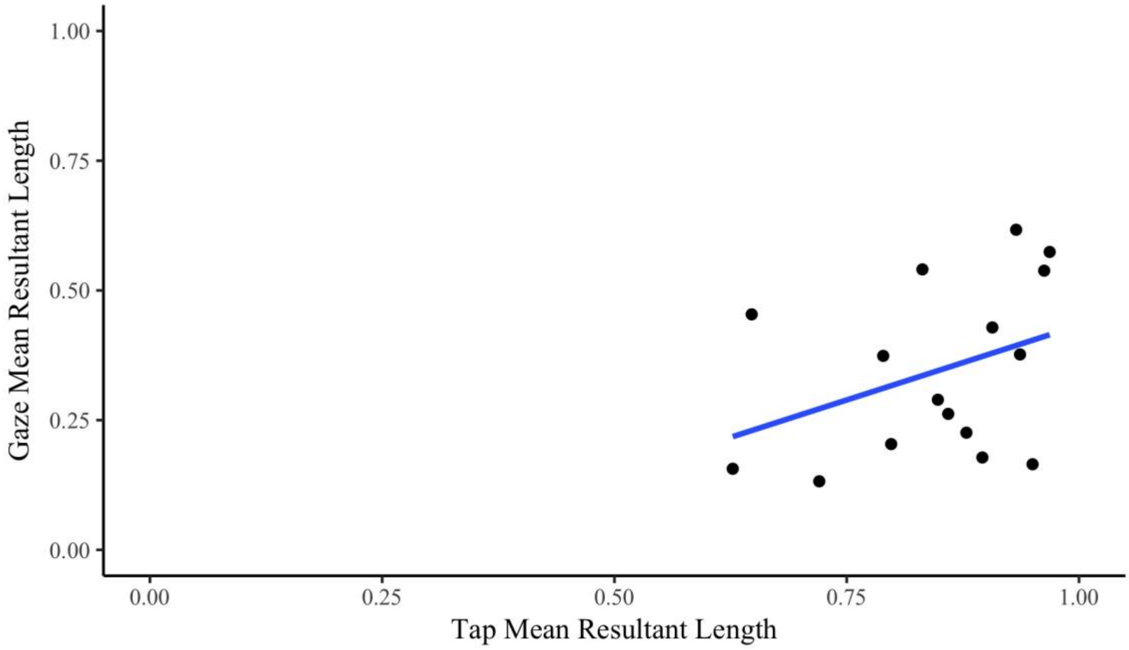


Figure 5-12 A scatterplot with linear regression line of the relationship between the Mean resultant length of finger-tap responses and the mean resultant length of eye movements

A final analysis considered both the lambda of distribution two and the finger-tap MRL values as predictors of gaze MRL performance. The addition of the finger-tap MRL as a predictor did not significantly increase the variance accounted for by lambda alone, $X^2(1) = .020$, $p = .322$ (NS).

5.2.3 Discussion.

The intention of this study was to quantify the volitional control of eye movements when synchronising sequential eye movements to an auditory IOI. The predicted bimodal pattern of results was clearly evident in the fixation duration distributions. Much like the SOD paradigms, the second distribution varied as a product of the referent IOI and the first was relatively invariant (Henderson & Pierce, 2008; Morrison, 1984). The proportion of fixation durations in the second distribution increased linearly as the IOI slowed, accounting for 49.99% off all fixations at the slowest IOI. The mean of the second distribution did vary linearly across the IOI, but the mean values were not one-to-one matches to the referent IOI (the period correction was imprecise), especially at the slowest IOI levels which had average durations that were too quick for the IOI. In contrast to the second distribution, the first varied much more conservatively across the IOI levels. The deviation of the first distribution was also much smaller than the second. These observed *mu* and *sigma* parameters from the first distribution are predicted to be the timing distribution of the visually invariant stochastic saccade timer (with a mean across all IOI levels of 250ms and a deviation of 65ms) proposed in the SWIFT, CRISP and ICAT models (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014). Whilst the mean fixation duration of the first distribution did increase as the IOI slowed, the deviation of this distribution also increased, which may suggest that the IOI difference is the product of multimodal fit limitations (it may be more invariant than the modelled data show). Although, an alternate explanation from the ICAT model is that a global component modulates both the saccade timer's mean and deviation as a product of task demands (Trukenbrod & Engbert, 2014). The lambda of the second distribution increased as the IOI increased, which is evidence for the prediction that temporal demands on fixations will increase with slower IOI.

The synchronisation of eye movements to these much slower IOI (compared to E3 and E4) show a notably improved MRL as the IOI level slowed, much like the 500ms threshold described in the saccadic SMS studies (Hornof & Vessey, 2011; Joiner, Lee, Lasker, & Shelhamer, 2007a). Few participants achieved MRL scores that approached tapping values (over .73), and anecdotal feedback during testing was that the task was very difficult. At all levels of IOI the circular means tended toward a predictive response, although the distance from the IOI became increasingly early as a timed response in relation to the IOI, as it slowed from 300 to 800ms, peaking a response that was 210ms before the onset.

The exploratory regression analysis confirmed a logical assumption, that those individuals who can inhibit more of their eye movements are better at synchronising (more control gives more precision). There are some limitations in the inference of the regression considering the sample size in the analysis (N = 16), but as the finger-tap performance did not increase the variance predicted by lambda alone it is plausible that the control of eye movements utilises a shared resource with fine-motor control. Alternatively, both could simply be the product of the auditory processing of temporal detail (Tierney & Kraus, 2013b). These different interpretations could be tested by providing auditory feedback on eye movements, both increasing the temporal awareness of eye movements and potentially enhancing the temporal precision of the eye movement itself to align with the auditory IOI.

The study provides clear evidence for two mechanisms underlying eye movement timing. Firstly, automatic (uncontrollable) eye movements are well characterised by the first distribution. This distribution was relatively invariant across IOI levels and had a small deviation, which is likely representative of the distribution underlying the thresholds for the random-walk. Secondly, this study provides evidence for the capacity for direct control of fixation durations, and the temporal precision of this control. Some participants were better at directly controlling the movements in time than others, and this was directly related to how many fixations they were able to control. Their skill at tapping in time did not account for any further variance. When considering the nature of eye movement timing, the persistent presence of (automatic) off-task saccades during this task are evidence for notable limitation on the moment-by-moment direct control of eye movement timing.

5.3: Experiment Six: Auditory Eye Movements and Direct Control

When explicitly synchronising movement, for example a finger-tap, with an external beat, the asynchronies between the two perceptions inform corrective behaviours that realign future movements to correspond more closely with the beat. This process is conceptualised as a feedback loop within error correction models, that informs both period and phase correction (Repp, 2005; Repp & Su, 2013). Gaze-contingent visual displays can inform this feedback process, as evidenced in the previous study (E5) which had many participants synchronising their eye movements and predicting the auditory IOI. There are also known limitations in awareness and temporal representation across saccades.

Self-awareness of when and where vision fixates and saccades is limited (Clarke et al., 2016). The underlying reason for this is logical; visual perception presents a broad and detailed percept of the world from incomplete information (Findlay & Gilchrist, 2003). To increase awareness of the fundamental limitations of vision, for example what is being processed in detail when, would detract from this coherent percept. An example of perceptual smoothing is how the visual system accounts for an incomplete stream of temporal information across saccades. A series of studies that flashed temporal visual information prior to saccades identified a phenomena coined chronostasis (Yarrow et al., 2001). The temporal information flashed just before the saccade was perceived for longer than it was presented. This perceptual blurring of time around saccades, and limitations in extracting temporal information from fixation onsets, may have provided an ineffective or temporally insensitive feedback in the previous experiments.

In contrast with vision, auditory perception is highly sensitive to temporal information. This difference is apparent in the previously discussed finger-tap studies to visual and auditory referents, which find less variable responses, faster tapping rates, and more synchronised behaviour to auditory over visual stimuli (Patel et al., 2005; Repp, 2003). One plausible argument is that SMS to visual stimuli engages auditory processing for temporal estimation (Repp, 2003), and that individual differences in SMS are directly related to temporal sensitivities to sound (Tierney & Kraus, 2013b). A saccadic SMS study by Hornof and Vessey (2011) used an auditory handclap to sonify eye movements when musicians were tasked to orient between two fixed horizontal targets at 250ms, 500ms and 1000ms. Performance was poor at 250ms but improved as the tempo slowed. The authors also compared sonification of the saccade onset and the fixation onset, and concluded the fixation onset handclap produced the more synchronised response. A sound cue aided the eye movement SMS performance.

A key motivation for this study is to extend experiment 5, to measure the inhibitory control and synchronisation of eye movements across different feedback types. Experiment 5 used a gaze-contingent visual referent only, and this study will extend this to employ three types of feedback, firstly visual only (as in E5). Secondly, a sonified representation of each fixation start time (saccadic tone) in addition to the gaze-contingent visual presentation. Thirdly, a sonified representation of fixation start times following a spatial shift (spatial tone) in addition to the gaze-contingent visual presentation. The important difference between the two auditory feedback conditions is that the saccadic feedback represents the beginning of every fixation irrespective of location. Considering the previous data which showed that 50% of eye movements occur outside of direct control, these automatically executed saccades will be a form of noise to the feedback signal. The spatial feedback only sounds when gaze shifts from one target to the next, and better presents intention during the task, potentially mirroring attentional shifts.

The predictions for the study are, first, that the saccadic feedback condition will enhance awareness of each eye movement, which will increase the proportion of inhibited fixations (an increased lambda) than both the visual and spatial conditions. The second prediction is that the spatial auditory feedback will produce the most synchronised eye movements, as the sound will best represent the intention to move (direct control) and is a more accurate temporal feedback cue for correction (a more useful auditory signal for alignment). The final prediction is that both auditory conditions will have circular means that are closer aligned to the IOI, as the auditory feedback will facilitate a more temporally precise response.

5.3.1 Methods.

Participants.

The effect sizes in Experiment 5 were generally very high (for example $\eta^2 = .738$ for the interaction between the distributions and IOI for the mean fixation duration). Power analysis (Faul et al., 2007) of these effects determined sample sizes less than 5. As the effect of auditory feedback is unknown, the sample size was pre-set to 24, which we deemed sufficient to identify a medium effect size. A total of 38 naïve participants completed a paid pre-test for the study, and 14 were unable to produce an in-time tapping response (to a 300ms IOI beat). This left 24 participants, (17 Female), with an age range from 18 to 33 (M = 23.5, SD = 3.97), who completed the entire study for payment. Musical training scores (M = 20.2, SD = 11.9) ranged from 7 (no training) to 40 (more training than 75% of the population).

Design modifications.

With the exception of the auditory representation tone, which was a 40ms (with minimal fading to reduce clipping) 500 Hz sine-tone for eye movements feedback, the only other change from experiment five was a reduction in IOI levels to three (300ms, 600ms and

900ms). These IOI were chosen as they represented the widest spread of three IOI values based on the range from E5. All other equipment and stimuli remained the same.

Procedure.

The participants were recruited to an ‘On Beat Eye Movement Study’, with the encouragement to sign up if they felt confident in their ability to produce consistent on-beat responses. The signup information also detailed that a pre-test of their beat tapping ability would be conducted prior to the eye-tracking study, the outcome of which would determine either the continuation to the eye tracking experiment or termination and a diminished payment. The tapping task mirrored that of E5 with a new tempo set to 300ms IOI (200 beats per minute). Those who produced one trial within 10ms of the target IOI continued to the eye movement study. Performance on this pre-test was generally bimodal. Those who completed the experiment were generally within 2ms of the target IOI, and those who did not had three highly variant averages far from the target IOI.

A 9-point eye calibration preceded the experiment validated to within 1°, and was repeated whenever the participant moved from the chin-rest or if the drift correction between trials exceeded 1°. The first trial was a gaze-contingent familiarisation trial. The participants were instructed to look from circle to circle clockwise, clicking the right trigger when the colour sequence changed. Each trial required a sequential fixation of 180 circles, or timed out after 3-minutes. The second trial was a further practice that had an additional instruction to synchronise when the eyes moved with the music (600ms IOI). During this second trial participants were instructed to hold fixation on each circle and to control each movement to match the beat. They were encouraged during this practice trial to ensure every transition synchronised with the beat and to click the right trigger when the colour sequence changed. Following these two practice trials the participants completed 12 trials, each trial randomly assigned to one level of feedback condition (visual, spatial tone, and saccadic tone) and musical IOI (300, 600, 900ms and silent control). When completing a visual trial the instructions were to match the circle transitions with the musical beat. When the trial was saccadic or spatial the instructions were to use the eye movement beep to move between the circles in time with the musical beat. Each condition level had the same repeated instruction to click the right trigger when the colour sequence changed, this was the only task when the IOI was silent. Following completion of these eye movement trials, the final tasks were to tap in time to each IOI (each trial recorded 60 taps) and to fill in a short musical training questionnaire.

Data analysis plan.

Eye movement classifications, data removal and data modelling approaches mirrored that in Experiment 5. The ‘mixtools’ package in R (Benaglia et al., 2009) was employed to identify the lambda, mu and sigma of each distribution, separated by feedback condition (visual, saccadic tone, and spatial tone) and musical IOI (300, 600 and 900ms). All model fits converged on bimodal fit to the data.

5.3.2 Results.

Feedback condition in silence.

