A Multimodal Paradigm for Investigating the Perisaccadic Temporal Inversion Effect in Vision

Leo G. Trottier (leo@cogsci.ucsd.edu)
Virginia R. de Sa (desa@cogsci.ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive
La Jolla, CA 92093-0515

Abstract

Despite the brief and discrete inputs the visual system receives from frequent sudden eye movements, we are able to act smoothly and perceive the visual world as temporally and spatially continuous. While the spatial part of this discrete-to-continuous transformation has received considerable research attention, the temporal part has only of late been given serious consideration. That we know relatively little is evidenced by the recent discovery that during a period around 50 ms before an eye movement, a pair of flashed stimuli will be perceived as having happened in the 'wrong' order (Morrone *et al.*, 2005). In this study we establish experimentally that unimodal auditory stimuli are not subject to this distortion and then use an audio-visual temporal order judgment paradigm to determine why and how this illusion occurs. By examining visual-system-caused changes to the multimodal point of subjective simultaneity we are able to discern how the processing of a single visual stimulus is altered as a function of when during eye movement generation the stimulus occurs. This research moves us closer to establishing a schedule of perceptual distortions during eye movements, which may help in answering questions of how perceptual continuity is maintained and serve to inspire and constrain computational models of this aspect of visual processing

Keywords: vision; multimodal; temporal inversion; perisaccadic perception; temporal perception; saccades; eye movements

Introduction

The stability and continuity of visual perception has puzzled thinkers at least since Plato (Grusser, 1994), and a large and productive research program has been organized around solving the problem's spatial aspect – how it is that disparate sources of information are knit into a coherent and useful whole. But this knitting takes place in and over time, and how we construct a *temporally* contiguous (or continuous) perception of the world is a question that has received considerably less attention. While it is not obvious that this issue is even logically coherent, this does not mean that it should go uninvestigated, but instead that in researching it we should concentrate on the seemingly obvious and the obviously puzzling. The 'temporal inversion illusion' of Morrone *et al.* (2005a) provides just such a problem, and in this paper we propose to examine it by using a second sensory modality (audition) as a 'temporal landmark'.

One presumed self-evident characteristic of our perception of time is that it is one-way and one-dimensional. It would thus be puzzling, to say the least, if we experienced the world 'in the wrong order'. It has been found, however, that exactly this takes place for two sufficiently brief and rapid visual stimuli if they are perceived immediately prior to making a saccade (the darting and ballistic eye movement we perform a few times each second). Morrone *et al.* (2005a) found that if both of two 8 ms flashes in different locations in the visual field were presented between 70 and 30 ms ('the inversion time window') before a saccade they would be consistently perceived as having happened in the order opposite from the one in which they were presented.

The illusion was elicited by having subjects stare at a fixation point (FP1) and perform an evoked saccade to a second fixation point (FP2) 30 degrees to the right. Some random period of time after FP2, two green bars near-equiluminant to a red background appeared in an unpredictable order. The subject was then to indicate which bar appeared first. When the interval between the two bars was less than 80 ms, and when they occurred around the inversion time window, subjects consistently reported perceiving them in the 'wrong' order (Figure 1). These results are ambiguous with regard to possible explanations; observing the effect on paired stimuli has made it difficult to discover what has happened to the processing of the stimuli individually. We believed the illusion was caused by different latencies of perception for stimuli at different stages of saccade preparation and we designed the following experiment to evaluate this hypothesis. Based on pilot data, we predicted that visual stimuli presented early in saccade preparation would be perceived more slowly than such stimuli presented later during saccade preparation. In particular, we predicted that distinct visual latencies of the early and late saccade preparation periods would cause a different inter-modality point of subjective simultaneity, since visual perceptual latency would be altered while auditory perceptual latency remained constant.

Methodological considerations

While studies of the perception of the relative timing of pairs of events have typically constrained themselves to one modality, investigating the timing of a single stimulus requires the use of a 'temporal yardstick'. In the case of brain imaging and EEG this is simply the precise timing of the equipment used. However, we can also use another sensory modality as our yardstick, provided we are confident that it will not be warped by our experimental manipulations. This confidence is difficult to come by in our proposed paradigm, as the complex integration of auditory and visual stimuli is the norm rather than the exception. In Morrone *et al.* (2005a) they test for auditory 'temporal compression' effects caused by saccades and find none. In our use of a multimodal paradigm we built on this result and determined experimentally that saccade preparation also had no effect on the perceived order of a pair of auditory stimuli (experiment one). We took care in the design of our multimodal experiment to consider and if necessary control for 'prior entry' effects, possible effects of audio-visual recalibration caused by a constant audio-visual temporal offset, the temporal 'cross-capture' of audition and vision, spatial source effects, and effects of additional auditory stimuli on saccadic latencies. These possible confounds are discussed in the appendix.

