Music-reading training alleviates crowding with musical notation

Yetta Kwailing Wong

Faculty of Education, University of Hong Kong, Pokfulam, Hong Kong
Cognition and Neuroscience Lab, Department of Applied Social Sciences, City University of Hong Kong, Kowloon Tong, Hong Kong ✉

Alan C.-N. Wong

Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong 🏠 ✉

Crowding refers to the disrupted recognition of an object by nearby distractors. Prior work has shown that real-world music-reading experts experience reduced crowding specifically for musical stimuli. However, it is unclear whether music-reading training reduced the magnitude of crowding or whether individuals showing less crowding are more likely to learn and excel in music reading later. To examine the first possibility, we tested whether crowding can be alleviated by music-reading training in the laboratory. Intermediate-level music readers completed 8 hr of music-reading training within 2 weeks. Their threshold duration for reading musical notes dropped by 44.1% after training to a level comparable with that of extant expert music readers. Importantly, crowding was reduced with musical stimuli but not with the nonmusical stimuli Landolt Cs. In sum, the reduced crowding for musical stimuli in expert music readers can be explained by music-reading training.

Introduction

Fluent music reading is crucial for musicians to expose themselves to a wide range of music pieces and activities. Empirical evidence shows that music-reading experts can encode short music sequences, each with four to five notes, at a rate of three sequences per second, which is three times faster than novices (Y. K. Wong & Gauthier, 2010a, 2010b, 2012; Y. K. Wong, Peng, Fratus, Woodman, & Gauthier, 2014). How do music-reading experts achieve such an amazing reading speed?

Here we focus on the relationship between reading speed and crowding. *Crowding*, which refers to the disruption of target recognition when distractor objects are located near the target, is considered a major bottleneck in object recognition and conscious perception (Levi, 2008; Pelli & Tillman, 2008; Whitney & Levi, 2011). Recent work has revealed that crowding can also occur when the distractor objects are placed farther away from the target (Harrison, Retell, Remington, & Mattingley, 2013; Herzog & Manassi, 2015; Sayim & Cavanagh, 2013). Crowding is especially robust in the visual periphery (Bouma, 1970), and the causes of crowding remain controversial (S. He, Cavanagh, & Intriligator, 1996; Levi & Waugh, 1994; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli & Tillman, 2008). In the reading literature, crowding is considered one of the major factors limiting the visual span (i.e., the number of letters one can reliably recognize without moving the eyes), which in turn limits one's reading speed (Legge et al., 2007; Pelli & Tillman, 2007).

Importantly, reduced crowding is associated with experience in various domains, such as video game playing (Green & Bavelier, 2007) and language learning (Williamson, Scolari, Jeong, Kim, & Awh, 2009). Crowding for faces is stronger with upright face flankers than with inverted face flankers, which could be explained by our richer experience with upright faces (Louie, Bressler, & Whitney, 2007). Several hours of uncrowding practice in the lab (i.e., learning to recognize target stimuli crowded by flankers) led to reduced crowding with Roman letters (Chung, 2007; Chung & Truong, 2013; Huckauf & Nazir, 2007; Sun, Chung, & Tjan, 2010), directly demonstrating that experience alleviates crowding.

Recently, reduced crowding has been observed in music-reading experts (Y. K. Wong & Gauthier, 2012). In this study, real-world music-reading experts experienced less crowding than novices when a simplified note (a dot presented on one horizontal line without the stem of the note) was crowded either by extra flanker dots or by extra lines. However, both groups experienced a similar level of crowding with the control nonmusical stimuli, Landolt Cs. These findings suggest that music-reading experience may reduce crowding specifically for musical stimuli. The findings, however, could be explained by a preselection bias: Those experts may have been born with better abilities for processing any stimuli resembling musical notation. Because of this inborn perceptual advantage, they could have found music reading easier and pursued musical training, thereby becoming efficient music readers. In other words, the reduced crowding in the experts could be the cause rather than the effect of the extensive music-reading experience.

The current study aimed to clarify this causal relationship between music-reading experience and crowding. Participants were trained to read four- to five-note sequences for 8 hr with an increasing speed. Their music-reading speed and crowding were measured before and after training. If music-reading experts showed a smaller crowding than novices due to the difference in music-reading experience, then we should expect reduced crowding with notes after training. Alternatively, if the reduced crowding in music-reading experts was simply caused by a preselection bias, the magnitude of crowding should not change after training.

Method

Participants

Twenty-eight intermediate-level music readers participated in the pretest. Twelve were excluded because their perceptual fluency for notes (see below) was high before training (below 500 ms), limiting the room for improvement. One was excluded because the contrast threshold for crowded stimuli (see below) was lower than 10%, preventing the observation of any potential reduction after training. A total of 15 participants (11 females, four males; average age = 22.3 years, SD = 3.51, range = 18–33 years) proceeded to the training. The average age of the onset of music training was 8.92 years (SD = 3.48, range = 4–15 years). All reported piano as their major instrument except for one flute player and one horn player, and all have passed the exam board of the Royal Schools of Music with the highest grade of 4 to 8 with their major instrument (M

= 6.07, SD = 1.58). Their self-rated ability in music reading and sight reading (playing the music score on the first read) was 5.13 (SD = 0.92) and 5.07 (SD = 1.49), respectively (1 = *novice level*, 10 = *expert level*). All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of the Chinese University of Hong Kong. They were paid HK\$50/hr and received up to HK\$600 for finishing the training (see below).