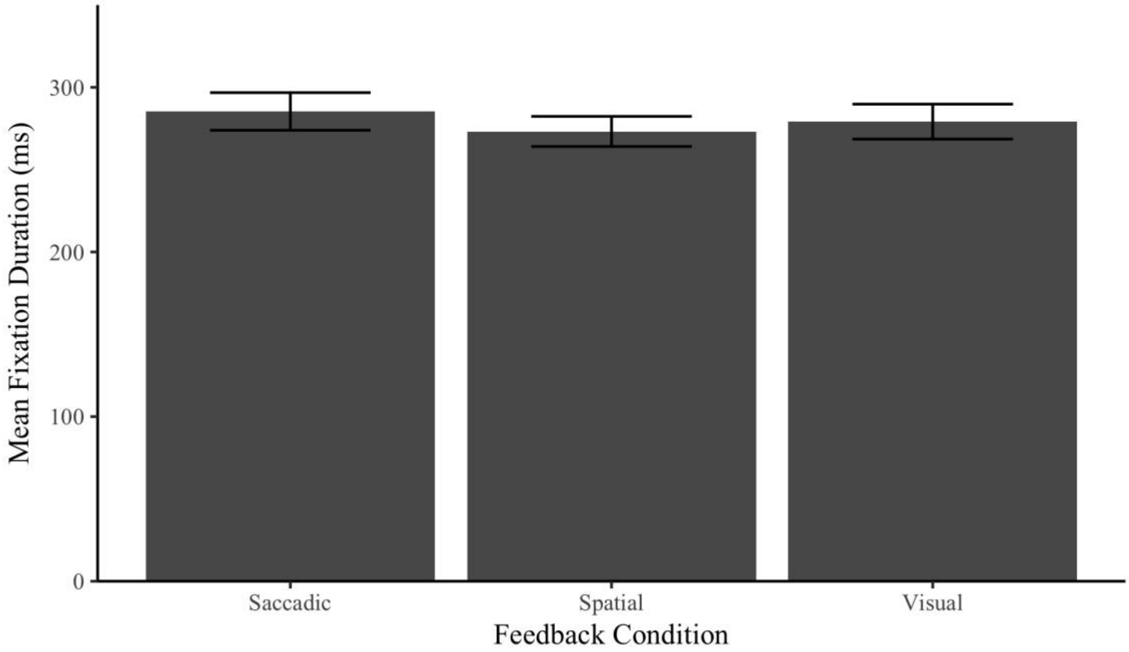


Figure 5-13 Mean Fixation Duration (±1SE) by Feedback Condition for the Silent Condition (no Musical IOI; the feedback tones were present in spatial and saccadic conditions)

As observed in Figure 5-13, there was a marginally significant main effect of feedback condition on fixation durations without music, $F(2,46) = 3.09$, $p = .06$, $\eta^2 = .118$ (NS). Bonferroni corrected pairwise comparisons showed a significant difference between the visual feedback condition and the spatial feedback condition ($p = .033$). A $\log_{10}$ correction of the fixation durations was also not significant, $F(2,46) = 2.26$, $p = .116$, $\eta^2 = .090$, although as reported without correction, the difference between visual and spatial feedback was significant ($p = .033$).

Distribution parameters: Lambda.

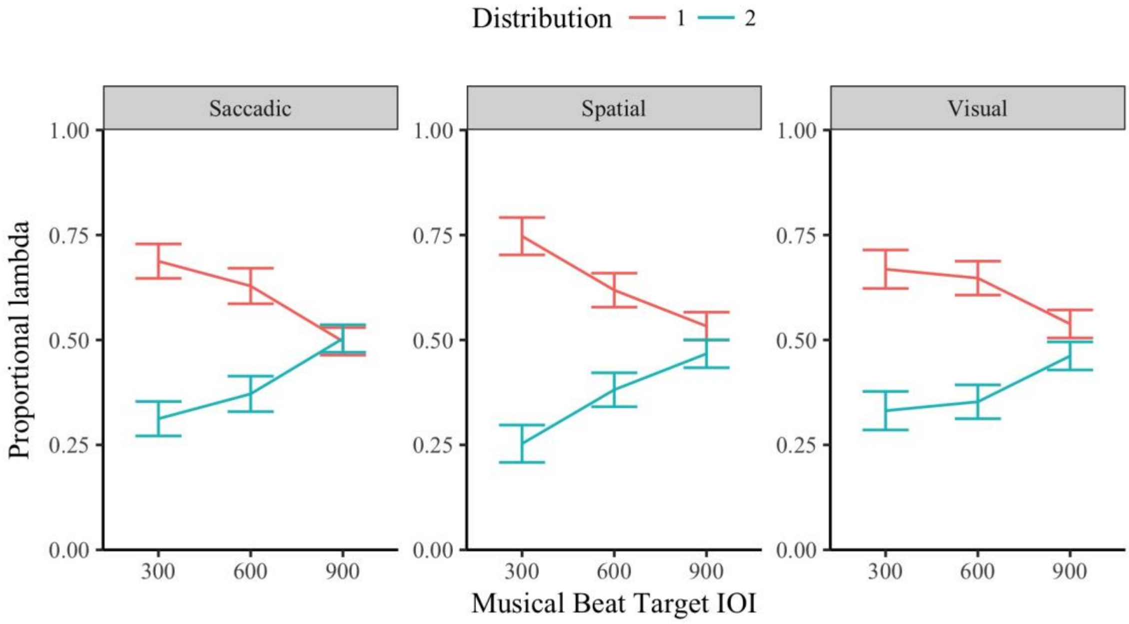


Figure 5-14 Lambda Proportion of the Two Distributions by Feedback Condition and Musical IOI (±1SE)

A repeated measures ANOVA of the second distribution *lambda* proportion (analysis of the *lambda* for only one distribution is required as together they equal 1) by feedback condition and musical IOI, had no significant main effect of feedback, $F(2,46) = .049$, $p = .62$, $\eta^2 = .021$ ($BF_{01} = 13.37$), see Figure 5-14. The main effect of musical IOI was significant $F(2,46) = 15.40$, $p < .001$, $\eta^2 = .401$; the proportion of *lambda* in the second distribution increased linearly as the IOI slowed ($p < .001$). The interaction between feedback condition and musical IOI was not significant, $F(4,92) = .79$, $p = .54$, $\eta^2 = .033$.

Distribution parameters: Mu.

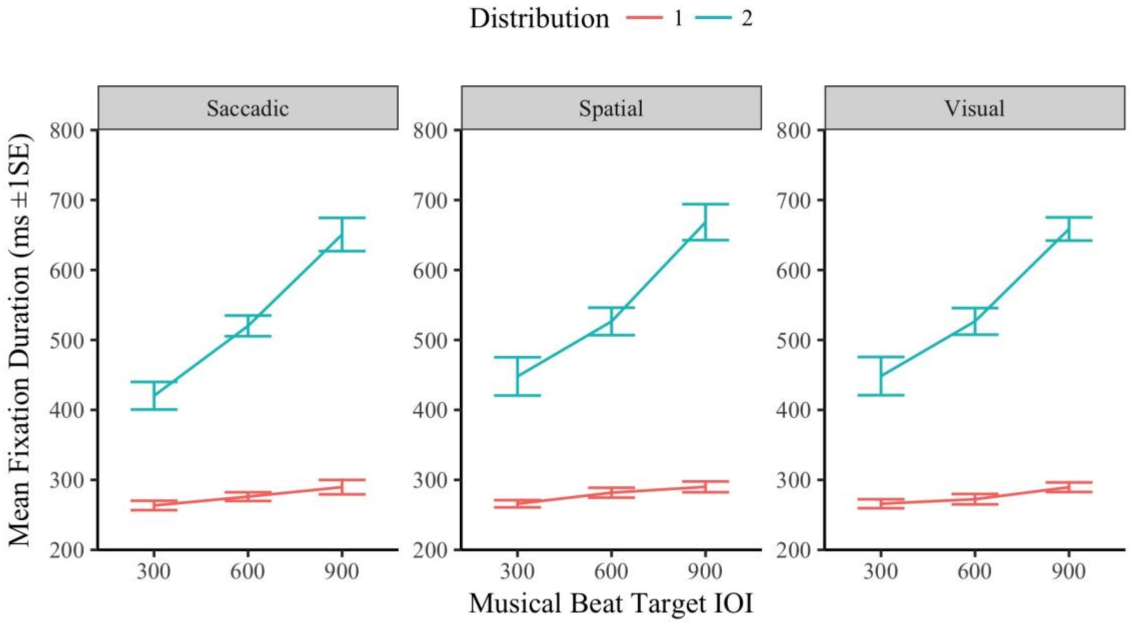


Figure 5-15 The Mean Fixation Durations (mu) of the Two Distributions by Feedback Condition and Musical IOI (±1SE)

A repeated measures ANOVA compared fixation duration means for the two distributions (i.e. *mu*), three feedback conditions and three IOI levels. The main effect of distribution was significant, $F(1,23) = 561.85$, $p < .001$, $\eta^2 = .961$; the fixations in the first distribution (M = 277.19, SD = 36.22) were significantly shorter than the second (M = 540.83, SD = 139.71). The main effect of feedback was not significant $F(2,46) = .76$, $p = .48$, $\eta^2 = .032$ ($BF_{01} = 34.11$). The main effect of musical IOI was significant, $F(2,46) = 64.31$, $p < .001$, $\eta^2 = .737$. As observed in Figure 5-15, the interaction between the two distributions and the musical IOI was significant, $F(2,46) = 68.98$, $p < .001$, $\eta^2 = .750$. The mean fixation durations in the second distribution lengthened significantly as the IOI slowed, $F(2,46) = 70.88$ $p < .001$, $\eta^2 = .755$. The first distribution also slowed as the IOI slowed, $F(2,46) = 10.14$, $p < .001$, $\eta^2 = .164$, although the linear trend was notably more conservative ($p = .05$). The interactions between distribution and feedback, feedback and IOI and the three-way interaction were all not significant ($F < 1.3$, $p > .3$).

Distribution parameters: Sigma.

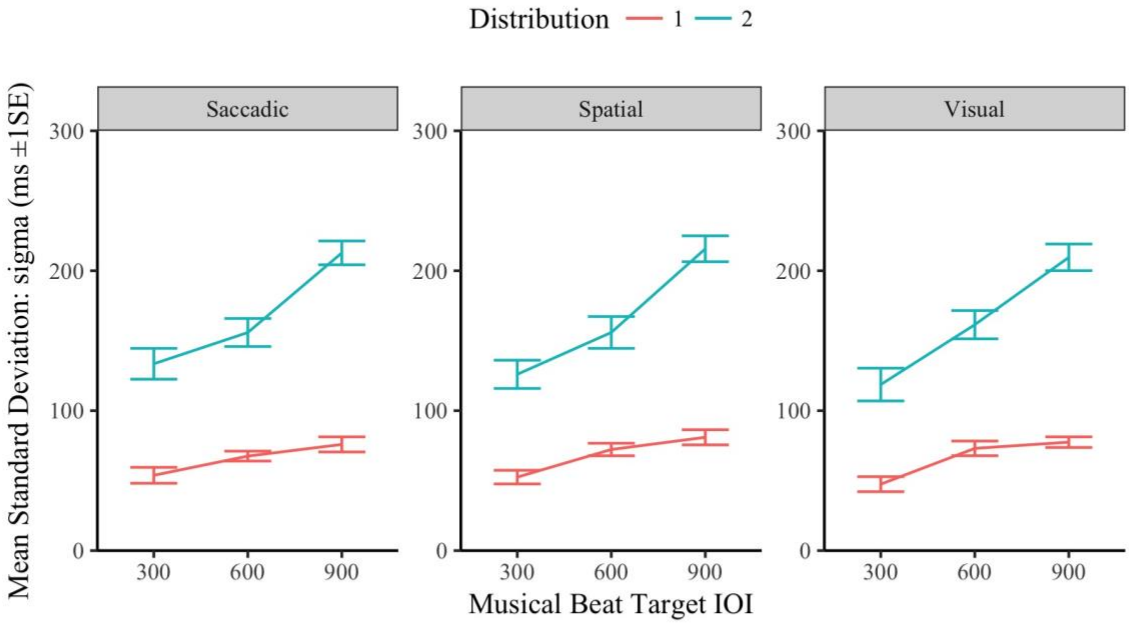


Figure 5-16 Standard Deviation (*sigma*) Means of Fixation Durations (±1SE) for the Two Distributions by Feedback Condition and Musical IOI (±1SE)

A repeated measures ANOVA compared fixation duration deviations (sigma) for the two distributions, three feedback conditions and three IOI levels. The main effect of distribution was significant, $F(1,23) = 309.38$, $p < .001$, $\eta^2 = .931$; the deviations in the first distribution (M = 66.78ms, SD = 26.34) were significantly smaller than the second distribution (M = 165.49, SD = 60.98). The main effect of feedback was not significant $F(2,46) = .27$, $p = .77$, $\eta^2 = .011$ ($BF_{01} = 37.13$). The main effect of musical IOI was significant, $F(2,46) = 52.92$, $p < .001$, $\eta^2 = .697$. As observed in Figure 5-16, the interaction between the two distributions and the musical IOI was significant, $F(2,46) = 26.64$, $p < .001$, $\eta^2 = .537$. The deviation of fixation durations in the second distribution increased significantly as the IOI slowed, $F(2,46) = 46.59$ $p < .001$, $\eta^2 = .669$. The first distribution deviation also increased as the IOI slowed, $F(2,46) = 24.05$, $p < .001$, $\eta^2 = .511$. The interaction is because the slope of the second distribution was much larger than the first. The interactions between distribution and feedback, feedback and IOI and the three-way interaction were all not significant ($F < 1$, $p > .4$).

