INVESTIGATIONS INTO THE ROLE OF EARLY VISUAL CORTEX IN EXPERTISE READING MUSICAL NOTATION

By

Yetta Kwailing Wong

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
In partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in
Psychology
December, 2010
Nashville, Tennessee

Approved:
Professor Isabel Gauthier
Professor Randolph Blake
Professor Frank Tong
Professor Geoffrey F. Woodman
Professor James W. Tanaka

ACKNOWLEDGMENTS

I feel extremely grateful to be able to work with my advisor, Isabel Gauthier. She has guided me to the world of science, which is full of excitement, challenges and fun, and has been fully supportive and encouraging through these years. She has shown me her commitment and enthusiasm with science, and the qualities of a great mentor and educator. I always feel blessed to be her student.

I would like to thank each member of my committee, who has offered invaluable advice and guidance for my work. I would like to especially thank Geoff Woodman, who has shown me the beauty of the ERP technique. Also, I would like to thank members in the Perceptual Expertise Network for the inspiration and training in my graduate study. In particular, I would like to thank Tim Curran for his constructive comments on my dissertation work.

I would also like to thank the past and present lab members for their help and many insightful discussions. Special thanks to Eunice Yang and Min-Suk Kang for their help during the preparation of my dissertation.

I am grateful to the support of my family in my graduate study. In particular, I would like to thank Alan Wong, who has filled my graduate school journey with sunshine and laughter, and shared with me the up and down moments in scientific discovery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

OVERVIEW OF THE DISSERTATION

This dissertation aims at investigating the role of early visual cortex in music reading expertise. This work was motivated by the surprising finding of neural selectivity for musical notes in early visual cortex with music reading expertise, which is not predicted by current theories about the role of early visual cortex in object recognition or in perceptual expertise. In this dissertation, I investigated the mechanisms underlying the recruitment of early visual cortex for musical notes by examining the temporal dynamics of the neural selectivity for musical notation using scalp electrophysiological recordings. I found that expertise effects for musical notes could be observed as early as 40-60ms after stimulus onset, suggesting that the initial visual processes for notes have been altered with experience in music reading. This early selectivity for notes is predicted by degrees of crowding and holistic processing within music reading experts, supporting the functional significance of this early effect. These results imply that the recruitment of the early visual cortex is, at least partially, a feedforward effect, and suggest that early visual cells become selective for musical notes with the acquisition of music reading expertise.

This dissertation begins (CHAPTER I) with a review of perceptual expertise studies and my previous work in music reading that motivated this dissertation work. Then I discuss current views on the role of early visual cortex in object recognition and perceptual expertise, followed by describing two possible mechanisms underlying the recruitment of early visual cortex for musical notes, and how temporal dynamics of the neural selectivity for notes can help to tease apart these two possibilities. After that, I briefly review the literature on crowding, which served as behavioral correlates for the ERP effects.

CHAPTER II reports the methods and results of the ERP experiment. Music reading experts and novices were recruited and performed a simple one-back task with musical notes, letters or pseudo-letters, with a design following that of the prior fMRI study. I observed ERP expertise effects for musical notation with various ERP components, including the C1 component bilaterally (40-60ms), the N170 component bilaterally (120-200ms), and the CNV component (-200-0ms).

Next, I describe the study on crowding and music reading expertise (CHAPTER III). I found that experts experienced less crowding for musical stimuli but not for non-musical novel stimuli (Landolt C). Correlation analyses in CHAPTER IV revealed the behavioral significance of the expertise effects obtained with the C1, N170 and CNV components. Both the C1 and N170 expertise effects were predicted by all behavioral measures, including music reading ability (measured by perceptual fluency), crowding and holistic processing, while the CNV expertise effect was predicted by perceptual fluency and crowding.

I conclude my dissertation with CHAPTER V, in which I discuss the implications of the expertise effects obtained with various ERP components and crowding, including the role of early visual cortex in music reading expertise, and general implications on studies in perceptual expertise, object recognition and visual crowding.

CHAPTER I

PERCEPTUAL EXPERTISE, MUSIC READING EXPERTISE AND EARLY VISUAL CORTEX

Perceptual expertise studies investigate how experts achieve excellent recognition performance at individuating objects within a category and study the visual mechanisms supporting their recognition performance. The relationship between behavioral and neural differences in experts and novices can also be used as a window to understand how the visual system works.

Perceptual expertise has been studied in real-world object domains, such as faces (Farah, Wilson, Drain, & Tanaka, 1998; Kanwisher, McDermott, & Chun, 1997), birds (Tanaka & Curran, 2001), cars (Gauthier, Skudlarski, Gore, & Anderson, 2000), Roman letters and words (Cohen et al., 2000; James, James, Jobard, Wong, & Gauthier, 2005), fingerprints (Busey & Vanderkolk, 2005), Chinese characters (Hsiao & Cottrell, 2009; Wong, Gauthier, Woroch, Debuse, & Curran, 2005), and also with training studies with computer-generated novel objects such as Greebles (Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1998), Spikies, Cubies and Smoothies (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006), Blobs (Yue, Tjan, & Biederman, 2006), block-like objects (Moore, Cohen, & Ranganath, 2006) and Ziggerins (Wong, Palmeri, & Gauthier, 2009a) (Fig. 1).

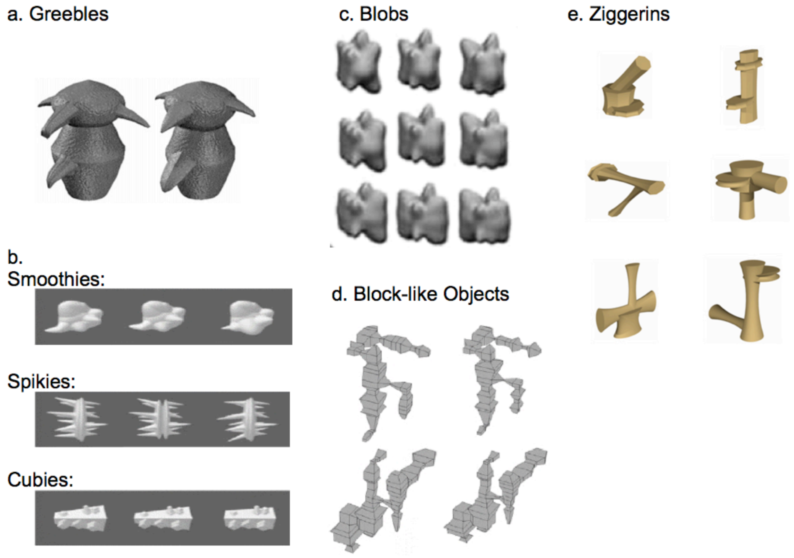

Figure 1. Novel objects used in perceptual expertise studies. (a) Greebles (Gauthier et al., 1997; 1998). (b) Smoothies, Spikies and Cubies (Op de Beeck et al., 2006). (c) Blobs (Yue et al., 2006). (d) Block-like objects (Moore et al., 2006). (e) Ziggerins (Wong et al., 2009a).

These perceptual expertise studies have identified various behavioral and neural differences that mark how experts and novices process visual objects differently in the visual system. For example, at the behavioral level, experts develop sensitivity to various dimensions of objects of expertise, including orientation (Diamond & Carey, 1986; Yin, 1969), configuration of object parts (Gauthier et al., 1998; Maurer, Grand, & Mondloch, 2002) and style regularity across objects (Gauthier, Wong, Hayward, & Cheung, 2006). Experts can also develop a tendency to process objects of expertise holistically (Wong et al., 2009a), and a fixation pattern slightly biased towards the left side of the objects (possibly related the increased reliance on the right hemisphere with some domains of expertise; Hsiao & Cottrell, 2009). At the neural level, fMRI studies show that objects of expertise selectively recruit ventral temporal regions, including the mid-fusiform gyrus (Downing, 2001; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; James et al., 2005;

Kanwisher et al., 1997; Moore et al., 2006) and the lateral occipital complex (Jiang et al., 2007; Op de Beeck et al., 2006; Yue et al., 2006). ERP studies also reveal that the N170 and N250 components are related to perceptual expertise in different object domains (Busey & Vanderkolk, 2005; Gauthier, Curran, Curby, & Collins, 2003; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002; Scott, Tanaka, Sheinberg, & Curran, 2006; Tanaka & Curran, 2001). The behavioral significance of these neural markers is established by their ability to predict individual behavioral performance in various domains of objects (Gauthier et al., 2003; Gauthier et al., 2000; Gauthier & Tarr, 2002; Wong, Palmeri, Rogers, Gore, & Gauthier, 2009b; Xu, 2005).

Expertise in reading musical notation

My previous work introduced music reading into perceptual expertise studies (Wong & Gauthier, 2010, in press). Musical notation is an interesting domain to study and compare with other domains of expertise for several reasons. First, while most objects are defined by their shape information, musical notation is defined by shape and spatial position such that notes with identical shapes but different spatial positions on the five-line staff are considered different. With such emphasis on spatial information in object individuation, visual processes underlying music reading are likely different from the studied domains of objects. Second, expert visual skills in reading music are typically associated with multimodal processes, such as pitch and timbre, motor execution, somatosensory feedback, emotion and other semantic information related to musical notes. The multimodal nature of music reading allows us to investigate how brain areas associated with different modalities respond to visual stimuli after the acquisition of perceptual expertise. Third, expert music reading typically requires years of extensive training to develop, making it relatively easy to find participants with different levels of expertise.

Music reading has received little attention in music-related studies, which have been heavily focused on auditory, motor and somatosensory modalities (Deutsch, 1998; Munte, Altenmuller, & Jancke, 2002; Peretz & Zatorre, 2003; Spiro, 2003). A few studies reported that musicians recognized note patterns better than novices visually, with a presentation duration ranging from 50ms to several seconds (Sloboda, 1976, 1978; Waters, Underwood, & Findlay, 1997). On the neural level, prior work has reported different neural substrates recruited for music reading. For example, passive viewing of a music score led to activity in early visual areas bilaterally and in an occipito-parietal area (Sergent, Zuck, Terriah, & MacDonald, 1992). After training with music reading and keyboard playing, a visual task with musical notation resulted in increased neural responses in parietal and frontal areas (Stewart et al., 2003). Finally, a study contrasting passive viewing of musical scores to Japanese or English texts revealed higher neural activity for musical notes than text in the right transverse occipital sulcus in all of eight musicians, but in none of the eight non-musician controls, suggesting that the right transverse occipital sulcus is recruited by expert music reading (Nakada, Fujii, Suzuki, & Kwee, 1998). However, the visual processes and mechanisms behind the superior performance of experts and the recruitment of the neural substrates remain largely unexplored.

In this dissertation, I investigated the role of early visual cortex in music reading expertise, motivated by findings in my prior work in music reading expertise (Wong &

Gauthier, 2010, in press). The following sections provide some relevant background information, including the general approach I chose to study music reading expertise, followed by a brief review of my two previous studies. Then, I discuss how this general approach to study music reading expertise enriches our understanding of different expertise-related phenomena.

General approach to study music reading expertise

The general approach I chose to study music reading expertise is to start with some behavioral and neural effects that have been well established in other domains of expertise and investigate how these expertise effects are similar or different with musical notation. This approach has several advantages. First, the presence (or even absence) of an expertise effect provides further constraints for the conditions under which the expertise effect can be obtained, given that music reading expertise has both common and unique characteristics compared to other domains. Second, the wide range of music reading abilities in the population enables us to study how a behavioral or neural effect is associated with levels of expertise. This is not easily addressed in other real-world expertise domains. For example, participants with no or intermediate-level expertise are hard to find for faces and letters, and experts are relatively rare for cars or fingerprints. Also, larger and more long-term expertise effects compared to lab-trained perceptual expertise can be observed with expert music reading that requires many years of deliberate practice.

Following this approach, my previous work explored music reading expertise in two studies that I review in the next sections, one focused on the neural substrates recruited by musical notation with expertise (Wong & Gauthier, 2010), and the other on whether holistic processing, a behavioral marker for expertise in numerous domains of objects, can be obtained with music sequences and how it compares between experts and novices (Wong & Gauthier, in press).

Previous study I: The fMRI study

This experiment aimed at identifying brain regions selective for musical notation with the acquisition of music reading expertise (Wong & Gauthier, 2010). Ten music reading experts and 10 novices were recruited for this fMRI experiment. In the scanner, participants were presented with blocks of single stimuli (single notes, single letters or single symbols) or string stimuli (5-note sequences, 5-letter strings or 5-symbol strings), and they were required to perform a simple visual task (to detect immediate repetition of images or to detect whether a gap was present on one of the five lines, Fig. 2). Although these tasks were not music related, they were appropriate for our search for brain regions that are automatically recruited for musical notation (rather than a well-practiced task). Also, both experts and novices can perform well in these simple visual tasks, so that differences in neural responses were not confounded by performance differences.

To search for brain regions selective for musical notes as a function of expertise, statistical parametric maps were generated for the interaction between Category (single notes vs. single letters and single symbols) and Group (experts vs. novices) for each voxel in the whole brain. A widespread multimodal neural network was found selective

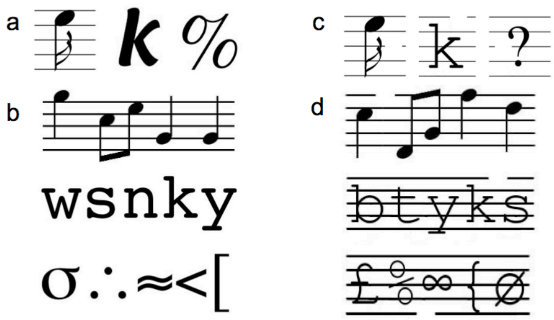

Figure 2. Example of the stimuli used in the scanner. (a-b) show the single and string stimuli used for the one-back task. (c-d) show the single and string stimuli for the gap-detection task.

for single musical notation for experts (Fig. 3a). As would be expected for expertise with visual objects, various high-level visual areas were identified as selective for single notes, including bilateral fusiform gyrus and an area along the right inferior temporal sulcus. Interestingly, musical notation also recruited early visual areas bilaterally, covering a large part of the calcarine fissure, which was never reported to be selective for objects of expertise in previous studies (Fig. 3b). In addition, an area in the left occipitotemporal junction showed higher selectivity for musical notation for novices than experts (Fig. 3a). The face-, letter- and letter string-selective regions, defined with separate localizer runs, were not selective for musical notation for either group, suggesting that the areas recruited by expert music reading are different from those recruited by expertise for faces, letters and letter strings.

In addition to these visual regions, a widespread multimodal network of other areas revealed higher selectivity for musical notation in experts than in novices, including (1) parietal regions such as bilateral occipitoparietal junction, bilateral intraparietal sulcus (IPS), the left angular gyrus and the left supramarginal gyrus; (2) primary and associative

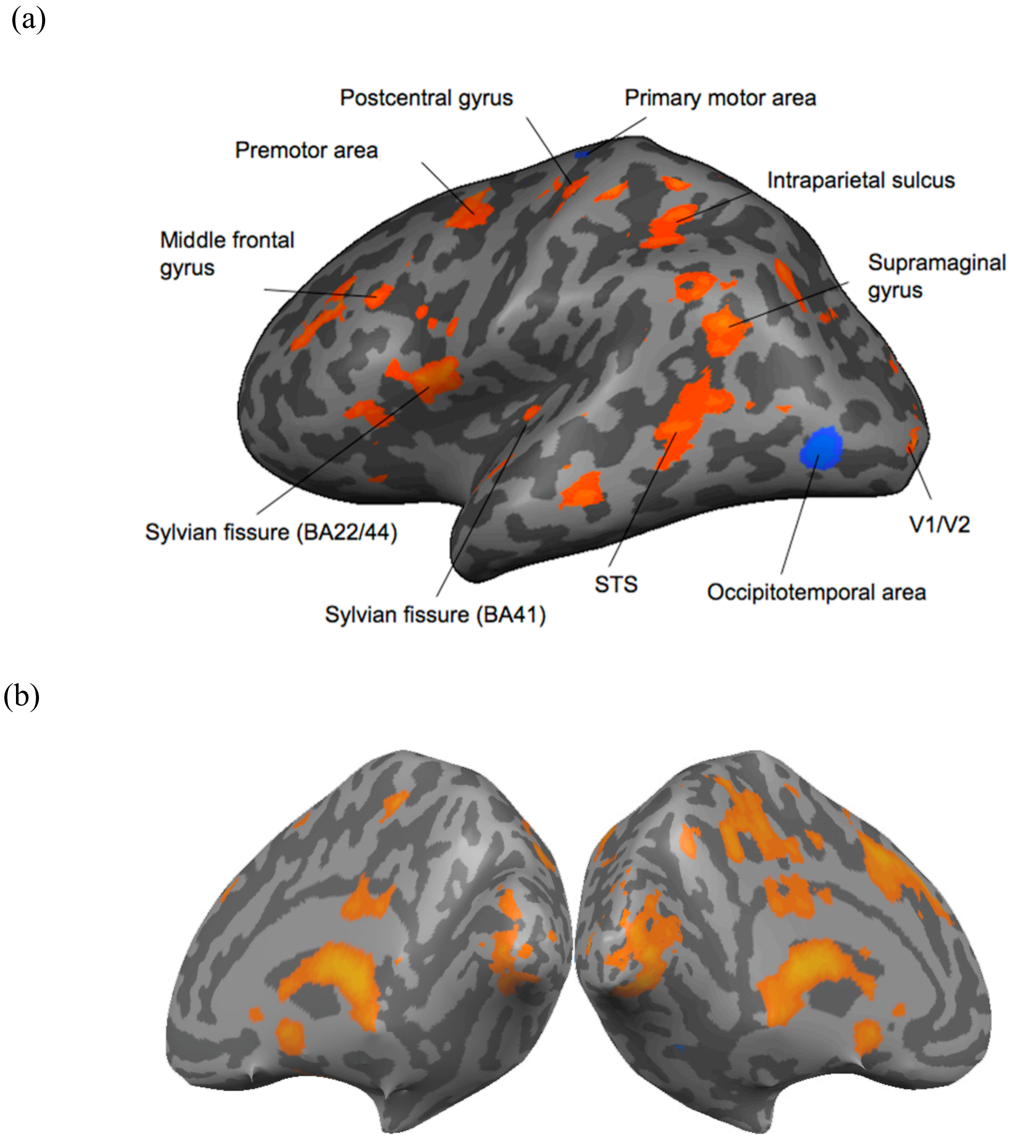

Figure 3. The multimodal network recruited for single notes for music reading experts. (a) A lateral view of the network, presented on one of the experts' inflated brain (left hemisphere). Orange clusters and blue clusters represent higher and lower selectivity for single notes for experts compared to novices respectively. (b) A medial view of the same network showing the extensive selectivity for single notes found in early visual cortex (along the calcarine fissure). The statistical parametric maps were generated at the threshold of $p < .05$, after correction for multiple comparisons using false discovery rate (FDR; Genovese, Lazar & Nichols, 2002).

auditory areas along the sylvian fissure bilaterally; (3) somatosensory areas in the postcentral gyrus bilaterally; (4) superior temporal gyrus for audiovisual processing bilaterally; (5) premotor areas bilaterally; (6) other frontal areas covering different parts of the inferior frontal gyrus, middle frontal gyrus and superior frontal sulcus; (7) other regions including the cingulate gyrus, precuneus, cerebellum and corpus callosum (Fig. 3a-b). Data analyses contrasting the string stimuli revealed a similar multimodal network recruited for music sequences, though less extensive than that for single notes, similar to Roman letters in which the network recruited for letter strings is less extensive compared to that for single letters (James et al., 2005). This widespread multimodal network showed selectivity for musical notes in experts in simple visual tasks, demonstrating the strong and automatic association between visual processing of notes and processing in other modalities with the acquisition of musical expertise.

To investigate whether neural activity in these areas predicted individual music reading ability, we examined the correlation between neural activity in these regions and individual music reading ability (measured as perceptual fluency with notes, see below). The correlation was significant in several brain areas associated with different modalities, including the right sylvian fissure, left superior temporal sulcus, right premotor area, right middle frontal sulcus, right superior frontal gyrus, and cingulate gyrus. Interestingly, a significant correlation was also found in the occipitotemporal area, in which the activity for musical notation was lower with better music reading skill. In contrast, the correlation did not reach significance in the face-, letter-, or letter string-selective areas.

In conclusion, experts at music reading recruit a widespread multimodal network when they see stimuli as simple as single musical notes, and some of the multimodal areas predict individual perceptual fluency for music sequences, confirming the behavioral relevance of the neural network. The visual specialization for musical notation is distinct from that for faces, letters and letter strings, which is consistent with the process-map hypothesis that expert perception of objects with different task demands should recruit different brain areas (Gauthier, 2000). Importantly, more work is needed to understand these task demands and how they relate to the specific brain areas recruited.

Previous study II: Holistic processing

A second project investigated holistic processing of music sequences and how it is modulated by music reading expertise (Wong & Gauthier, in press). Holistic processing, the tendency to process objects as wholes rather than as parts, is regarded as a hallmark of face recognition (Farah et al., 1998; Maurer et al., 2002; Young, Hellawell, & Hay, 1987). One operational definition of holistic processing is that observers are shown to be unable to selectively attend to part of an object (as in the composite effect; Young et al., 1987). Such failures of selective attention are associated with perceptual expertise for various non-face object categories including cars (Gauthier et al., 2003), fingerprints (Busey & Vanderkolk, 2005) and novel objects such as Greebles and Ziggerins (Gauthier & Tarr, 2002; Gauthier et al., 1998; Wong et al., 2009b). Holistic processing effects are also stronger for those faces with which we have the most experience, such as faces of one's own race (Michel, Rossion, Han, Chung, & Caldara, 2006; Tanaka, Kiefer, & Bukach, 2004) or one's own age (de Heering & Rossion, 2008).

Although holistic processing is found to increase with perceptual experience, other work suggests that holistic effects can be found with unfamiliar objects. For example, holistic effects were obtained in participants with no expertise with Chinese characters (Hsiao & Cottrell, 2009), when novel objects (Greebles) were presented in the context of faces, or if they were first encoded as two misaligned halves rather than as a whole object (Richler, Bukach, & Gauthier, 2009a). Based on the observation that holistic effects for novices are largely dependent on specific contexts (e.g. Richler et al., 2009a) while holistic effects for experts are relatively automatic and remain stable across contexts or task constraints (e.g. Michel et al., 2006; Richler et al., 2009a; Richler, Cheung, Wong, & Gauthier, 2009b), we hypothesized that holistic processing for novices is more strategic-based while that for experts is more automatic. We tested this hypothesis with music sequences in two behavioral experiments, in which we created different contexts prompting different processing strategies. These contextual manipulations were expected to change the pattern of strategic-based holistic effects (e.g. for novices) but not the relatively automatic holistic effects (e.g. for experts).

In two experiments, a sequential matching task was used (Fig. 4). On each trial, two four-note sequences were presented sequentially, and participants were asked to judge whether the target note (one of the four notes cued by two arrows in the second sequence) was the same or different from the equivalent note in the first sequence. Holistic processing is indexed as the difference between congruent and incongruent trials (a congruency effect), in which shifting an irrelevant note (adjacent to the target note) either led to an identical or conflicting response to that for the target part respectively (Cheung, Richler, Palmeri, & Gauthier, 2008; Richler, Gauthier, Wenger, & Palmeri, 2008; Richler, Tanaka, Brown, & Gauthier, 2008; Wong et al., 2009a). Different behavioral performance for congruent and incongruent conditions indicates that participants are affected by the irrelevant non-target part of the sequence, i.e., they process the sequence holistically.

In Experiment 1, our hypothesis was tested by manipulating target distribution, in which targets appeared in central positions (the two center notes) in 75% of the trials and in the periphery (the leftmost and rightmost notes) in 25% of the trials (25p75c). This manipulation was intended to bias participants to pay more attention to notes in the center positions and relatively ignore those in the periphery. This attentional strategy should in turn affect participants' susceptibility to incongruent information in different positions, i.e. the magnitude of holistic processing. If holistic processing for novices is based on this attention strategy, the magnitude of holistic processing should be modulated by target position, while that for experts should be relatively stable if it is more automatic. Consistent with our hypothesis, the congruency effect for novices was larger for

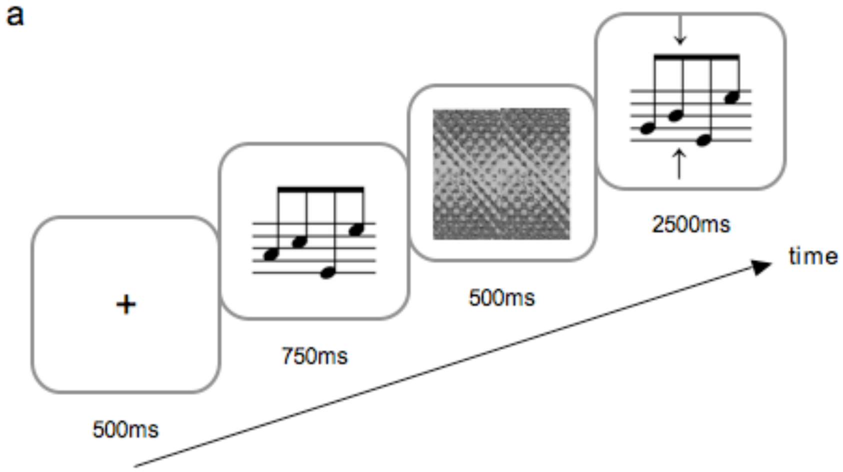

Figure 4. The experimental paradigm used for the sequential matching task in the holistic processing study.

periphery-target trials than center-target trials, while that for experts and intermediate readers were similar across target positions.

Experiment 2 further explored the nature of holistic processing in experts and novices by a parametric manipulation of target distribution (25p75c, 50p50c, 75p25c) in different blocks. Participants were explicitly informed of the target distribution in each block. Results indicated that the congruency effect for novices was modulated by target likelihood, i.e., the congruency effect increased when targets appeared in relatively unexpected positions (periphery-target trials for 25p75c, and center-target trials for 75p25c; Fig. 5b), supporting our hypothesis that holistic processing for novices is affected by attentional strategies. In contrast, whether targets appeared in likely or unlikely positions did not influence the congruency effect for experts (Fig. 5a-b), suggesting that holistic processing for experts is more automatic.