General Method

In both experiments subjects were seated approximately 30 cm from a CRT display running at 120 Hz. Eye movements were monitored using an SR Research Eyelink II camera-based eye tracker sampling the most accurate eye at 500 Hz. Each trial began with the subject staring at a 1° in diameter black fixation point (FP1) 15 degrees left of the midline projected against a 20 cd/m^2 red background (Figure 3). After a random interval of between 225 and 425 ms FP1 was extinguished and an identical fixation point was presented 15 degrees right of the midline (FP2). Subjects were instructed to look to FP2 as quickly as possible. Auditory stimuli synchronization to visual stimuli was verified and calibrated using an oscilloscope, and auditory stimuli were presented to participants using headphones.

Experiment one

Stimuli and apparatus

In the first auditory-stimuli-only control experiment, two 4 ms auditory 'clicks' (at 1000 and 1750 Hz) were presented with a mean onset time between 67 and 117 ms after the appearance of FP2. These stimuli timings were chosen to maximize the number of stimuli that occurred during the critical 70 to 30 ms pre-saccadic period identified by Morrone *et al.* under the assumption that saccadic reaction times would average 140 ms. The time between stimuli ranged evenly between 8 and 92 ms. After every trial the subject (LT) indicated using a button press which stimulus (the high pitched click or the low pitched click) occurred first. One block of 440 trials was run and analyzed.

Analysis

After data collection the mean of the two click onset times relative to the eye-tracker-provided saccade onset time was calculated. We refer to this derived measure as 'Relative Onset Time'. We then fitted psychometric curves to the data according to the subject's indication of which stimulus appeared to have happened first. Data were binned to ensure unambiguous and sufficient data existed for each plot point.

Results

Performance was perfect under all stimulus timings and all orderings, irrespective of Relative Onset Time. Most importantly, performance remained perfect even during the critical inversion time window (Figure 2).

Experiment two

Stimuli and apparatus

The target experiment was an audio-visual version of experiment one. Trial sequences were nearly identical, with the subject staring at FP1 until its offset and the onset of FP2. To mitigate problems of motor-sensory recalibration (Stetson *et al.*, 2006) and prior entry (Vibell *et al.*, 2007; Zampini *et al.*, 2005) we presented a auditory cue (4 ms, 1000 Hz) in conjunction with the visual saccade cue (i.e., onset of FP2 coincided with a auditory 'click'). After the audio-visual 'go' signal subjects were presented with one auditory (4 ms, 1750 Hz) and one visual stimulus in randomly varying orders and at randomly varying inter-stimulus intervals (uniformly among 0 to 200 ms). The visual stimulus was evenly either at the top or bottom of the screen. While it is impossible to predict exactly when the subject was going to begin his eye movement, stimuli were optimized for each subject's saccadic reaction time so that flash stimuli occurred around 50 ms before saccade onset. Each subject's task was to indicate using a button press which modality was presented first.

The task was a difficult one and potential subjects were first tested at fixation (without an eye movement) in their ability to perform multimodal temporal order judgments. Of the five potential subjects tested, only three could perform the task at fixation. We only accepted data when the subject could in testing, on the day of data collection, achieve better than 75% correct over 20 trials with inter-stimulus intervals (ISIs – time between the click and the flash) ranging evenly between 17 and 92 ms inclusive. In this paper we present two subjects' data (two males, one naïve to the purposes of the experiment); the third indicated he was not performing saccade-like eye movements and his data have been excluded. The first subject (LT) performed 16 blocks of 96 trials over multiple days, and the second (AT) performed 8 blocks of 96 trials over multiple days.

Analysis

As in experiment one we calculated offline the saccade onset time of every trial. Whereas in our first experiment and in Morrone *et al.* (2005a) *mean* onset time of both stimuli relative to saccade onset time was the value of interest, in experiment two we were only interested in the relative onset time of each trial's visual stimulus, since we assumed the timing of auditory perception was not systematically affected by its time of presentation relative to saccade onset. For experiment two, then, Relative Onset Time refers exclusively to the timing of the presentation of the visual stimulus.