Mean resultant length

Prior to reporting the MRL values, the circular uniformity of the data was tested with the Rayleigh's test (Wilkie, 1983). Only one participant had significantly non-uniform responses to all conditions. Inversely no participant had uniform responses too all conditions. As observed in Table 5-2. and Figure 5-17, performance at the 300ms IOI was notably poor although some participants did achieve modest MRL

Table 5-2 Proportion of significantly non-uniform responses by Feedback Condition and Musical IOI

	Saccadic	**Spatial**	**Visual**
300ms	.25	.16	.33
600ms	.46	.58	.58
900ms	.75	.67	.79

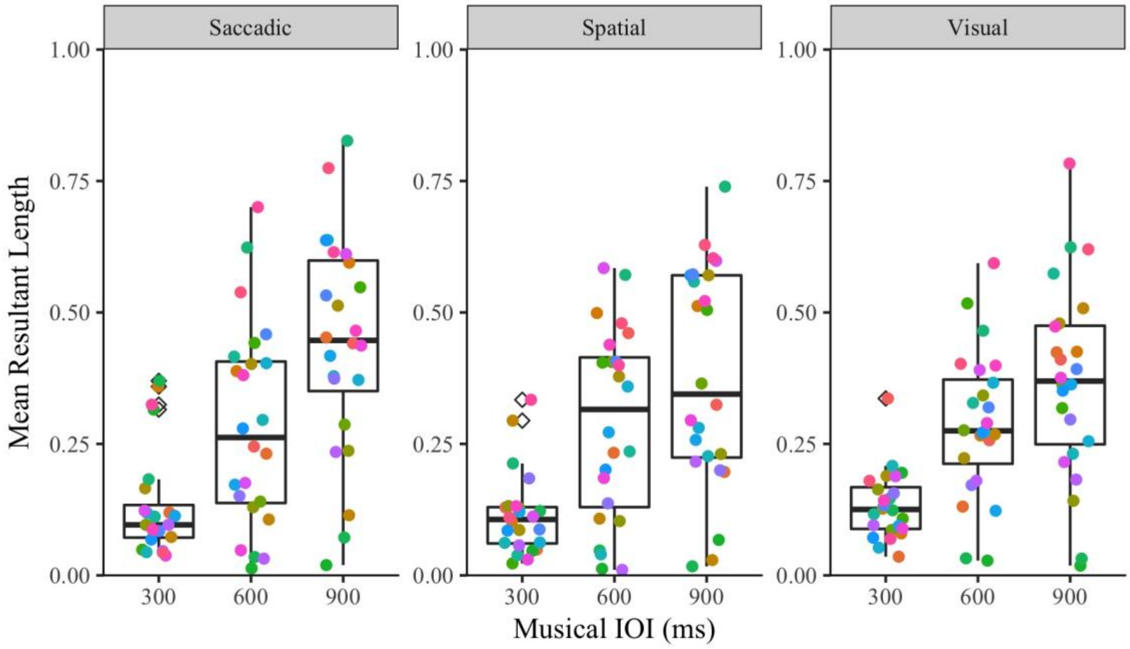


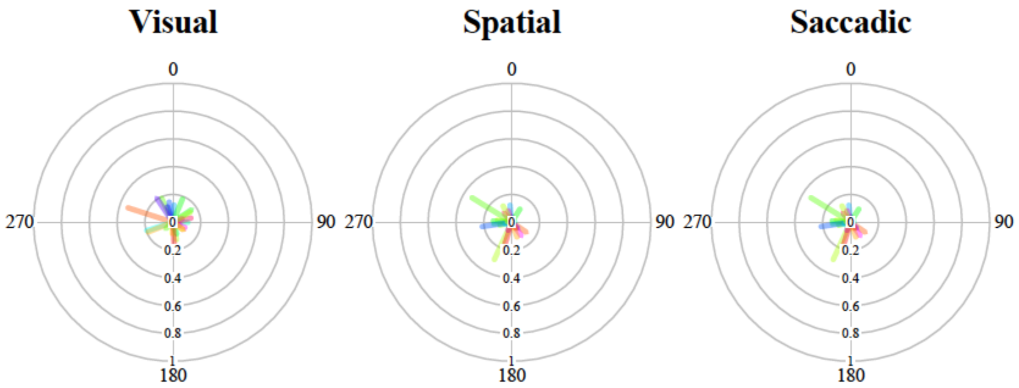
Figure 5-17 A boxplot of MRL values by Musical IOI and Feedback Condition, colours represent participants (diamonds are potential outliers, in this instance better performers)

A repeated measures ANOVA of the MRL values by feedback condition and musical IOI found no significant main effect of feedback, $F(2,46) = 2.04$, $p = .14$, $\eta^2 = .082$ ($BF_{01} = 9.638$). The main effect of IOI was significant, $F(2,46) = 11.89$, $p < .001$, $\eta^2 = .341$; the MRL increased linearly as the IOI slowed ($p < .001$). The interaction between feedback and IOI was not significant, $F(4,92) = 1.66$, $p = .17$, $\eta^2 = .067$.

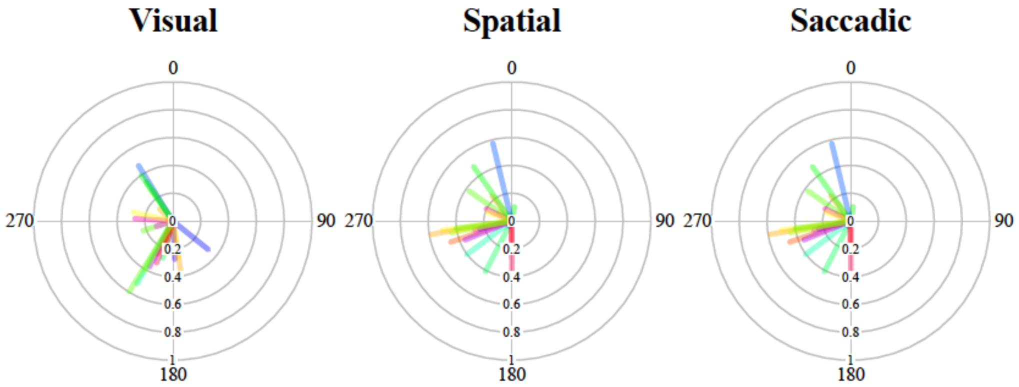
Circular mean.

As in experiment 5, the circular mean values were converted into an absolute linear distance measure of the circular mean from the target IOI. In Figure 5-18, the pattern of predictive responses clustering prior to 0 in phase emerged at the 600 and 900ms conditions. The MRL values and phase location were much less meaningful at 300ms considering the number of significantly uniform responses (Rayleigh test). Therefore, the analysis of the mean distance omitted the 300mm condition (the circular mean does not represent the data for most participants).

300ms Circular Plots



600ms Circular Plots



900ms Circular Plots

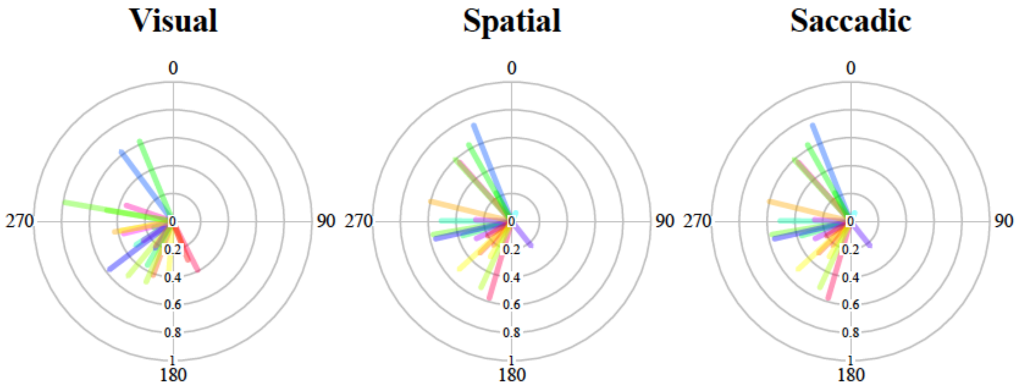


Figure 5-18 The circular mean (vector orientation) and mean resultant length (vector length, the radius is MRL =1) by feedback condition and musical IOI level. Colours represent participants

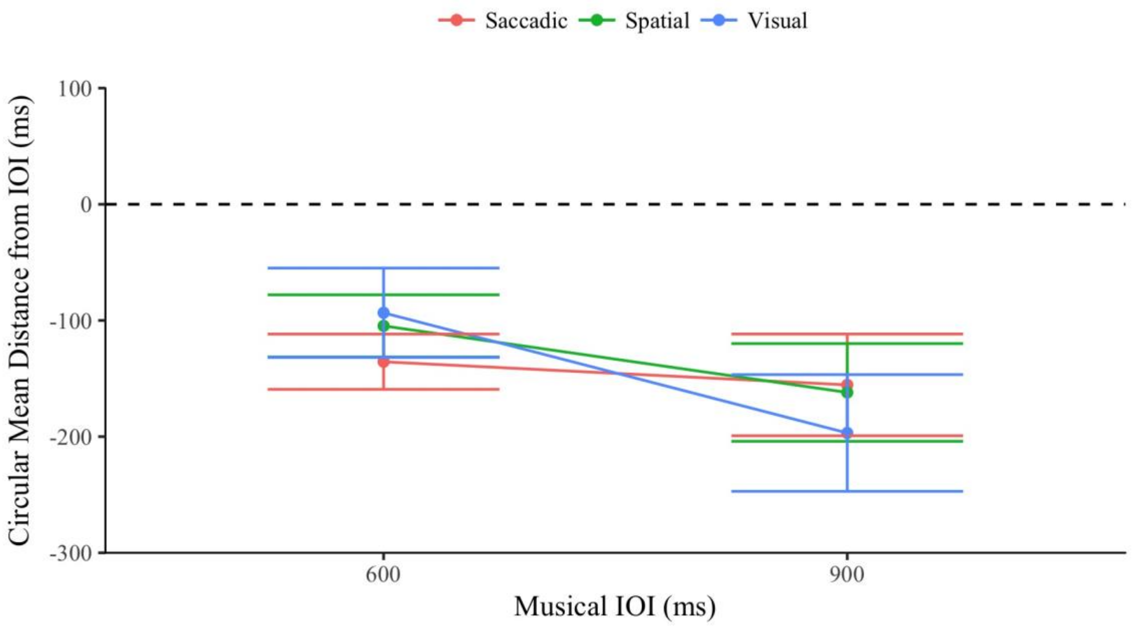


Figure 5-19 Mean Distance (ms) of the Circular Mean from the Musical IOI by Feedback Condition (±1SE)

A repeated measures ANOVA of the mean distance was contrasted across feedback condition and the two remaining levels of IOI (600 and 900ms). As seen in Figure 5-19, the main effect of musical IOI was significant, $F(1,23) = 6.02$, $p = .022$, $\eta^2 = .207$; the 900ms condition was further from the IOI (earlier in relation to the beat) than the 600ms condition. The main effect of feedback type was not significant, $F(2,46) = .092$, $p = .913$, $\eta^2 = .004$ ($BF_{01} = 13.85$). The interaction between musical IOI and feedback condition was also not significant, $F(2,46) = .463$, $p = .633$, $\eta^2 = .020$.

Correlations between direct control and synchronisation.

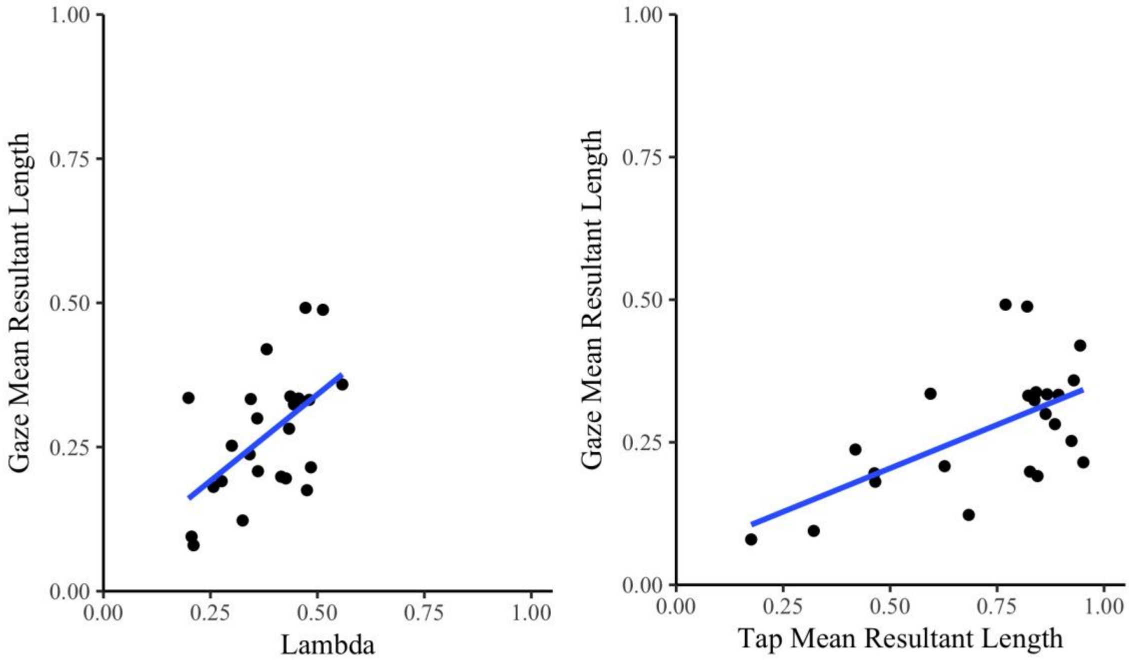


Figure 5-20 Scatterplots with linear regression line of the relationship between the proportion of lambda in distribution two and gaze mean resultant length (left) and of finger-tap response mean resultant length and the mean resultant length of eye movements (right)

A linear regression was calculated to predict gaze MRL based on the proportion of lambda in the second distribution (eye movements subject to direct control). A significant association was found between lambda and MRL, $F(1,22) = 9.574, p < .01$, $R^2 = .303$. The participants predicted MRL is .042 + .599*lambda. Participants' MRL increased by .599 when lambda increased by 1. A linear regression was also calculated to predict Gaze MRL based on the Tap MRL scores. A significant association was found between the two MRL values, $F(1,21) = 13.75, p < .01$, with an R^2 of .3257. The participants' predicted gaze MRL is .056 + .300*tap MRL. Participants' gaze MRL increased by 0.300 when their tap MRL increased by 1 (see Figure 5-20). A final piece of analysis was to consider both the tap MRL and the lambda as predictors of gaze MRL performance. The addition of the tap MRL as a predictor did significantly increase the amount of variance accounted for by lambda alone, $X^2(1) = .044$, $p = .023$.

5.3.3 Discussion.

This study predicted that the auditory feedback conditions would enhance the synchronisation and temporal control of eye movements. The results of the study provide Bayesian evidence that the feedback conditions did not influence any of the distribution parameters (*lambda, mu, sigma*) or synchronisation measures (MRL and mean distance). None of the predictions were supported, although all of the musical IOI effects identified in E5 were replicated. The addition of an auditory referent to the gaze-contingent visual cue of eye movements did not aid or detract from the temporal control of eye movements, which seems contrary to the evidence in support of the predictions that SMS to visual referents engages auditory processing (Repp, 2003), and SMS performance is directly related to temporal representation of sound (Tierney & Kraus, 2013b). The evidence from this study finds that the temporal representation of synchrony utilises only visual feedback, and was convincingly invariant to informative (spatial) and noisy (saccadic) auditory eye movement cues.

The main findings in this study replicate experiment five. The second modelled fixation duration distribution increased linearly as the IOI slowed. The proportion of fixations in the second distribution (subject to direct control) peaked at 50% (as in E5). The synchronisation performance at 300ms was again very poor, which suggests that the IOI is too fast (close to the saccade timer) to facilitate sufficient control. At the slower IOI levels eye movements were predictive of the IOI, becoming increasingly early (~210ms) at the slowest 900ms condition.