In addition, our correlation analyses suggest that holistic processing of music sequences for experts and novices arise from different underlying mechanisms. First, a higher perceptual fluency for music sequences predicted a larger congruency effect for experts but a smaller congruency effect for novices (Fig. 6a). Second, we analyzed the correlation between individual holistic effects and neural selectivity for musical notation for those participants who participated in both the present study and the fMRI study for music reading expertise (Wong & Gauthier, 2010). Neural selectivity for musical notes in the right fusiform face-selective area (rFFA) was predicted by individual holistic effects in opposite directions for the two groups (Fig. 6b-c). The finding is consistent with the prior findings that the rFFA is associated with holistic processing, including faces (Rotshtein, Geng, Driver, & Dolan, 2007; Schiltz & Rossion, 2006) and other objects of

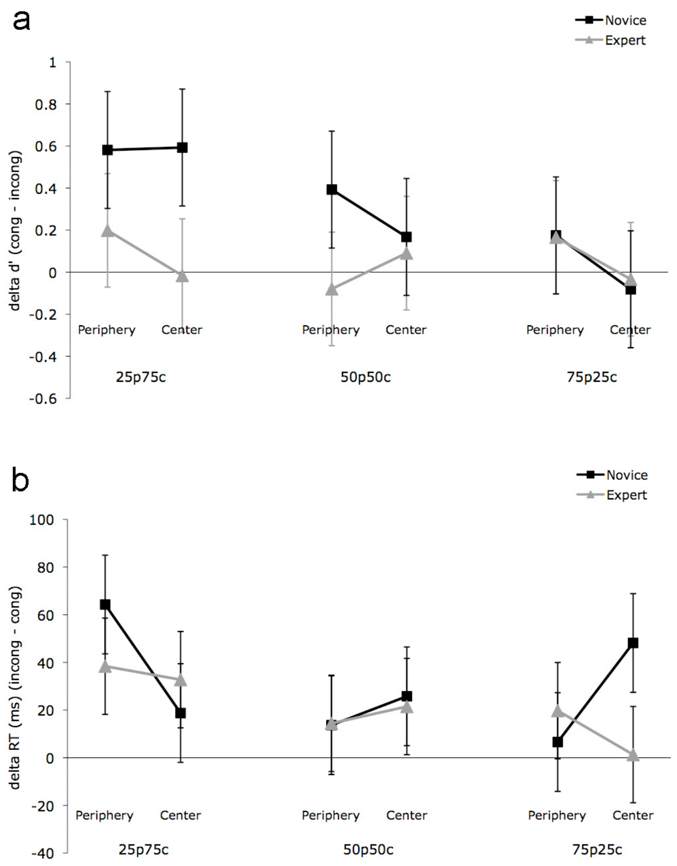

Figure 5. The mean congruency effect measured with delta d’ (a) and delta RT (b) for Experiment 2. Error bars show the 95% CI for the within-subject effects for the Group x Position x Distribution interaction.

expertise (Gauthier & Tarr, 2002; Wong et al., 2009b), and supports the hypothesis that mechanisms for holistic processing for experts and novices are different.

In conclusion, our results suggest that holistic effects in experts and novices are of a different nature. In experts, holistic effects were relatively stable across contexts prompting different attentional strategies, consistent with a stable and automatic perceptual tendency of perceiving objects as wholes, a hallmark of object and face expertise. In novices, holistic effects were also obtained. Instead of reflecting a perceptual tendency, however, the effects were more strategic and were subject to influence from tasks and instructions. Individual holistic effects were predicted by our behavioral and neural measures for the two groups in opposite directions, further supporting the hypothesis that different mechanisms underlie holistic effects in the two groups. This work revealed that observing holistic effects is not sufficient evidence for holistic processing. It is important to examine both the magnitude of holistic processing and whether it varies across task and contextual manipulations.

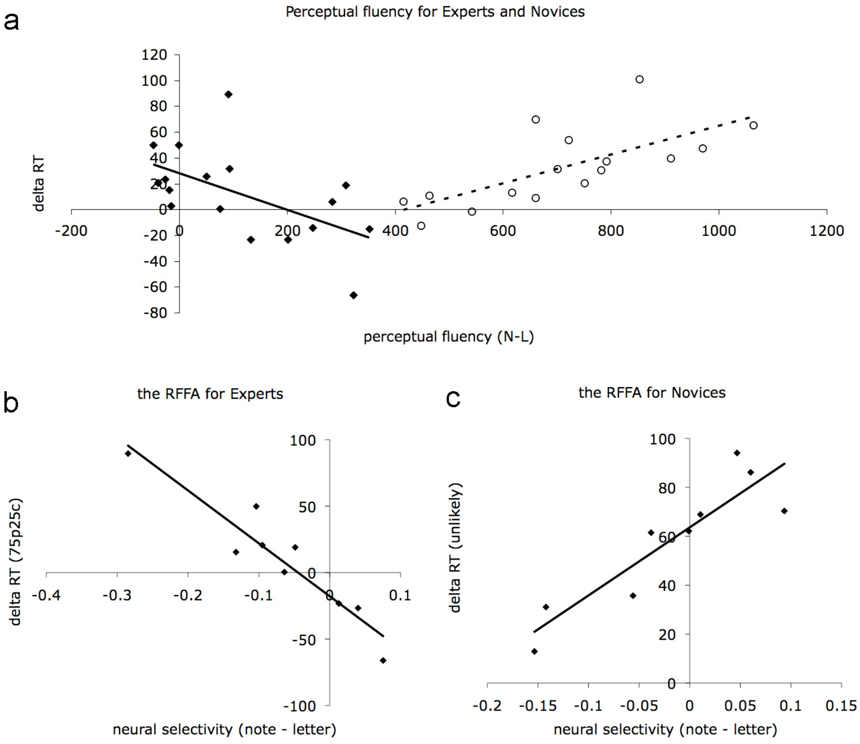

Figure 6. Correlation between holistic processing and other behavioral or neural measures for notes, including (a) perceptual fluency for experts (black dots and solid line) and novices (open circles and dotted line) in 75p25c; Neural selectivity for music sequences in the rFFA for experts in 75p25c (b) and that for novices in ‘unlikely’ condition, which combined the periphery-target trials in 25p75c and central-target trials in 75p25c (c).

Summary and implications

These two studies provided useful information in understanding the neural and behavioral effects associated with perceptual expertise. For the functional organization of visual cortex, while previous studies have focused on higher visual areas, our study demonstrated that a large part of the visual cortex can be recruited with objects of expertise, from early to late visual areas bilaterally. The visual selectivity for musical notes for experts, which is distinct from that for faces, single letters and letter strings, supports the role of perceptual experience in determining, at least partially, the regions recruited for objects of expertise (Gauthier, 2000). Our results also address an interesting paradox in the study of holistic processing, in which holistic processing is associated with perceptual expertise but can also be observed in novices. We showed that holistic effects can arise in novices through different mechanisms, which do not necessarily indicate holistic processing as a stable perceptual tendency of processing objects as wholes.

These studies also generated interesting and unexpected results that are worthy of further investigation. In particular, the recruitment of bilateral early visual cortex for expert perception of musical notation is surprising because it is not predicted in the literature on object recognition or on perceptual expertise. In the following sections, I briefly review the role of early visual cortex in object recognition and perceptual expertise in the literature, and discuss possible mechanisms for the recruitment of early visual areas for expert notation perception.

Role of early visual cortex in perceptual expertise

For perceptual expertise in general

From the literature in object recognition or perceptual expertise, it is not expected that bilateral early visual cortex would be recruited for expert perception of musical notation. First, V1 cells are typically considered local feature detectors based on the response properties of the cells. For example, V1 neurons are tuned to simple features such as bars in different orientations and are partially monocular (at least in layer 4 in V1, Hubel & Wiesel, 1968). They have small receptive fields and are retinotopically organized such that different regions of V1 correspond to different parts of the visual field (see review in Grill-Spector & Malach, 2004). Therefore, in various theories and computational models of object recognition, V1 cells are active for all kinds of visual judgments, and local and featural information from early visual cells is combined in later stages of the visual hierarchy for object recognition (DiCarlo & Cox, 2007; Grill-Spector & Malach, 2004; Kourtzi & DiCarlo, 2006; Riesenhuber & Poggio, 1999).

Also, early visual areas are not object selective. For instance, they respond similarly to noise patterns, textures and objects (Grill-Spector & Malach, 2004; Malach et al., 1995). Early visual activity does not predict object recognition performance, as visual activity remains high even when recognition performance drops to chance level (Grill-Spector, Kushnir, Hendler, & Malach, 2000). In contrast, higher visual areas, including the lateral occipital cortex, respond selectively to different categories and forms of objects compared to noise patterns or scrambled objects (Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995), and visual activity in these areas corresponds well with behavioral performance (Grill-Spector et al., 2000). In more anterior areas along the ventral pathway, such as the fusiform gyrus and the parahippocampal gyrus, various small and focal regions are functionally specialized for certain object categories, including faces (Kanwisher et al., 1997), body parts (Downing, 2001; Peelen & Downing, 2007), buildings and scenes (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998), and letters and words (Cohen et al., 2000; James et al., 2005). Therefore object recognition is thought to be achieved in later processing stages along the visual hierarchy (DiCarlo & Cox, 2007; Grill-Spector & Malach, 2004; Kourtzi & DiCarlo, 2006; Riesenhuber & Poggio, 1999).

Consistent with the object selectivity found in higher visual areas but not in early visual areas, perceptual expertise studies that localized the early visual cortex as a region-of-interest (ROI) did not find any significant training effects after perceptual expertise training (Lerner, Epshtein, Ullman, & Malach, 2008; Op de Beeck et al., 2006). Instead, regions in the fusiform gyrus are often recruited for objects of expertise, and neural activity in higher-level visual areas predicts behavioral performance for objects of expertise (Gauthier, Curby, Skudlarski, & Epstein, 2005; Gauthier et al., 2000; Gauthier & Tarr, 2002; Wong et al., 2009b; Xu, 2005), consistent with the idea that object recognition is achieved in higher visual areas.

For music reading expertise

Although early visual cortex is thought to contain simple local feature detectors and is not selective for objects, our results converge to suggest that early visual cortex may be important for music reading expertise. First, the early visual selectivity for musical notation is extensive, covering a large part of the calcarine fissure bilaterally

(Fig. 3b). Second, neural selectivity for notes in the early visual cortex predicts the degree of holistic processing for music reading experts in both hemispheres, even though the task in the scanner was unrelated to any congruency manipulations. Experts who show larger holistic effects tend to recruit the right early visual areas more (*r* = .606, *p* = .08) and the left early visual areas less (*r* = -.70, *p* = .036; unpublished data; Fig 7). These correlations in different directions were unexpected but this pattern is reminiscent of the idea that the right hemisphere is more related to holistic processing while the left is more related to analytic processing (Levy-Agresti & Sperry, 1968; Patterson & Bradshaw, 1975). These results suggest that the engagement of the early visual cortex in music reading expertise may have an important functional role.

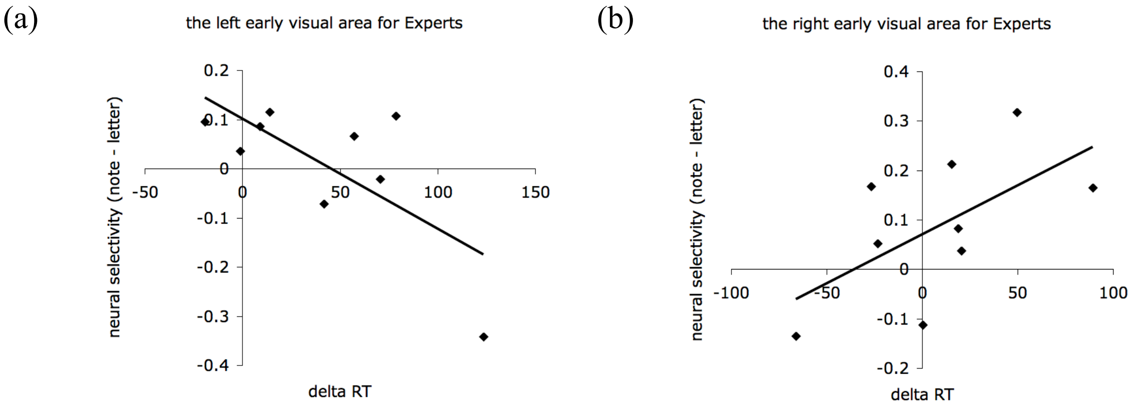

Figure 7. Correlation between neural selectivity for musical notes and holistic processing in bilateral early visual cortex. The congruency effect (delta RT) was in 25p75c for the left early visual cortex (a) and in 75p25c for the right visual cortex (b).

It is unlikely that the recruitment of early visual areas for musical notes is merely a result of experts directing more attention to musical notes, based on two pieces of evidence. First, the similar behavioral performance in the simple visual judgments between groups suggests that the tasks engaged the two groups comparably. Also, further analyses revealed that the early visual recruitment was found separately for the one-back task and the gap-detection task (though not as extensively for the gap-detection task), i.e. the early visual recruitment occurred even when attention was directed to the five-line staff instead of the notes. It suggests that, with the acquisition of music reading expertise, early visual cortex is automatically recruited for reading musical notation. In addition, it should be noted that if the recruitment of early visual cortex was merely the result of the attention-grabbing nature of objects of expertise, the same result should have been observed in many of the studies comparing expert to novice perception in other domains, which was not the case.

Possible mechanisms for recruitment of early visual cortex

What are the mechanisms behind this recruitment of early visual cortex for objects of expertise? There are at least two possibilities suggested by the literature, strengthened feedback for musical notation with long term experience and altered response properties of V1 cells through perceptual learning.

Strengthened feedback with long-term experience

The first possible mechanism is that early visual selectivity for musical notation is a result of strengthened feedback from higher areas. V1 receives feedback projections from higher visual areas and other brain regions including the lateral intraparietal area (LIP), superior temporal polysensory area (STP), frontal eye fields (FEF) and auditory cortex (Barone, Batardiere, Knoblauch, & Kennedy, 2000; Falchier, Clavagnier, Barone,

& Kennedy, 2002; Salin & Bullier, 1995). These recurrent connections are thought to modulate V1 activity based on top-down knowledge, and are closely related to visual awareness in various tasks such as visual detection, perceptual grouping, selective attention, binocular rivalry, etc. (see review in Tong, 2003). In a recent study using a multivoxel pattern analysis technique, object category information was found in the foveal region of V1 even though objects were never physically presented in the fovea, and the information was correlated with behavioral discrimination accuracy (Williams et al., 2008), suggesting that V1 receives feedback with detailed visual information at the level of object categories.

Other studies further suggest that feedback to early visual cortex can be strengthened through visual training. In a training study on detecting shapes from shading, V1 developed sensitivity to shading information of the objects after training, which was argued to be strengthened feedback from V2 with experience (Lee, 2002; Lee, Yang, Romero, & Mumford, 2002). More direct evidence comes from a visual search training study, which reported that neural activity during visual search for a shape in the trained orientation (a rotated “T”) was larger than untrained orientations in early visual areas (Sigman et al., 2005). It was likely a result of top-down modulation during visual search, because the increased activity was found during the search for shapes in the trained orientation, regardless of whether the target shape was physically present or not in the trials. It was hypothesized that the representation of the trained objects had been shifted from higher visual areas to early regions, where multiple representations of the objects corresponding to different visual field positions can be built, in response to the task demand of recognizing multiple objects within a brief presentation time (Sigman &

Gilbert, 2000; Sigman et al., 2005). The task demand of searching for a trained target sets up the appropriate interaction between feedback connections and local circuits to provide relevant contextual information, and such top-down modulation is a possible result of learning (Gilbert & Sigman, 2007).

In music reading, experts are trained to rapidly read multiple musical notes that are distributed in different parts of the visual field. It is possible that building multiple representations for musical notation in the retinotopic regions underlies the rapid and accurate recognition for music sequences for experts. Although the top-down modulation of early visual cortex is thought to be task-specific (Gilbert & Sigman, 2007), training effects in V1 have been reported to transfer to a task different from that of the training (Lee et al., 2002). Also, it is possible that a real-world domain of expertise like music reading, which has many more years of deliberate visual training than lab-trained studies, can establish a feedback network that is strong enough to be triggered in other simple visual tasks. In other words, the early visual selectivity for musical notation may be a result of strengthened feedback from higher areas due to the task demand of recognizing multiple musical notes quickly during music reading.

Altered response properties of early visual cells

A second possible mechanism is that response properties of early visual cells are altered to achieve better discrimination for musical notation. Previous studies have suggested that neural activity in the primary visual cortex can be altered with perceptual learning training. For instance, fMRI studies reported that V1 activity either decreased (Mukai et al., 2007; Schiltz et al., 1999) or increased (Furmanski, Schluppeck, & Engel, 2004; Maertens & Pollmann, 2005; Schwartz, Maquet, & Frith, 2002) following perceptual learning training with various training tasks, such as texture discrimination and orientation discrimination. Although these findings are susceptible to being explained by feedback effects (given the poor temporal resolution of fMRI technique), other studies provided more direct evidence that response properties in V1 cells are altered with perceptual learning. For example, the shape of the orientation-tuning curve of V1 cells was altered after an orientation identification training with gratings (Schoups, Vogels, Qian, & Orban, 2001). A training effect as early as 40ms was reported in an ERP study after a short texture discrimination training, suggesting that changes in V1 activity occur after perceptual learning (Pourtois, Rauss, Vuilleumier, & Schwartz, 2008). As this type of visual training typically leads to long-term learning effects that last for several months or even years, it has been suggested that the improved discrimination performance is supported by long-term structural changes (Karni & Sagi, 1993). It is possible that years of experience of discriminating different musical notes lead to local and long-term changes in response properties of early visual cells (such as the shape or amplitude of the tuning curves towards musical notation), in a manner similar to perceptual learning, so that early visual cortex becomes more selective to musical notation for experts compared to novices. Consistent with this possibility, a recent study revealed that the cortical thickness is significantly larger for musicians than non-musicians in V1 (Bermudez, Lerch, Evans, & Zatorre, 2009).

Using temporal dynamics to study mechanisms for early visual recruitment

Is early visual selectivity for musical notation a result of strengthened feedback to early visual cortex, or altered response properties of early visual cells, or both? These two mechanisms are not mutually exclusive, as it is also possible that the early visual recruitment is a result of more complicated interactions between feedback and changes in local response properties (Gilbert & Sigman, 2007). It is not easy to differentiate the effects of feedback signals from higher areas or those of feedforward signals generated from visual areas locally, because feedback can happen very quickly. For example, visual information reaches various brain regions including the motor and other frontal areas, within 100ms (Lamme & Roelfsema, 2000). In addition, feedback that is critical to visual awareness occurs as early as 5-45ms across visual regions (Pascual-Leone & Walsh, 2001).

Despite possible early onsets of top-down effects, the relative contributions between feedforward and feedback signals differ across time. Visual processing that happens earlier in time is thought to be more dependent on feedforward signals and is less susceptible to top-down effects. For example, it has been argued that a feedforward sweep of information takes around 100ms to complete (Lamme & Roelfsema, 2000), and the first 100ms of information processing is mainly feedforward flow of information (Zhang & Luck, 2009). In contrast, other cognitive processes heavily involved in top-down modulations, such as expectancy or semantic influences, occur later in time (Kutas & Hillyard, 1980; Sutton, Bararen, Zubin, & John, 1965). In general, feedback plays a more important role as time elapses after stimulus onset, when visual information reaches higher levels of processing (Hopf, Vogel, Woodman, Heinze, & Luck, 2002; Luck, 2005).

Therefore, one way to explore the mechanisms underlying the early visual recruitment for music reading expertise is to study the temporal dynamics of the visual expertise effects for notes with the ERP technique (CHAPTER II). ERP is a useful technique for addressing this question for two reasons. First, ERP analyses can detect differences in visual processes between experts and novices. Perceptual expertise in different domains has been linked to differences in two ERP components, namely the N170 and the N250. For example, the amplitude and/or the latency of the N170 are different between experts and novices in various object categories (Busey & Vanderkolk, 2005; Gauthier et al., 2003; Tanaka & Curran, 2001; Wong et al., 2005), and this difference arises with novel object categories only after perceptual expertise training (Rossion et al., 2002; Rossion, Kung, & Tarr, 2004). The N250 is also related to training in subordinate-level recognition, i.e., individuating objects within a category (Scott et al., 2006). Few studies have investigated the ERP components related to expert perception of musical notation. Only two ERP studies focused on visual perception of musical notes (Gunter, Schmidt, & Besson, 2003; Schön & Besson, 2002). However, neither study included any novice group or non-musical visual stimuli as controls. Therefore their results are not informative to the questions of interest about early visual recruitment with music reading expertise.

Second, ERPs, with their high temporal resolution, can help dissociate sensory or perceptual mechanisms from higher processes such as semantic effects (Rossion et al., 2004; Wong et al., 2005). An early ERP effect is more likely attributed to feedforward than feedback signals (Hopf et al., 2002; Luck, 2005). Moreover, if the early visual recruitment for musical notes is the result of feedback, possible sources of the feedback signal can be identified by looking at the earliest expertise effect for notes. In sum, studying the temporal dynamics provides constraints for the possible mechanisms behind the early visual recruitment for music reading expertise.

Behavioral significance for recruiting early visual cortex

Regardless of the temporal dynamics of the early visual selectivity for musical notation, it is likely that the extensive recruitment of this region has some behavioral significance in expert music reading. To address this question, the correlation between different ERP expertise effects (if any) and three behavioral measures were analyzed. The first measure was a measure of perceptual fluency with notes, as a measure of individual ability in music reading. The second was a measure of holistic processing of music sequences, which predicted activity for musical notes in bilateral early visual cortex (see above). Apart from these two measures used in my previous work, I report a behavioral study on visual crowding. It is an effect that has been associated with early visual cortex (Arman, Chung, & Tjan, 2006; Fang & Sheng, 2008; Tjan & Nandy, 2010), therefore it is possibly related to early visual cortex recruitment for music reading expertise, and it serves as an interesting expertise marker for music reading expertise by itself. Next, I briefly review the crowding literature, the relationship between crowding and perceptual expertise, and discuss why crowding effects may be modulated by music reading expertise.

Crowding

Crowding refers to the impaired discrimination performance of an object by inappropriately integrating the object with nearby contours or features (Levi, 2008; Pelli & Tillman, 2008). It occurs in all positions of the visual field, but is particularly robust in the peripheral visual field where an isolated object can be identified easily, while recognition is disrupted or becomes impossible once flankers (distractor objects) are added close to that object. Crowding occurs when flankers are put too close to the target object. The critical crowding distance between flankers and the target is roughly half of the eccentricity of the target in the visual field (the Bouma law; Bouma, 1970; Pelli & Tillman, 2008). It is often regarded as a bottleneck for object recognition (see recent reviews in Levi, 2008; Pelli & Tillman, 2008).

The causes of crowding remain controversial, with the proposed theories ranging from the limitation of neural structures (e.g. large receptive fields or long-range horizontal connections in visual periphery; Levi, 2008; Levi & Waugh, 1994), inappropriate featural integration of the targets and flankers within a spatial region (Pelli & Tillman, 2008), to insufficient resolution of spatial attention (He, Cavanagh, & Intriligator, 1996; Tripathy & Cavanagh, 2002). The neural regions involved in crowding are also unclear (Levi, 2008). Recently, fMRI studies suggest that crowding is related to the primary visual cortex, as early visual activity for spatial regions close to the targets is altered across crowded and uncrowded conditions, but the activity for the targets remains similar across conditions (Arman et al., 2006; Fang & Sheng, 2008). Also, it has been argued that crowding is related to improper encoding of image statistics in peripheral V1, since models built on such anatomical structure (and other saccade-related properties) replicate various crowding phenomena, including the Bouma's law (Tjan & Nandy, 2010). Other modeling studies suggest that crowding is related to higher visual areas. For example, a study suggests that crowding is related to V2, as images similar to the jumbled images perceived during crowding (with overlapping target and flanker features) can be created with image statistics models applying receptive field sizes of V2 of macaque monkeys (Freeman & Simoncelli, 2010). Another study suggests that crowding is related to V4, since their population code model that explains multiple crowding phenomena with orientation bars is best fitted with receptive field properties that are similar to a type of V4 cells (van den Berg, Roerdink, & Cornelissen, 2010). Finally, it has been suggested that crowding occurs in area TE, based on a re-entrant model that explains the effect of target-flanker distance in crowding and the visual adaptation of crowded stimuli (Jehee, Roelfsema, Deco, Murre, & Lamme, 2007).

Crowding and perceptual expertise

The relationship between crowding and perceptual experience has not been studied until recently. Crowding studies are typically conducted with participants well practiced or even pre-trained with the tasks and with small sets of stimuli (e.g. Louie, Bressler, & Whitney, 2007; Martelli, 2005; Petrov, Popple, & McKee, 2007; Saarela, Sayim, Westheimer, & Herzog, 2009; Tripathy & Cavanagh, 2002; Zhang, Zhang, Xue, Liu, & Yu, 2009), which perhaps reflects the implicit hypothesis that crowding is regarded as a sensory bottleneck and is not qualitatively affected by practice. Also, the critical distance for crowding to occur is thought to be independent of object category, regardless of whether we are experienced with the objects (e.g. letters or faces) or not (e.g. chairs or Gabor filters; Pelli & Tillman, 2008). However, evidence from recent studies suggests that perceptual experience affects the magnitude of crowding. For example, crowding for upright face recognition is stronger when flankers are upright faces compared to inverted faces (Louie et al., 2007). It is not simply caused by the higher similarity between upright face target and upright face distractors, because inverted face targets were crowded similarly with upright or inverted face flankers. As configural processing is more effective with upright than inverted faces (Farah et al., 1998; Maurer et al., 2002) and is linked to several domains of perceptual expertise (Bukach, Gauthier, & Tarr, 2006), this suggests that crowding is modulated by perceptual experience. More direct evidence comes from training studies in which recognition of a crowded letter can be improved with several hours of practice with the same task (Chung, 2007; Huckauf & Nazir, 2007), and the improvement generalized to untrained spacing between targets and flankers (Chung, 2007), suggesting that perceptual training can alleviate crowding. In addition, crowding effects in different visual quadrants were different for native English speakers compared to native Asian language speakers (Japanese, Chinese, Korean, etc.), which occurred when the flankers were Roman letters but not when flankers were false font characters or geometric shapes, again demonstrating experience-dependent modifications in visual crowding (Williamson, Scolari, Jeong, Kim, & Awh, 2009).

Crowding and music reading expertise

In music reading, the five-line staff is always presented in the same spatial region as the musical notes to serve as a spatial reference, and multiple notes are typically presented close to each other. Therefore the staff and the adjacent notes essentially create a ‘crowded’ image and make visual discrimination difficult, similar to the stimuli used to study crowding. Recognizing musical notes in multiple visual positions simultaneously is perceptually challenging, especially when the notes are almost always crowded by other notes and the staff lines, and yet music reading experts have acquired the skill to support rapid music reading. For example, experts can recognize four-note music sequences three times faster than novices, as revealed in the perceptual fluency measure in the previous studies (a mean of 265ms for experts and 857ms for novices, averaged from the two prior studies; Wong & Gauthier, 2010; in press). As crowding can be alleviated by perceptual experience (Chung, 2007; Huckauf & Nazir, 2007), music reading experts may have learned to ‘uncrowd’ the note patterns from the five-line staff and/or from adjacent notes compared to novices.