Our hypothesis was that stimuli early in the saccadic window (160 to 50 ms pre-saccade) would be processed slower than those in the later window (50 to 0 ms pre-saccade). To test our hypothesis we analyzed data separately in these two windows. Trials were divided by Relative Onset Time (of their flash stimulus) into those occurring either early or late in saccade preparation. Data were then binned (13 bins, -200 to 200 ms mid-bin) by ISI in order to construct a logistic-based psychometric function. This was done using *psignifit* toolbox version 2.5.6 for Matlab (see http://bootstrap-software.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann and Hill (2001a). Confidence intervals were similarly found by the BCa bootstrap method implemented by *psignifit*, based on 4999 simulations (Wichmann and Hill 2001b). Our value of interest, the point of subjective simultaneity (PSS), was the signed time (based on stimulus ordering – negative indicated auditory stimulus preceded visual stimulus) between the two stimuli when the subject would have been equally likely to indicate that either stimulus occurred first (0.5 on the ordinate axis).

Results

Using the above-described bootstrap software, subject LT's point of subject simultaneity was calculated to be approximately -38 ms for early period (-160 to -50 ms) flash stimuli and -83 ms for late (-50 to 0) period flash stimuli, a difference of 45 ms (Figure 4). The two-tailed 95% confidence interval on the difference between the early-period and late-period PSS was significant at the 0.05 level. Subject AT's data showed the same pattern as LT's and was near-significant (but not significant at the 0.05 level). His early period PSS was +10 ms and his late period PSS was -32 ms (Figure 4), a difference of 42 ms, similar to that of LT. These results suggest that on average, visual stimuli occurring *early* in saccade preparation (between 160 and 50 ms pre-saccade) have a perceptual latency around 40 ms longer than visual stimuli occurring *late* in saccade preparation (50 to 0 ms pre-saccade).

Discussion

Our data indicate that Morrone *et al.*'s temporal inversion illusion is likely caused by a difference in the perceptual latencies of stimuli presented earlier versus later in saccade preparation. Because of the 40 ms latency difference, pairs of visual stimuli occurring 20 ms apart and in different preparation time-windows would be seen as reversed. The first stimulus would have a latency of *x* ms, the second a latency of $[x + 20] - 40 = [x - 2]$ ms, leading to the perception that the latter stimulus occurred 20 ms *before* the earlier one.

There are a number of explanations for why such saccade-related delays or accelerations could be happening. Morrone *et al.* (2005a) originally proposed that the compression may have been caused by the slowing of a 'neural clock', but themselves admitted that this did not explain the inversion effect. They later suggested (Morrone *et al.*, 2005b) that postdiction in combination with relativistic effects resulting from the speed of information transfer in the updating of spatiotopic fields may explain the temporal inversion illusion. This explanation, however, has also not been given in any detail.

Interestingly, follow-up neurophysiological work (Ibbotson *et al.*, 2006) found that MT neurons respond at different latencies depending on when during saccade preparation the stimulus is presented, however the only effect described was significantly faster latencies for stimuli immediately preceding the physical onset of the saccade. This result is compatible with our finding of shorter perceived latencies in the period 40 to 50 ms before saccade onset. All the research just discussed, however, assumes the existence of a 'neural clock' upstream of the affected processing. Timing using such a clock is usually assumed to be done through the use of an oscillator and an accumulator. State dependent networks (SDNs) are an alternative account which may be more appropriate for the short intervals in which we are interested. SDNs represent time as the development of a dynamic physical system (Karmarkar & Buonomano, 2007), and are more flexible since they need not represent time 'linearly' as oscillation-accumulator systems must.

To conclude, though the original Morrone *et al.* (2005a) study set out to investigate the perception of timing around the onset of eye movements, this reversal of perceived order is qualitatively different from a simple error in time measurement. Whereas it is easy to imagine a distinct neural mechanism for judging the passage of time and the duration of intervals, we would expect the kinds of errors this mechanism made would be products of biases and a lack of precision, errors which are consonant with a notion of unidirectional time. Errors of the ordering of perception do not agree in this way, suggesting that they are not the product of a mechanism for measuring the perception of timing. Instead, these errors point to a different field of study: the timing of perception.