The auditory tone feedback did not influence behaviour as predicted, or show any presence effect in the distribution parameters. The simple 500 Hz sign-wave tone was distinct from the drum sounds, but this design decision may have had an effect on its usefulness. As the tone was distinct from the music, it is possible that the tone was segregated as an uninformative stream of auditory information and not attended to (Bregman, 1990). To overcome this limitation a future study should utilise a tone representation that matches the beat referent within the music. For example, using a kick-drum sound over a musical track with kick drum sounds on the beat would make segregation of the two streams impossible and reinforce the perceptions of asynchrony. An alternate approach that could make the task more interactive would be presenting temporally incomplete music, where the task is to fill the gaps with eye movement sounds. This would provide sub-division temporal referents that could increase synchronisation performance.

In conclusion, the voluntary (direct) control of eye movements was limited to around 50% of all movements, and relied solely on visual cues to align in time. Participants can synchronise their eye movements, but there are notable individual differences in ability, which may be linked to a wider capacity to inhibit or control movement. The ability to control when your eyes move (i.e. overt shifts of attention) is an essential skill for tasks that require focus at specific locations and attention to detail. Considering the clear differences in capacity for control of movement between oculomotor and other SMS movements, the implications of these findings as evidence for differences in the temporal control of supra-modal attention as opposed to early auditory processing ability (Tierney and Kraus, 2013a), would require further studies that that test this prediction. Finally, voluntary control as a form of direct control of eye movements has convincingly slowed up to 50% of all fixations, yet is not represented within any of the current eye movement timing models.

Chapter 6

Transitional Probabilities Modulate the Precision of Direct-Control

6.1 Chapter Overview

This final empirical chapter extends the prior volitional control findings from chapter 5, to measure whether manipulating the probability of a known inhibitory influence (saccade reversals) will cumulatively increase the proportion of eye movements that can be controlled, and increase the temporal precision of direct control. The results show that increasing the probability of a saccade direction change does increase the proportion of fixations subject to inhibition, but this increase is not represented in synchronisation or closer approximations of the beat interval. Furthermore, a speeding in the first distribution (representative of the saccadic timer) as the probability of direction change increases, is argued as evidence against the global component proposed in ICAT (Trukenbrod & Engbert, 2014).

6.2: Experiment Seven: Task Complexity and the Voluntary Control of Eye Movements

The recent eye movement timing models account for top-down influences, including visual factors and task (for example, visual search or memorisation of the same scene) as direct modulators (through inhibition, rate of rise evaluations in stay-go decisions, and saccade cancelation) of oculomotor timing (Nuthmann et al., 2010; Tatler et al., 2017; Trukenbrod & Engbert, 2014). Within experiments 5 and 6 (chapter 5), this thesis has extended the conceptualisation of direct control to evidence the capacity for, and precision of, volitionally controlled eye movements during a highly constrained sequential visual search task. A key finding of these experiments was the predictive relationship between the proportion of fixations in the second distribution (*lambda*: fixations subject to direct control) and the synchronisation of eye movements (MRL). An increase in the proportion of fixations that are subject to delay (*lambda*) increased the level of synchronisation. An important caveat is the lambda proportion in the second delayed distribution was relatively low, accounting for only half of the fixations within the distribution at the most ideal conditions (e.g. at very slow intervals).

Demands on the level of cognitive processing can influence the duration of fixations. Reading research investigating eye movement timings have shown both linguistic and low-level visual factors delay fixation durations; for example, when fixating less frequently occurring words or longer word lengths (Rayner, 1998). As previously discussed within the introduction to this thesis, the viewing task can also influence fixation durations; for example, fixations durations are longer when tasked to memorise a scene than when searching for an object within the same scene (Henderson et al., 1999). A feature of the visual search paradigm employed throughout this thesis has been the intention to limit visual and task influences on eye movement timing during the experiments. By creating a continuous search task that required minimal foveal processing and alterations in saccade programming decisions (e.g. where to orient to), the task may have inadvertently over-emphasised the role of the saccadic timer and minimised opportunities for direct control (i.e. there may be more opportunity to delay a saccade that is already being inhibited). This overtly simplified visual environment may not be representative of the normal decision making taking place under most viewing tasks, so this study will test whether changing the probability of a repeated saccade program alters the proportion of saccades under direct control.

When visual attention reorients back to a previously attended location, either covertly or overtly (with a corresponding eye movement), the return is subject to a delay (Posner & Cohen, 1984), termed inhibition of return (IOR). This delay is also present in eye movement durations that re-fixate at prior fixation locations, commonly referred to as saccadic-IOR (Klein, 1988; 2000). A prominent early theory that conceptualised the IOR delay is the foraging facilitator hypothesis, that the allocation of attention to new locations is promoted and the likelihood of returning to prior locations actively demoted

(Posner & Cohen, 1984; Posner, Rafal, Choate, & Vaughan, 1985). There is notable evidence that contradicts this account (for example: Hooge, Over, van Wezel, & Frens, 2005), as when visual onsets occur at previously fixated locations the probability of attending these locations increases, described as a ‘facilitation of return’ (Smith & Henderson, 2009; 2011). The delay in fixation durations is somewhat attributable to saccadic momentum (Smith & Henderson, 2009), or a systematic tendency for saccades to orient in the same direction as the one prior (Tatler & Vincent, 2008). In other words it is the direction change that incurs delay (Smith & Henderson, 2011). The important behaviour relevant to this study is that saccades that reorient to the immediately prior location (a saccade reversal) show a robust delay in fixation duration (Ludwig, Farrell, Ellis, & Gilchrist, 2009; Smith & Henderson, 2011). This delay will be utilised in this experiment as an inhibiting (direct control) influence on the timing of eye movements. The probability of a change in direction, as a return to the previously fixated location (transitional probabilities), will be manipulated across three levels, a low probability of direction change (*Low*), a medium probability (*Medium*) and a high probability of direction change (*High*).

This study will extend the previous studies (E5 and E6) to consider volitional control of eye movements to musical IOI (at speeds that can be synchronised to: 500, 700 and 900ms) across the three transitional-probability (TP) conditions (*Low, Medium* and *High*). The predictions of the study are firstly, that the increasing probability of direction change will disrupt saccadic momentum (Smith & Henderson, 2009), increasing the proportion of fixations (*lambda*) subject to inhibition. Secondly, increasing in the TP of direction changes is predicted to interact with the musical IOI demands (controlling saccades to move in time with musical IOI). Specifically, an increase in TP and slowing the target IOI will increase both direct control and temporal control (higher *lambda,* higher MRL). Thirdly, the predictions for the *mu* and *sigma* components of the first distribution are two sided. If, as proposed in the ICAT model (Trukenbrod & Engbert, 2014), there is a global component that varies the saccade timer as an estimator of demand, the mean of the first distribution (*mu* 1) will shift as the TP increases (the *sigma* should remain constant). Alternatively, if the saccade timer is not subject to a global component (as described in the CRISP model, Nuthmann et al., 2010), an increase in TP or IOI demands will not influence the mu and sigma of the first distribution.

6.3 Method

Participants.

The study required completion of an online pre-screen which measured beat-tapping ability prior to attending the study. The main study tested 30 participants who all completed both the eye-tracking and a tapping task, which confirmed their ability to produce an in-time tapping response. Six were either unable to produce a sufficiently synchronised response to all three IOI levels when simply tapping a track-pad (their average MRL was below 0.5, or more random than in synch: 4/6), or the eye-tracking data was incomplete (2/6). As this task required analysis of on-beat responses the poor performance and lost data excluded them from analysis. Power analysis of the previous effects (and those of IOR) were large (indicating small sample sizes), therefore the sample size was set to 24, sufficient to detect a medium effect. The study analysed 24 participants, (20 Female), with an age range from 18 to 39 (M = 24.29, SD = 5.5), who contributed data for analysis. The musical training scores (on a scale from 7: no training to 49: more training than 95% of the population); none of the participants were professional musicians.

Design Modifications.

The eye-tracking setup, monitor, and headphones remained the same as the previous experiments (E5, E6). The musical beat loop remained the same as E5, with three levels selected, at levels that could previously be synchronised to (500, 700 and 900ms). The major change was the visual stimuli, which utilised a 12-circle elliptical shape with the same dimensions as the prior two experiments, see Figure 6-1. Within the gaze-contingent triggered circle was a small white arrow (.5º along the longest edge). The arrow either pointed clockwise (directed toward the centre of the next circle) or anticlockwise. Only one circle displayed an arrow at any moment, and correct sequential orientations were required to show the next circles arrow direction.

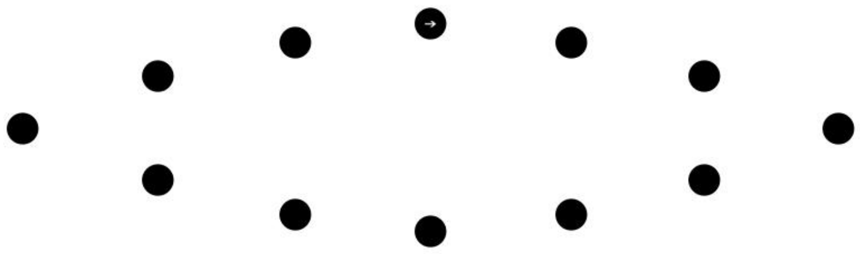


Figure 6-1 An example image of the gaze-contingent directional arrow within the elliptical shape

The data analysis, data cleaning steps, and the statistical modelling parameters employed in E5 and E6 were maintained for this study. As with the previous experiments, the mixture modelling converged on a bi-modal fit for all participants in each condition. Visualisation of the fit was particularly close, as the second mode is more pronounced at IOI levels slower than 400ms.

Procedure.

All participants gave informed consent to take part in a ‘Moving Your Eyes to Music’ study. A 9-point eye calibration preceded the experiment validated to within 1º, and was repeated whenever the participant moved from the chin-rest or if the drift correction between trials exceeded 1º. The first trial was a gaze-contingent familiarisation trial. The participants were instructed to look from circle to circle as directed by the arrow. This practice trial (and the next trial) had a low probability of direction changes. Each trial required the sequential orientation of 120 circles (60 less than the previous experiments to reduce fatigue due to the increased task demands), or timed out after 3-minutes. The second trial had the additional instruction for participants to synchronise their eye movements with the beats in the music (set to 500ms IOI). During this second trial participants were instructed to hold their eyes on each circle and to only move to the next circle in time with beat. They were encouraged to ensure every movement synchronised with the beat and to always move in the direction indicated by the arrow.

The transitional probability conditions were counterbalanced across participants (groups A and B). Those in group A completed each of the four trials (each IOI and silent) firstly in the low (changes in direction between 28 and 31 circles), then medium (changes in direction between 15 and 18 circles) and ended with the high probability condition (changes in direction between 2 to 5 circles). The IOI order (and silent condition) was randomised within each condition type. Those in group B completed the conditions in the reverse order: high, medium, and then low. When the TP changed (after four trials) the participants were given a break and were informed of the new TP (e.g. “The frequency of direction changes in the next four trials will be: High”). Following completion of these eye movement trials, the final task was to tap in time to each IOI on an Apple trackpad to confirm their ability to produce a synchronised response (each trial recorded 60 taps). The experiment ended with a short musical training questionnaire (Müllensiefen et al., 2014).

6.4 Results

Transitional probabilities (TP) in silence.

A repeated measures ANOVA compared the fixation durations across the three TP conditions in silence. There was no main effect of TP, $F(2,46) = 1.14$, $p = .330$, $\eta^2 = .047$ ($BF_{01} = 3.63$). The mean fixation durations in silence were invariant when the only task was to navigate the visual direction. A further analysis of the standard deviation of fixation durations in the silent condition showed a significant main effect of TP, $F(2,46) = 7.40$, $p = .002$, $\eta^2 = .243$. The standard deviations increased linearly (more variable) as the TP increased (more changes in direction). The distribution of fixations broadened around a constant mean as a function of cognitive demand.

Distribution parameter: Lambda

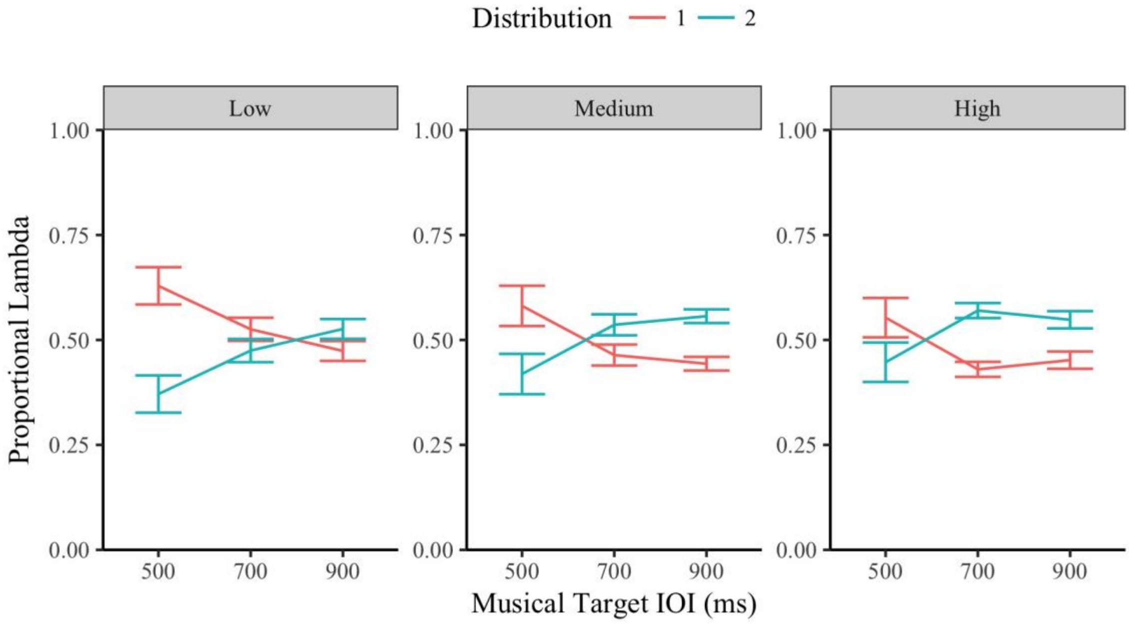


Figure 6-2 The Proportion of Fixations within each of the two Bimodal Distributions (lambda) by Transitional Probability Conditions and Musical IOI

A repeated measures ANOVA compared the proportion of *lambda* in the second distribution (slower fixations subject to direct control) across TP conditions and musical IOI. There was a significant main effect of TP, $F(2,46) = 3.87$, $p = .030$, $\eta^2 = .144$, the proportion of fixations subject to direct control increased linearly as the TP increased ($p = .04$), see Figure 6-2. The main effect of musical IOI was also significant, $F(2,46) = 12.36$, $p < .001$, $\eta^2 = .349$; as in previous studies there was a linear effect, where the slowest IOI had the largest proportion of fixations in the second distribution ($p < .001$). The interaction between TP and Musical IOI was not significant, $F(4,92) = .50$, $p = .73$, $\eta^2 = .021$.