Significance of the crowding study

In the crowding experiment, I investigated the influence of the staff lines and flanker notes on recognition performance for musical notation in music reading experts and novices. It is interesting in its own right as a possible perceptual expertise marker for music reading expertise. This work can also contribute to the crowding literature, especially given that the relationship between crowding and perceptual expertise is largely unexplored. Furthermore, previous work suggests that crowding is, at least partly, associated with the primary visual cortex (Arman et al., 2006; Fang & Sheng, 2008; Tjan & Nandy, 2010). Examining the correlations between behavioral crowding effects and various ERP components can test the associations between crowding and early visual cortex.

Overview of the studies

In this dissertation, I tested the underlying mechanisms of such early visual recruitment with an ERP study (CHAPTER II), and investigated the effect of music reading expertise on crowding (CHAPTER III). The crowding study, together with other behavioral measures (perceptual fluency and holistic processing), served as behavioral correlates to any ERP expertise effects obtained in the ERP study (CHAPTER IV). I conclude my dissertation with CHAPTER V in which I discuss the implications of the expertise effects obtained with ERPs and crowding study, in terms of studies in music reading expertise, perceptual expertise and object recognition, and crowding in general.

CHAPTER II

THE ERP EXPERIMENT

The goal of the ERP experiment is to examine the temporal dynamics of the visual selectivity for musical notes with music reading expertise. In this study, single musical notes, single Roman letters or single pseudo-letters were presented briefly one after another, and participants were required to detect immediate repetitions of a stimulus (a one-back task). The use of single stimuli and one-back task, which involves simple visual judgments and has been shown to be effective in revealing visual expertise effects, allows us to link the findings in this ERP experiment to that of the prior fMRI study (Wong & Gauthier, 2010).

All the stimuli were either on a five-line staff or not (Fig. 8). The no-staff conditions were included to investigate whether any ERP selectivity for notes is dependent on associations with the identity of the notes (e.g. letter name, pitch, motor execution, etc.). By removing the staff, the pitch information of the notes is also removed, such that musical notes with different pitches become visually identical. As a result, it is no longer possible to individuate musical notes according to the pitch, but it still allows categorization of notes according to different rhythmic values (e.g. a fourth note or an eighth note). If an ERP effect is dependent on the individual identity of the notes (associated with different pitches and names), the effect should be observed for the on-staff conditions, but diminished or abolished for the no-staff conditions. Alternatively, if an effect arises because of the shape of the notes regardless of their spatial positions (or pitch information), the effect should be observed in both on-staff and no-staff conditions.

To look for expertise effect for musical notes, different ERP components were compared between musical notes and pseudo-letters across groups. An expertise effect would be reflected by a significant interaction between the group and stimulus conditions.

Method

Participants

The criteria for recruiting music reading experts and novices followed previous studies (Wong & Gauthier, 2010; in press). Participants were recruited from Vanderbilt University and the Nashville community for cash payment. All participants reported their amount of experience in music reading and rated their music-reading ability (1 = do not read music at all; 10 = expert in music reading) and their handedness was assessed with the Edinburgh Handedness Inventory (Oldfield, 1971). Eleven participants (including the author) who have at least 10 years of music reading experience and/or consider themselves music reading experts were recruited in the expert group (7 females and 4 males; mean age = 21.7, s.d. = 3.0; 9 right-handed and 1 left-handed), with 14.5 years of music reading experience and a self-rating score of 9.45 on average. Eleven participants who reported being unable to read music were recruited in the novice group (6 females and 5 males; mean age = 25.0, s.d. = 6.1; 9 right-handed and 1 left-handed), with 0.45 years of music reading experience and a self-rating score of 1.54 on average. All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University. They were paid $12 per hour of behavioral testing and $35 for the EEG experiment.

Stimuli and Design

The experiment was conducted on Mac Mini using Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). There were 18 black-and-white images in each of 3 object categories (musical notes, Roman letters and pseudo-letters; Fig. 8). The 18 musical notes were generated in Matlab and were 9 different notes (ranging from the 'E' on the bottom line to the 'F' on the top line) in two different time values, including quarter notes (a closed circle) and sixteenth notes (a closed circle with two tails). The Roman letters included 18 uppercase letters (excluding A, E, I, J, O, T, X and Z) in the Courier font. The 18 pseudo-letters were created by various combinations of the parts from the Roman letters with comparable complexity (Wong et al., 2005). The stimuli in all categories were shown either on a five-line staff or not (Fig. 8). For no-staff stimuli, 6 musical notes were used, including a quarter note (a closed circle), an eighth note (a closed circle with one tail) and a sixteenth note (a closed circle with two tails), either pointing upward or downward. Six Roman letters and 6 pseudo-letters were drawn from the set to keep the stimulus variability similar across stimulus conditions, and the chosen letters and pseudo-letters were counterbalanced across participants. All stimuli were presented with a visual angle of approximately 1.28° x 1.28° and a viewing distance of about 114 cm from the monitor.

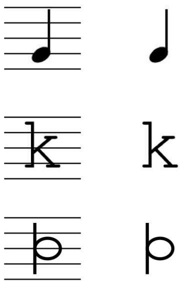

Figure 8. Examples of the single notes (top), Roman letters (middle) and pseudo-letters (bottom) in the ERP study either on an identical five-line staff (left column) or not (right column).

The mean luminance and mean contrast (Weber contrast) were matched across the three object categories. The mean luminance values were calculated by taking the mean of all pixel values from 0-255 (RGB values). The Weber contrast was calculated using the formula [(255 – mean luminance) / 255], where 255 refers to the RGB value of the white background of all stimuli. For the with-staff condition, the notes, letters and pseudo-letters had a mean luminance of 222.3, 222.3 and 222.7 (with s.d. = 2.10, 2.08 & 2.39) and a mean contrast of 0.128, 0.128 and 0.127 (with s.d. = .0080, .0082 & .0096) respectively. For the no-staff condition, the notes, letters and pseudo-letters had a mean luminance of 243.5, 242.7 and 243.9 (with s.d. = 2.51, 2.72 & 2.74) and a mean contrast = 0.045, 0.048 and 0.044 (with s.d. = .0096, .011 & .011) respectively. One-way ANOVAs on Category revealed that the luminance and the contrast between different object categories for the on-staff conditions or for the no-staff conditions were similar, with all Fs(1,51) < 1.

Following the design of the fMRI study that motivated this experiment (Wong & Gauthier, 2010), a one-back task was used, in which each of the six object categories (notes, letters and pseudo-letters, with or without staff) was presented in blocks of 6 trials

(Fig. 9). Each block began with a black fixation dot at the center of the screen for 500ms, followed by six trials, each with a stimulus presented for 700ms and then the black fixation dot presented for a randomized period of 250-450ms. Then the fixation dot turned grey for 2s and turned black for 200ms cuing the start of the next block. Participants were required to press a key on a gamepad (with the right thumb) as fast as possible when they detected a repeat of the stimulus. Participants were instructed to maintain fixation throughout the whole block and were encouraged to blink only during the grey dot period.

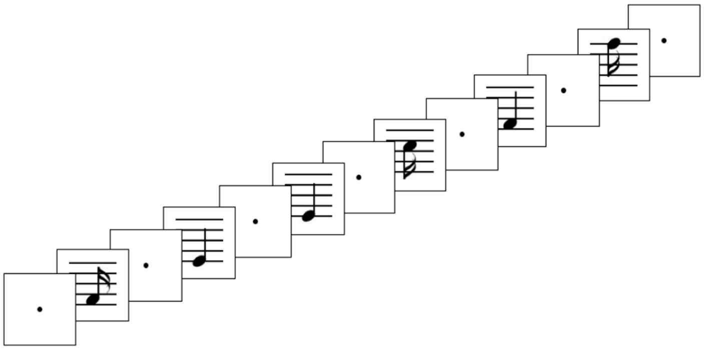

Figure 9. The one-back task used in the ERP study.

For the note condition, the stimulus order was constrained such that participants could perform the task based on the parts of the notes. That is, for notes on the staff, consecutive stimuli always pointed to different orientations or had different number of tails unless they repeated. For notes with no staff, consecutive stimuli always pointed to different orientations unless they repeated. These constraints ensured that the task was sufficiently easy for novices and that the task was not affected by removing the staff from the notes. Participants were explicitly informed about these constraints. Similar to the fMRI study (Wong & Gauthier, 2010), the stimuli were spatially jittered five pixels in all four directions randomly to reduce visual adaptation (especially for the staff conditions, in which the separation between the staff lines was about 11 pixels). There were 720 trials for each of the six object categories, including 60 repeated trials (repeat rate = 8.3%). The order of the blocks for different object categories was counterbalanced. The trials were divided into 8 runs, and participants were encouraged to take frequent rests (provided about every 4.5 min of testing). Participants received feedback on accuracy and response time on the screen every 30 blocks of trials, and were given constant verbal feedback on eye movements within the blocks. They were given 72 trials for practice before the test. The whole experiment took around 3 hours including the setup of the sensor cap for EEG recording.

Recording and analysis

The electroencephalogram (EEG) was recorded from tin electrodes held on the scalp by an elastic cap (Electrocap International, Eaton, OH). A subset of the International 10/20 System sites was used (F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, PO3, PO4, O1, and O2) in addition to non-standard sites OL (halfway between O1 and T5) and OR (halfway between O2 and T6). The right mastoid electrode served as the reference site. The signals were re-referenced off-line to the average of the left and the right mastoids (Nunez, 1981). The electrooculogram (EOG) was recorded using electrodes positioned 1 cm lateral to the external canthi to measure horizontal eye movements and with an electrode beneath the left eye, referenced to the right mastoid, to measure vertical eye movements and blinks. The EEG and EOG were amplified by an SA

Instrumentation amplifier with a gain of 20,000 and a band-pass filter of 0.01–100 Hz. The amplified signals were digitized at 250 Hz by a PC-compatible computer and averaged off-line. Trials associated with behavioral responses (all repeated trials) or those accompanied with artifacts (eye movement or blocking) were excluded from the averages.

Data analysis was performed with ERPSS (The Event-Related Potential Software System; 1993), Matlab, and with the EEGLAB toolbox (version 8.0). For ocular artifact rejection, a two-step procedure was used that has been described previously (Woodman & Luck, 2003). Briefly, the cross covariance between the single-trial EOG waveform and a 100ms step function was computed, and trials with maximum covariance exceeding a certain threshold were rejected. The threshold was adjusted for each participant according to visual inspection of the EOG waveforms. The averaged horizontal EOG (HEOG) waveforms were used to reject any participants with systematic unrejected eye movements. Trials with blocking (saturated activity) were rejected if the signal was at the maximum or minimum value for 20ms or consistently hovering around the maximum or minimum value (> 40 data points in a 1s time window). One expert and one novice with more than 25% of the trials rejected were excluded from all analyses. On average, 9.9% and 10.7% of the trials were rejected for the expert group and novice group, respectively. The ERPs were baseline-corrected with respect to 200ms pre-stimulus interval (except the analyses for the Contingent Negative Variation, CNV; see below).

Measure of perceptual fluency

The perceptual fluency for music sequences was measured for each participant to quantify music-reading expertise. A sequential matching paradigm with 4-note music sequences was used (Wong & Gauthier, 2010; in press). On each trial, a fixation cross was presented at the center of the screen for 200ms, followed by a 500ms pre-mask, a target four-note sequence for a varied duration and, after a 500ms post-mask, two four-note sequences appeared side-by-side, one identical to the first sequence, and the other with one of the notes shifted by one step (randomly chosen out of the four notes, with the up/down shifts counterbalanced). The task was to select the matching sequence by key press. The perceptual threshold was estimated using QUEST (Psychtoolbox; Watson & Pelli, 1983), and was quantified as the duration of the target sequence required to keep performance at 80% accuracy. Sequences were randomly generated using notes ranging from the note below the bottom line (a 'D' note) to the note above the top line (a 'G' note). Contrast for all the stimuli was lowered by about 60% to avoid a ceiling effect. The threshold was measured four times, each with 40 trials, and the thresholds were averaged.

To control for individual differences not specifically tied to expertise with notes, perceptual fluency for four-letter strings was also measured in an identical procedure. The four-letter strings were randomly generated with 11 letters: b,d,f,g,h,j,k,p,q,t,y. These letters were selected because they contain parts extending upward or downward, similar to musical notation. To create the distractor string, one of the four letters was chosen (counterbalanced across stimuli) and replaced by a different letter randomly drawn from the set. The string stimuli were also shown with the same lowered contrast as the note sequences.

Predictions for the ERP results

In the following sections, I briefly discuss the predictions for observing expertise effects at different time windows corresponding to various ERP components, including the C1, N170, P3 and the Contingent Negative Variation (CNV), and their implications to the mechanisms underlying the recruitment of early visual areas for music reading expertise (Wong & Gauthier, 2010).

Expertise effect for C1

If the early visual recruitment is (at least partly) a feedforward effect generated by altered response of V1 cells, an expertise effect should be observed with the C1 component. The C1 component is the first major visual ERP component that typically onsets 40-60ms post-stimulus and peaks 80-100ms post-stimulus, and is largest in the midline posterior electrode sites (Luck, 2005). The polarity and the topography of C1 change as a function of stimulus positions on the visual field. In particular, C1 is positive when stimuli are presented in the lower visual field and is negative when stimuli are presented in the upper visual field (Clark, Fan, & Hillyard, 1995), consistent with the anatomy of the calcarine fissure, in which the lower visual field is represented on the upper bank of the fissure and the upper visual field on its lower bank. For this reason, and based on other source localization analyses, the C1 component is thought to be generated in V1 (Clark et al., 1995; Clark & Hillyard, 1996; Foxe et al., 2008; Jeffreys & Axford, 1972; Kelly, Gomez-Ramirez, & Foxe, 2008; Martinez et al., 1999; Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Pourtois et al., 2008; Proverbio & Adorni, 2009; Stolarova, Keil, & Moratti, 2006). However, such organization where the upper and lower visual field are represented in the ventral and dorsal region respectively is also shared by V2 and at least part of V3 (Sereno et al., 1995; Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998), such that obtaining a component with reversed polarity in the upper or lower visual field does not necessarily imply that the component is generated in V1 (Foxe & Simpson, 2002).

Although it is debated whether the C1 is solely generated in V1, the early timing of the C1 component still largely constrains the source of this effect. In one study, scalp current density analyses revealed that the early portion of C1 (before 56ms) had only one single generator at the posterior midline site, while multiple source generators were observed at the occipitoparietal regions by 75ms. Based on this finding, it has been proposed that only the early portion of C1 (around 45-60ms) was dominated by V1 activity (Foxe & Simpson, 2002). This time window is before the typical onset of the next component, P1 (around 60-90ms), which is thought to be generated in extrastriate cortex (Luck, 2005), and is consistent with single-cell recording data with macaque monkeys suggesting that most of the cells active 60ms post-stimulus are in the LGN and V1 (Schmolesky et al., 1998).

The C1 is highly sensitive to stimulus properties such as contrast and spatial frequency (Luck, 2005), and is modulated by spatial attention (Kelly et al., 2008), and emotional content of the stimulus (Pourtois et al., 2004). In addition, the C1 is modulated by multisensory processes, such as audiovisual stimuli compared to unimodal stimuli in a discrimination task (Giard & Peronnet, 1999) and a speeded detection task (Fort, Delpeuch, Pernier, & Giard, 2002), or even when the target was presented in conjunction with task-irrelevant sensory information in a non-target modality (Karns & Knight, 2008). Importantly, the C1 can be modulated by perceptual experience. For example, perceptual learning with texture discrimination led to a reduced C1 component in the trained compared to the untrained visual field quadrant (the left or right upper visual field; Pourtois et al., 2008). Also, associating a grating with threat-related stimuli modulated the C1 response for the grating compared to that associated with neutral pictures, suggesting that short-term conditional learning modulates the C1 for conditioned stimuli (Stolarova et al., 2006).

The C1 is small for stimuli on horizontal midline, as in the case for the current ERP study, such that it is easily combined with the P1 wave and becomes difficult to observe (Luck, 2005). While a clear C1 may not be obtained with stimuli presented at fovea, it may be possible to capture its modulation by expertise, as suggested by the robust expertise effect for notes in V1 obtained in fMRI (Wong & Gauthier, 2010; for C1 effects with midline stimulus presentation, see also Foxe et al., 2008; Proverbio et al., 2009).

However, unlikely typical C1 studies, a clearly identifiable C1 component (e.g. with a clear onset and offset of the component) might not be observed from the waveform under foveal presentation of the stimuli. Therefore, it may be arguable whether any expertise modulation during this time window (if found) is indeed a C1 effect or not. Note that calling such modulation a C1 effect would be largely based on the early timing and a topographic distribution of the effect that are consistent with the typical characteristics of the C1, and referring this to the C1 component is consistent with the literature (Foxe et al., 2008; Proverbio et al., 2009). Although calling it a C1 effect may be controversial, the naming of such effect does not affect the interpretation of this effect since any effect observed as early as 40-60ms would correspond well to V1 activity, and hence largely constrained the source of the expertise effect.

In sum, it is reasonable to look for expertise effects for musical notes in the C1 components. If expertise effects with musical notes can be observed as early as the 40-60ms time window, the early visual recruitment is at least partly feedforward and is likely to be heavily contributed by V1, consistent with the fMRI results that V1 is recruited in music reading expertise (Wong & Gauthier, 2010).

Expertise effect for N170

As mentioned above, perceptual expertise has been associated with changes in the amplitude and/or latency of the N170 component in various object categories, including faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996), dogs and birds (Tanaka & Curran, 2001), cars (Gauthier et al., 2003), fingerprints (Busey & Vanderkolk, 2005) and Chinese characters (Wong et al., 2005), and this difference arises with novel object categories (e.g. Greebles) only after perceptual expertise training (Rossion et al., 2002; Rossion et al., 2004). Therefore, it is predicted that a similar N170 expertise effect would be obtained with musical notes compared to novel categories like pseudo-letters in ventral temporal recording sites. Also, letter selectivity for the N170 for letters is expected in the left hemisphere only, similar to that observed in the previous study with a similar design (Wong et al., 2005).

The N170 is often thought to originate from the inferior occipitotemporal cortex (Allison, Puce, Spencer, & McCarthy, 1999; Bentin et al., 1996; Horovitz, Rossion, Skudlarski, & Gore, 2004), and is attributed to perceptual mechanisms rather than higher processes such as semantics (Rossion et al., 2004; Wong et al., 2005). If the N170 is the earliest expertise effect observed for musical notation at ventral temporal recording sites, the engagement of the early visual cortex for notes may be related to a feedforward-feedback loop between early visual areas and ventral temporal areas, similar to the case of spatial attention (Martinez et al., 1999).

Expertise effect for P3

Perceptual expertise effects are not often associated with the P3, perhaps because late components are more susceptible to influences of earlier components such as the N170 effects, which make it hard to interpret differences observed at the P3. A recent study reported an expertise effect for the P3 with Chinese characters, which was not a carry-over effect from earlier N170 differences (Wong et al., 2005). Similarly, it is possible to observe expertise effects for the P3 with music reading expertise.

The P3 is heavily modulated by top-down influences, including expectation, task relevancy and predictability (Coles & Rugg, 1995; Luck, 2005; Sutton et al., 1965). If the P3 is the earliest expertise effect observed for musical notation, the early visual recruitment is likely to be a result of strengthened feedback from higher areas.

Expertise effect for CNV

Since a block design was used in the current ERP study, participants were able to anticipate the category of the upcoming stimulus (except the first stimulus of each block) in a relatively short time window (250-450ms inter-stimulus interval). Therefore a slow negative potential before the presentation of the each stimulus was expected, which is called the Cognitive Negative Variation (the CNV; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The CNV is a slow negative component that develops during the anticipatory period of an upcoming event (typically a sensory stimulus such as a tone or a light) over a few hundred milliseconds to a few seconds in the fronto-central region, and is terminated by the presentation of the anticipated event (Luck, 2005; Walter et al., 1964). It has both a cognitive component and a motor component, and depends on expectancy, predictability, task relevancy, and whether there is any expected motor response to the upcoming event (Leuthold, Sommer, & Ulrich, 2004; Walter et al., 1964). If the anticipatory period is as long as a few seconds, several sub-components can be observed, in which the early component is more cognitively-related and the late component is more related to motor preparation and execution (Brunia, Van Boxtel, & Bocker, in press; Leuthold et al., 2004).

The CNV increases monotonically with response time in a task requiring speeded key press responses (Loveless, 1973), and is task-dependent, such as a shallow or deep processing of a word (Leynes, Allen, & March, 1998) or a verbal or spatial judgment of a stimulus (McEvoy, Smith, & Gevins, 1998). Importantly, the CNV is modulated by expertise or learning experience. For example, the CNV was of different magnitudes when musicians performed different judgments on auditory-presented chords, but not for non-musicians (Muller, Hofel, Brattico, & Jacobsen, 2010). In a go/no-go task simulating driving conditions, the CNV difference between go trials and no-go trials was larger for professional taxi drivers compared to controls (Belkic, Savic, Djordjevic, Ugljesic, & Mickovic, 1992). In a speeded-response task, the CNV was more negative for participants who had low compared to intermediate meditation experience, and intermediate to high meditation experience (Travis, Tecce, & Guttman, 2000). In the same study, when a distraction task was added during the anticipatory period before the speeded response, the CNV magnitude for people with higher meditation experience were less affected, suggesting that the CNV reflects allocation of attentional resources and can be modulated by meditation experience. More direct evidence comes from studies showing practice or learning effects of the CNV with the same group of participants. In one study (Rose, Verleger, & Wascher, 2001), a speeded-choice response (with the left or right hand) was required, and the choice of hands was first cued before the anticipatory period with a 100%, 50% or 0% informative cue stimulus. The magnitude of the CNV decreased with time as participants learned the associative meaning of the cue stimulus. The learning effect was only found for the 50% and 100% informative cue stimulus, suggesting that the CNV became smaller with associative learning. In another working memory study (McEvoy et al., 1998), the magnitude of the CNV increased in the third testing session compared to the first session (accompanied with improved behavioral performance), suggesting that practice on the same task leads to changes in the CNV component. These studies suggest that it is plausible to observe expertise effect with the CNV in the current ERP study. Therefore the group differences of the CNV component were also analyzed.

Behavioral Results

Perceptual fluency

As expected, experts demonstrated a higher perceptual fluency than novices for music sequences but not for letter strings. A one-way ANOVA for Group (Experts /

Novices) was performed on the perceptual threshold for matching four-note sequences. The main effect of Group was significant, $F(1,18) = 47.0$, $p \leq .0001$, such that the perceptual threshold for experts (mean = 341.6 ms) was faster than that for novices (mean = 1098.0 ms). In contrast, the main effect of Group for matching four-letter strings was not significant, $F(1,18) < 1$, with a mean perceptual threshold of 194.4 ms and 232.9 ms for experts and novices respectively. This confirms that our criterion identifies experts who have superior perceptual fluency for reading music sequences, which cannot be explained by a general perceptual advantage.

Behavioral result of the ERP study

For the one-back task, only repeated trials that prompted a behavioral response were included in data analysis, and the analysis of RT was performed on correct trials only. For the staff conditions, a 2x3 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Letters / Pseudo-letters) on accuracy revealed a main effect of Stimulus, $F(2,36) = 6.34$, $p = .004$ (Fig. 10). LSD tests ($p < .05$) revealed that the accuracy was better for Roman letters than for notes. The interaction between Group and Stimulus was not significant, $F(2,36) < 1$. No main effects or interactions reached significance for response time.

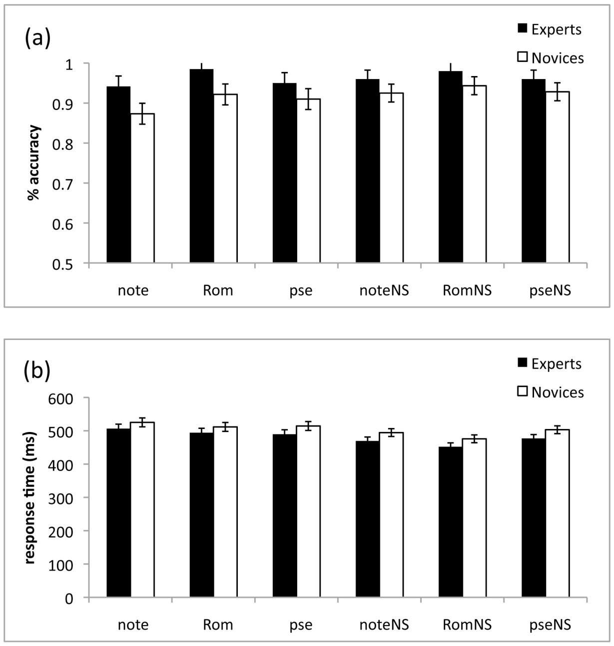

Figure 10. Accuracy (a) and response time (b) for all the stimulus categories in the one-back task. Error bars show the 95% CI for the Group x Stimulus interaction for the staff condition and no-staff condition respectively. 'NS' refers to no-staff conditions.

For no-staff conditions, a 2x3 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Letters / Pseudo-letters) on accuracy did not reveal any significant effects (all *p*s > .14; Fig. 10). For response time, the main effect of Stimulus was significant, *F*(2,36) = 10.8, *p* = .0002. LSD tests (*p* < .05) revealed that the response time for Roman letters was faster than that for notes or pseudo-letters. The interaction between Group and Stimulus was not significant, *F*(2,36) < 1.

In sum, the performance on the one-back task was similar across the two groups, replicating the findings in the similar fMRI study (Wong & Gauthier, 2010).

ERP results

In the following sections, I first report the ERP expertise effects with musical notation, in which different ERP components were compared between musical notes and pseudo-letters across groups. The next section reports the results for letters, using the pseudo-letters as a novel category to look for letter selectivity for various ERP components that are common for both groups. These two sections start with results involving stimuli presented on the staff ('on-staff' conditions), followed by results with stimuli presented on a white background ('no-staff' conditions). The on-staff conditions and the no-staff conditions were analyzed separately such that the staff background would not be a stimulus confound (either present in all stimulus categories or in none of the categories). For each subsection, four types of analyses are reported: (1) the expertise effects with the early portion of the C1 component (40-60ms); (2) the N170 component (120-200ms); (3) the P3 component (300-600ms); and (4) the CNV component (-200 to 0ms before stimulus onset). Lastly, since some of the ERP effects were different across the on-staff and no-staff conditions, I directly compare the findings between the on-staff and no-staff conditions to investigate if any of the ERP expertise effects are dependent on the staff.

Musical notation (on-staff)

To look for expertise effects for musical notes, the neural selectivity for musical notes was examined for each of the ERP components. The average scalp voltage was computed for each stimulus condition within the corresponding time window. Then, the scalp voltage was compared between musical notes and pseudo-letters across the expert and novice groups. An expertise effect would be reflected by an interaction between the group and stimulus conditions.

Expertise effect for C1

The early portion of the C1 component (40-60ms) was examined to look for expertise effects contributed by feedforward visual processes. The topographic distribution of this Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) interaction was maximal along the posterior parietal midline recording sites (Fig. 11), consistent with that of the C1 component (Clark et al., 1995; Luck, 2005). The C1 analyses were focused on PO3/PO4 and Pz where the interaction effect was maximal (Fig. 11).