Figures

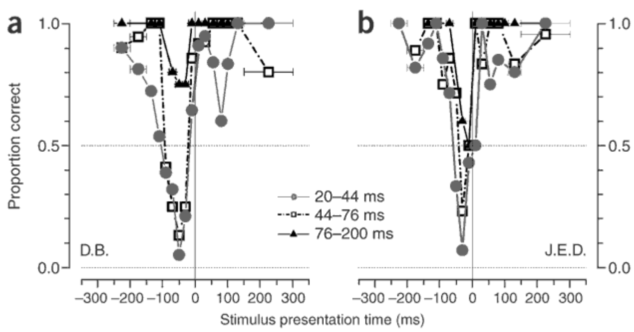

Figure 1: From Morrone *et al.* (2005). For two subjects, proportion correct for order judgments as a function of presentation time. Data are plotted for three different ISIs (20-44, 44-76, 76-200). Note that at around -40 ms subjects are nearly always 'wrong', especially for the shorter ISI trials, as is evident in the curves' dip at the point.

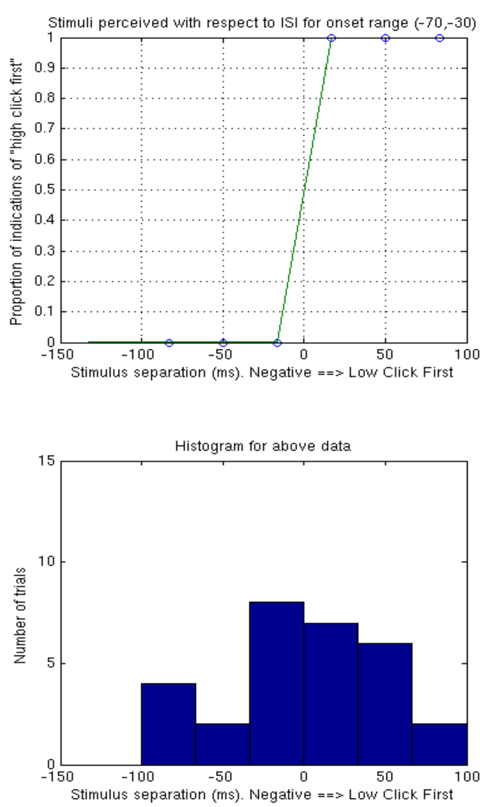

Figure 2: Above plots show that auditory stimuli are not subject to saccade-induced temporal inversion. Performance remains perfect throughout the pre-saccade time window.

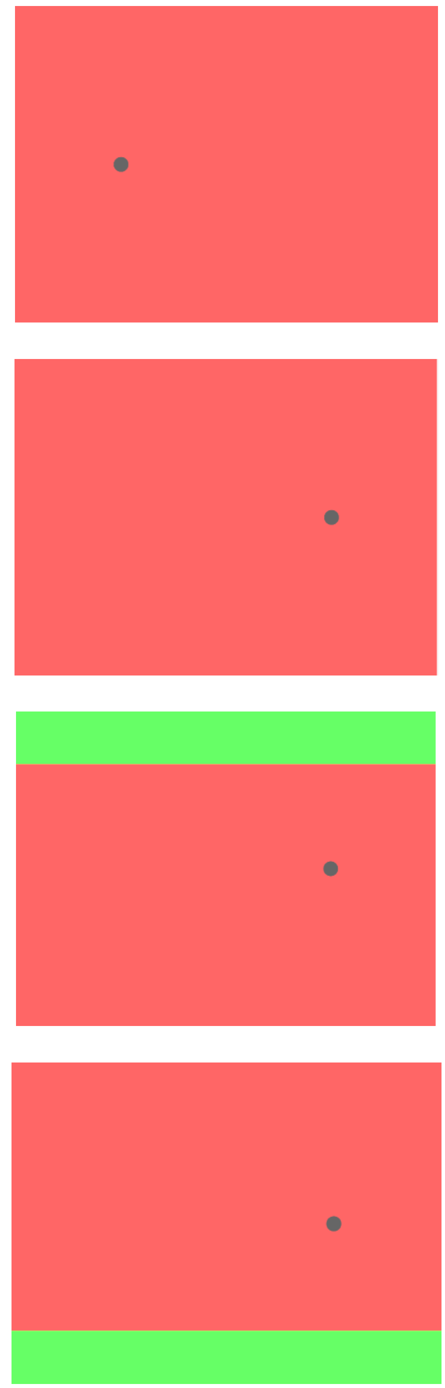

Figure 3: The four events in the experimental sequence of Morrone et al (2005). The subject fixated a point on the left side of the screen. A target appeared to the right side of the screen. While the subject was preparing to saccade, two 8 ms flashes with varying ISIs were presented at the top and bottom of the screen. Our multimodal paradigm (experiment two) is identical, save that one of the flashes is replaced by an audible 4 ms 'click'. Experiment one is identical save that both flashes are replaced by different frequency 'clicks'.