Distribution parameter: Mu.

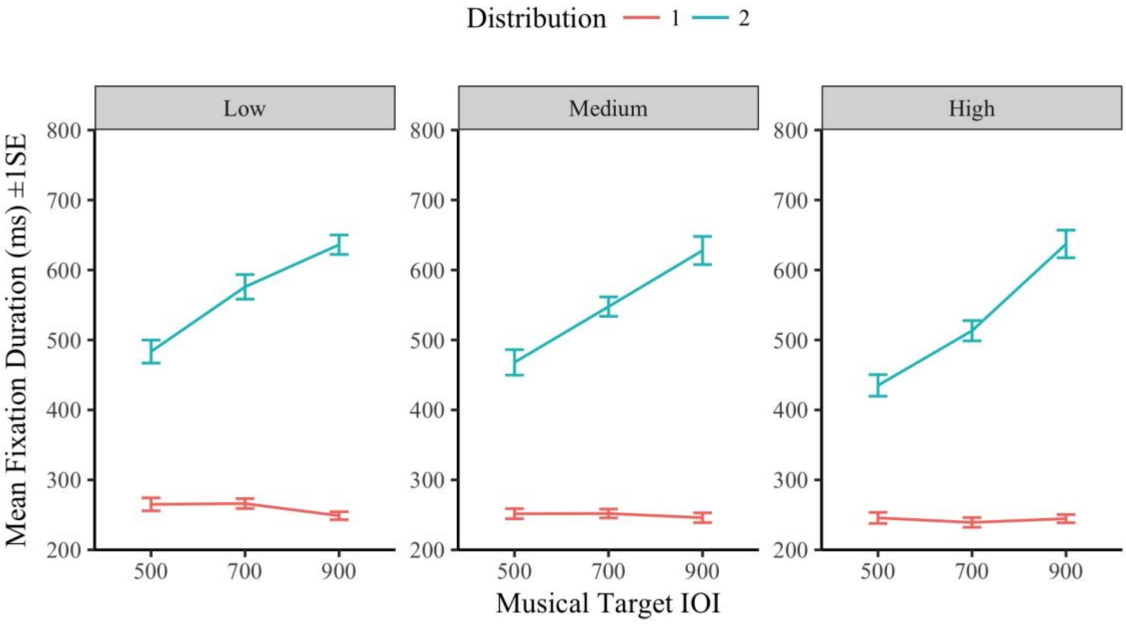


Figure 6-3 Mean Fixation Duration (ms ±SE) by Distribution, Transitional Probability Condition and Musical IOI (±1SE)

A repeated measures ANOVA of the mean fixation durations (*mu*) were contrasted between the distributions, TP conditions, and musical IOI. As is clearly evident in Figure 6-3, the main effect of distribution was significant $F(1,23) = 1960.30$, $p < .001$, $\eta^2 = .988$. The first distribution had significantly shorter fixation durations (M = 251.00ms, SD = 35.16) than the second distribution (M = 547.15ms, SD = 108.61). The main effect of TP was significant, $F(2,46) = 7.94$, $p < .001$, $\eta^2 = .257$. The main effect of musical IOI was also significant, $F(2,46) = 48.37$, $p < .001$, $\eta^2 = .678$.

The interaction between TP and distribution was significant, $F(2,46) = 3.29$, $p = .05$, $\eta^2 = .125$. Fixation durations in the first distribution significantly shortened as the TP level increased, $F(2,46) = 10.89$, $p < .001$, $\eta^2 = .321$. The second distribution had a similar significant reduction in fixation duration as the TP level increased, $F(2,46) =$

6.25, $p < .001$, $\eta^2 = .214$. The interaction was significant because the reduction was larger in the first distribution.

The interaction between musical IOI and distribution was also significant, $F(2,46) = 111.11$, $p < .001$, $\eta^2 = .829$. The fixation durations in the second distribution increased significantly as the IOI level slowed, $F(2,46) = 76.01$, $p < .001$, $\eta^2 = .768$. In contrast the fixation durations in the first distribution did not, $F(2,46) = 1.86$, $p = .182$, $\eta^2 = .075$ ($BF_{01} = 1.67$). The interactions between TP and IOI, and the three-way interaction between distribution, TP and musical IOI were both not significant ($F < 2.1$, $p > .1$).

Distribution parameter: Sigma.

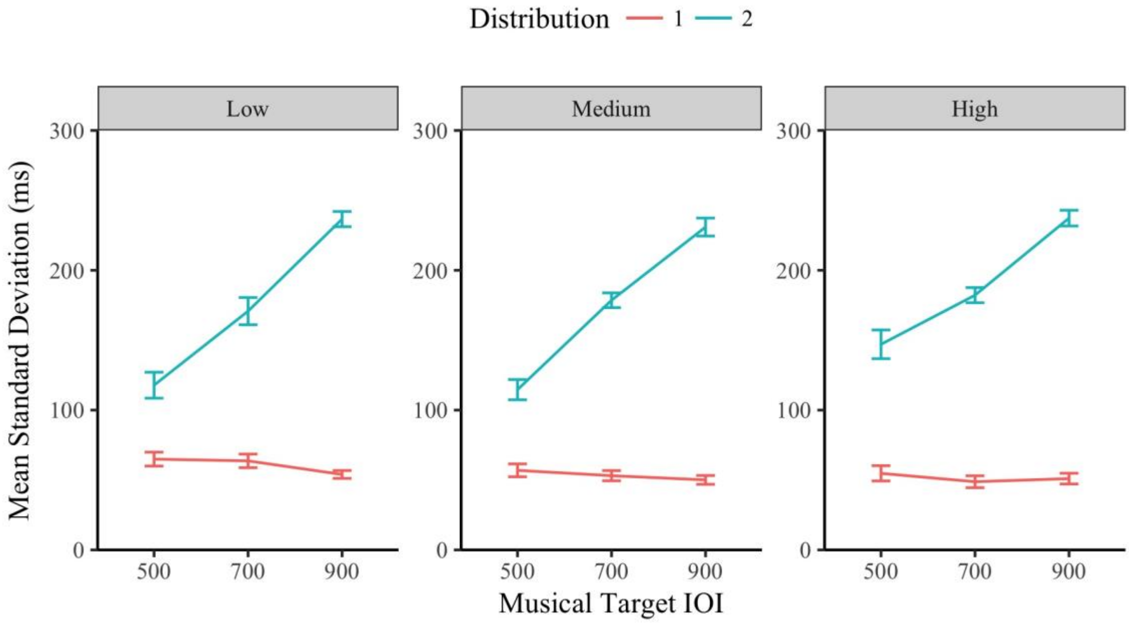


Figure 6-4 Mean Deviation (*sigma*) of Fixation Durations by Distribution, Transitional Probability and Musical IOI levels

A repeated measures ANOVA of the deviation of fixation durations (*sigma*) were contrasted between distribution, TP, and Musical IOI. The main effect of distribution was significant $F(1,23) = 677.93$, $p < .001$, $\eta^2 = .967$. The first distribution had a significantly smaller deviation (M = 55.26ms, SD = 21.17), than the second distribution (M = 179.54ms, SD = 57.73). The main effect of TP was not significant, $F(2,46) = 2.06$, $p = .14$, $\eta^2 = .082$. The main effect of musical IOI was significant, $F(2,46) = 135.85$, $p < .001$, $\eta^2 = .855$.

The interaction between TP and distribution was significant, $F(2,46) = 5.48$, $p = .01$, $\eta^2 = .192$. The first distribution deviations decreased linearly as the TP increased, $F(2,46) = 5.63$, $p = .01$, $\eta^2 = .197$. The second distribution deviations in fixation also significantly varied across TP conditions, $F(2,46) = 3.64$, $p = .03$, $\eta^2 = .137$, although only the medium and high TP conditions different significantly (Bonferroni corrected pairwise comparisons $p < .05$); see Figure 6-4.

The interaction between Distribution and Musical IOI was also significant, $F(2,46) = 191.11$, $p < .001$, $\eta^2 = .893$. The deviation of fixation durations in the second distribution increased significantly as the IOI level increased, $F(2,46) = 249.31$, $p < .001$, $\eta^2 = .916$, while the deviations in the first distribution did not, $F(2,46) = 2.02$, $p = .14$, $\eta^2 = .081$, $BF_{01} = 1.63$. The interactions between TP and IOI and the three-way interaction between distribution, TP and musical IOI were both not significant ($F < 1.6$, $p > .17$).

Mean resultant length.

As in all prior experiments, the circular uniformity of the circle transition fixation start times were tested with the Rayleigh's test (Wilkie, 1983). The proportion of participants with a significantly non-uniform response are described in Table 6-1. Fifteen participants had significantly non-uniform responses to all conditions, all other participants had either 1 (8/24) or 2 (1/24) non-circular conditions, often at the High TP 500ms IOI condition.

Table 6-1 Proportion of significantly non-uniform responses by Transitional Probability Condition and Musical IOI

	Low	**Medium**	**High**
500ms	.92	.92	.75
700ms	.96	1	.92
900ms	1	.96	.96

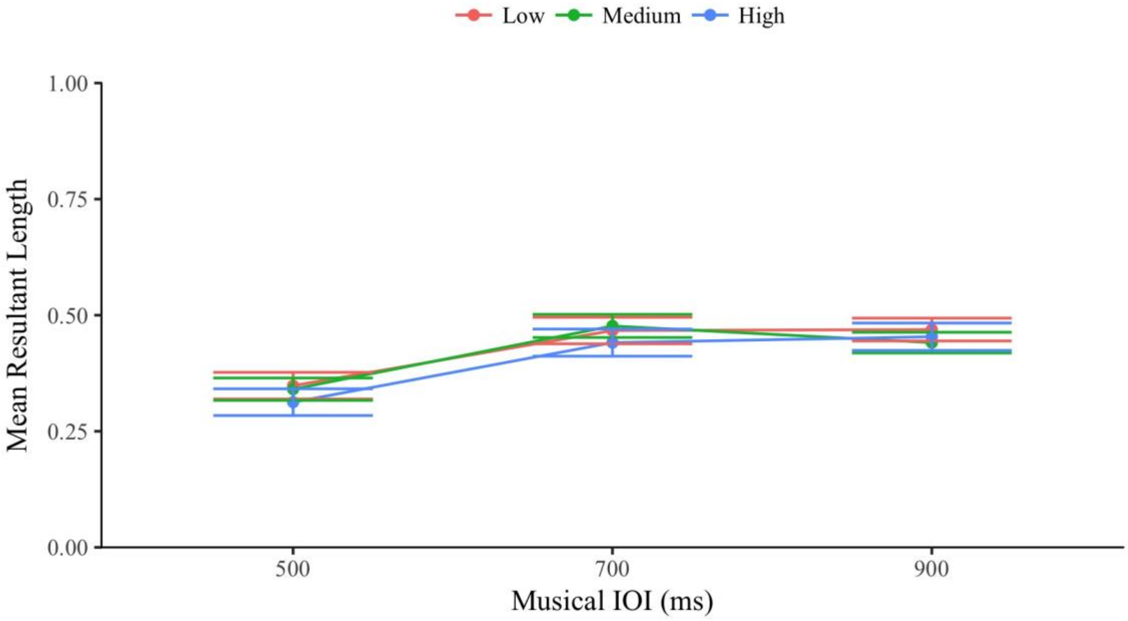


Figure 6-5 Mean Resultant Length (±1SE) values by Transitional Probability and Musical IOI

A repeated measures ANOVA of the MRL values contrasted TP and musical IOI. The main effect of TP was not significant, $F(2,46) = 1.36, p = .267, \eta^2 = .056$ ($BF_{01} = 8.253$). The main effect of IOI was significant, $F(2,46) = 57.31, p < .001, \eta^2 = .714$. As observed in Figure 6-5, the 500ms condition had significantly lower MRL values than the 700 and 900ms conditions (Bonferroni corrected, $p < .001$). The interaction between TP and IOI was not significant $F(4,92) = .557, p = .695, \eta^2 = .024$.

Circular mean.

The mean distance linear measure employed in E5 and E6 was again utilised in this experiment, in order to contrast the predictive nature of eye movements between the TP and IOI conditions. As observed in the circular plots of Figure 6-6, the majority of participants had a circular mean prior to the IOI onset (before 0). As the TP levels increase some participants emerge as reactive rather than predictive in response.