For PO3/PO4, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) x Hemisphere (Left / Right) on the C1 revealed a main effect of Group, $F(1,18) = 4.69$, $p = .044$, with a more positive response for experts than novices

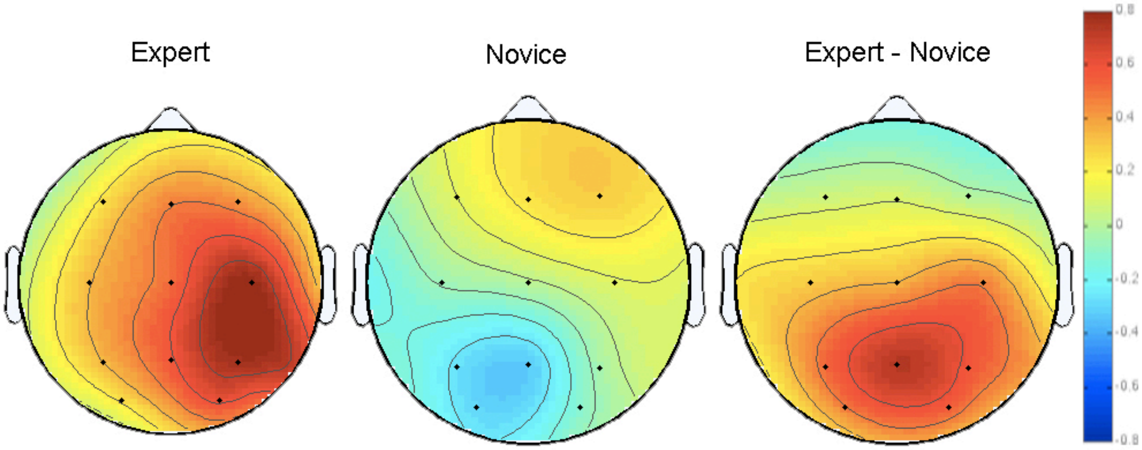

Figure 11. Topographic distributions of ERP differences with the contrast of [notes - pseudo-letters] for the C1 for the on-staff conditions in experts (left), novices (middle) and the difference between the two groups (right).

in general. The Stimulus x Hemisphere interaction was significant, $F(1,18) = 8.10$, $p = .011$, such that the voltage was more positive for notes than that for pseudo-letters on the right hemisphere, but voltages were similar on the left hemisphere (Scheffé tests, $p < .05$; Fig. 12 a-b; 13a). Importantly, the Group x Stimulus interaction was significant, $F(1,18) = 7.05$, $p = .016$. Scheffé tests ($p < .05$) revealed that the C1 was more positive for notes than for pseudo-letters in experts, but not in novices. This suggests that the C1 is selective for notes with expertise. This C1 expertise effect did not interact with Hemisphere ($p > .6$).

For Pz, the Group x Stimulus interaction was significant, $F(1,18) = 10.5$, $p = .0045$ (Fig, 12c; 13b). Scheffé tests ($p < .05$) revealed that, in experts, the C1 was more positive for notes than for letters. In contrast, in novices, the C1 was more negative for notes than for letters, again suggesting that the C1 is selective for notes with music reading experience.

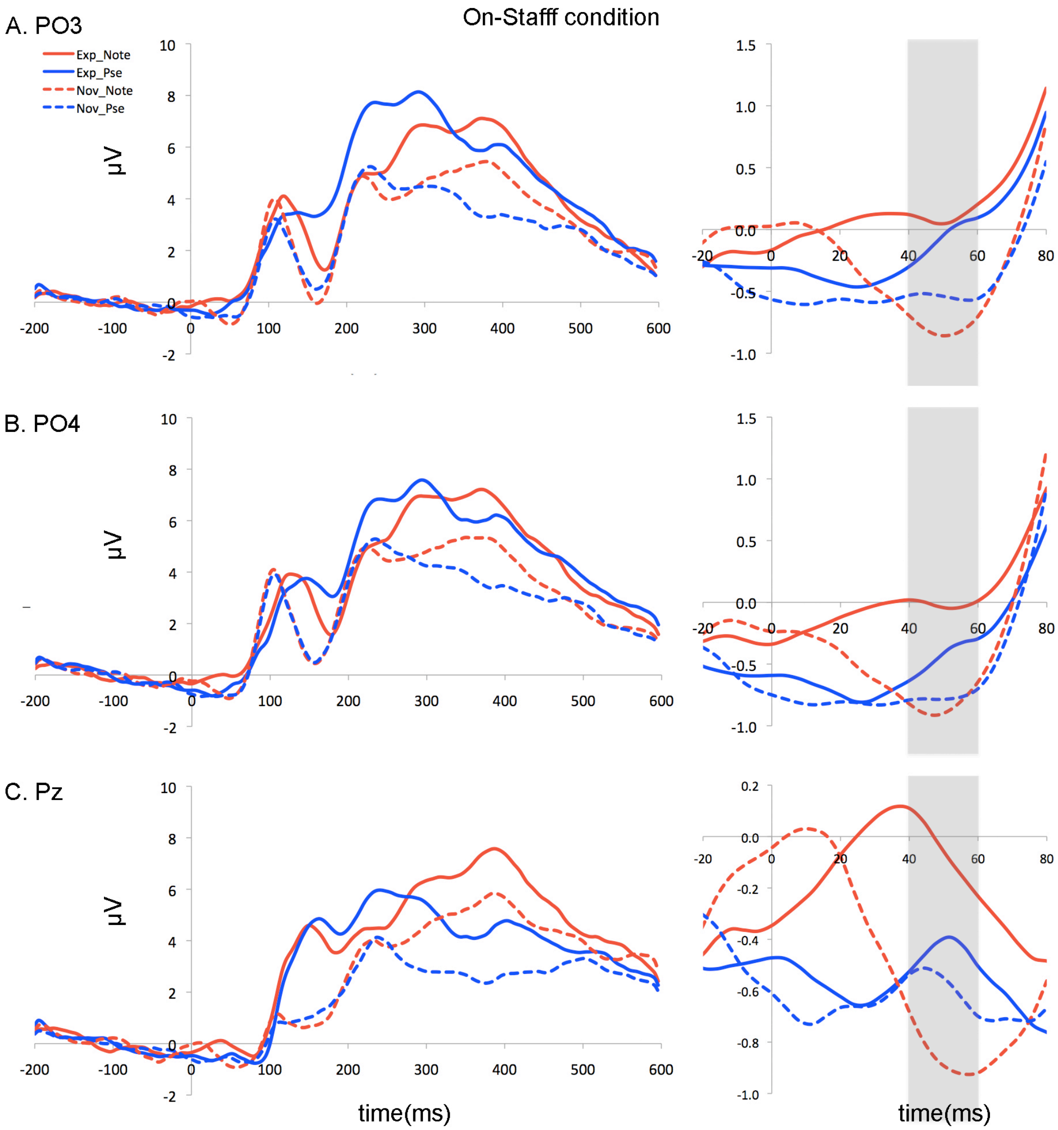

Figure 12. ERPs for the on-staff conditions on the posterior parietal channels, including (A) PO3; (B) PO4; and (C) Pz. Solid lines plot the activity for experts and dashed lines plot that for novices, with notes in red and pseudo-letters in blue. Left graphs show the ERPs from -200ms to 600ms, with the ERPs during the first 80ms highlighted on the right. The grey bars represent the early portion of the C1 (40-60ms).

C1 effect replicated with split-half data

To further test the reliability of the results, in particular the C1 expertise effects that occurred early in time, all the unrejected trials were split into the 1st half and the 2nd half for the same analyses. Results were generally replicated with this split-data method for PO3/PO4 and Pz. For PO3/PO4, the Group x Stimulus interaction approached significance for both the 1st half, *F*(1,18) = 3.29, *p* = .08 and for the 2nd half, *F*(1,18) = 3.52, *p* = .07. Similarly, for Pz, the Group x Stimulus interaction also approached significance for the 1st half, *F*(1,18) = 3.87, *p* = .065 and was significant for the 2nd half, *F*(1,18) = 7.08, *p* = .016.

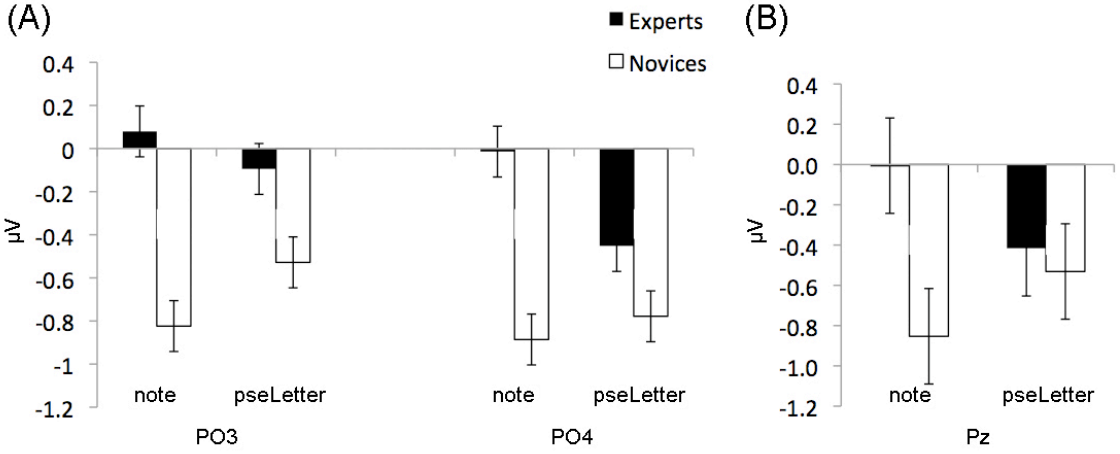

Figure 13. Averages of the scalp voltages for the C1 for the on-staff conditions in (A) PO3/PO4 and (B) Pz. Error bars plot the 95% CI for the highest order interaction for each graph.

In sum, the C1 expertise effects were robust effects since they were replicated across both halves of the data on PO3/PO4 and Pz. Next, I explore whether the C1 expertise effects could be explained by pre-stimulus noise, eye movements or an artifact caused by baseline correction.

C1 effect not caused by pre-stimulus noise

The C1 effects was tested in the time bin between -40 and -20ms, -20 and 0ms, 0 and 20ms, and 20 to 40ms in all the above channels (PO3/PO4 and Pz). Results revealed that none of the Group x Stimulus interaction effect was significant, with all *p*s > .2, except two time windows where the p-values were close to .1 (-40 to -20ms for Pz, *p* = .15; 20 to 40ms for PO3/PO4, *p* = .13), suggesting a trend for an interaction effect. However, none of these trends were replicated in the two split-half datasets (-40 to -20ms for Pz: *p* = .19 for the 1st half and *p* = .38 for the 2nd half; 20 to 40ms for PO3/PO4: *p* = .46 for the 1st half and *p* = .10 for the 2nd half). This suggests that the early visual effects did not exist at the stimulus onset or during the pre-stimulus baseline.

C1 effect not explained by eye movements

To test whether eye movements can explain the early visual effects, the EOG in both the vertical eye channels (VEOG) and the horizontal eye channels (HEOG) were examined. If the early visual effects were caused by systematic eye movements, the Group x Stimulus differences should be found in either the VEOG or HEOG channels in the same time window (40 – 60ms). Results revealed that the Group x Stimulus interaction on the C1 was not significant for VEOG (*p* = .13) or HEOG (*p* = .28). The trend for interaction for the VEOG was not replicated across the two split-half datasets (*p*

= .11 for the 1st half; p = .77 for the 2nd half), suggesting that eye movements cannot account for the C1 effects.

C1 effect not caused by baseline correction

Since the waveforms were baseline-corrected by the average voltage of the pre-stimulus period between -200 and 0ms (i.e., this average voltage was assigned to be ‘0 μV’ for that trial), one may worry that the C1 effects were artifacts created by baseline correction. To test this alternative hypothesis, data analyses were performed again with minimal baseline correction using the average of 4 data points before stimulus onset (-12ms to 0ms). Four data points before 0ms was used instead of using the single time point at 0ms (i.e., no baseline correction) because comparing the voltage difference against one data point is susceptible to high frequency noise. Results showed that all the Group x Stimulus interactions approached significance using this measure (p = .075 for PO3/PO4; p = .078 for Pz), with a similar pattern such that the voltage for notes but not for pseudo-letters was different across groups for all channels (Scheffé tests, $p < .05$). These results suggest that the C1 effects were not caused by baseline correction.

In sum, early expertise effects were obtained as early as 40ms for notes on the staff. The timing and the topographic distribution of these effects were consistent with that of the C1 component, suggesting that V1 is responding differently to notes compared to other objects because of extensive music reading experience.

Expertise effect for N170

Is the N170 modulated by expertise for musical notes? To address this question, the N170 for notes was compared to the N170 for pseudo-letters on the OL/OR and T5/T6 channels (Wong et al., 2005). The topographic distribution for the selectivity for notes was consistent with the occipital-temporal distribution of the typical N170 effects (Fig. 14).

For OL/OR, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) x Hemisphere (Left / Right) on the N170 revealed a Group x Stimulus interaction, $F(1,18) = 5.05$, $p = .037$ (Fig. 15 top; 16a). Scheffé tests ($p < .05$) revealed that the N170 for notes was more negative than the N170 for pseudo-letters for experts but not for novices. Also, the Stimulus x Hemisphere interaction was significant, $F(1,18) = 6.13$, $p = .024$, with the N170 for notes more negative than the N170 for pseudo-letters for the left, but not the right hemisphere. The Group x Stimulus interaction did not interact with Hemisphere ($p = .17$).

For T5/T6, a similar pattern was found: the Group x Stimulus interaction was significant, $F(1,18) = 5.99$, $p = .025$; and the Stimulus x Hemisphere interaction was significant, $F(1,18) = 7.11$, $p = .016$ (Fig. 15 bottom; 16b).

Since the N170 expertise effect might be a carry-over effect from the previous P1 component (60-120ms), the same analyses were performed on the P1 in these channels. For both OL/OR and T5/T6, the only significant effect was a main effect of Stimulus (for OL/OR, $F_{1,18} = 28.1$, $p \leq .0001$; for T5/T6, $F_{1,18} = 39.9$, $p \leq .0001$), and no effects involving Group reached significance (all ps > .2). This suggests that the N170 effect was not caused by differences that were already observed earlier in time.

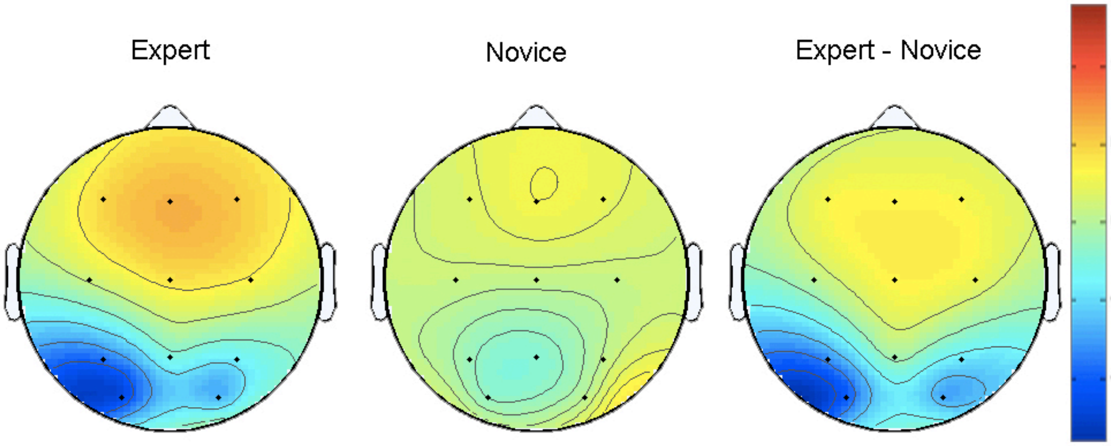

Figure 14. Topographic distributions of ERP differences with the contrast of [notes - pseudo-letters] for the N170 for the on-staff conditions in experts (left), novices (middle) and the difference between the two groups (right).

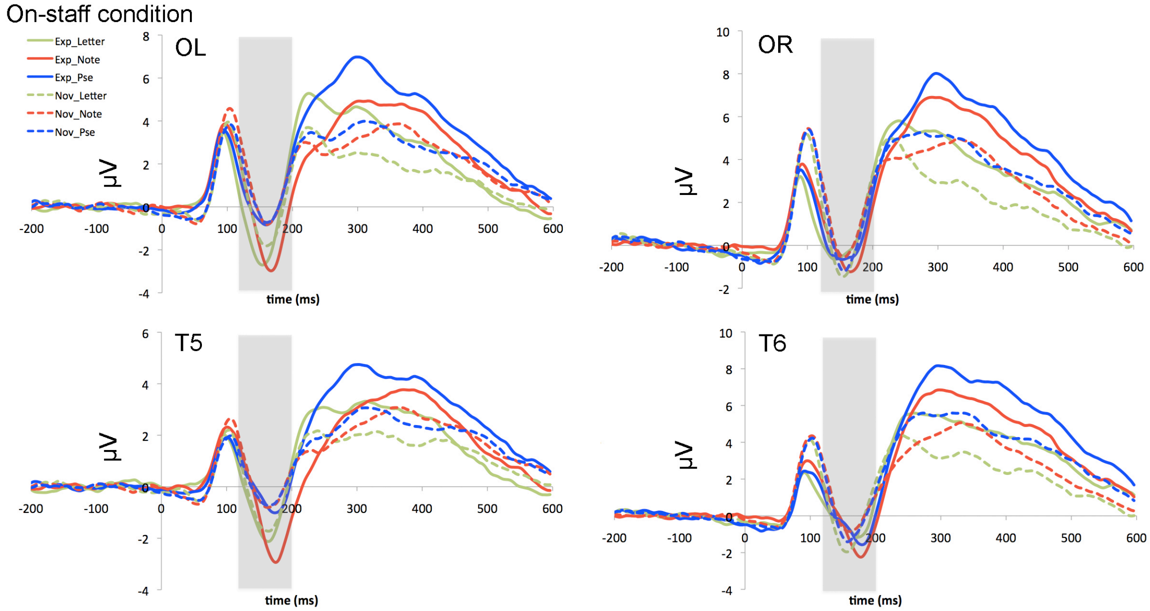

Figure 15. ERPs for the on-staff conditions for the N170 components in OL/OR (top) or T5/T6 (bottom). Solid lines plot the activity for experts and dashed lines plot that for novices, with notes in red, letters in green and pseudo-letters in blue. The grey bars represent the time window for the N170 component (120-200ms).

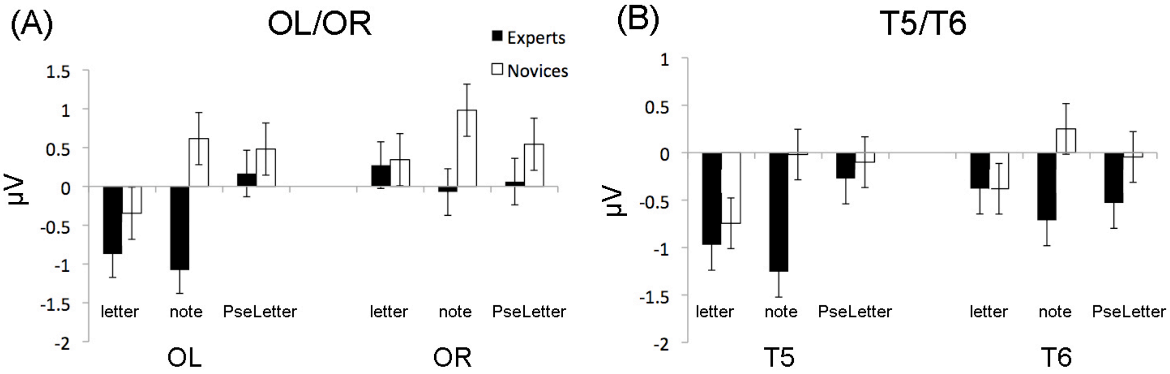

Figure 16. Averages of the scalp voltages for the N170 for the on-staff conditions in (A) OL/OR and (B) T5/T6. Error bars plot the 95% CI for the Group x Stimulus x Hemisphere interaction for each graph.

In sum, the N170 expertise effects for notes were obtained in both hemispheres for notes on the staff, and these effects were similar to that obtained for the other kinds of perceptual expertise (Bentin et al., 1996; Busey & Vanderkolk, 2005; Gauthier et al., 2003; Rossion et al., 2002; Tanaka & Curran, 2001; Wong et al., 2005).

Expertise effect for P3

To examine if the P3 component was modulated by music reading expertise, a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) was performed on the P3 component on Pz, Cz and Fz (Wong et al., 2005). All channels revealed a significant main effect of Stimulus, such that the P3 for notes was larger than that for pseudo-letters (for Pz, $F_{1,18} = 49.2$, $p \leq .0001$; for Cz, $F_{1,18} = 89.4$, $p \leq .0001$; for Fz, $F_{1,18} = 69.7$, $p \leq .0001$). However, no Group x Stimulus interaction was found in any channels (all *p*s > .3). In other words, no expertise effect was found for notes on the staff for the P3.

Expertise effect for CNV

To test if the anticipatory effect for the CNV component was modulated by music reading expertise, average scalp voltage (-200 to 0ms) was examined before baseline correction such that any anticipatory negativity accumulated before the onset of each stimulus could be examined. The topographic distribution of the Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) interaction revealed that the effects were centered at Cz (Fig. 17), consistent with the typical distribution of the CNV component in the central-frontal region (Coles & Rugg, 1995; McEvoy et al., 1998; Rose et al., 2001; Travis et al., 2000). The effect was distributed towards the left hemisphere, consistent with the motor-preparation component of the CNV, since participants responded by the right thumb on a gamepad (Leuthold et al., 2004; Walter et al., 1964). The CNV analyses focused on the Cz where the effect was maximal (Fig. 17).

For Cz, a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) on the scalp voltage between -200 and 0ms revealed a main effect of Stimulus, $F(1,18) = 19.2$, $p = .0004$, in which the CNV for notes was more negative than the CNV for pseudo-letters. The Group x Stimulus interaction was significant, $F(1,18) = 4.65$, $p = .045$ (Fig. 18-19). Scheffé tests ($p < .05$) revealed that the CNV for notes was more negative than that for pseudo-letters for experts but not for novices. This effect was not a carry-over effect from the previous P3 component, since there was no Group x Stimulus effect obtained at Cz for the P3 (see above). The expertise effect with the CNV component for musical notes suggests that the anticipation for notes is altered by music reading expertise.

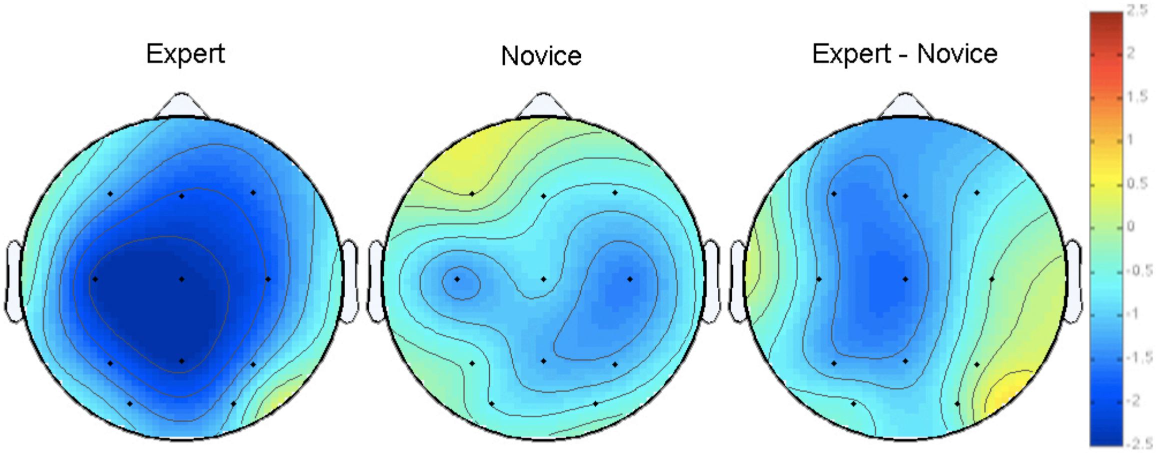

Figure 17. Topographic distributions of ERP differences with the contrast of [notes - pseudo-letters] for the CNV for the on-staff conditions in experts (left), novices (middle) and the difference between the two groups (right).

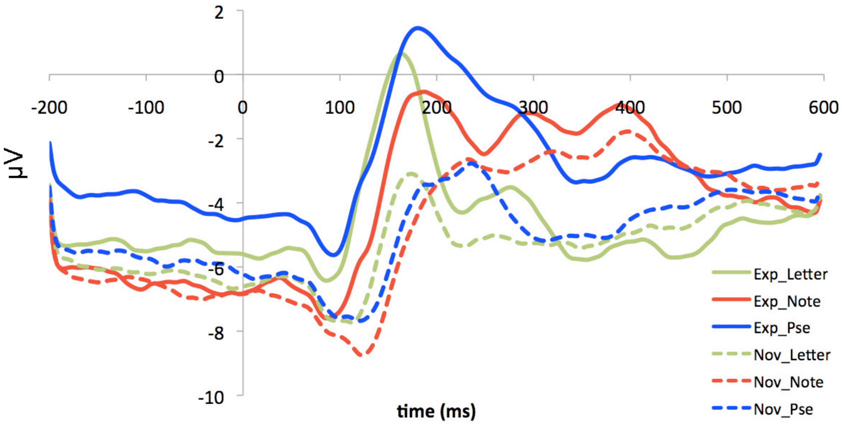

Figure 18. ERPs for the on-staff conditions for the CNV at Cz. Solid lines plot the activity for experts and dashed lines plot that for novices, with notes in red, letters in green and pseudo-letters in blue. The waveforms show ERP activity before baseline correction.

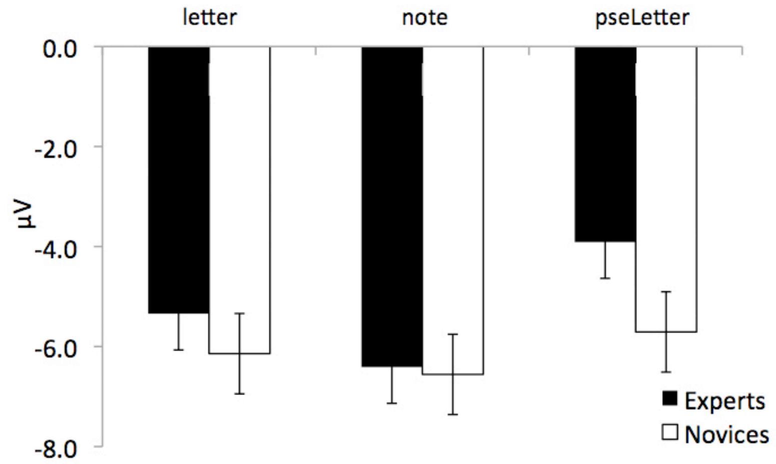

Figure 19. Group means for the scalp voltages for the CNV component for the on-staff conditions. Error bars plot the 95% CI for the Group x Stimulus interaction.