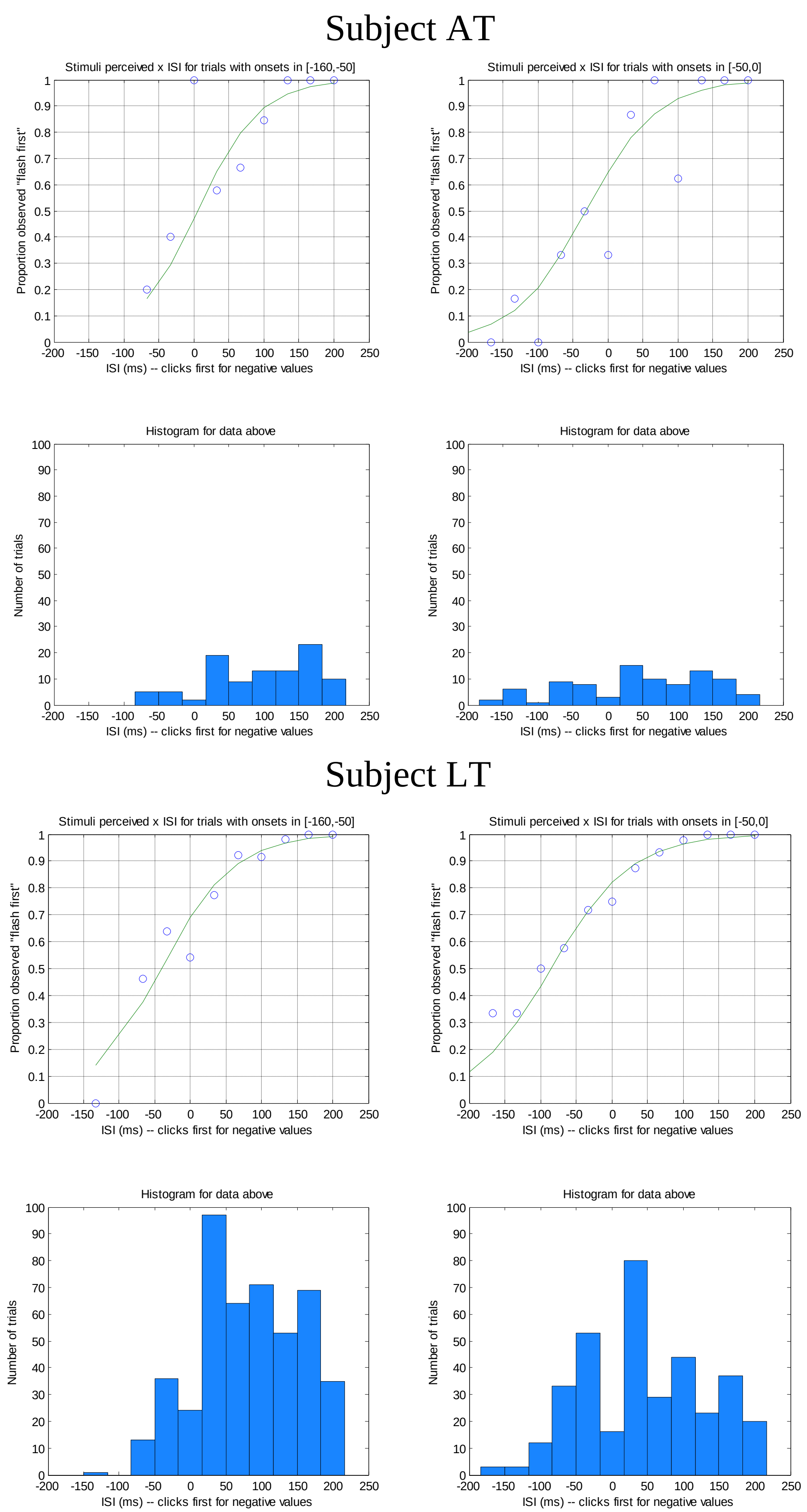

Figure 4: Data and fitted psychometric functions (logistic) for subjects AT (early PSS: +10 ms, late PSS: -32 ms) and LT (early PSS: -38 ms, late PSS: -83 ms). On the left and right are data for stimuli presented early and late in saccade preparation, respectively. Where the function crosses 0.5 is the PSS.