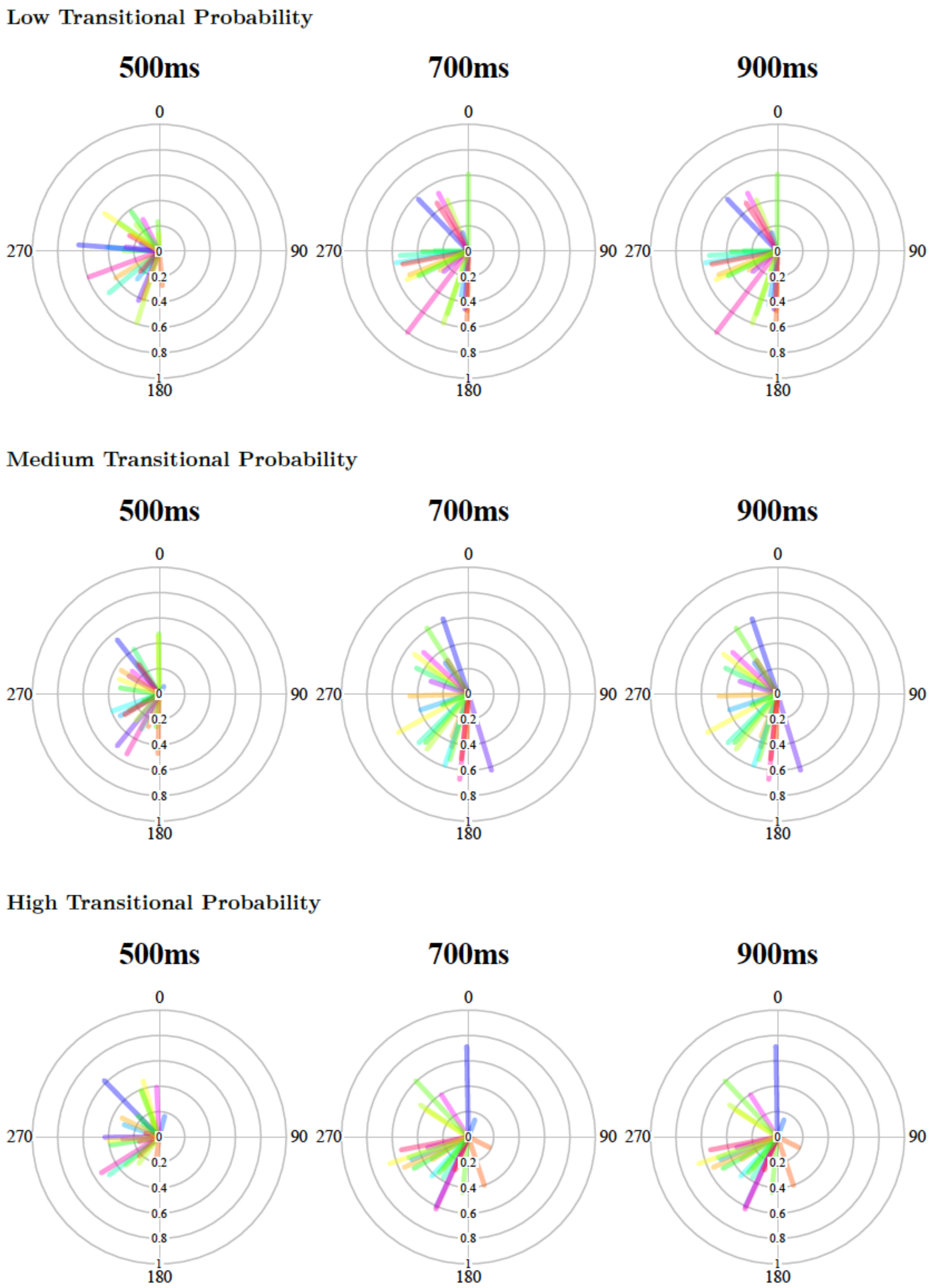


Figure 6-6 The Circular Mean (vector orientation) and Mean Resultant Length (vector length, the radius is MRL =1) by Transitional Probability Condition and Musical IOI

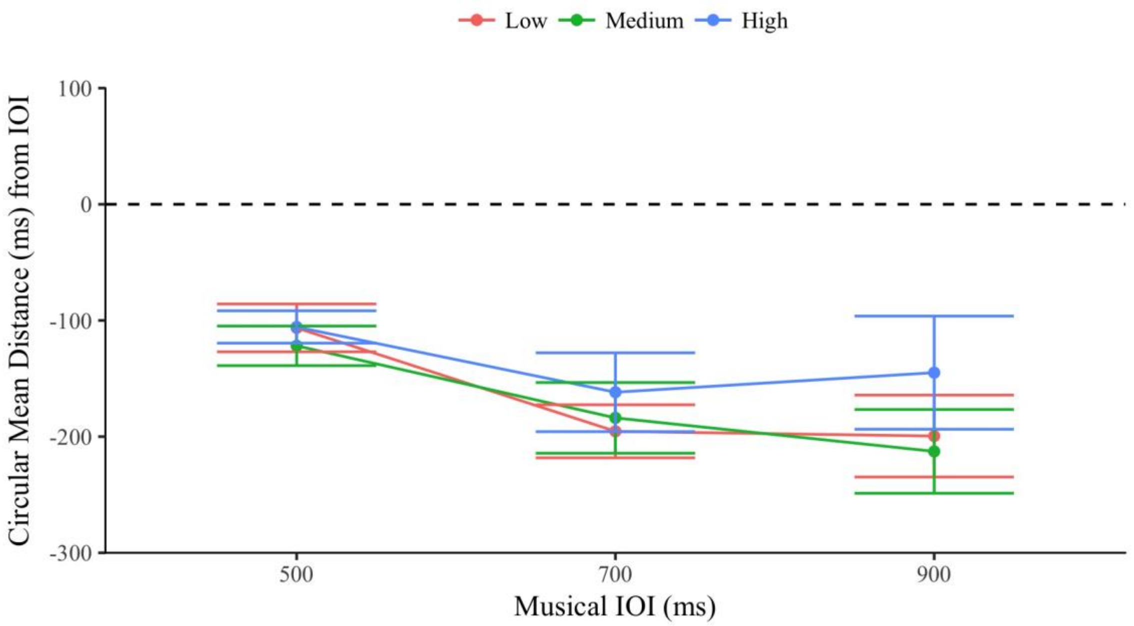


Figure 6-7 Mean Distance (ms) of the Circular Mean from the IOI by Transitional Probability and Musical IOI (±1SE)

A repeated measures ANOVA of the mean distance was contrasted across TP and musical IOI conditions. The main effect of TP was not significant, $F(2,46) = 1.88$, $p = .164$, $\eta^2 = .076$ ($BF_{01} = 5.604$). The main effect of IOI was significant, $F(2,46) = 5.607$, $p = .007$, $\eta^2 = .196$. The 500ms IOI condition was significantly closer to the IOI than the 700 and 900ms conditions (Bonferroni corrected, $p < .003$). Despite the high TP at 900ms showing a reduced distance (see Figure 6-7), the interaction between TP and IOI was not significant, $F(4,92) = .452$, $p = .771$, $\eta^2 = .019$.

Correlations between direct control and synchronisation.

In this experiment, there are two factors acting on the direct control of eye movements: the transitional probability differences (i.e. saccade reversal uncertainty) and the volitional intention to control when the eyes move. The previous experiments (E5 and

E6) found a significantly predictive relationship between the *lambda* in the second (directly controlled) fixation duration distribution and MRL when the only factor was volitional control.

A linear regression was calculated to predict MRL performance of eye movements with *lambda*. The association was not significant, $F(1,22) = 2.30$, $p = .144$, $R^2 = .095$. As previously mentioned there are two factors acting on the direct control of eye movements. When the TP conditions are included in the regression model as an interacting predictor with lambda on MRL, the regression was again not significant, $F(5,68) = 1.49$, $p = .204$, $R^2 = .102$. The TP levels did not significantly influence MRL performance, although when their interaction with *lambda* (a negative influence) is accounted for in the model, the main effect of *lambda* was a significant predictor ($p = .016$). Participants' MRL increased by 0.555 when their lambda increased by 1.

6.5 Discussion

A motivation for this study was to measure whether increasing the probability of changes in saccade direction, a known delaying effect on eye movement timing (Klein, 2000; Smith & Henderson, 2009), also increased the capacity for, and temporal control of, volitional eye movements. The initial prediction was that an increase in transitional probability (TP) would increase the proportion of fixations subject to inhibition. The results support this prediction, as there was a linear increase in the proportion of fixations in the delayed second distribution as the probability of direction change increased (i.e. saccade direction uncertainty increased inhibition). As the lambda measure represented the second distribution, which had a linear delay in response to the musical IOI levels, an increase in lambda subject to TP demands is also evidence that increasing inhibition facilitated some further delay to these movements (e.g. volitional movements that delay to match a musical IOI). Although, the deviation of these distributions increased as TP increased, i.e. the temporal control was not the same. In sum, the period of inhibition initiated by uncertainty in saccade direction was further inhibited by volitional influence.

The second prediction was that the TP and musical IOI inhibitory influences would interact, cumulatively increasing both the proportion of fixation durations in the second distribution (lambda), and increasing the temporal control of saccades (higher MRL values). The results did not support this prediction. Whilst both a slowing of the musical IOI and an increase in TP did increase the proportion of lambda independently, the two effects did not interact. Contrasting the lambda values with the previous experiments, this study had a slightly higher maximum mean lambda (55%) than E5 and E6 (50%). Interestingly, the mu of the second distribution was further from the target IOI and the sigma was larger as the TP increased. The TP demands did increase the number of fixations subject to inhibition, but the increased control did not increase the level of synchronisation achieved, either in the predictive response (mean distance) or in the synchronisation level (MRL). Therefore, the TP direction uncertainty increased the proportion of inhibited eye movements, but detracted from rather than enhanced the temporal control of them.

The final prediction was regarding the first distribution parameters (mu and sigma) as a representation of the saccadic timer. The ICAT model (Trukenbrod & Engbert, 2014) proposes a global component that shifts the underlying distribution of the saccadic timer to accommodate task demands, utilising accumulated evidence from previous fixations (an indirect process), i.e. the global shifts in distribution are argued to begin the random-walk further from threshold as task demands increase. This global component ultimately serves as a secondary inhibition processes (in addition to inhibition in the random-walk), as the speeding of eye movements is limited by physiological constraints (e.g. saccade programming and eye-brain lag). The final prediction was that the mu of the first distribution (as an estimation of the distribution of the saccade timer) would increase as the TP increased if the global component was modulating the distribution of the saccade timer. The mean (i.e. mu) of the first fixation duration distribution significantly shortened as the TP task demands increased (but not in relation to musical IOI). The deviation (i.e. sigma) of this first distribution also decreased as the task demands increased. This distribution shift was in the opposite direction predicted by an adaptive global component as described in the ICAT model. A simple interpretation of the shortening and narrowing of the first distribution, is that these narrowed values better represent the profile of the invariant saccade timer. The reduction in sigma and speeded mu suggest that more fixations from the positive tail of the first distribution (slower than the mean) were inhibited (incorporated in the second distribution) as the TP uncertainty increased. Therefore, whilst the first distribution did shift as a response to saccade direction changes, the reduction in mu and sigma when TP increased is arguably a further refinement of the estimation of the distribution underlying the saccade timer, which rather than being adaptive appears optimally fast and invariant as described in CRISP (Nuthmann et al., 2010). Although, some caution in this interpretation should be maintained as these are relatively small shifts of model-estimated parameters.

These results present two findings of relevance for eye movement timing models. Firstly, as with the SOD paradigms (Henderson & Pierce, 2008; Henderson & Smith, 2009), these data support mixed-control. The subpopulation of quick eye movements that were not inhibited by volitional intent or TP uncertainty is relatively large (45%), with a mean of 245ms and a deviation of 51ms (in the High TP, 900ms condition). These estimates of visually invariant fixations are much larger than those proposed by the competing maverick component in LATEST (Tatler et al., 2017), and does not support the global component proposed in ICAT (Trukenbrod & Engbert, 2014). This paradigm is arguably an excellent candidate for modelling direct-control limitations in active vision. Secondly, direct-control as an inhibitory process differs in temporal precision between that exerted by voluntary control and that from visual factors (e.g. saccade direction uncertainty). The synchronisation of eye movements was predicted by volitional control, and not by the sum of inhibition from both visual and voluntary factors.

In conclusion, the proportion of fixations that are subject to direct control, either volitionally controlled or in response to task demands, were notably limited. This suggests that there is a relatively large subpopulation of eye movements are automatic, and not related to task demands (as described by mixed-control). The novel TP task component in this study did increase the number of inhibited fixations, but this increased proportion subject to direct control did not facilitate better synchronisation or closer estimation of the IOI with eye movements. This may be because the uncertainty in saccade direction detracted from a limited resource. Alternatively, those movements subject to volitional control may be selectively limited and the increased proportion of inhibited movements produced by TP uncertainty were not as accessible to volitional influence.

Chapter 7

General Discussion

The overarching aim of this thesis was to measure auditory influences on eye movement timing. Prior research has identified robust integration effects of audiovisual information influencing perception (e.g. Driver & Spence, 1998), yet research that investigates how auditory information influences the timing of eye movements as mediators of this integration, are sparse. The specific auditory influence investigated within this thesis was musical beats at different inter-onset intervals (IOI). By employing clearly accented and predictable onsets in time, the experiments within the thesis were able to measure more than simple presence or absence effects, but also the temporal correspondence (synchronisation and prediction of onsets) and period shifts (latency and fixation duration variance). The core questions addressed within the thesis considered firstly, whether musical beats implicitly influence the timing of eye movements, as a change in period or alignment in phase. Secondly, whether eye movements can explicitly synchronise with simple musical beats when aligning self-produced (gaze-contingent) visual onsets. Finally, whether direct control of eye movements is limited to a sub-population of eye movements, and if those movements that can be controlled are temporally precise.

The literature to motivate this research (reviewed in Chapter one), was broadly from two research fields: eye movement timing and sensorimotor synchronisation (SMS). The SMS literature detailed the capacity for humans to entrain movements to external rhythms, the rate and modality limitations of both motor and oculomotor movements to music, and what behaviours indicate and facilitate synchronisation. The overview of eye movement timing research focused on what influences *when* the eyes move; not only the physiological constraints, but approaching oculomotor timing as an adaptive system that represents the perceptual complexities of the environment. In identifying the effectors of eye movement timing (e.g. task differences, visual complexity, and visual salience), as identified in eye movement timing models, two key factors stand out. Firstly, how these effects are accounted for within proposed mechanisms of the current eye movement timing models differs greatly. Secondly, distinctions in direct control mechanisms (e.g. volition, visual factors, task demands) within these models account only for the eye movement behaviour they model (e.g. reading or static scene viewing), but have limited scope for volitional control irrespective of visual content or the temporal dynamics in visual onsets. Within this discussion, the main findings of the previously discussed research will be summarised as a provocation for adaption or refinement to the current eye movement timing models, but also as evidence for weak direct control of eye movements i.e. a dominant automatic and invariant saccadic timer that limits the capacity for moment to moment direct-control of eye movements. Auditory influences and temporal precision of volitional control will be discussed in addition to limitations of the current research. The future directions for this research will focus on dynamic viewing behaviour and the temporal dynamics of attention and eye movements.

7.2 Summary of Core Findings

The first two studies identified no implicit effect of music on the timing of gaze-contingent eye movements. The circular visual search paradigm did show systematic oculomotor effects of saccade direction (downward saccades were slower), but compelling evidence of invariance to both the presence of music and changes in the interval between the musical beats. Following further refinements to the visual search paradigm, the elliptical visual search (E3 and E4a) was highly successful in producing systematically quick eye movements that were subject to similar delays throughout the circles in the sequence. Yet again there was no evidence of implicit entrainment of eye movements to music. Even when the period of the music was very close to the period of eye movements and only minor phase alignments were required to synchronise, the gaze-contingent eye movements did not entrain. Musical beats did not implicitly entrain gaze-contingent eye movements.