Dissociating the CNV effect from the C1 effect

Is it possible that the C1 effect was accounted for by the pre-stimulus CNV differences across groups? The two components appear to be dissociable based on several pieces of evidence. First, the topographic distributions of the two components are different. The CNV was distributed on the central sites and lateralized towards the left hemisphere (Fig. 17). In contrast, the C1 effect was distributed bilaterally at the posterior parietal and midline sites (Fig. 11). Second, the effects have different time signatures. The CNV was a slow and steady negativity observed in a relatively broad time window before stimulus onset (-200 to 0ms), while the C1 effect was a transient effect found at 40 to 60ms, not before 40ms (from -40ms to 40ms) and not after 60ms (the P1 component; *p*s > .2 for the Group x Stimulus interaction at PO3/PO4 or Pz). Third, the C1 effect can still be observed after filtering that removes the expertise effect in the CNV component. Since the CNV component is a slow waveform, the slow change across time can be removed by using a high pass filter of 2 Hz. After the filtering, the expertise effect in the

CNV was no longer significant at Cz (Group x Stimulus interaction, $F < 1$) or at any other channels (all *p*s > .2), while the C1 effect was still significant at the posterior parietal or midline sites ($p = .0092$ for PO3/PO4; $p = .019$ for Pz). In other words, the stimulus-evoked C1 effect and the CNV effect are dissociable spatially and temporally, and the CNV could be independently removed without affecting the C1 effects, suggesting that the two effects are different.

Summary of the findings

To summarize, expertise effects were obtained for notes on staff for the C1 bilaterally, as early as 40ms, in a time window and with a topographic distribution that are consistent with what is typically observed for the C1 component (Clark et al., 1995; Foxe & Simpson, 2002; Luck, 2005). These effects were replicated in split-half data sets, and could not be accounted for by pre-stimulus noise, eye movement or baseline correction. In addition, expertise effects were also observed for the N170 bilaterally and the CNV on a frontal-central site. Unlike expertise with Chinese characters (Wong et al., 2005), no expertise effect for the P3 was found.

Musical notation (no-staff)

Expertise effect for C1

For no-staff conditions, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) x Hemisphere (Left / Right) for the C1 revealed a main effect of Hemisphere for PO3/PO4, $F(1,18) = 12.2$, $p = .0026$, with a more positive voltage for the left compared to the right hemisphere. However, no other main effect or interaction reached significance (all *p*s > .15; Fig. 20a-b; 21a).

For Pz, the Group x Stimulus interaction was significant, $F(1,18) = 5.26$, $p = .034$ (Fig. 20c; 21b). Scheffé tests ($p < .05$) revealed that the C1 for notes was more positive than that for pseudo-letters for experts but not for novices. However, a similar effect was already observed before and right after stimulus onset ($p = .035$ for time 0 to 20ms; $p = .082$ for time -20 to 0ms), suggesting that this group difference could be the result of pre-stimulus differences. Also, the topographic distribution was frontal-central towards the right (Fig. 22), which was different from the posterior parietal distribution typically observed for the C1 (Clark et al., 1995; Foxe & Simpson, 2002; Luck, 2005). These suggest that this effect may be different from the C1 component.

In other words, an expertise effect for the C1 was observed for notes without staff. However, this effect might be susceptible to pre-stimulus noise and had a different topographic distribution compared to the typical C1 distribution that call for careful interpretation.

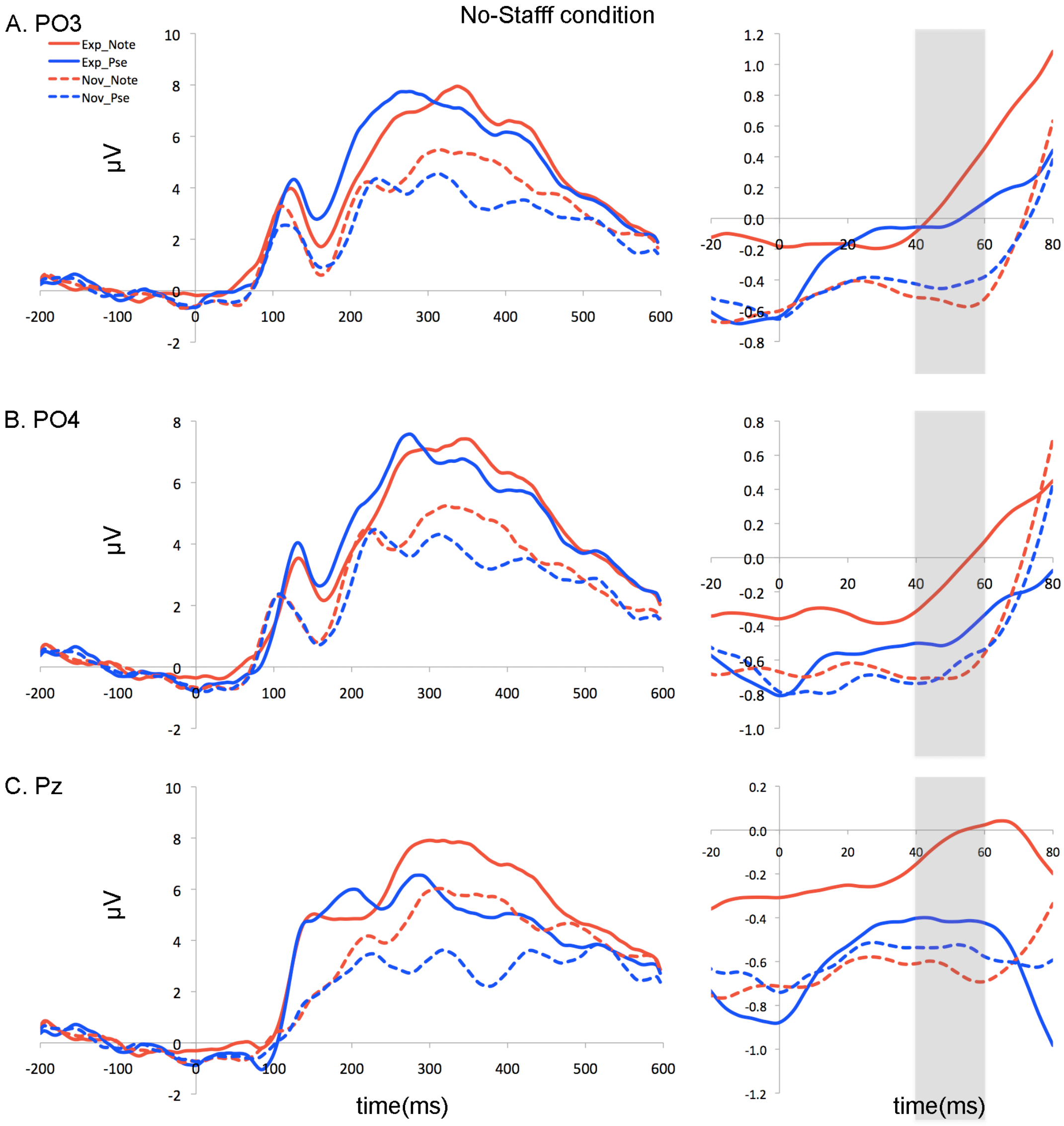

Figure 20. ERPs for the no-staff conditions on the posterior parietal channels, including (A) PO3; (B) PO4; and (C) Pz. Solid lines plot the activity for experts and dashed lines plot that for novices, with notes in red and pseudo-letters in blue. Left graphs show the ERPs from -200ms to 600ms, with the ERPs during the first 80ms highlighted on the right. The grey bars represent the early portion of the C1 (40-60ms).

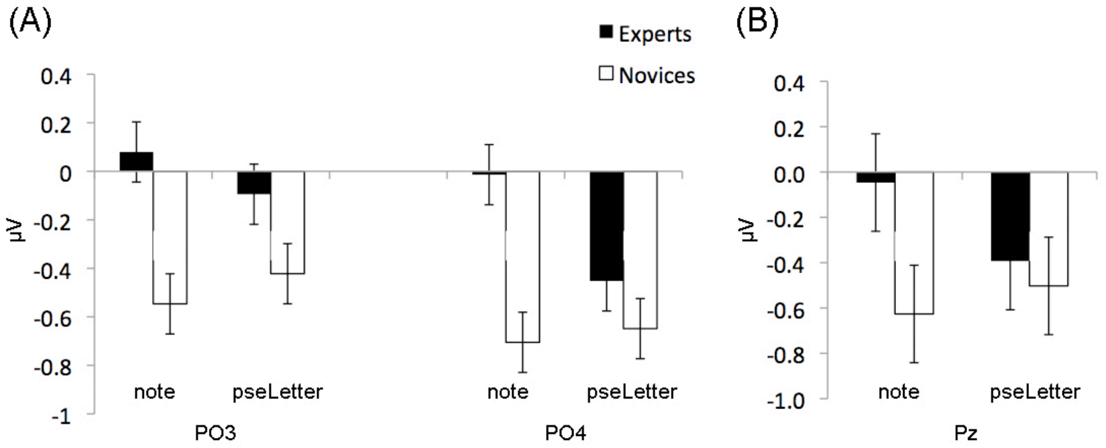

Figure 21. Averages of the scalp voltages for the C1 for the no-staff conditions in (A) PO3/PO4 and (B) Pz. Error bars plot the 95% CI for the highest order interaction for each graph.

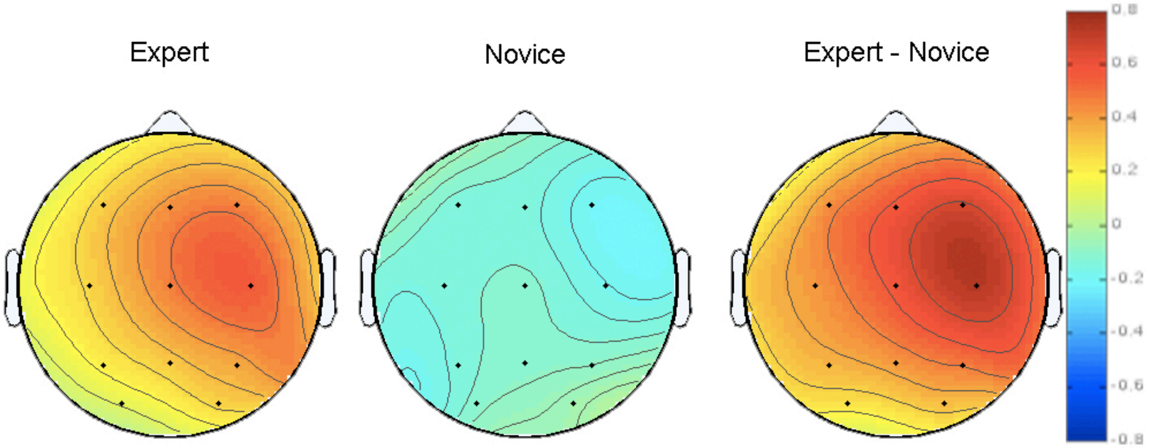

Figure 22. Topographic distributions of ERP differences with the contrast of [notes - pseudo-letters] for the C1 for no-staff conditions in experts (left), novices (middle) and the difference between the two groups (right).

Expertise effect for N170

For notes without staff, an expertise effect for the N170 was also obtained. For OL/OR, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) x Hemisphere (Left / Right) on the N170 revealed a Group x Stimulus interaction, $F(1,18) = 6.24$, $p = .022$ (Fig. 23 top; 24a). Scheffé tests ($p < .05$) revealed that the N170 was more negative for notes than pseudo-letters for experts but not for novices. No other main effects or interaction reached significance. The topographic distribution of the N170 expertise effect was bilateral ventral-temporal, consistent with the typical N170 effects (Fig. 25).

For T5/T6, a pattern similar to that seen for the OL/OR channels was obtained, (Fig. 23 bottom; 24b): the Group x Stimulus interaction was significant, $F(1,18) = 5.93$, $p = .026$.

The same analyses were performed on the P1 in these channels to test if the N170 effects were carry-over effects from the P1. For both OL/OR and T5/T6, the Group x Stimulus interaction did not reach significance (all ps > .2). The only effect that approached significance was the interaction between Group and Hemisphere (for OL/OR, $F_{1,18} = 3.87$, $p = .065$; for T5/T6, $F_{1,18} = 4.33$, $p = .052$), which did not differ across stimulus conditions (all ps > .3). This suggests that the N170 effect was not caused by differences that were already present earlier in time.

In sum, an expertise effect for the N170 for notes without staff was obtained in both hemispheres, suggesting that the higher sensitivity for notes is not limited to individuation of the notes or the pitch processing of the notes, and may be related to perceptual fluency with the shape of the notes.

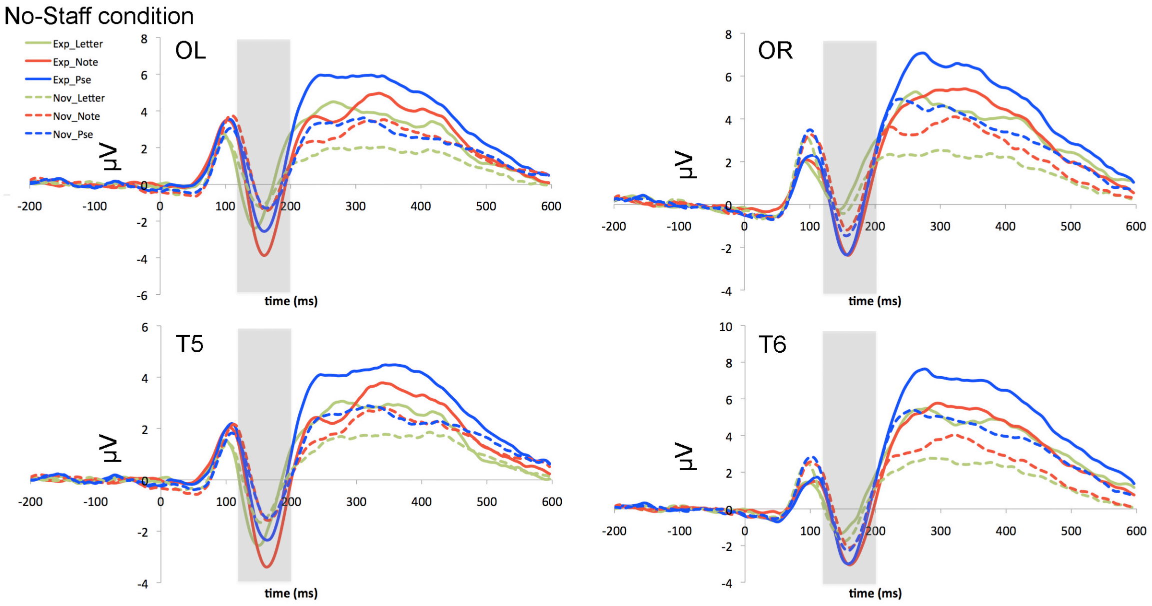

Figure 23. ERPs for the no-staff conditions for the N170 components in OL/OR (top) or T5/T6 (bottom). Solid lines plot the activity for experts and dashed lines plot that for novices, with notes in red, letters in green and pseudo-letters in blue. The grey bars represent the time window for the N170 component (120-200ms).

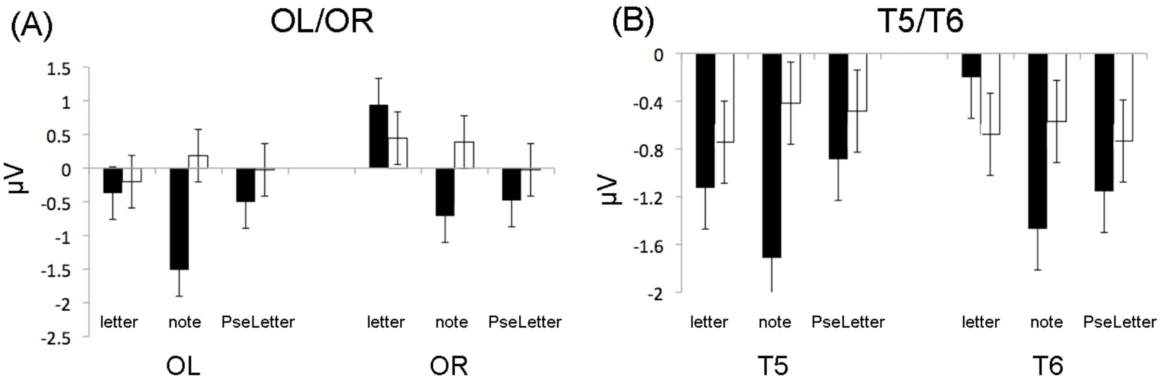

Figure 24. Averages of the scalp voltages for the N170 for the on-staff conditions in (A) OL/OR and (B) T5/T6. Error bars plot the 95% CI for the Group x Stimulus x Hemisphere interaction for each graph.

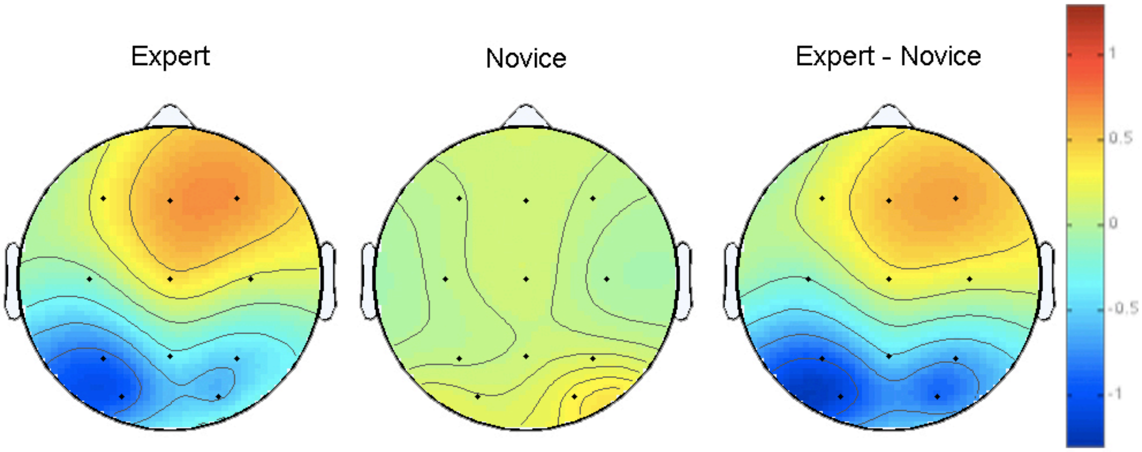

Figure 25. Topographic distributions of ERP differences with the contrast of [notes - pseudo-letters] for the N170 for no-staff conditions in experts (left), novices (middle) and the difference between the two groups (right).

Expertise effect for P3

To examine if the P3 component was modulated by music reading expertise, a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) was performed on the P3 component on Pz, Cz and Fz. All channels revealed a significant main effect of Stimulus, such that the P3 for notes was larger than that for pseudo-letters (for Pz, $F_{1,18} = 35.3$, $p \leq .0001$; for Cz, $F_{1,18} = 71.2$, $p \leq .0001$; for Fz, $F_{1,18} = 46.0$, $p \leq .0001$). However, no Group x Stimulus interaction was found in any channels (all *p*s > .2). In other words, no expertise effect was found for notes without staff for the P3 component.

Expertise effect for CNV

For notes without staff, no Group x Stimulus effect was observed for the CNV at Cz ($p > .15$), or at other central-parietal sites (C3, C4, P3, P4 or Pz; all *p*s > .2) or the frontal sites (F3, F4 or Fz; all *p*s > .2).

Summary of the findings

For notes without staff, expertise effects were obtained for the C1 and the N170 bilaterally, while no expertise effect was found for the P3 and the CNV. However, the C1 effect was susceptible to pre-stimulus noise, and had a different topographic distribution from the typical C1 effect, suggesting that this early visual effect might have a different source other than early visual cortex.

Letters (on-staff)

In this study, all participants were experts with Roman letters (either as native English speakers or being highly proficient in English). Without a novice group for letters, it is not possible to investigate the expertise effects with letters in the same manner as what was performed for musical notes. However, it is still possible to examine the selectivity for letters by comparing the voltage difference between letters and pseudo-letters in all participants. While this contrast is susceptible to effects driven by stimulus differences alone, it allows us to explore how the brain responds to this expert object category compared to a novel category.

To look for selectivity for letters, the scalp voltage was compared between letters and pseudo-letters in all participants. Although the factor of Group (Experts / Novices) was still included, no group difference was predicted since the expertise defining the groups was about musical notes but not about letters. A significant main effect of stimulus suggests that letter selectivity is obtained with that ERP component.

Letter selectivity for C1

Is the C1 selective for letters compared to pseudo-letters? A 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) on the C1 revealed no significant main effect of Stimulus for PO3/PO4 (p = .13). The only significant effect obtained was the main effect of Hemisphere for PO3/PO4, in which the voltage for the left hemisphere was more positive than that for the right hemisphere (for PO3/PO4, $F_{1,18} = 9.75$, p = .0059). For Pz, the main effect of Stimulus did not reach significance either (p = .09). Thus, no selectivity for letters (compared to pseudo-letters) was observed in this early time window.

Letter selectivity for N170

In a previous study with similar stimuli and a similar design, letter selectivity for the N170 was found in the left hemisphere but not in the right hemisphere (Wong et al., 2005). To test whether the result was replicated in the current study, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) was performed on the N170 on OL/OR and T5/T6.

For OL/OR, results revealed a significant main effect of Stimulus, $F(1,18) = 8.68$, p = .009, with a more negative N170 for letters than for pseudo-letters. The Stimulus x Hemisphere interaction was significant, $F(1,18) = 18.1$, p = .0005. Scheffé tests ($p < .05$) revealed that the N170 for letters was more negative than that for pseudo-letters in the left hemisphere but not in the right hemisphere (Fig. 15 top; 16a), replicating previous findings for letters (Wong et al., 2005).

For T5/T6, a pattern similar to that seen for the OL/OR channels was obtained, (Fig. 15 bottom; 16b): the Group x Stimulus interaction was significant, $F(1,18) = 8.09$, $p = .011$. Similar to the OL/OR, Scheffé tests ($p < .05$) revealed that the N170 for letters was more negative than that for pseudo-letters on the left hemisphere but not on the right hemisphere.

Analyses on the earlier P1 component suggest that these N170 effects were not simply carry-over effects from the P1. For OL/OR, the Stimulus x Hemisphere interaction did not reach significance ($p > .1$). For T5/T6, the Stimulus x Hemisphere interaction on the P1 was significant, $F(1,18) = 4.57$, $p = .047$. However, the P1 for letters and pseudo-letters were similar on the left hemisphere but were marginally different on the right hemisphere (Scheffé tests, $p < .05$). This pattern was qualitatively different from that for the N170 results, in which a more negative N170 for letters than pseudo-letters on the left hemisphere but not on the right.

In sum, letter selectivity for the N170 was found in the left but not the right hemisphere, replicating prior results for letters (Wong et al., 2005).

Letter selectivity for P3

To test if the letter selectivity for the P3 was found here as in the previous study (Wong et al., 2005), a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) was performed on the P3 component on Pz, Cz and Fz.

For Fz and Cz, none of the effects reached significance (all *p*s > .05). For Pz, a main effect of Stimulus was significant, $F(1,18) = 73.9$, $p \leq .0001$, with a smaller P3 for letters than pseudo-letters. The Group x Stimulus interaction was also significant, $F(1,18)$

= 8.78, *p* = .0083, in which the P3 for letters was similar across groups, but the P3 for pseudo-letters was larger for experts than novices.

In general, a less positive P3 was found for letters than pseudo-letters, replicating the trend obtained in the prior study (Wong et al., 2005).

Letter selectivity for CNV

For Cz, a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) on the CNV component revealed a main effect of Stimulus, *F*(1,18) = 8.02, *p* = .011, in which the CNV for letters was more negative than that for pseudo-letters, and this effect did not interact with Group (*p* > .1; Fig. 18 & 19). This effect was not a carry-over effect from the previous P3 component, since the main effect of Stimulus was not significant on Cz for the P3 (see above).

Summary of the findings

In sum, for letters on staff, letter selectivity was observed for the N170 and the P3, replicating the findings in a prior study (Wong et al., 2005). Letter selectivity was obtained for the CNV, suggesting that the CNV differences may be common for both music reading expertise and letter expertise. However, letter selectivity was not found for the C1 for letters on staff.

Letters (no-staff)

Letter selectivity for C1

For the C1 component, no letter selectivity was found for letters without staff. A 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) revealed no main effect of Stimulus for PO3/PO4 ($p > .6$). For Pz, the main effect of Stimulus was not significant either ($p > .6$). No other effects reached significance (all *p*s > .05).

Letter selectivity for N170

Letter selectivity for the N170 was expected for the no-staff conditions, as it was first reported with letters on a white background (Wong et al., 2005). However, such an effect was not observed for the no-staff conditions in the present study.

For OL/OR, a 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) on the N170 revealed a main effect of Stimulus, $F(1,18) = 14.6$, $p = .0012$, with the N170 for letters being more *positive* compared to that for pseudo-letters. Surprisingly, the interaction between Group and Stimulus was significant, $F(1,18) = 6.76$, $p = .018$. Scheffé tests ($p < .05$) revealed that the N170 for letters was more positive than pseudo-letters for experts but not for novices (Fig. 23 top; 24a). The Stimulus x Hemisphere interaction was also significant, $F(1,18) = 8.27$, $p = .010$, and did not interact with Group ($F < 1$). Scheffé tests ($p < .05$) revealed that the N170 for letters was more positive than pseudo-letters in the right hemisphere but not in the left hemisphere.

The T5/T6 channels resulted in similar N170 effects as that in OL/OR. A 2x2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) on the N170 revealed a significant Stimulus x Hemisphere interaction, $F(1,18) = 6.24$, $p = .022$ (Fig. 23 bottom; 24b). Similar to OL/OR, the N170 for letters was more positive than pseudo-letters in the right hemisphere but not in the left hemisphere (Scheffé tests, $p < .05$).

It is surprising that the letter selectivity for the N170 was only observed for the on-staff conditions but not for the no-staff conditions. The weakened N170 for letters in the left hemisphere was found for both groups (Fig. 24), suggesting that it is not a specific consequence of musical training. One possible explanation for the difference between the on-staff and no-staff conditions is that 18 stimuli were used for the on-staff conditions but only 6 stimuli were used for the no-staff conditions. So, on average, each stimulus was presented for 40 times for on-staff conditions but for 120 times for the no-staff conditions. The higher number of repetition of the stimuli may lead to relatively more visual adaptation, which may have reduced the N170 expertise effect for letters for the no-staff conditions.