Appendix: Treatment of possible confounds

The 'prior-entry' effect remains the subject of some debate, but essentially it is the speeded processing and lower latency of one stimulus caused by increased attention to its modality or source. This effect could have biased the perceived latency of visual stimuli. To eliminate prior-entry effects we thus used both a visual and auditory saccade 'go' cue (FP2).

It has been found that additional auditory stimuli after the presentation of a saccadic target can modulate the saccadic reaction time (Colonius & Arndt, 2001). The use of a FP2 auditory onset cue should mitigate this because the effect of the auditory onset cue is to make the two auditory stimuli into one 'double click' with variable inter-click timing. Additionally, every trial employed the same amplitude saccade to the same target in visual space, which helps in saccade planning, reducing the effects of 'distractor' auditory stimuli.

The plasticity of the relationship between auditory and visual stimuli was also the cause of some concern, and the repeated simultaneous multimodal auditory 'go' signal was also helpful in preventing the recalibration of auditory-visual relationships (Stetson et al., 2006).

High precision multimodal order judgments are difficult to perform and require constant attention. To help maintain high levels of subject performance we performed a small amount of without-saccade audio-visual temporal order judgment training each day before running subjects (more details in the experiment two stimuli and apparatus section). Also, during the experiment order judgment feedback was given when the timing of the eye movement resulted in stimuli being presented far before or far after saccade onset (200 ms before or 70 ms after).

The cross-capture of audition and vision refers to the tendency for asynchronous audio-visual stimuli to be drawn to each other in their perceived timing (Fendrich & Corballis, 2001). While this might well affect our stimulus presentation, the effect should remain constant irrespective of the timing of the eye movement (and it is only in relation to the eye movement's timing that we are interested in the different relationship of the audio-visual stimuli). Consequently, cross-capture may weaken effects of saccade preparation on perceived stimulus latency but should not confound them.

Some results suggest that saccadic latencies can be affected by the spatial relationship between the perceived location of the auditory stimulus and the location of the visual one (Zampini *et al.*, 2003), a possible effect we controlled for by presenting both visual and auditory stimuli centrally.

Lastly, the difficulty of the task restricted the useful subject pool somewhat, and has been raised as a possible confound. We have found that individuals with significant musical experience are ideal candidates for this study as they have good multimodal temporal order resolution abilities, and both subjects used had such experience. Our study, however, is focused on visual processing, and it has been found that musicians do not possess superior eye movement abilities (Gruhn *et al.*, 2006) or improved saccadic reaction times.

Acknowledgments

This work is supported by an NSERC PGS-M to L.G. Trottier and NSF Career Grant 0133996 to V.R. de Sa and NSF Science of Learning Center grant to G.W. Cottrell. An earlier version of this work was presented at the graduate student conference COGNITIO (Trottier, 2006) and as a poster at COSYNE 2007.

References

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. Perception & Psychophysics, 63(1), 126-147.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. Perception & Psychophysics, 63(4), 719-725.

Gruhn, W., Litt, F., Scherer, A., Schumann, T., Weiss, E. M., & Gebhardt, C. (2006). Suppressing reflexive behaviour: Saccadic eye movements in musicians and non-musicians. Musicae Scientiae, 10(1), 19-32.

Grusser, O. J. (1994). Early concepts on efference copy and reafference. Behavioral and Brain Sciences, 17(2), 262-265.

Ibbotson, M. R., Crowder, N. A., & Price, N. S. C. (2006). Neural basis of time changes during saccades. Current Biology, 16(19), R834-R836.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. Neuron, 53(3), 427.

Morrone, M. C., Ross, J., & Burr, D. (2005a). Saccadic eye movements cause compression of time as well as space. Nature Neuroscience, 8(7), 950-954.

Morrone, M. C., Ross, J., & Burr, D. C. (2005b). Keeping vision stable: Rapid updating of spatiotopic receptive fields may cause relativistic-like effects.Unpublished manuscript.

Stetson, C., Xu, C., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. Neuron, 51(5), 651-659.

Vibell, J., Klinge, C., Zampini, M., Spence, C., & Nobre, A. C. (2007). Temporal order is coded temporally in the brain: Early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. Journal of Cognitive Neuroscience, 19(1), 109-120.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. Perception & Psychophysics, 63(8), 1293-1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: Ii. Bootstrap-based confidence intervals and sampling. Perception & Psychophysics, 63(8), 1314-1329.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. Experimental Brain Research, 152(2), 198-210.

Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. Neuroscience Letters, 381(3), 217-222.