In the light of the lack of any implicit musical influence on eye movements, experiment 3 also tasked explicit synchrony and further contrasted eye movements with a commonly measured SMS movement, the finger-tap. This modification was to confirm that sequences of eye movements could synchronise with intention (Leow et al., 2017). Explicit control of eye movements at very fast intervals was exceptionally poor (although the musical IOI were within one standard deviation eye movement latencies during the task), neither making the required small shifts in period to match the musical IOI, or synchronising in phase (unlike finger-taps completing the same task which were more easily controlled in time). The inability to resolve small temporal asynchronies between gaze-continent onsets and an auditory rhythm suggest a limitation in either (or both) the perception of the asynchrony, or in the temporal control of the eye movements. There is evidence that both factors may have influenced performance (as discussed in Chapters 3 and 4): temporal information is imprecise across saccades (Yarrow et al., 2001), self-awareness of fixation behaviour is limited (Clarke et al., 2016), and not all eye movements can be directly controlled (Henderson & Pierce, 2008). In contrast, the eye movements that responded to the visual onsets produced by the synchronised tap responses (motor-produced onsets), were both more synchronised and closer to the referent IOI that when explicitly controlling them. It appears the visual system is built to predict and react to visual onsets in the world and produced by motor actions (visual transients), not produce them with guided fixations, i.e. if the relationship was a dance, eye movements would follow. The effect of corresponding musical beats to exogenous visual onsets was tested in experiment 4B. Interestingly, audiovisual correspondences in exogenous onsets significantly increased eye movement latencies by three milliseconds, an effect not present when music accompanied random visual sequences. This effect is arguably not a presence effect, but rather a selective response to AV correspondence, likely a product of increased cognitive load. Importantly, it is evidence of auditory influence on eye movement timing, even in the absence of entrainment to the auditory beat.

The final three experiments discussed in the thesis measured the ability to directly control eye movements precisely in time to much slower musical IOI than the prior experiments. The adaption of the visual search paradigm to enforce timed shifts between the circles, produced a bi-modal fixation duration distribution much like that described in the SOD paradigms (Henderson & Pierce, 2008; Morrison, 1984). These studies showed that a sub-population of fixations (50-55% of all fixations) were inhibited with a mean delay increasing as the referent musical IOI slowed. The remaining sub-population of fixations were relatively short in duration (245-250ms), relatively invariant across condition, and arguably represent the distribution of the stochastic saccade timer, for example that described in CRISP (Nuthmann et al., 2010). Finally, individual differences in ability to voluntarily inhibit fixations (not all direct-control inhibition) predicted synchronisation performance. In the following subsection, these results will be discussed as evidence for the contrasting eye movement timing models.

7.3 Eye Movement Timing Models

Within the introduction to this thesis prominent eye movement timing models were contrasted to identify different mechanisms as candidates for auditory influence. Summarised within section 1.2.8 the key models differed across four key factors. First, what mechanism accounts for apparent randomness underlying eye movement distributions (e.g. a saccade timer or maverick component). Secondly, the existence of a two-stage saccade program immediately preceding a saccade, although the existence of this program was not considered within the thesis. Thirdly, how top-down influences are conceptualised to modulate fixations (e.g. immediately on all fixations with *direct* control, a delayed estimation of demand with *indirect* control, and as a limited immediate response with *mixed* control). Finally, the existence of any accommodation of a volitional influence or capacity to move the eyes irrespective of visual or task demands. In the following section these differences (with the exception of the saccade program) will be discussed in light of the findings from the thesis.

7.3.1 Source of randomisation.

The emergence of a minimally variant and speeded subpopulation within fixation duration distributions in E5 to E7, is arguably strong evidence for a stochastic saccade timer as described in the SWIFT and CRISP models (Engbert et al., 2005; Nuthmann et al., 2010). The parameters of this distribution observed in this thesis (a mean of 245ms, a standard deviation of 50ms, and accounting for around 45% of fixations), suggest a pervasive automatic process that competes with direct control mechanisms. This pervasive presence of the saccade timer at shorter intervals likely explains the absence of explicit synchrony in the earlier experiments. An initial prediction was that the saccade timer could be susceptible to auditory entrainment. The evidence from the later studies in the thesis show that this distribution may shift in response to external factors, i.e. the idea that musical beats could be a factor is not implausible. The saccade timer rather than a constant or an isochronous program that is entrained by predictable events, appears inherently random around an optimised interval (e.g. 245ms) and likely an adaptive biological process, as described by Carpenter (1999), that protects vision from such entraining effects.

The LATEST model (Tatler et al., 2017) also has a representation of a visually invariant competing sub-component, but limits its scope to around 10% of fixations and those involving much shorter durations (less than 100ms). Whilst it is plausible that the distribution parameters during the gaze-contingent tasks employed in this thesis over-represented the proportion of fixations and the mean duration of the saccade timer, the estimation is arguably not out by such a large margin (e.g. 35%). The invariant and stochastic timer was observed as a persistent and dominant presence in the fixation data.

As discussed in chapter 6, the ICAT model (Trukenbrod & Engbert, 2014) proposes a global component that shifts the distribution of the saccade timer as an indirect estimation of demand. The fixation durations estimated within the first distribution in the final experiment (E7), did not slow as the uncertainty over saccade direction increased as predicted by a global estimator, rather the fixations in the first distribution shortened in duration and deviation. This was interpreted as increased inhibition demands further refining the parameters of the invariant saccade timer. Although, in the two earlier volitional control experiments the first distribution did shift slowing as the IOI level slowed. Although this shift is likely an estimation error and the deviations of the first distribution also increased, which would not occur if the distribution shifted. Finally, the saccade timer was impervious to the entraining influence of musical beats.

7.3.2 Top-down influence.

With the exception of the indirect global estimation component in ICAT model (Trukenbrod & Engbert, 2014), top-down influence on fixation durations is reflected across the reviewed oculomotor timing models as a direct process (immediately altering fixation durations subject to the currently processed information). How this process is conceptualised and the limitations of control differ between the models. In SWIFT, CRISP and ICAT (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod & Engbert, 2014), direct control is limited to a subpopulation of fixations (mixed control) and is described as an inhibitory process, i.e. top-down factors can only *delay* (not speed) fixation durations. In experiment 3, participants were able to significantly speed their fixations when attempting to synchronise to the fastest IOI condition (Explicit 264ms), producing fixation durations faster than observed during the implicit task. This data suggests that top-down influence (e.g. intent) can speed fixations, arguably through some down-regulating of the level of cognitive processing (although this was not tested behaviourally, e.g. reduced accuracy). The ability to speed fixation behaviour is more easily conceptualised in the LATEST model, where the ‘go’ decision (in a stay-go evaluations) could be enhanced by demoting the emphasis to stay. As described in the previous section on the saccade timer, the results in this thesis evidence clear limitations in the capacity of direct control to affect each fixation duration, contrary to the direct control proposed by the more cognitive models (Rayner et al., 1998; Reichle et al., 1999; Reingold et al., 2012). There is some capacity to not only increase fixation durations (e.g. by engaging deeper cognitive processing), but also to intentionally speed some fixations.

Musical beats as a top-down (direct control) influence was limited to a small but constant increase in fixation durations when responding to audiovisual onsets (as opposed to producing them with gaze-contingent movements). This auditory effect was produced in a highly reductive environment, where the relationship between the auditory and visual onsets was a coincidental temporal one (letter onsets do not produce music). The perceptual information of real-world representations far exceeds that of basic temporal correspondence of motion and sound. For example, the creaking of a tree branch swaying during a storm adds a rich context to the visual percept of its motion. Were audiovisual correspondences of real-world objects incorporated into eye movement timing models, much greater top-down influence would be predicted than that identified in the simplified stimuli of this thesis. Sources of sound are by nature of their production, intertwined with motion. The interaction of audiovisual information can both modulate the ambiguity and relevance of objects in the world (Coutrot & Guyader, 2013), which would likely be represented in the timing to them (both *when* they fixate and the *duration* of the fixations). In omitting visual dynamics (motion) and sound correspondences the temporal dynamics of attention as a predictive force guiding eye movements through time are lost, or simply unaccounted for. A notable example of audiovisual influence is the 'attentional synchrony' of multiple viewers gaze behaviour, clustering to key points in time when viewing audiovisual dynamic scenes (Mital et al., 2010; Smith & Mital, 2013). Temporal dynamics of the audiovisual stimuli both constrain and increase predictable viewing behaviours, effects currently unaccounted for in eye movement timing models.

7.3.3 Volitional control.

A distinction in eye movement behaviour identified in the Findlay and Walker's framework (Findlay & Walker, 1999), was the capacity to volitionally exert control on eye movements, orienting or holding them at fixation irrespective of visual information to be perceived or the cognitive processing demands (level 5 in the information flow, see section 1.2.2). Arguably, it was this ability (deemed 'uncommon' in the Findlay and Walker model) that produced much of the eye movement data analysis in this thesis. A novel finding identified in this thesis was that the individual differences in the ability to volitionally inhibit sequential eye movements predicted synchronisation ability. It was surprising that the addition of auditory feedback tones did not further increase this association, considering the proposed close relationship between the temporal processing of sound and SMS ability (Tierney & Kraus, 2013a; 2016). A plausible interpretation is that the synchronisation of eye movements was the product of individual differences in attentional control, i.e. the ability to inhibit shifts of attention and the addition of auditory tones did not aid this. Importantly for future eye movement timing models, volitional control as shown in this thesis can have a sizable effect on the timing of eye movements, both speeding and slowing them irrespective of the visual information demands. This capacity if not factored for in future oculomotor timing models will misclassify fixation durations as responding to currently fixated visual information when they are not.

7.4 Limitations and Future Directions

Much of this thesis has considered eye movement timing and control during gaze-contingent tasks. The advantage of gaze-contingent visual onsets is that the temporal dynamics are entirely self-paced; the limitation is the absence of visual transients. There were notable differences in the timing of eye movements to exogenous onsets (compared to endogenous), the visual system clearly utilises visual transients to predict temporal demands as synchronisation was enhanced in these conditions. In quantifying the temporal control of eye movements during the final three reported experiments it is plausible that the informative nature of visual transients was omitted, inadvertently decreasing some of the participants' synchronisation ability. It is also plausible the auditory binding of temporal correspondences engages visual transients in the parafoveal visual field, rather than the centre of fixation. The prevalence of motion sensitive photoreceptive cells (rods) increases with eccentricity from the foveal region (Polyak, 1941). Considering the close perceptual relationship between motion and sound, it is plausible that temporal integration of audiovisual information is informed by the (increasingly) motion sensitive regions within the visual field. As such the auditory influence may have been limited in the gaze-contingent studies, due to the lack of visual transients. Future studies could contrast the temporal control of attention shifts both as exogenous and endogenous control to auditory onsets.

The dominance of a visual or task invariant saccade timer questions the close temporal relationship between attention and eye movements. A common misconception in vision research is that the locus of visual attention and where (and when) eye movements orient is synonymous (rather than correlated). This assumption is based on research that shows covert and overt shifts of attention correspond with eye movements (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1996; Martinez-Conde et al., 2013). Considering the dominance of the saccade timer in this data, i.e. eye movements that are relatively short and not representing visual or task demands on the current fixation, the temporal relationship between shifts of visual attention (covert or overt) and gaze shifts may be limited. The saccade timer likely is an efficient way of keeping attention moving, but as these visually invariant movements do not reflect cognitive evaluations of the environment, this subpopulation of quick fixations may at times be in opposition to the attentional command to stay or shift elsewhere (e.g. anti-saccade tasks and the double-step paradigm). Future work that co-registers electroencephalography (EEG) measures of attentional shifts, for example the n2pc component (Eimer, 1996), with eye movements could address the temporal dynamics and limitations of this relationship.

7.5 Conclusion

The intention of this thesis was to measure auditory influence on eye movement timing. In addressing this topic, a novel visual search paradigm was generated, which facilitated a high level of control over visual and task influences on eye movement timing. A notable success of the research within the thesis was the effectiveness of the eye movement control exerted within this paradigm, and the flexibility it afforded to manipulate visual factors (e.g. saccade direction). In utilising variations in this paradigm this thesis has identified a novel quantification of the limitations of direct control, evidence for a predictive relationship between volitional inhibition of eye movements and temporal precision, and some influence of audiovisual correspondence on eye movement timing.

These findings have broader implications for psychological research. Eye movement timing is currently studied as an insight into broader cognitive traits, including executive function (e.g. go-no-go tasks and delayed orientation responses) and inhibitory control (e.g. anti-saccade tasks). Yet these applications and the measures they employ are arguably omitting two competing control mechanisms acting on eye movements (the saccade timer and direct control). This thesis argues that the current eye movement timing models are incompletely accounting for oculomotor behaviour. The existing models inform current predictions regarding eye movement behaviour when reading or viewing a scene, yet their interpretation would likely misrepresent how people shift their gaze. The evidence for individual differences in direct-control within the thesis has important implications for the broader study of individual differences, including developmental studies, clinical studies (e.g. attention deficit hyperactivity disorder, anxiety, autism spectrum disorder), and within the typical population. Future studies could explore neurological and genetic factors that account for differences in the ratio of direct control exerted on fixation durations by utilising this paradigm.

References

Adeli, H., Vitu, F., & Zelinsky, G. J. (2017). A Model of the Superior Colliculus Predicts Fixation Locations during Scene Viewing and Visual Search. *Journal of Neuroscience*, *37*(6), 1453–1467.

Agostinelli, C., & Lund, U. (2017). R package "circular": Circular Statistics (version 0.4-9.3). Retrieved from: https://www.rdocumentation.org/packages/circular

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, *59*(4), 3677–3689.

Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, *204*(1), 91–101.

Batten, J. P., & Smith, T. J. (2018). Looking at Sound: Sound Design and the Audiovisual Influences on Gaze. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), *Seeing into Screens* (pp. 85–102). New York: Bloomsbury Publishing USA.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967–983.

Benaglia, T., Chauveau, D., Hunter, D., & Young, D. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, *32*(6), 1–29.

Bégel, V., Benoit, C.-E., Correa, A., Cutanda, D., Kotz, S. A., & Bella, S. D. (2017). "Lost in time" but still moving to the beat. *Neuropsychologia*, *94*(C), 129–138.

Borji, A., & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. USA: MIT Press.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722.