To test this hypothesis, the N170 effects for the first 200 trials were examined, in which the stimuli were presented for approximately 40 times. Results revealed a similar pattern as above for OL/OR and T5/T6. For OL/OR, a main effect of Stimulus, $F(1,18) = 15.9$, $p = .0009$; an interaction between Group and Stimulus, $F(1,18) = 5.96$, $p = .025$; and a Stimulus x Hemisphere interaction, $F(1,18) = 5.98$, $p = .025$. For T5/T6, the interaction between Stimulus and Hemisphere approached significance, $F(1,18) = 3.77$, p

= .068, with the same pattern as described above. These results suggest that the absence of the N170 letter selectivity cannot be explained by more visual adaptation.

Letter selectivity for P3

The P3 results for letters without staff were similar to those for on-staff conditions. A 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) was performed on the P3 component on Pz, Cz and Fz.

For Cz and Pz, a main effect of Stimulus was significant (for Cz, $F_{1,18} = 10.0$, $p = .0053$; for Pz, $F_{1,18} = 39.7$, $p \leq .0001$), with a smaller P3 for letters than pseudo-letters. Similar to that for letters with staff, the Group x Stimulus interaction approached significance for Pz, $F(1,18) = 3.72$, $p = .07$, in which the P3 for letters was similar across groups, but the P3 for pseudo-letter was larger for experts than novices.

In general, a less positive P3 was found for letters than pseudo-letters, replicating the trend obtained by Wong et al. (2005).

Letter selectivity for CNV

For Cz, a 2x2 ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) on the CNV component did not reveal a main effect of Stimulus ($F < 1$). The Group x Stimulus interaction was marginally significant, $F(1,18) = 4.10$, $p = .058$, in which the CNV for pseudo-letters was more negative for novices than experts (Scheffé tests, $p < .05$). Therefore no letter selectivity was found for the CNV for letters without staff.

Summary of the findings

For letters without staff, letter selectivity was only observed for the P3, but not for the C1, N170, or the CNV.

Comparing on-staff and no-staff conditions

The results reported above suggest that some of the ERP effects are modulated by whether the stimuli are presented on the staff background, including the C1 and the CNV for notes, and the N170 and the CNV for letters. To further explore this effect, the influence of staff was directly evaluated by adding the factor of Staff (on-staff / no-staff) to the Group x Stimulus x Hemisphere ANOVA in each of these cases. The results reported in this section are focused on significant effects involving the staff.

The C1 for notes

For musical notes, the C1 expertise effect appears to be stronger for the on-staff conditions than the no-staff conditions at the posterior parietal sites (Fig. 11 & 22). However, the higher order ANOVA on the C1 did not reveal any significant effect on PO3/PO4 (all *p*s > .3) or Pz (all *F*s < 1). Therefore, no significant staff modulation on the C1 expertise effect was obtained.

The N170 for letters

For letters, the N170 letter selectivity was opposite for the on-staff and no-staff conditions, with the N170 for letters more negative than pseudo-letters for on-staff conditions, but the N170 for letters more positive than pseudo-letters for no-staff conditions. This difference was confirmed by performing a higher order ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Hemisphere (Left / Right) x Staff (on-staff / no-staff) on the N170 on OL/OR. Results revealed a significant interaction between Stimulus and Staff, $F(1,18) = 39.7$, $p \leq .0001$, and Scheffé tests ($p < .05$) confirmed the trend described above.

Also, the Group x Stimulus x Staff interaction was marginally significant, $F(1,18) = 3.26$, $p = .089$. The analyses were then performed separately for the two groups. Within experts, the Stimulus x Staff interaction was significant, $F(1,9) = 67.8$, $p \leq .0001$, with the N170 letter selectivity showing opposite patterns for the on-staff and no-staff conditions as described above (Scheffé tests, $p < .05$). Within novices, the Stimulus x Staff interaction was also significant, $F(1,9) = 6.67$, $p = .030$. Scheffé tests ($p < .05$) revealed that the N170 was more negative for letters than pseudo-letters for the on-staff conditions, but the N170 was similar for the two categories for the no-staff conditions.

At electrodes T5/T6, the main effect of Staff was significant, $F(1,18) = 4.75$, $p = .043$, with the N170 more negative for the no-staff than on-staff conditions. Moreover, the interaction between Staff and Stimulus was significant, $F(1,18) = 15.2$, $p = .001$, with the N170 letter selectivity significant only for the on-staff conditions (Scheffé tests, $p < .05$). Unlike the OL/OR, the Group x Stimulus x Staff interaction was not significant, $F(1,18) < 1$.

In sum, the N170 letter selectivity was modulated by the staff background. Both experts and novices showed an N170 letter selectivity for the on-staff conditions, while this letter selectivity disappeared for no-staff conditions for novices and was even reversed for experts.

The CNV for notes

The CNV expertise effect for musical notes was only found for the on-staff conditions but not for the no-staff conditions. To directly test the effect of staff, an ANOVA with Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) x Staff (on-staff / no-staff) was performed on the CNV at Cz. The interaction between Stimulus and Staff was marginally significant, $F(1,18) = 3.98$, $p = .062$, with the CNV for notes more negative than that for pseudo-letters only for the on-staff conditions (Scheffé tests, $p < .05$). However, the Group x Stimulus x Staff interaction was not significant, $F(1,18) < 1$, so there is little evidence here to conclude that the CNV expertise effect depends on the presence of the staff.

The CNV for letters

The letter selectivity for the CNV was found for the on-staff conditions but not for the no-staff conditions. However, an ANOVA with Group (Experts / Novices) x Stimulus (Letters / Pseudo-letters) x Staff (on-staff / no-staff) was performed on the CNV at Cz revealed that the effect of staff did not reach significance ($p > .2$). This suggests that the CNV letter selectivity was not significantly different between the on-staff and no-staff conditions.

Summary of findings

In sum, only the N170 letter selectivity was significantly affected by the staff, while neither the C1 for notes nor the CNV for notes or letters was significantly modulated by the staff background.

General Discussion

In this ERP experiment, the temporal dynamics of music reading expertise effects were investigated. I tested whether the expertise effects observed in early visual cortex in a prior fMRI study (Wong & Gauthier, 2010) were the result of altered cell response in early visual cortex (a feedforward effect), or strengthened feedback from higher areas (a feedback effect) or both. Music reading experts and novices were recruited, and the neural selectivity for musical notes was compared to a novel category of pseudo-letters for various ERP components with a simple one-back task. Results for each ERP component are discussed in the following sections (Table 1).

Table 1. Summary table for the ERP effects obtained in the ERP study, in which only electrode sites with significant ERP effects for notes or letters are shown. Expertise effects for notes refer to the Group (Experts / Novices) x Stimulus (Notes / Pseudo-letters) interaction. Letter selectivity refers to the main effect of Stimulus (Letters / Pseudo-letters).

		C1		**N170**	**P3**	**CNV**
		Left/ Right	Midline	Left/ Right	Midline	Midline
Expertise effects for Notes						
	on-staff	PO3/ PO4	Pz	OL/ OR T5/ T6	---	Cz
	no-staff	---	Pz	OL/ OR T5/ T6	---	---
Letter selectivity						
	on-staff	---	---	OL T5	Pz	Cz
	no-staff	---	---	OR T6	Pz, Cz	---

The C1 effect

Expertise effects were obtained with musical notes (on-staff) as early as 40ms, with a timing and topographic distribution consistent with the C1 component. The C1 expertise effects were robust as they were replicated in split-half datasets, and could not be explained by pre-stimulus noise, eye movements and baseline correction. Visual-evoked effects obtained in the early portion of the C1 component (40-60ms) are likely to be heavily contributed by the primary visual cortex (Foxe & Simpson, 2002; Schmolesky et al., 1998). It suggests that the initial visual processing of notes in the early visual cortex is different with extensive music reading experience, and that the expertise effect obtained in V1 in the fMRI study (Wong & Gauthier, 2010) is, at least partly, a feedforward effect.

No letter selectivity was found for the C1, regardless of whether letters were on a staff or not. It is possible that the expertise effect for letters cannot be revealed because no letter novices were included in this experiment. Therefore, analyses were performed for the notes within experts only to see if, in the case for musical notation, a novice group is required to obtain expertise effects in this early time window. In experts, a 2x2 ANOVA with Stimulus (Notes / Pseudo-letters) x Hemisphere (Left / Right) on the C1 revealed a significant main effect of Stimulus for PO3/PO4, $F(1,9) = 6.40$, $p = .032$; and for Pz, $F(1,9) = 11.8$, $p = .0074$, with the voltage for notes more positive than that for pseudo-letters. This suggests that the C1 effect may be obtained by comparing two different categories of objects for which participants have different amounts of expertise, as demonstrated here with the case of notes. However, especially in the case of retinotopic cortex, a contrast where the stimuli are perfectly matched is preferable. Although no early expertise effect was observed for letters in the current experiment, further studies are required to investigate whether such early visual effects can be observed with letters or other types of perceptual expertise stimuli.

The N170 effect

For the N170, expertise effects were observed for both notes with staff and notes without staff in both hemispheres, suggesting that these expertise effects do not depend on the pitch information of the notes, and are possibly related to the shape discrimination of the notes. The N170 selectivity for musical notes was observed bilaterally, consistent with the fMRI findings that bilateral ventral temporal areas (e.g. the fusiform gyrus) are selective for musical notes (Wong & Gauthier, 2010). The N170 results add to the literature that visual processes associated with perceptual expertise with different object categories occur in the same time window (around 170ms after stimulus onset), similar to that for faces, birds, dogs, cars, fingerprints and letters (Bentin et al., 1996; Busey & Vanderkolk, 2005; Gauthier et al., 2003; Tanaka & Curran, 2001; Wong et al., 2005), even though some of these categories (at least for faces, letters and musical notes) recruit different brain regions as revealed in prior fMRI studies (James et al., 2005; Wong & Gauthier, 2010).

Letter selectivity for the N170 was found in the left hemisphere only (for on-staff letters), replicating previous findings (Wong et al., 2005). However, similar letter selectivity was not obtained for letters without staff, which was unexpected. Both experts and novices showed an N170 letter selectivity for the on-staff condition, while this letter selectivity disappeared for no-staff conditions for novices and was reversed for experts. The lack of N170 selectivity for letters without staff was not due to the repeated use of a smaller set of stimuli, and was found for both experts and novices, suggesting that it is not simply a result of music reading expertise. The experimental design and the stimuli for the no-staff condition were almost identical to those used in for the previous study

(Wong et al., 2005). The only major difference in the current study is that the no-staff conditions were presented interleaved with the on-staff conditions. It is possible that processing stimuli with a staff background may affect the subsequent processing of the same stimuli on a blank background, but the mechanisms of such effects remain unclear.

The P3 effect

The letter selectivity for the P3 was replicated in the current study, for both on-staff or no-staff conditions (Wong et al., 2005). No expertise effect for musical notes was found for the P3, either for notes with or without staff, which does not support the account for strengthened feedback to early visual areas with late components that are heavily modulated by top-down effects, such as the P3 (Sutton et al., 1965) or the N400 (Kutas & Hillyard, 1980).

Since the P3 component is related to many cognitive processes (Luck, 2005), it remains unclear what process is engaged by letters but not by notes that is captured by the P3 component. One possibility is the linguistic processing (e.g. phonological processes) that may be automatically engaged for letters but not for musical notes or pseudo-letters. This hypothesis needs to be tested with future studies designed to tap onto linguistic factors.

The CNV effect

Expertise effects were observed for the CNV for musical notes only when the notes were presented on a staff. Letter selectivity was also observed for the CNV, only when the letters were on the staff background. This is consistent with previous findings that the CNV component is modulated by experience (Belkic et al., 1992; Muller et al., 2010; Travis et al., 2000). The CNV is a component that is modulated by a wide range of factors. In this case, the CNV expertise effects were unlikely a result of performance differences (given the similar accuracy and response time for the two groups), and were not related to any task differences or task relevancy of the stimuli (all participants performed the same task). It is unlikely that the CNV differences were driven by the predictability of the upcoming event, since the 1st stimulus was always unpredictable for each block, but the object category was 100% predictable for the rest of the block. The CNV is slightly lateralized towards the left, corresponding to the use of the right hand for the speeded responses, and suggesting that the CNV expertise effects may be at least partly related to motor preparation. One possibility is that, with musical training, the motor system of experts automatically prepares for the upcoming notes, even though such motor preparation is task-irrelevant. Another speculation is that experts are simply anticipating more compared to novices, such as the relative position of the notes or the auditory interval of the note sequences, given their richer knowledge with musical notation. It will require further investigation to understand the factors driving the expertise effects of the CNV.

CHAPTER III

CROWDING AND EXPERTISE WITH MUSICAL NOTATION

The goal of this experiment was to investigate whether music reading expertise alleviates crowding in the parafoveal visual field, and to relate the crowding effect to the ERP expertise effects reported above. In this study, participants were required to judge whether a black dot was presented on a line or on the space (above or below the line). This is a visual task that can be performed by novices without any musical training, but is also critical to music reading expertise since a note on or off a line has a different identity. Two kinds of crowding were examined. First, the target note and its line could be flanked vertically by four extra lines (two above and two below the original line; Fig. 26a-c). Second, the target note and its line could be flanked horizontally by two extra notes. I expected both groups to show a crowding effect, i.e. their performance should decrease for crowded stimuli, and that crowding should affect novices more strongly than experts. To examine whether a smaller crowding effect for experts (if found) is specific to musical stimuli, crowding was also measured with a set of control stimuli (Landolt C; Fig. 26d-e). I also measured far and near acuity and contrast sensitivity for each individual to test if group differences in basic visual functions accounted for any expertise effect in crowding. In addition, perceptual fluency and holistic processing of musical notes were measured in all participants. These measures and crowding served as behavioral correlates for the ERP components (CHAPTER IV).

Method

Participants

All the participants in the ERP study participated in the crowding experiment (except the author), and additional participants were recruited from Vanderbilt University and the Nashville community for cash payment. Apart from those who participated in the ERP study, 22 experts and 11 novices were recruited. Only 14 experts and 10 novices completed both the crowding experiment and the perceptual fluency test (identical to that of the ERP experiment) such that their music reading ability could be measured. Therefore, including the ERP participants, 24 experts and 20 novices completed these behavioral studies.

All participants were recruited according to the same criteria as for the ERP study, and all participants reported their amount of experience in music reading and rated their music-reading ability (1 = do not read music at all; 10 = expert in music reading), and their handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). The expert group included 12 females and 12 males (mean age = 22.9, s.d. = 6.2; 22 right-handed, 1 left-handed and 1 ambidextrous), with 13.7 years of music reading experience and a self-rating score of 9.08 on average. The novice group included 9 females and 11 males (mean age = 25.0, s.d. = 6.4; 19 right-handed and 1 left-handed), with 0.41 year of music reading experience and a self-rating score of 1.35 on average. All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University. They were paid $12 per hour of behavioral testing.

Stimuli and Design

The experiment was conducted on Mac Mini using Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Stimuli were presented on a CRT monitor at 1024x768 pixel resolution and 100Hz refresh rate, with a mean luminance of 28.2 cd/m^2 in a dimly lit room. All the stimuli were generated with Matlab and were black in color presented on a grey background. The stimuli were 60 x 60 pixels in size, subtending about 1.3° x 1.3° of visual angle, centered at about 2.6° to the left or right of the central fixation point with 90cm viewing distance (fixed with a chin rest). The stimuli were presented for 100ms randomly on the left or right, so that the stimuli disappeared before any saccade could be made towards them.

For all musical stimuli, a black elliptical dot similar to the bottom part of a musical note was used for all targets and flankers (Fig. 26a-c). The target dot was either on, above or below the middle horizontal line. For the 5-line condition, 2 extra lines were added above and below the middle line with a spacing of 10 pixels. For stimuli with flanker dots, a dot was added on the left and right of the target dot. The flanker dots were either on, above or below the middle horizontal line, with all the possible combinations counterbalanced throughout the experiment. The three dots were always asymmetrical in space (the two flanker dots always had different distance from the target) such that detecting the position of the flankers (which was much easier than the crowded targets) was not informative about the correct response of the trial. The eccentricity differences between target and flankers, or that between target and extra lines, ranged from 0.22° to 0.43°, which was well within the critical spacing between targets and flankers (roughly half of the eccentricity of the stimuli, which was 1.3°; Bouma, 1970; Pelli & Tillman,

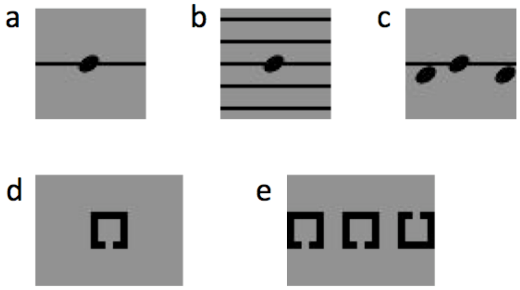

Figure 26. Examples of the stimuli used in the crowding experiment, showing a baseline musical note (A), and when the note is crowded with extra lines (B) or extra dots (C). (D) and (E) show the Landolt C used as control stimuli with baseline and crowded condition respectively.

2008). Therefore crowding was expected to occur for all extra lines and all flanker positions.

For the control stimuli, a set of stimuli, Landolt C, was generated with Matlab (Fig. 26d-e). Each Landolt C was 20 x 20 pixels in size, with a 6-pixel gap either at the top or bottom of the square. For the crowded condition, two Landolt Cs were added on the left and right of the middle target. The spacing between the center of the target and flankers was 30 pixels, translating to about 0.64° visual angle, so crowding was also expected in this condition (visual angle < 1.3°; Bouma, 1970). The gap of the flankers was either at the top or bottom, with all the possible combinations counterbalanced in the experiment.

On each trial, a central fixation dot was shown for 500ms, followed by a stimulus for 100ms (Fig. 27). For musical stimuli, the task was to judge whether the dot (or the central dot for stimuli with flankers) was on a line or on the space and to respond by key press. For Landolt Cs, the task was to judge whether the gap was at the top or bottom of the Landolt C (or the central Landolt C in the crowded condition). Accuracy was emphasized, and participants were encouraged to take their time to decide if needed. The dependent measure was the Weber contrast, calculated with the equation [(background luminance – target luminance) / background luminance], with the background luminance always be grey (RGB value = 128). The contrast threshold for 75% accuracy was estimated four times using QUEST (Psychtoolbox; Watson & Pelli, 1983), each estimated with 40 trials, and the average contrast threshold was used. Participants were first tested with the musical stimuli followed by the control stimuli. The trials were blocked for each condition (uncrowded, crowded with notes or crowded with lines for musical stimuli; uncrowded or crowded for control stimuli), and the order of the blocks was counterbalanced.

Two factors were manipulated, with Group as a between-subject factor (experts / novices), and Crowding (crowded / uncrowded) as within-subject factors for the 3 types of crowding (crowding with notes, lines or control stimuli). Participants were provided 24 trials for practice with feedback before testing, and no feedback was provided for the test.

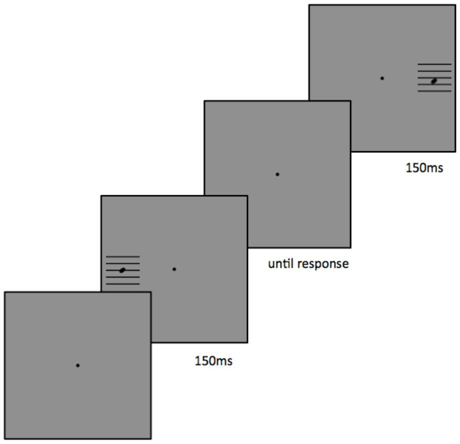

Figure 27. The paradigm used for the crowding experiment.

Measure of basic visual functions

To compare the two groups in terms of basic visual functions, far and near acuity values and functional contrast sensitivity were measured (Stereo Optical Vision Tester; Chicago, IL). The acuity test involved reading out uppercase letters presented in different sizes, and the accuracy of which corresponded to a certain acuity level. Functional contrast sensitivity was measured by asking participants to judge whether the presented gratings were tilted to the left, right or straight up. The gratings were presented with different spatial frequencies (1.5, 3, 6, 12 or 18 cpd) in different contrasts. All the tests were performed with both eyes and with corrected vision (if needed).

Measure of perceptual fluency

Perceptual fluency for music sequences was also measured. An identical sequential matching paradigm with four-note music sequences was used as in the ERP experiment.

Measure of holistic processing

The measure of holistic processing was a short version of the previous study (Wong & Gauthier, in press) by including the key conditions only. The stimuli were four-note music sequences generated in Matlab, and all notes were connected with a horizontal line (eighth notes). All stimuli were black and were shown on a white background at 7.2° x 4.8° degrees of visual angle.

A sequential matching paradigm was used. On each trial, a fixation cross at the center of the screen was presented for 500ms, the first stimulus for 750ms, a mask for

500ms, and then the second stimulus for 2500ms. One of the four notes on the 2nd sequence was indicated as the target note with two arrows. Participants were asked to judge whether the target note was the same or different from the equivalent note in the first sequence. Half of the trials were ‘same’ trials with the target note unchanged, the other half were ‘different’ trials with the target note shifted one step up or down. Participants were instructed to respond only according to the matching status of the target note by key press. Both speed and accuracy were emphasized and responses were required within 2500ms after the onset of the 2nd sequence, or were counted as errors (< 1% of the trials).

Four factors were manipulated, with Group as a between-subject factor (experts / novices), and Congruency (congruent / incongruent), Target Position (center / periphery), Target Distribution (25p75c / 75p25c; see below) as three within-subject factors. Targets either appeared in the two center positions of the sequence (the 2nd or the 3rd note) or in the peripheral positions (the 1st or the 4th note). Target Distribution was manipulated across two blocks of trials. Within each block, targets were distributed either 25% in periphery and 75% at center (25p75c) or 75% in periphery and 25% at center (75p25c). The order of the target distributions was counterbalanced across participants within each group. Participants were told about the target distribution immediately before each block. By manipulating target distribution, different contexts were created that encouraged relatively more attention to notes in some positions more than others, such that the contextual dependency of the holistic processing could be examined. Compared to the previous study, the target distribution of 50% in periphery and 50% at center was dropped, considering that the main group differences were revealed in the other conditions. To manipulate Congruency, a note adjacent to the target was considered the "distractor" (left or right counterbalanced if the target was one of the central two notes). In the 2nd sequence, the distractor note could be shifted one step up or down, resulting in different congruency conditions. Specifically, on congruent trials, the distractor note remained unchanged (compared to the 1st sequence) on 'same' trials while it changed on 'different' trials. For incongruent trials, the distractor note changed on 'same' trials and remained unchanged on 'different' trials. Dependent measures were sensitivity (d') and response time (RT) for correct responses. Holistic processing was defined as the congruency effect, using the difference in performance (d' or RT) between congruent trials and incongruent trials. There were a total of 512 trials, with 64 trials for each of the three within-subject conditions. Twenty practice trials with feedback were included, followed by test trials without feedback.

Results

Perceptual fluency

Four experts were excluded from data analyses because their perceptual fluency for notes or letters was > 3 s.d. away from the mean of the rest of the group. Therefore, 20 experts and 20 novices were included in the following analyses.

As expected, experts had a higher perceptual fluency than novices for music sequences but not for letter strings. A one-way ANOVA for Group (Expert / Novice) was performed for the perceptual threshold for matching four-note sequences. The main effect of Group was significant, $F(1,38) = 44.6, p \leq .0001$, with the perceptual threshold for experts (mean = 463.0 ms) lower than that for novices (mean = 1335.0 ms). In contrast, the main effect of Group for matching four-letter strings was not significant (p = .2), with a mean perceptual threshold 206.8 ms and 259.9 ms for experts and novices respectively. This confirms that experts have a higher perceptual fluency for reading music sequences, which cannot be explained by a general perceptual advantage.

Basic visual functions

All participants had normal far and near acuity (20/20 or 20/30). All participants (except one novice) had a normal functional contrast sensitivity, but excluding that novice (with a far and near acuity of 30/30 and a functional contrast sensitivity of 20/100) from the analyses did not change the pattern or the significance of the results of the crowding experiment. These results suggest that any group differences observed in the crowding experiment cannot be accounted for by a difference in basic visual functions.

Crowding

For crowding with extra lines, a 2x2x2 ANOVA with Group (Experts / Novices) x Crowding (baseline / crowded) on contrast threshold revealed a main effect of Group, $F(1,38) = 11.7$, $p = .0015$, with a lower contrast threshold for experts than novices. The main effect of Crowding was significant, $F(1,38) = 96.5$, $p \leq .0001$, which interacted with Group, $F(1,38) = 10.7$, $p = .0023$ (Fig. 28a). Scheffé tests ($p < .05$) revealed that experts performed better than novices for both the baseline and the crowded conditions, and the performance difference between the baseline and crowded condition was smaller for experts than novices.

For crowding with flanker notes, a 2x2x2 ANOVA with Group (Experts / Novices) x Crowding (baseline / crowded) on contrast threshold revealed a main effect of Group, $F(1,38) = 12.9$, $p = .0009$, with a lower contrast threshold for experts than novices. The main effect of Crowding was significant, $F(1,38) = 122.6$, $p \leq .0001$, and it interacted with Group, $F(1,38) = 10.1$, $p = .003$ (Fig. 28a). Scheffé tests ($p < .05$) revealed that experts and novices performed similarly for the baseline condition, but the crowding effect was smaller for experts than novices.

For control stimuli (Landolt C), a 2x2x2 ANOVA with Group (Experts / Novices) x Crowding (baseline / crowded) on contrast threshold revealed a main effect of Crowding, $F(1,38) = 759.8$, $p \leq .0001$, with the contrast threshold smaller for the baseline than the crowded condition (Fig. 28b). Importantly, no main effect or interaction involving Group reached significance (all ps > .3), suggesting that the amount of crowding experienced by the two groups was similar for non-musical stimuli.

To compare crowding created by extra lines and by flanker dots, a 2x3 ANOVA with Group (Experts / Novices) x Crowding (baseline / extra lines / flanker dots) was performed on contrast threshold. The main effect of Group was significant, $F(1,18) = 16.4$, $p = .0002$, with a lower contrast threshold for experts than novices. The main effect of Crowding was significant, $F(2,76) = 63.6$, $p \leq .0001$, and it interacted with Group, $F(2,76) = 5.62$, $p = .0053$ (Fig. 28a). Scheffé tests ($p < .05$) revealed that experts and novices performed similarly for the baseline condition, and the crowding effect was smaller for experts than novices for both types of crowding. Crowding created by extra lines and flanker dots was similar for experts. However, for novices, the contrast threshold for flanker dots was higher than that for extra lines.

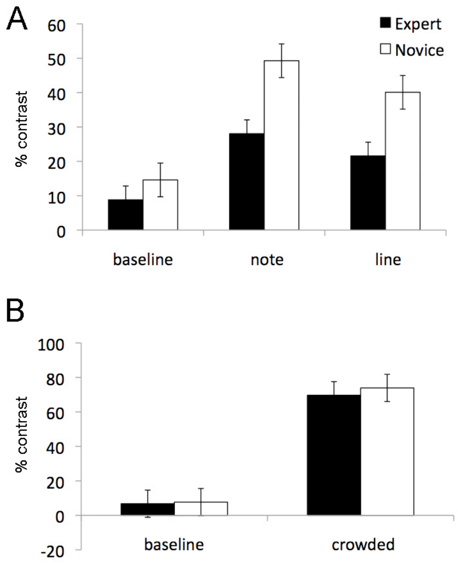

Figure 28. The contrast threshold for crowding with musical stimuli (A) and that for crowding with control stimuli (B). Error bars plot the 95% CI for the Group x Crowding interaction for all conditions.

In sum, experts experienced less crowding than novices when crowding elements were staff lines or flanking notes. However, the amount of crowding was similar across groups for control stimuli, suggesting that music reading experience helps alleviate crowding specifically for musical stimuli.

Predicting crowding with perceptual fluency

To examine whether the amount of crowding decreases with music reading ability, the correlations between individual perceptual fluency (note – letter) and the amount of crowding (contrast threshold of crowded – baseline condition) were considered. Perceptual fluency predicted both crowding with notes (r = .40, p = .01; Fig. 29a) and crowding with lines (r = .34, p = .033; Fig. 29b) when all participants were

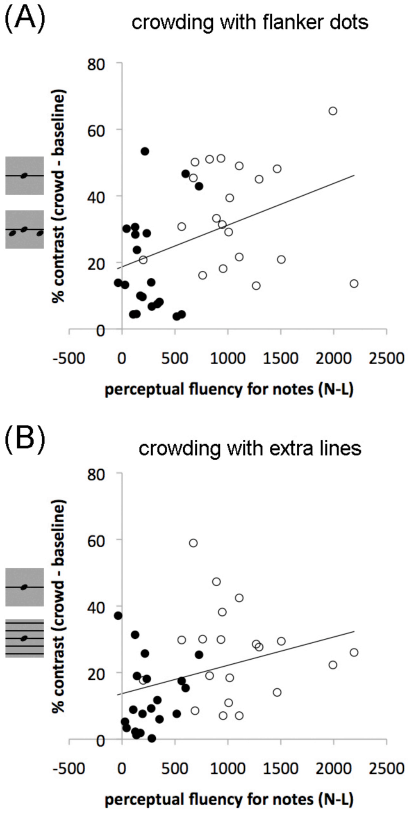

Figure 29. Correlations between perceptual fluency with notes and crowding with flanker notes (A) or crowding with extra lines (B). Data points for experts are the black circles while that for novices are the open circles.

included, but not within novices or experts separately. In contrast, perceptual fluency did not predict the amount of crowding for control stimuli (*p*s > .1).

Holistic processing

It was expected that this shortened holistic measure would produce similar patterns of the congruency effect to that of the previous study, in particular, a different pattern of the congruency effect for different contexts for novices but not for experts (Wong & Gauthier, in press). Two novices were excluded from analyses because of a general accuracy less than 60%. Therefore 20 experts and 18 novices were included in all the following analyses involving this holistic measure.

For delta d' (congruent d' – incongruent d'), a 2x2x2 ANOVA with Group (Experts / Novices) x Target Position (center / periphery) x Target Distribution (25p75c / 75p25c) was performed. The Target Position x Target Distribution interaction was significant, $F(1,36) = 6.09$, $p = .019$ (Fig. 30a). Scheffé tests ($p < .05$) revealed that the congruency effect was larger for center-target trials than periphery-target trials for 25p75c but not for 75p25c. No other main effect or interaction reached significance.

For delta RT (incongruent RT – congruent RT), there was a main effect of Target Distribution, $F(1,36) = 8.61$, $p = .006$, in which the congruency effect was larger for 25p75c than 75p25c (Fig. 30b). Also, the main effect for Target Position was significant, $F(1,36) = 8.42$, $p = .006$, with a larger congruency effect for periphery-target trials than center-target trials. However, no other effects reached significance, indicating that the group differences in holistic processing that were observed mainly in delta RT (Wong & Gauthier; in press) were not found in the present study. Relative to this previous study, the pattern for experts was similar in that their congruency effect was largely independent of context, except an increased congruency effect for the center-target trials for delta d' (that was also obtained in the previous study with delta RT). However, the pattern for novices was different. In the previous study, the congruency effect for novices was driven by target likelihood, such that the congruency effect increased for target positions at which the target was unlikely. In the current study, however, such a target likelihood effect was only observed in one target distribution (25p75c) but not another (75p25c). To speculate, one possible reason for the different findings across studies is that the contextual manipulation (by target distribution and target position) may not be as effective when the experiment is shortened (2/3 as long as the previous version).

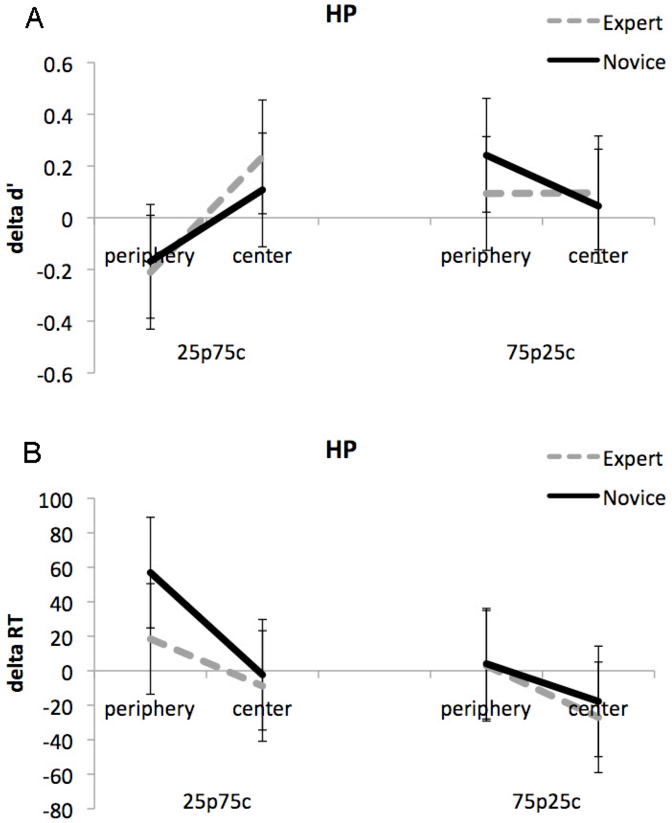

Figure 30. Congruency effects in the holistic processing experiment. (A) shows results with delta d' and (B) shows that with delta RT. Solid lines and dashed lines plot the performance for novices and experts respectively. Error bars plot the 95% CI for the Group x Target Position x Target Distribution interaction.

General Discussion

In this crowding experiment, music reading experts and novices were required to judge the position of a dot with respect to a line, which is a central task in music reading. The influence of adding extra lines or flanker dots on task performance was tested as two forms of crowding effects. Music reading experts experienced less crowding with extra lines or flanker dots compared to novices, and this effect cannot be accounted for by differences in basic visual functions. This alleviation of crowding was specific to musical stimuli, since both groups experienced similar crowding effect with unfamiliar object category such as Landolt C.

Although perceptual fluency with musical notes predicted the degree of crowding for musical stimuli (either induced by extra lines or flanker dots) but not for the control stimuli, the correlation was only found across the two groups but not within each group. This suggests that the correlation results can be interpreted in several ways. First, it is possible that better music readers have a better ability to uncrowd musical stimuli. Second, the correlation may merely reflect group differences in perceptual fluency. Finally, perceptual fluency may be related to a third variable that better predicts the degree of crowding across individuals. The linear relationship between perceptual fluency and crowding remains to be established.

In sum, the results suggest that perceptual experience enhances the ability to uncrowd objects of expertise specifically, in contrast with a recent proposition that crowding is independent of object category, with perceptual expertise or not (e.g. Pelli & Tillman, 2008). Also, the expertise effects with crowding were obtained without directly practicing on the task (e.g. Chung, 2007; Huckauf & Nazir, 2007), suggesting that crowding can be reduced by practicing on a task different from the testing task (e.g. Green & Bavelier, 2007).

CHAPTER IV

BEHAVIORAL SIGNIFICANCE OF THE ERP EFFECTS

To explore the behavioral correlates of the ERP expertise effects, the correlation between various ERPs (the C1, N170 and CNV) and several behavioral measures were considered, including perceptual fluency, the crowding effect and the degree of holistic processing. The ERP effects were computed as the selectivity for notes (scalp voltage for notes – that for pseudo-letters). The author was excluded from all correlation analyses since she did not participate in some of these behavioral studies. Also, one expert with an exceptionally large N170 effect (> 3 s.d. from the mean of the rest of the group for the occipito-temporal channels, both on-staff and no-staff conditions) and another expert with an exceptionally large C1 effect (> 3 s.d. from the mean of the rest of the group for the PO channels for on-staff conditions) were excluded from the correlation analyses with the N170 and the C1 respectively. The correlations were either analyzed with all the participants or within the expert group.

Correlation Results

Predicting ERPs with perceptual fluency

Are ERP expertise effects predicted by a quantitative measure of expertise in music reading, that of perceptual fluency specifically for notes (the difference between perceptual fluency for notes minus that for letters)?

Across all participants, perceptual fluency predicted the C1 and N170, and was correlated with the CNV with marginal significance. For the C1, the correlation between perceptual fluency and neural selectivity for notes (with staff) was significant for Pz (r = -.55, p = .019; Fig. 31c) and was at trend for PO4 (r = -.43, p = .073).

For the N170, perceptual fluency predicted the selectivity for notes either with staff (r = .54, p = .020 for OL; r = .58, p = .012 for T5; Fig. 31a) or without staff (r = .49, p = .038 for OL; r = .49, p = .037 for T5; Fig. 31b), and such correlations were not observed for the right hemisphere (all ps > .18).

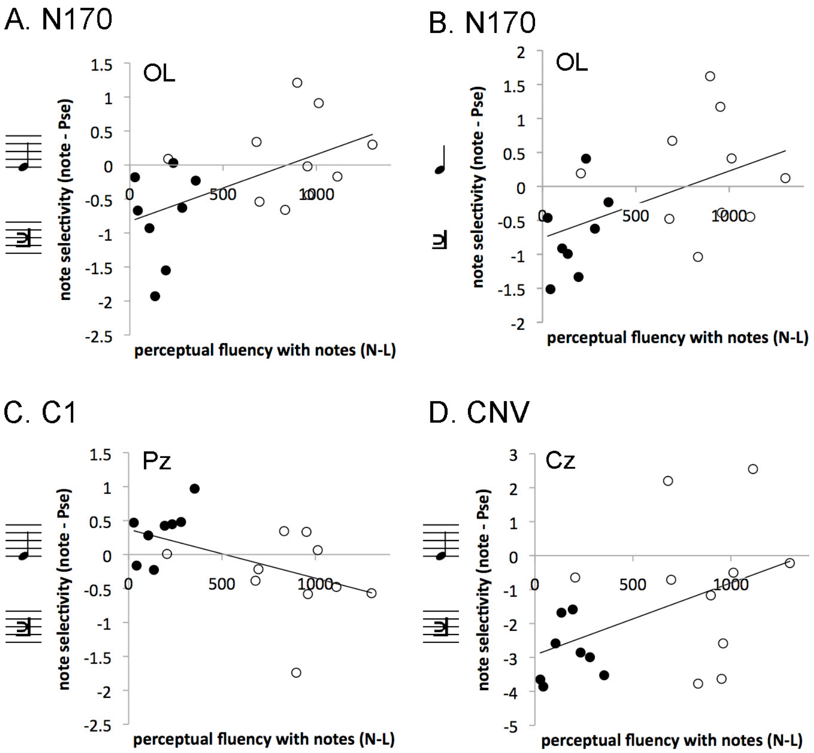

Figure 31. Perceptual fluency predicts the selectivity for musical notes measured in ERPs, including the N170 for on-staff conditions in OL (A), the N170 for no-staff conditions in OL (B), the C1 effect in Pz (C) and the CNV in Cz (D) for on-staff conditions. Data points for experts are the black circles while that for novices are the open circles.

For the CNV, perceptual fluency predicted the neural selectivity for notes (with staff) with marginal significance at Cz ($r = .45$, $p = .053$; Fig. 31d). The trends for early visual effects and the CNV were not observed for notes with no staff (*p*s > .2).

Since the range of the perceptual fluency measure for the expert group was narrow (from 28ms to 353ms), the correlation was not analyzed within the expert group.

Predicting ERPs with crowding

It is of interest to see whether various ERP effects can predict the amount of crowding experienced by the individuals, especially for the C1 effects, since some work associates crowding with the early visual cortex (Arman et al., 2006; Fang & Sheng, 2008; Tjan & Nandy, 2010). Correlations between ERPs and crowding with lines and that between ERPs and crowding with notes were examined separately.

For all participants, the C1 selectivity for notes predicted the amount of crowding with flanker notes but not with extra lines. For notes with staff, the C1 predicted the crowding with notes at PO3 ($r = -.51$, $p = .031$; Fig. 32a). For notes without staff, the early visual effects predicted the crowding with notes at Pz ($r = -.67$, $p = .003$; Fig. 32b), and marginally at PO3 ($r = -.49$, $p = .055$) and at PO4 ($r = -.49$, $p = .051$). No correlation for crowding with lines was observed (all *p*s > .15).

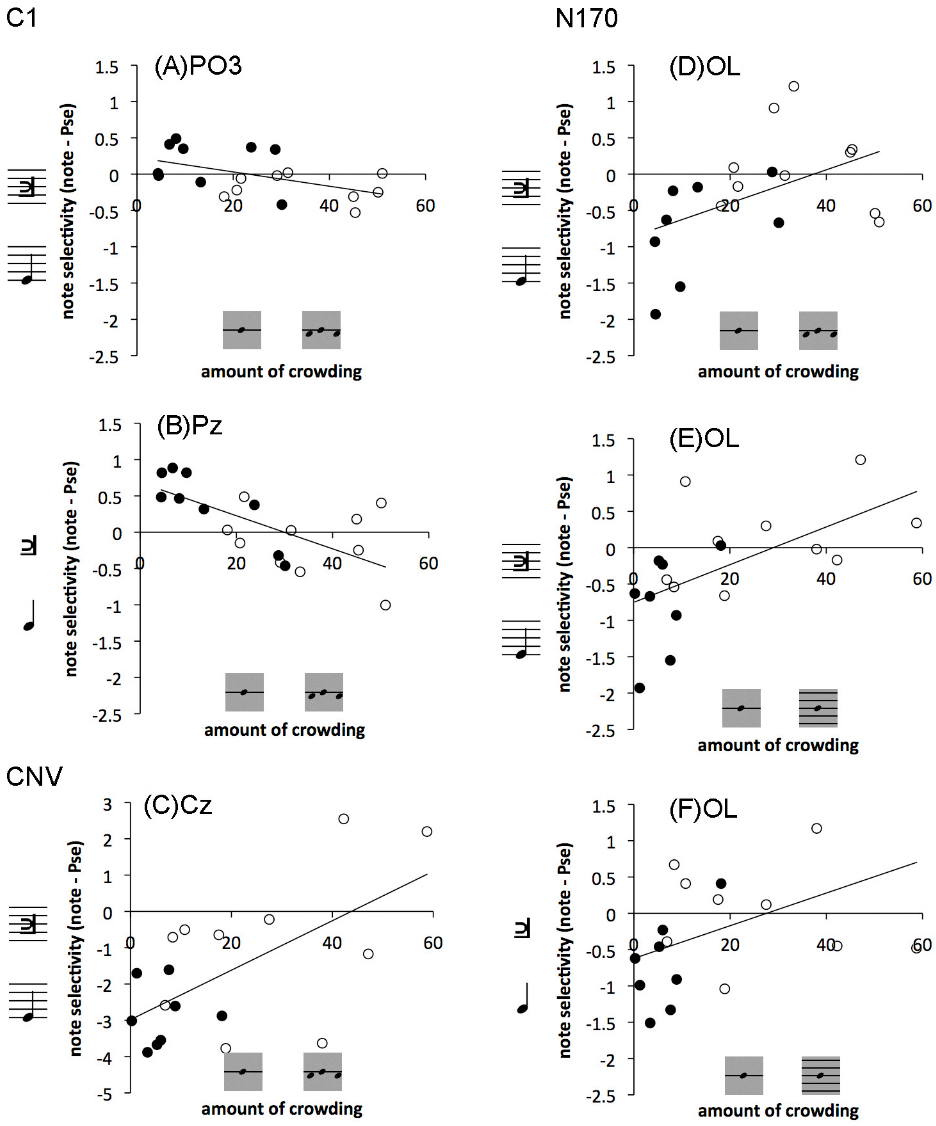

Figure 32. Crowding predicts the selectivity for musical notes with various ERP components. Examples showing that crowding with notes was predicted by the C1 for on-staff conditions at PO3 (A); no-staff conditions at Pz (B); the CNV for on-staff conditions at Cz (C); and the N170 for on-staff conditions at OL (D). Crowding with lines was predicted by the N170 at OL for on-staff conditions (E) or no-staff conditions (F). Data points for experts are the black circles while that for novices are the open circles.

The N170 selectivity for notes predicted both the amount of crowding with flanker notes and that with extra lines. For on-staff conditions, the N170 selectivity for notes predicted the crowding with notes at OL (*r* = .47, *p* = .049; Fig. 32d) and at T5 (*r* =

.57, *p* = .014), and predicted the crowding with lines at OL (*r* = .59, *p* = .010; Fig. 32e) and at T5 (*r* = .70, *p* = .001). For notes without staff, the N170 selectivity for notes predicted the crowding with lines at both sites (for OL, *r* = .46, *p* = .053; Fig. 32f; for T5, *r* = .48, *p* = .045).

For the CNV, selectivity for notes with staff predicted crowding with lines at Cz (*r* = .61, *p* = .005; Fig. 32c) but not for notes without staff.

Within the expert group only, the only significant correlation was that between crowding by flanker notes and the C1 for no-staff conditions at Pz, *r* = -.88, *p* = .002. The correlations between crowding and the N170 or CNV were not significant.

Predicting ERPs with holistic processing

For holistic processing, the correlation analyses were performed within the expert group since the congruency effect truly reflects a perceptual tendency only for experts (Wong & Gauthier, in press). Analyses for the congruency effect focused on the delta RT measure (since it was the measure that revealed the largest group differences in the previous study). All four conditions (center/periphery target positions x two target distributions 25p75c / 75p25c) were tested.

The only condition that produced significant correlations was the center-target trials in 75p25c. Among experts, the congruency effect predicted the C1 and the N170. For the C1, the congruency effect was positively correlated with notes without staff bilaterally, including PO3 (*r* = .83, *p* = .006; Fig. 33a) and PO4 (*r* = .92, *p* = .001; Fig. 33b). The congruency effect in the same condition was negatively correlated with the

N170 at T5 ($r = -.73$, $p = .039$; Fig. 33c). These suggest that experts who have a larger holistic effect tend to have a larger C1 and N170 selectivity for notes.

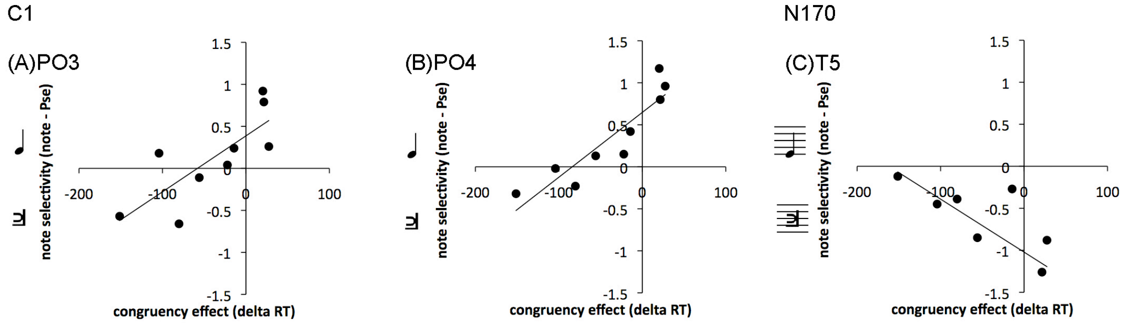

Figure 33. Holistic processing predicts the selectivity for musical notes with various ERP components. All plots show the congruency effect for the same condition (center-target trials in 75p25c). Within experts, the congruency effect was positively correlated with the C1 for no-staff conditions on PO3 (A) and PO4 (B), and was negatively correlated with the N170 for on-staff conditions at T5 (C).

General Discussion

The ERP selectivity for musical notes was predicted by all of the behavioral measures included in this study (Table 2). Perceptual fluency with notes predicts selectivity for notes with the C1, N170 and CNV, in which better music readers tend to have a larger C1, N170 and CNV selectivity for notes. Crowding with extra notes predicts the C1 and the N170, and crowding with extra lines predicts the C1, N170 and CNV, with participants showing a smaller crowding effect also show a larger C1, N170 and CNV selectivity for notes. Finally, the holistic processing of music sequences (among experts) is correlated with both the C1 and the N170, in which experts showing a larger holistic effect tend to have a larger C1 and N170 selectivity. These results are consistent with the fMRI findings that the holistic processing of music sequences is related to early visual processes bilaterally (unpublished data, CHAPTER I).

Table 2. Summary of the result of the correlation analyses. Only significant results or results with marginal significance ($p < .08$, indicated with ‘#’) are included. The ‘+’ signs indicate correlations performed within the expert group.

	C1		**N170**		**CNV**	
	on-staff	no-staff	on-staff	no-staff	on-staff	no-staff
perceptual fluency	Pz r = -.55 PO4 r = -.43$^{\#}$		OL r = .54 T5 r = .58	OL r = .49 T5 r = .49	Cz r = .45$^{\#}$	
crowding						
flanker notes	PO3 r = -.51	PO3 r = -.49 PO4 r = -.49 Pz r = -.67 Pz r = -.88^{+}	OL r = .47 T5 r = .57			
extra lines			OL r = .59 T5 r = .70	OL r = .46$^{\#}$ T5 r = .48	Cz r = .61	
holistic processing		PO3 r = .83^{+} PO4 r = .92^{+}	T5 r = -.73^{+}			

From the correlation results, it appears that perceptual fluency does not predict ERP effects as well as crowding or holistic processing. Specifically, since experts perform better than novices in all these behavioral measures, correlations across groups may merely reflect group differences instead of a linear relationship. Instead, obtaining correlation within the expert group is more informative about the linear relationship between behavioral measures and ERP effects. Such evidence was obtained for crowding and holistic processing but not for perceptual fluency (Table 2). Indeed, the scatter plots from Figure 31 suggest that correlations with perceptual fluency are driven by group differences, since little linear trend can be observed within each group. Interestingly, the weaker relationship between perceptual fluency and neural selectivity for notes (as compared to that with other behavioral measures) was not only observed in ERPs, but also observed in the prior fMRI results, in which perceptual fluency did not predict visual selectivity for notes in the visual cortex, but holistic processing did (Wong & Gauthier, 2010; unpublished data, CHAPTER I).

Why is perceptual fluency less useful in predicting neural selectivity for notes as compared to other behavioral measures? One explanation is that neural selectivity for notes is simply not mediated by one's perceptual skill for musical notes, but rather by other group differences that are acquired in musical training, such as verbal naming, auditory memory of the relative pitch differences across notes, or motor execution. While this may be the case, this would not explain why other visual perceptual measures predict neural selectivity for notes in ERPs or in BOLD signals (such as crowding and holistic processing), which suggests that visual perceptual ability with musical stimuli does capture some variability in the neural measures for note selectivity.

Another plausible explanation is that perceptual fluency is a crude test for general music reading ability. Perceptual fluency was indexed using a threshold to measure how quickly one can perceive a four-note music sequence with enough details such that they can accurately match the presented sequence among two highly similar choices (one of the notes had one step off in the distractor sequence). On the one hand, at this early stage of investigating expertise effects with music, we have not yet evaluated the reliability of this measure. It is possible that this measure of perceptual fluency is not sufficiently reliable to capture anything other than the largest differences between groups. On the other hand, while fluency is a basic component in reading music, and there is no doubt that experts have acquired skills to perceive musical notes more fluently than novices (Wong & Gauthier, 2010; in press; CHAPTER II & III), higher fluency may be achieved by various means. For example, an expert who has developed sensitivity to the relative position of the notes (related to the holistic processing measure) can better discriminate between two highly similar sequences since the relative position of the notes is altered in the distractor sequence. Another expert who has developed a precise representation of a note on a line versus a note between two lines (related to the crowding measure) can also better discriminate between two highly similar sequences, since the shifted note in the distractor sequence is moved either from a note on a line to between two lines or vice versa. Other experts who have acquired a highly automatic multimodal representation for notes may better discriminate between two similar sequences because different sequences prompt different auditory, somatosensory or motor representations of the notes (as suggested by the correlation between perceptual fluency and multimodal areas, Wong & Gauthier, 2010). In other words, perceptual fluency for notes may be supported by multiple visual abilities or even multimodal abilities, making it less suitable for predicting the specific functions underlying the recruitment of specific brain areas. In contrast, measures of crowding or holistic processing may be specific components that can contribute to perceptual fluency for notes, and at the same time precise enough to pinpoint specific functional recruitment of different brain regions, as suggested by the correlation results with ERPs and fMRI.

In sum, better music readers tend to have a smaller crowding effect with musical stimuli (CHAPTER III) and a larger holistic effect (Wong & Gauthier, in press), and all these results converge to suggest a coherent picture: Better music readers tend to have a larger C1, N170 and CNV selectivity for notes, a smaller crowding effect created by extra lines or flanker notes, and a larger holistic effect. The relationships between these factors are potentially more complex since these factors share variances. Future studies may use multivariate methods with a larger sample of experts to better reveal how these factors are related to each other.

The correlation results also confirm the behavioral significance of the ERP components. In particular, although the C1 selectivity for musical notes without staff was susceptible to pre-stimulus noise and had a different topographic distribution, the correlation between the C1 and the crowding effect (with flanker notes) suggested that the C1 effect was not merely random noise.

CHAPTER V

CONCLUDING REMAKRS AND FUTURE DIRECTIONS

Summary and overview

This dissertation was motivated by the surprising finding of neural selectivity for musical notes in early visual cortex (Wong & Gauthier, 2010), and the fact that note selectivity predicted individual degrees of holistic processing within music reading experts (unpublished data). As discussed in CHAPTER I, selectivity for objects of expertise in early visual cortex is not expected from theories of object recognition, from previous findings about object recognition, or from previous findings about the brain regions recruited for perceptual expertise. In the current study, the temporal dynamics of the neural selectivity for musical notation were examined using scalp electrophysiological recordings, taking advantage of the high temporal resolution of ERPs to test whether the early visual selectivity observed in fMRI was more likely the result of feedforward processes with altered V1 cell responses, or the result of strengthened feedback processes from higher areas. Several behavioral measures were included as behavioral correlates to explore the behavioral significance of the ERP expertise effects, including perceptual fluency of notes that quantifies individual expertise, holistic processing of notes that predicted the fMRI early visual selectivity, and crowding with musical stimuli that was included because crowding has been associated with early visual cortex.