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., & Toiviainen, P. (2014). Hunting for the beat in the body: on period and phase locking in music-induced movement. *Frontiers in Human Neuroscience*, *8*(183), 903.

Carpenter, R. H. (2001). Express saccades: is bimodality a result of the order of stimulus presentation? *Vision Research*, *41*(9), 1145–1151.

Carpenter, R. H. S. (1999). A neural mechanism that randomises behaviour. *Journal of Consciousness Studies*, *6(1),* 13-13.

Carpenter, R. H. S., Reddi, B. A. J., & Anderson, A. J. (2009). A simple two-stage model predicts response time distributions. *The Journal of Physiology*, *587*(16), 4051–4062.

Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, *21*(4), 515–532.

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision research*, *102*, 41-51.

Clarke, A. D., Mahon, A., Irvine, A., & Hunt, A. R. (2016). People are unable to recognize or report on their own eye movements. *Quarterly Journal of Experimental Psychology (2006)*, 1–20.

Coutrot, A., & Guyader, N. (2013, August). Exploration of dynamic natural scenes: influence of unrelated soundtracks on eye movements. In *17th European Conference on Eye Movements (ECEM 2013).*

Coutrot, A., Guyader, N., Ionescu, G., & Caplier, A. (2014). Video viewing: do auditory salient events capture visual attention? *Annals of Telecommunications*, *69*(1-2), 89–97.

Day, R.-F., Lin, C.-H., Huang, W.-H., & Chuang, S.-H. (2009). Effects of music tempo and task difficulty on multi-attribute decision-making: An eye-tracking approach. *Computers in Human Behavior*, *25*(1), 130–143.

De Freitas, J., Liverence, B. M., & Scholl, B. J. (2014). Attentional rhythm: a temporal analogue of object-based attention. *Journal of Experimental Psychology: General*, *143*(1), 71–76.

de Urabain, I. R. S., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, *134*, 43–59.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Drewing, K. (2013). Delayed auditory feedback in repetitive tapping: A role for the sensory goal. *Quarterly Journal of Experimental Psychology*, *66*(1), 51-68.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254–262.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*(4), 777–813.

Escoffier, N., Herrmann, C. S., & Schirmer, A. (2015). Auditory rhythms entrain visual processes in the human brain: Evidence from evoked oscillations and event-related potentials. *NeuroImage*, *111*(C), 267–276.

Escoffier, N., Sheng, D. Y. J., & Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica*, *135*(1), 12–16.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Findlay, J. M., & Gilchrist, I. D. (2003). Active Vision. Oxford University Press, USA.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(04), 661–674.

Fischer, B., & Boch, R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Research*, *260*(1), 21–26.

Fisher, N. I. (1995). Statistical Analysis of Circular Data. New York: Cambridge University Press.

Flach, R. (2005). The transition from synchronization to continuation tapping. *Human Movement Science*, *24*(4), 465–483.

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: effects of image type and image features. *Vision Research*, *50*(8), 779–795.

Fraisse, P. (1982). The Psychology of Music. In D. Deutsch (Ed.), (pp. 149–180). New York: Academic Press Inc.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495.

Fuchs, A. F., Kaneko, C. R. S., & Scudder, C. A. (1985). Brainstem control of saccadic eye movements. *Annual review of neuroscience*, *8*(1), 307-337.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, *15*(3), 566–573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, *17*(6-7), 1055–1082.

Henderson, J. M., Olejarczyk, J., Luke, S. G., & Schmidt, J. (2014). Eye movement control during scene viewing: Immediate degradation and enhancement effects of spatial frequency filtering. *Visual Cognition*, *22*(3-4), 486–502.

Henderson, J. M., Weeks, P. A., Jr, & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.

Honda, H., & Findlay, J. M. (1992). Saccades to targets in three-dimensional space: Dependence of saccadic latency on target location. *Perception & Psychophysics*, *52*(2), 167–174.

Hooge, I. T. C., Over, E. A. B., van Wezel, R. J. A., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*(14), 1901–1908.

Hooge, I. T., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception & Psychophysics*, *58*(7), 969–976.

Hooge, I., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, *38*(9), 1295–1302.

Hornof, A. J., & Vessey, K. E. V. (2011). The Sound of One Eye Clapping: Tapping an Accurate Rhythm With Eye Movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1225–1229.

Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. H. (2012). Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome, *77*(4), 388–398.

Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1525–1534.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience, 2(3),* 194.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Transactions on pattern analysis and machine intelligence, 20(11), 1254-1259.*

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75.

Joiner, W. M., & Shelhamer, M. (2006). An internal clock generates repetitive predictive saccades. *Experimental Brain Research*, *175*(2), 305–320.

Joiner, W. M., Lee, J.-E., Lasker, A., & Shelhamer, M. (2007a). An internal clock for predictive saccades is established identically by auditory or visual information. *Vision Research*, *47*(12), 1645–1654.

Joiner, W. M., Lee, J.-E., Lasker, A., & Shelhamer, M. (2007b). An internal clock for predictive saccades is established identically by auditory or visual information. *Vision Research*, *47*(12), 1645–1654.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kato, M., & Konishi, Y. (2006). Auditory dominance in the error correction process: a synchronized tapping study. *Brain Research*, *1084*(1), 115–122.

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, *68*(1), 17–26.

Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*(6181), 430–431.

Klein, R. M. (2000). Inhibition of Return. *Trends in Cognitive Sciences*, *4*(4), 138–146.

Koelsch, S., Fritz, T., & Schlaug, G. (2008). Amygdala activity can be modulated by unexpected chord functions during music listening. *Neuroreport*, *19*(18), 1815.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology (2006)*, *63*(11), 2220–2230.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The Role of Attention in the Programming of Saccades. *Vision Research*, *35*(13), 1897–1916.

Kraus, N. (2011). Musical training gives edge in auditory processing. *The Hearing Journal*, *64*(2), 10–12.

Krause, V., Pollok, B., & Schnitzler, A. (2010). Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychologica*, *133*(1), 28–37.

Kuribayashi, R., & Nittono, H. (2014). Speeding up the tempo of background sounds accelerates the pace of behavior. *Psychology of Music*, 1–10.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59.

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*(4), 527–566.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119.

Leow, L.-A., Waclawik, K., & Grahn, J. A. (2017). The role of attention and intention in synchronization to music: effects on gait. *Experimental Brain Research*, *38*, 830.

Loehr, J. D., Large, E. W., & Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1292–1309.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572.

Ludwig, C. J. H., & Gilchrist, I. D. (2003). Target similarity affects saccade curvature away from irrelevant onsets. *Experimental Brain Research*, *152*(1), 60–69.

Ludwig, C. J. H., Farrell, S., Ellis, L. A., & Gilchrist, I. D. (2009). The mechanism underlying inhibition of saccadic return. *Cognitive Psychology*, *59*(2), 180–202.

Luke, S. G., Smith, T. J., Schmidt, J., & Henderson, J. M. (2014). Dissociating temporal inhibition of return and saccadic momentum across multiple eye-movement tasks. *Journal of Vision*, *14*(14), 9–9.

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, *144*(2), 257–263.

Marr, D. (1982). Vision. San Francisco: W. H. Freeman and Company.

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*(2), 83–96.

Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence. *Biological Cybernetics*, *70*(5), 463–473.

Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, *6*(4), 332–340.

McAuley, J. H., Rothwell, J. C., & Marsden, C. D. (1999). Human anticipatory eye movements may reflect rhythmic central nervous activity. *Neuroscience*, *94*(2), 339–350.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 747–748.

McNeill, W. H. (1995). Keeping Together in Time. Cambridge: Harvard University Press.

Mendonça, C., Oliveira, M., Fontes, L., & Santos, J. (2014). The effect of instruction to synchronize over step frequency while walking with auditory cues on a treadmill. *Human Movement Science*, *33*(C), 33–42.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2010). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation*, *3*(1), 5–24.

Moelants, D. (2002). Preferred Tempo Reconsidered. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwich (Eds.), (pp. 580–583). Presented at the Music Perception and Cognition, Sydney, 2002.

Moens, B., Muller, C., van Noorden, L., Franěk, M., Celie, B., Boone, J., et al. (2014). Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music. *PLoS ONE*, *9*(12), e114234.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 667–682.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218–228.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, *9*(2), e89642.

Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews*, *64*, 229–251.

Nuthmann, A. (2016). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*(2), 370–392.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405.

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, *163*(2), 226–238.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*(183), 1-14.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*(3), 533–546.

Phillips-Silver, J., & Trainor, L. J. (2008). Vestibular influence on auditory metrical interpretation. *Brain and Cognition*, *67*(1), 94–102.

Poghosyan, V., & Ioannides, A. A. (2007). Precise mapping of early visual responses in space and time. *NeuroImage*, *35*(2), 759–770.

Polyak, S. L. (1941). The Retina. Oxford, England: Chicago Press.

Posner, M. I., & Cohen, Y. (1984). Components of Visual Orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X* (pp. 531–556). London.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience*, *18*(6), 2188–2199.

Quigley, C., Onat, S., Harding, S., & Cooke, M. (2008). Audio-visual integration during overt visual attention. *Journal of Eye Movement Research 1(2),* 1-17.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*(8), 829–837.

Rayner, K., Reichle, E. D., Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 243–393). Oxford: Elsevier.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*(6), 834–839.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using EZ Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological Review*, *119*(1), 155–185.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445-476.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, *39*(26), 4403–4411.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*(2), 177–206.

Repp, B. H. (2003). Rate Limits in Sensorimotor Synchronization With Auditory and Visual Sequences: The Synchronization Threshold and the Benefits and Costs of Interval Subdivision. *Journal of Motor Behavior*, *35*(4), 355–370.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992.

Repp, B. H. (2011). Rate Limits of On-Beat and Off-Beat Tapping With Simple Auditory Rhythms: 1. Qualitative Observations. *Music Perception*, *22*(3), 479–496.

Repp, B. H., & Doggett, R. (2007). Tapping to a Very Slow Beat: A Comparison of Musicians and Nonmusicians. *Music Perception*, *24*(4), 367–376.

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness. *The Quarterly*

Journal of Experimental Psychology. a, Human Experimental Psychology, *57*(3), 499–521.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28(5),* 1085-1099.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452.

Repp, B. H., London, J., & Keller, P. E. (2008). Phase Correction in Sensorimotor Synchronization with Nonisochronous Sequences. *Music Perception*, *26*(2), 171–175.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The Disappearance of Steadily Fixated Visual Test Objects*. *Josa*, *43*(6), 495–501.

Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, *22*, 241–259.

Schäfer, T., & Fachner, J. (2014). Listening to music reduces eye movements. *Attention, Perception & Psychophysics*, *77*(2), 551–559.

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755–759.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134.

Shelhamer, M., & Joiner, W. M. (2003). Saccades Exhibit Abrupt Transition Between Reactive and Predictive, Predictive Saccade Sequences Have Long-Term Correlations. *Journal of Neurophysiology*, *90*(4), 2763–2769.

Smith, T. J. (2014). Audiovisual Correspondences in Sergei Eisenstein's Alexander Nevsky: A Case Study in Viewer Attention. *Cognitive Media Theory*, 1–19.

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108.

Smith, T. J., & Henderson, J. M. (2011). Looking back at Waldo: Oculomotor inhibition of return does not prevent return fixations. *Journal of Vision*, *11*(1), 3–3.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 16–16.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics*, *73*(4), 971–995.

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: A TMS investigation. *Brain and Cognition*, *82*(2), 127–136.

Styns, F., van Noorden, L., Moelants, D., & Leman, M. (2007). Walking on music. *Human Movement Science*, *26*(5), 769–785.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2(2)*(5), 1–18.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5–5.

Thurlow, W. R., & Jack, C. E. (1973). Certain Determinants of the “Ventriloquism Effect.” *Perceptual and Motor Skills*, 1171–1184.

Tierney, A. T., & Kraus, N. (2013a). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231.

Tierney, A., & Kraus, N. (2013b). The ability to move to a beat is linked to the consistency of neural responses to sound. *Journal of Neuroscience*, *33*(38), 14981–14988.

Tierney, A., & Kraus, N. (2016). Getting back on the beat: links between auditory–motor integration and precise auditory processing at fast time scales. *European Journal of Neuroscience*, *43*(6), 782–791.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 8190–8195.

Toiviainen, P., Luck, G., & Thompson, M. (2009). Embodied Metre in Spontaneous Movement to Music. *Proceedings of the Conference of European Society for the Cognitive Sciences of Music ESCOM*, 526-530.

Traquair, H. M., & Dott, N. M. (1938). The Normal Field of Vision. In *An Introduction to Clinical Perimetry* (pp. 4–5). London.

Trukenbrod, H. A., & Engbert, R. (2014). ICAT: a computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065.

Veale, J. F. (2013). Edinburgh Handedness Inventory – Short Form: A revised version based on confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(2), 164–177.

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. W. Keele (Eds.), *Motor skills* (Vol. 2, pp. 181–262). San Diego: Elsevier.

Wallengren, A. K., & Strukelj, A. (2015). Film Music and Visual Attention: A Pilot Experiment using Eye-Tracking. *Music and the Moving Image*.

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: effects of auditory grouping on visual motion perception. *Psychological Science*, *12*(2), 109–116.

Wilkie, D. (1983). Rayleigh Test for Randomness of Circular Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *32*(3), 311–312.

Wing, A. M., & Kristofferson, A. B. (1973a). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, *14*(1), 5–12.

Wing, A. M., & Kristofferson, A. B. (1973b). The timing of interresponse intervals. *Perception & Psychophysics*, *13*(3), 455–460.

Wooding, D. S., Ruddock, K. H., & Mannan, S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*(3), 363–386.

Yang, S. N., & McConkie, G. W. (2001a). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, *41*, 3567–3585.

Yang, S. N., & McConkie, G. W. (2001b). Eye movements during reading: a theory of saccade initiation times. *Vision Research*, *41*(25-26), 3567–3585.

Yarbus, A. L. (1967). Eye Movements and Vision. (B. Haigh & L. A. Riggs, Trans.). New York: Springer Science.

Yarrow, K., Haggard, P., & Rothwell, J. C. (2008). Saccadic chronostasis and the continuity of subjective temporal experience across eye movements.

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, *414*, 302–305.

Zendel, B. R., Ross, B., & Fujioka, T. (2011). The Effects of Stimulus Rate and Tapping Rate on Tapping Performance. *Music Perception*, *29*(1), 65–78.