CHAPTER II to IV reported the findings of the ERP study, the crowding study, and the results of the correlation analyses between the ERP expertise effects and the three behavioral measures. As reported in CHAPTER II, expertise effect for notes were obtained with various ERP components, including the C1 component bilaterally (40-60ms), the N170 component bilaterally (120-200ms), and the CNV component (-200-0ms). The N170 effects were obtained for both musical notes with or without staff, while the C1 and CNV were only obtained with notes on staff. CHAPTER III reported an expertise effect for crowding, in which experts experienced less crowding for musical stimuli (created by adding extra lines or flanker dots) but not for non-musical novel stimuli (Landolt C). Correlation analyses in CHAPTER IV revealed the behavioral significance of the expertise effects obtained with the C1, N170 and CNV components. Both the C1 and N170 expertise effects were predicted by all behavioral measures, including perceptual fluency, crowding with extra notes, crowding with extra lines and holistic processing, while the CNV expertise effect was predicted by perceptual fluency and crowding with extra lines.

In sum, the ERP results suggest that the fMRI expertise effect observed in the early visual cortex (Wong & Gauthier, 2010) is, at least partly, a result of feedfoward visual processes. Since the N170 expertise effects for musical notes were found in both hemispheres, it is possible that the fMRI expertise effect was partly contributed by a feedforward-feedback loop between early and late visual areas. However, it is not easy to test this possibility with the current ERP technique. Even if similar N170 effects are observed at the posterior parietal channels where early visual activity is typically observed, it is hard to determine whether the N170 effects indeed come from the early visual cortex (given the inverse problem of source localization of ERPs). In contrast, no P3 expertise effect was observed with musical notes, suggesting that the fMRI expertise effect is unlikely a result of top-down effects such as expectancy- or semantics-related processes.

Implications and future directions

Music reading expertise and early visual cortex

Music reading expertise recruits V1

Obtaining an expertise effect in the early part of the C1, as early as 40 to 60ms, indicates that the initial feedforward processes of musical notes are different between experts and novices. Neural activity as early as this time window is considered to be sensory-evoked and is largely contributed by the primary visual cortex (Foxe & Simpson, 2002; Schmolesky et al., 1998), consistent with the observation in the prior fMRI study that early visual cortex is recruited for musical notes with the acquisition of music reading expertise (Wong & Gauthier, 2010). Note that V1 may not be the only source that generates the early C1 effect, as it is possible that a small portion of the cells in the next processing stages, such as V2 and V3, are already activated in this early time window and contribute to the early visual selectivity for musical notes by some feedforward-feedback loops. Taking the ERP and the fMRI findings together, it is highly likely that V1 is one of the major sources of the C1 expertise effect. Future experiments may consider using TMS to selectively affect the activity of the early visual cortex and see if music reading performance will be affected. Given the associations between the C1 expertise effect and the wide range of behavioral performances, including perceptual fluency, crowding and holistic processing, one should observe a larger decrease in performance in music reading experts compared to novices.

Response properties of V1 cells with music reading expertise

In what ways are the response properties of V1 cells changed with music reading expertise? It is possible to speculate based on the properties of the C1 effect obtained in the current study. First, the C1 expertise effect was obtained for the on-staff conditions. Since all stimulus categories shared an identical five-line background, the effect cannot be explained by the sensitivity of early visual cortex to the lines or to the spatial frequency of the lines. It also appears unlikely that participants paid more attention to the staff lines specifically for the note condition, since the position of the notes on the lines was largely task-irrelevant (the one-back task could be performed by judging whether the notes are pointing upward or downward, or by the number of tails on the stem of the notes). In addition, the early visual effect for the no-staff conditions is possibly different from a typical C1 effect (given its different topographic distribution), and is at least less as extensive as the on-staff conditions, which was found in all the tested channels (PO3/4 and Pz). Based on these findings, the C1 effect for notes with staff may be related to an interaction between the shape of the notes and the five-line staff. To speculate, one possibility is that some V1 cells that are selective to the staff may interact automatically with cells that are selective to the shape of the notes, which give rise to the selectivity for notes for the C1. Alternatively, with extensive experience, some V1 cells may become selective for the whole stimulus of musical notes, where the shape of musical notes is always considered with the staff lines to process the notes meaningfully. Therefore, selectively presenting either the staff (in combination with letters or pseudo-letters) or the shape of the notes without the staff (no-staff conditions) does not activate these cells, and thus does not result in a similar C1 effect. This hypothesis may be tested by adaptation studies to see how much the neural substrates responsible for the C1 effect can be adapted by the staff lines or the shape of the notes.

In addition, both the fMRI study (Wong & Gauthier, 2010) and the current ERP study converged to suggest that V1 selectivity for musical notes increases with a higher degree of holistic processing, even though neither of the ERP or fMRI measures are related to music sequences or have any congruency manipulations. Holistic processing in music reading experts may be caused by automatic encoding of relative positions of adjacent notes in music sequences (Wong & Gauthier, in press). Together with the finding that early visual cortex is selective for music sequences (Wong & Gauthier, 2010), it is possible that early visual cortex also codes the relative position of the notes. Future studies may test the C1 effect with music sequences and add congruency manipulations to test this hypothesis.

Is V1 specifically recruited for musical notes?

Is the recruitment of early visual cortex specific for the category of musical notes? It is possible that previous studies on other expertise domains have missed the early visual selectivity simply because it is not expected, or because there were not enough trials to gain enough statistical power to reveal the early visual effects (about 100-200 trials are typically included for each condition for a typical N170 studies related to perceptual expertise, while the current study had 660 trials for each condition). The analyses with letters did not reveal any C1 differences between letters and pseudo-letters.

Although a novice group is not always necessary for revealing C1 effects (at least in the case of musical notes), including a novice group would provide a more powerful contrast to reveal any expertise effects in this early time window. Further studies are required to investigate whether such early visual effects can be obtained with letters or other types of perceptual expertise.

Why is V1 recruited for musical notes?

What component(s) of music reading expertise drives the recruitment of early visual cortex? There are at least two possible hypotheses. One hypothesis is that the early visual cortex is recruited because of the task demands of music reading, including fast recognition and higher spatial resolution of the encoding. In music reading, one needs to recognize multiple musical notes simultaneously that are crowded with extra lines and other notes that are close together. It is perceptually very challenging, especially when some of the notes may fall outside of the fovea. Music reading experts are trained to read multiple musical notes accurately and efficiently within a very short time such that they can execute the designated movement accurately. One way to fulfill the task demand is to represent music sequences in early visual cortex, which would have several advantages. First, it is much faster for information to reach the early visual cortex compared to higher visual cortex. Therefore, representing musical notes in early visual cortex can speed up the visual processes. Second, the early visual cortex is retinotopically organized and has small receptive fields, and thus contains a high spatial resolution representation of the visual world (Lee, 2002; Mumford, 1991). Representing musical notes in early visual cortex, such as having a precise representation of whether the dot is on or off a line, or having a representation of the relative positions of the notes with high spatial resolution, allows multiple crowded musical notes to be processed simultaneously in parafoveal region. Indeed, it has been suggested that perceptual learning that requires simultaneous recognition of multiple briefly-presented objects can lead to the recruitment of the early visual cortex (Sigman & Gilbert, 2000; Sigman et al., 2005). In other words, the task demand of recognizing multiple crowded objects quickly outside of the fovea, which requires high spatial resolution to achieve object individuation, may drive the recruitment of early visual cortex. Consistent with this hypothesis, a recent training study using a visual search task that requires participants to search for a novel object (Ziggerins) in a certain target orientation (e.g. 0º) simultaneously presented with an array of seven identical distractors (plane-rotated in 90º, 180º or 270º) resulted in the recruitment of the early visual cortex (Wong, Folstein, & Gauthier, 2010; see also Sigman et al., 2005).

Alternatively, music reading is an essential part of music performance which requires multimodal integration of visual, auditory, somatosensory and motor processes. In the prior fMRI study, it is demonstrated that simple visual judgment with musical notes automatically recruits a widespread multimodal network, including auditory, somatosensory, motor and other frontal regions (Wong & Gauthier, 2010). Previous work has shown that simultaneously processing information presented in two modalities, regardless of whether the 2nd modality is task relevant or not, results in changes in the C1 response (Fort et al., 2002; Giard & Peronnet, 1999; Karns & Knight, 2008). Such modulation of the C1 may occur because of a sensory gain, i.e., an increased neural activity of the visual cells with additional sensory information from another modality, or because multimodal stimuli recruit neurons that are not activated solely by visual inputs in or near the striate cortex (Giard & Peronnet, 1999). It is possible that extensive experience in music reading that is coupled with multimodal processes have induced long-term changes in the cell response in early visual cortex towards the musical notes, such as by an increased neural response of the same visual cells or by automatically recruiting more cells for musical notes that are not normally activated by other visual stimulus.

Future Directions

One of the immediate questions that can be asked is whether the predictability of the category of the coming stimulus is important. In reading words or musical notes, the category of the stimulus is stable and predictable, and this characteristic of the reading task may be important to obtain the early visual selectivity for notes. In both the ERP and the fMRI study (Wong & Gauthier, 2010), a block design was used in which the category of the upcoming stimulus is 100% predictable. Such knowledge may help to set up appropriate interaction between feedback connections and local circuits that can most efficiently process the next musical note, as a result of extensive learning experience (Gilbert & Sigman, 2007). If predicting the category of the upcoming stimulus is important, it is expected that the C1 selectivity for musical notes cannot be observed when the stimulus category is randomized. A similar C1 selectivity for notes would suggest that the sensory-evoked selectivity in V1 does not require setting up a contextual neural network for processing musical stimuli.

Perceptual expertise and object recognition

Theories and models of object recognition hypothesize that object recognition and individuation of objects within a category are achieved in higher visual cortex (DiCarlo & Cox, 2007; Grill-Spector & Malach, 2004; Kourtzi & DiCarlo, 2006; Riesenhuber & Poggio, 1999). The present findings suggest that object selectivity can be obtained during the initial feedforward processes in early visual cortex, and the role of early visual cortex in object recognition is more than merely local and featural encoding (Hubel & Wiesel, 1968). These findings suggest that both early and higher visual cortex can be selective for objects of expertise, possibly depending on the task demand of the domain of expertise (Wong et al., 2009b). Future work should investigate what components of various visual perceptual skills are critical to determine whether early areas, late areas or both would be recruited for objects of expertise.

Crowding

In the literature, it has been assumed that crowding is independent of object category, with perceptual expertise or not (e.g. Pelli & Tillman, 2008). However, recent studies suggest that crowding can be modulated by prior experience, such as practice with the same task (Chung, 2007; Huckauf & Nazir, 2008), by one's native language (Williamson et al., 2009) or by experience with playing video games (Green & Bavelier, 2007). The present study provides more direct evidence that perceptual expertise can alleviate crowding specifically with objects of expertise without direct practice on the task. Crowding with musical stimuli can be predicted by individual expertise in music reading (quantified by perceptual fluency), suggesting that the ability to uncrowd musical stimuli is related to music reading ability. Furthermore, crowding with musical stimuli can be predicted by the C1 and the N170, consistent with the idea that crowding can be related to multiple levels of visual processes (Millin, Arman, & Tjan, 2010). The relationship between crowding and the C1 component also supports the hypothesis that crowding is related to early visual processes (Arman et al., 2006; Fang & Sheng, 2008; Tjan & Nandy, 2010).

The present findings highlight the relationship between crowding and perceptual expertise. Instead of being independent of object category (Pelli & Tillmann, 2008), experts experience less crowding compared to novices with objects of expertise. Future work should investigate the mechanisms by which perceptual expertise can help experts to uncrowd objects of expertise, including facilitation due to a better representation of the objects of expertise, or facilitation due to labels associated with the objects, or a reduction in the obligatory integration of the target or flankers in crowding (e.g. better selective attention for objects of expertise). Future studies are required to tease apart these possible mechanisms in reducing crowding. In addition, since uncrowding objects of expertise may recruit different mechanisms from that for novel objects, it would be important to consider the influence of perceptual experience with the stimulus set on crowding-related findings, especially in cases where a small set of stimuli was used or substantial prior practice was given before measurement (e.g. Louie et al., 2007; Martelli, 2005; Petrov et al., 2007; Saarela et al., 2009; Tripathy & Cavanagh, 2002; Zhang et al., 2009).

Final conclusions

This dissertation clarifies the mechanisms underlying the recruitment of early visual cortex for music reading expertise. It reveals that early visual cortex is recruited by music reading expertise during the initial feedforward processes, suggesting that the role of early visual cortex can sometimes be object selective with extensive perceptual experience. Neural selectivity for musical notes as early as 40-60ms is the earliest expertise effect observed to date. This work demonstrates that music reading expertise is a useful domain to study how the visual system works and how experience alters our visual mechanisms.

REFERENCES

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli. *Cerebral Cortex, 9*, 415-430.

Arman, A. C., Chung, S. T. L., & Tjan, B. (2006). Neural correlates of letter crowding in the periphery [Abstract]. *Journal of Vision, 6(6)*, 804a.

Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J Neurosci, 20*(9), 3263-3281.

Belkic, K., Savic, C., Djordjevic, M., Ugljesic, M., & Mickovic, L. (1992). Event-related potentials in professional city drivers: Heightened sensitivity to cognitively relevant visual signals. *Physiology & Behavior, 52*, 423-427.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience, 8*(6), 551-565.

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex, 19*, 1583-1596.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177-178.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Brunia, C. H. M., Van Boxtel, G. J. M., & Bocker, K. B. E. (in press). Negative slow waves as indices of anticipation: The bereitschaftspotential, the contingent negative variation, and the stimulus preceding negativity. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends Cogn Sci, 10*(4), 159-166.

Busey, T., & Vanderkolk, J. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*(4), 431-448.

Cheung, O. S., Richler, J. J., Palmeri, T., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *J Exp Psychol Hum Percept Perform, 34*(6), 1327-1336.

Chung, S. (2007). Learning to identify crowded letters: Does it improve reading speed? *Vision Research, 47*(25), 3150-3159.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human brain mapping, 2*, 170-187.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience, 8*(5), 387-402.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*(Pt 2), 291-307.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of Mind* (pp. 1-26). New York: Oxford University Press.

de Heering, A., & Rossion, B. (2008). Prolonged visual experience in adulthood modulates holistic face perception. *PLoS ONE, 3*(5), e2317.

Deutsch, D. (1998). *The Psychology of Music.* (2nd ed.). London: Academic Press.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of experimental psychology General, 115*(2), 107-117.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn Sci, 11*(8), 333-341.

Downing, P. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science, 293*(5539), 2470-2473.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron, 23*(1), 115-125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci, 22*(13), 5749-5759.

Fang, F., & Sheng, H. (2008). Crowding alters the spatial distribution of attention modulation in human primary visual cortex. *Journal of Vision, 8(9)*(6), 1-9.

Farah, M.-J., Wilson, K.-D., Drain, M., & Tanaka, J.-N. (1998). What is "Special" about Face Perception? *Psychological Review, 105*(3), 482-498.

Fort, A., Delpeuch, C., Pernier, J., & Giard, M.-H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex, 12*, 1031-1039.

Foxe, J. J., & Simpson, G. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining 'early' visual processing. *Experimental Brain Research, 142*, 139-150.

Foxe, J. J., Strugstad, E. C., Sehatpour, P., Molholm, S., Pasieka, W., Schroeder, C. E., et al. (2008). Parvocellular and magnocellular contributions to the initial generators of the visual evoked potential: High-density electrical mapping of the 'C1' component. *Brain Topogr, 21*, 11-21.

Freeman, J., & Simoncelli, E. P. (2010). Crowding and metamerism in the ventral stream [Abstract]. *Journal of Vision*.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol, 14*(7), 573-578.

Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends Cogn Sci (Regul Ed), 4*(1), 1-2.

Gauthier, I., Curby, K. M., Skudlarski, P., & Epstein, R. A. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive, Affective, & Behavioral Neuroscience, 5*(2), 222-234.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat Neurosci, 6*(4), 428-432.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci, 3*(2), 191-197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 431-446.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568-573.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*(15/16), 2401-2428.

Gauthier, I., Wong, A. C.-N., Hayward, W. G., & Cheung, O. S.-C. (2006). Font-tuning associated with expertise in letter perception. *Perception, 35*(4), 541-559.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490.

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron, 54*(5), 677-696.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res, 41*(10-11), 1409-1422.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci, 3*(8), 837-843.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu Rev Neurosci, 27*, 649-677.

Gunter, T. C., Schmidt, B. H., & Besson, M. (2003). Let's face the music: a behavioral and electrophysiological exploration of score reading. *Psychophysiology, 40*(5), 742-751.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334-337.

Hopf, J. M., Vogel, E. K., Woodman, G., Heinze, H. J., & Luck, S. (2002). Localizing visual discrimination processes in time and space. *J Neurophysiol, 88*(4), 2088-2095.

Horovitz, S. G., Rossion, B., Skudlarski, P., & Gore, J. C. (2004). Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *NeuroImage, 22*, 1587-1595.

Hsiao, J. H., & Cottrell, G. W. (2009). Not all visual expertise is holistic, but it may be leftist: The case of chinese character recognition. *Psychological Science, 20*(4), 455-463.

Hubel, D. H., & Wiesel, T. N. (1968). *Receptive Fields and Functional Architecture of Monkey Striate Cortex.* Paper presented at the J. Physiol., London.

Huckauf, A., & Nazir, T. (2007). How odgcrnwi becomes crowding: stimulus-specific learning reduces crowding. *JOV, 7*(2), 18.11-12.

James, K. H., James, T. W., Jobard, G., Wong, A. C., & Gauthier, I. (2005). Letter processing in the visual system: different activation patterns for single letters and strings. *Cognitive, affective & behavioral neuroscience, 5*(4), 452-466.

Jeffreys, D. A., & Axford, J. G. (1972). Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. *Experimental Brain Research, 16*, 1-21.

Jehee, J., Roelfsema, P., Deco, G., Murre, J., & Lamme, V. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons—Explaining crowding phenomena. *Brain Research, 1157*, 167-176.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53*(6), 891-903.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosc., 17*, 4302-4311.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250-252.

Karns, C. M., & Knight, R. T. (2008). Intermodal auditory, visual and tactile attention modeulates early stages of neural processing. *Journal of Cognitive Neuroscience, 21*(4), 669-683.

Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb Cortex, 18*, 2629-2636.

Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Curr Opin Neurobiol, 16*(2), 152-158.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Lamme, V., & Roelfsema, P. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci, 23*(11), 571-579.

Lee, T. (2002). Top-down influence in early visual processing: a Bayesian perspective. *Physiol Behav, 77*(4-5), 645-650.

Lee, T., Yang, C., Romero, R., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat. Neurosci., 5*(6), 589-597.

Lerner, Y., Epshtein, B., Ullman, S., & Malach, R. (2008). Class Information Predicts Activation by Object Fragments in Human Object Areas. *Journal of cognitive neuroscience*.

Leuthold, H., Sommer, W., & Ulrich, R. (2004). Preparing for action: Inferences from CNV and LRP. *Journal of Psychophysiology, 18*, 77-88.

Levi, D. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635-654.

Levi, D., & Waugh, S. J. (1994). Spatial scale shifts in peripheral vernier acuity. *Vision Res, 34*(17), 2215-2238.

Levy-Agresti, J., & Sperry, R. W. (1968). Differential perceptual capacities in major and minor hemispheres. *Proceedings of the National Academy of Sciences, 61*, 1151.

Leynes, P. A., Allen, J. D., & March, R. L. (1998). Topographic differences in CNV amplitude reflect different preparatory processes. *International Journal of Psychophysiology, 31*, 33-44.

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. *JOV, 7*(2), 24.21-11.

Loveless, N. E. (1973). The contingent negative variation related to preparatory set in a reaction time situation with variable foreperiod. *Electroencephalography and Clinical Neurophysiology, 35*, 369-374.

Luck, S. (2005). *An introduction to the event-related potential technique*, Cambridge, MA: MIT Press.

Maertens, M., & Pollmann, S. (2005). fMRI reveals a common neural substrate of illusory and real contours in V1 after perceptual learning. *Journal of cognitive neuroscience, 17*(10), 1553-1564.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A, 92*(18), 8135-8139.

Martelli, M. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *JOV, 5*(1), 58-70.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriatte visual cortical areas in spattial attention. *Nat Neurosci, 2*(4), 364-369.

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

McEvoy, L. E., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cereb Cortex, 8*, 563-574.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological science : a journal of the American Psychological Society / APS, 17*(7), 608-615.

Millin, R., Arman, A. C., & Tjan, B. S. (2010). Reduced neural activity with crowding is independent of attention and task difficulty [Abstract]. *Journal of Vision*.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *J Neurosci, 26*(43), 11187-11196.

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J Neurosci, 27*(42), 11401-11411.

Muller, M., Hofel, L., Brattico, E., & Jacobsen, T. (2010). Aesthetic judgments of music in experts and laypersions - An ERP study. *International Journal of Psychophysiology, 76*, 40-51.

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological cybernetics, 65*(2), 135-145.

Munte, T. F., Altenmuller, E., & Jancke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat Neurosci Rev, 3*, 473-478.

Nakada, T., Fujii, Y., Suzuki, K., & Kwee, I. L. (1998). 'Musical brain' revealed by high-field (3 Tesla) functional MRI. *Neuroreport, 9*(17), 3853-3856.

Nunez, P. L. (1981). *Electric fields of the brain.* New York: Oxford University Press.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *J Neurosci, 26*(50), 13025-13036.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510-512.

Patterson, K. E., & Bradshaw, J. L. (1975). Differential hemispheric mediation of nonverbal visual stimuli. *J Exp Psychol Hum Percept Perform, 1*, 246-252.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nat Rev Neurosci, 8*(8), 636-648.

Pelli, D., & Tillman, K. (2008). The uncrowded window of object recognition. *Nat Neurosci*, 1129-1135.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Peretz, I., & Zatorre, R. J. (2003). *The cognitive neuroscience of music.* New York: Oxford University Press.

Petrov, Y., Popple, A. V., & McKee, S. P. (2007). Crowding and surround suppression: Not to be confused. *Journal of Vision, 7(2):12*, 1-9.

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlatets of rapid spatial orienting towards fearful faces. *Cereb Cortex, 14*, 619-633.

Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Res, 48*(1), 55-62.

Proverbio, A. M., & Adorni, R. (2009). C1 and P1 visual responses to words are enhanced by attention to orthographic vs. lexical properties. *Neurosci Lett, 463*, 228-233.

Richler, J. J., Bukach, C. M., & Gauthier, I. (2009a). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception & Psychophysics., 71*(3), 530-540.

Richler, J. J., Cheung, O. S., Wong, A. C.-N., & Gauthier, I. (2009b). Doe response interference contribute to face composite effects? *Psychonomic Bulletin & Review, 16*, 258-263.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*(2), 328-342.

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*(6), 1356-1368.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Rose, M., Verleger, R., & Wascher, E. (2001). ERP correlates of associative learning. *Psychophysiology, 38*, 440-450.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left lateralized face-like electrophysiological responses. *Psychological Science, 13*(3), 250-257.

Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 14521-14526.

Rotshtein, P., Geng, J. J., Diriver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience, 19*(9), 1435-1452.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9(2):5*, 1-11.

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol Rev, 75*(1), 107-154.

Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: changes in human brain activity with training in orientation discrimination. *Neuroimage, 9*(1), 46-62.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage, 32*(3), 1385-1394.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *J Neurophysiol, 79*(6), 3272-3278.

Schön, D., & Besson, M. (2002). Processing pitch and duration in music reading: a RT-ERP study. *Neuropsychologia, 40*(7), 868-878.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412*(6846), 549-553.

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proc Natl Acad Sci USA, 99*(26), 17137-17142.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience, 18*, 1453-1465.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging [see comments]. *Science, 268*(5212), 889-893.

Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992). Distributed neural network underlying musical sight-reading and keyboard performance. *Science, 257*(5066), 106-109.

Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nat Neurosci, 3*(3), 264-269.

Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron, 46*(5), 823-835.

Sloboda, J. A. (1976). Visual perception of musical notation: registering pitch symbols in memory. *Q J Exp Psychol, 28*(1), 1-16.

Sloboda, J. A. (1978). Perception of contour in music reading. *Perception, 7*(3), 323-331.

Spiro, J. (2003). Music and the brain. *Nat Neurosci, 6*(7), 661.

Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage, 20*(1), 71-83.

Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: Evidence for sensory plasticity in early affective perception. *Cereb Cortex, 16*, 876-887.

Sutton, S., Bararen, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. *Science, 150*, 1187-1188.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*(1), 43-47.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1-B9.

Tjan, B., & Nandy, A. S. (2010). Saccade-distorted image statistics explain target-flanker and flanker-flanker interactions in crowding [Abstract]. *Journal of Vision*.

Tong, F. (2003). COGNITIVE NEUROSCIENCE: Primary visual cortex and visual awareness. *Nat Rev Neurosci, 4*(3), 219-229.

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences, 2*(5), 174-183.

Travis, F., Tecce, J. J., & Guttman, J. (2000). Cortical plasticity, contingent negative variation, and transcendent experiences during practice of the Transcendental Meditation technique. . *Biological Psychology, 55*, 41-55.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Res, 42*(20), 2357-2369.

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS ONE, 6*(1), e1000646.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 308-384.

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

Williams, M., Baker, C., Op De Beeck, H., Mok Shim, W., Dang, S., Triantafyllou, C., et al. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci, 11*(12), 1439-1445.

Williamson, K., Scolari, M., Jeong, S., Kim, M.-S., & Awh, E. (2009). Experience-dependent changes in the topography of visual crowding. *Journal of Vision, 9*(11), 1-9.

Wong, A. C.-N., Gauthier, I., Woroch, B., Debuse, C., & Curran, T. (2005). An early electrophysiological response associated with expertise in letter perception. *Cognitive, Affective, and Behavioral Neuroscience, 5*(3), 306-318.

Wong, A. C.-N., Palmeri, T., & Gauthier, I. (2009a). Conditions for face-like expertise with objects: Becoming a Ziggerin expert - but which type? *Psychological Science, 20*(9), 1108-1117.

Wong, A. C.-N., Palmeri, T., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009b). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS One, 4*(12), e8405.

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2010). Perceptual learning recruits both dorsal and ventral extrastriate areas [Abstract]. *Journal of Vision*.

Wong, Y. K., & Gauthier, I. (2010). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience, 22*(4), 695-713.

Wong, Y. K., & Gauthier, I. (in press). Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices. *Cognitive, Affective, & Behavioral Neuroscience*.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance, 29*(1), 121-138.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cereb Cortex, 15*(8), 1234-1242.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145.

Young, A. W., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception, 10*, 747-759.

Yue, X., Tjan, B., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*(22), 3802-3811.

Zhang, J., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research, 49*(1), 44-53.

Zhang, W., & Luck, S. (2009). Feature-based attention modulates feedforward visual processing. *Nat Neurosci, 12*(1), 24-25.