Stimuli

The experiment was conducted on a Mac Mini (Apple, Cupertino, CA) using Matlab (MathWorks, Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). A total of 17,600 four- to five-note sequences were randomly generated in Matlab with the constraints that (a) the notes were from D4 to G5, (b) there was no repeated sequence, and (c) the notes within each sequence did not repeat in pitch. The sequences were created with five different combinations of different dot sizes and stem lengths to simulate printing styles of different music scores such that the learned reading skill would not be specific to one specific printing style. The stimuli were black on a white background and subtended a visual angle of about 8.77° × 3.92°.

Training regimen

Participants underwent eight 1-hr training sessions within 2 weeks (Figure 1). The training contained 20-trial blocks with sequences randomly drawn from the stimulus set. For each target sequence, a distractor sequence was created with one of the four notes shifted by one step. The upward or downward shifts and the note position within the sequence that was altered were counterbalanced across sequences. In each trial, a mask appeared for 500 ms, followed by a target sequence for a varied duration and another mask for 500 ms (Figure 2A). Then the target sequence and a distractor sequence appeared side by side. Participants were required to select the target sequence by keypress within a response time window. Feedback was provided for each trial. For correct trials, the word *correct* and a bell-ringing sound were presented for 1s. For incorrect trials, the word *incorrect* and a door-banging noise were presented for 1s. The presentation duration of the target sequence was 1800 ms in the first block and was reduced after each block whenever participants attained 90% accuracy (i.e., 18 correct trials, or points, out of 20) in that block until the duration reached 60 ms at the end. The response time windows were 5, 3, and 2 s for presentation durations above 1000 ms,

Training Procedure

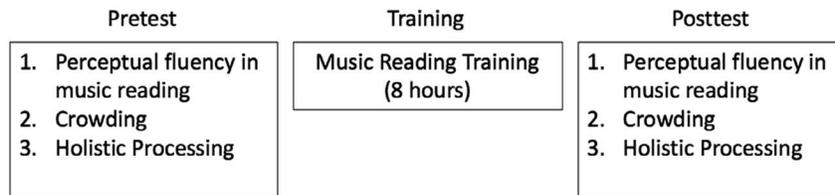| Pretest	Training	Posttest
1. Perceptual fluency in music reading 2. Crowding 3. Holistic Processing	Music Reading Training (8 hours)	1. Perceptual fluency in music reading 2. Crowding 3. Holistic Processing

Figure 1. The design of the training, pretests, and posttests.

between 1000 and 400 ms, and below 400 ms, respectively. To motivate participants, a special trial that was worth 3 points randomly appeared with a 1/80 chance. Also, participants were allowed to accumulate one, two, and three tokens with 60%, 75%, and 90% accuracy in a block, respectively. With 10 tokens, participants would obtain a chance to initiate the 3-point special trial when preferred. At most, three

(A) Fluency Task
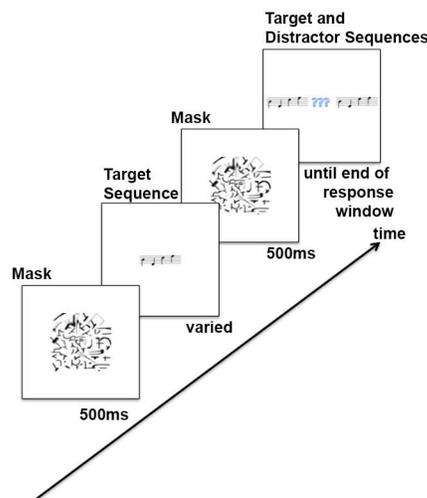
(B) Crowding task - uncrowded condition
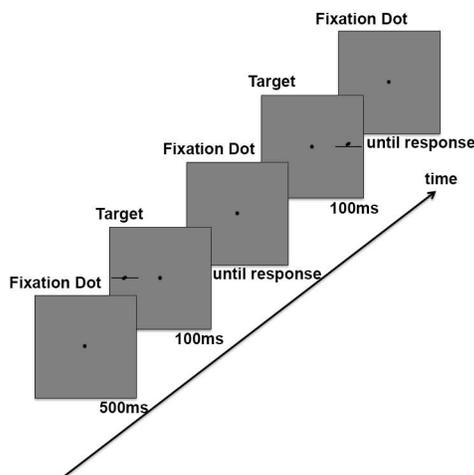
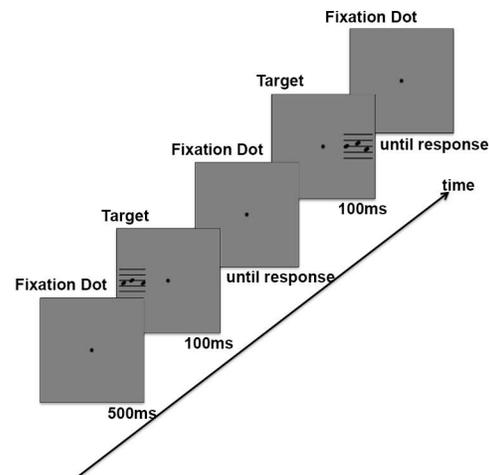**(C) Crowding task - crowded condition**

Figure 2. The experimental paradigm for (A) the training task, (B) the uncrowded condition of the crowding task, and (C) the crowded condition of the crowding task.

chances of initiating these special trials could be accumulated. Participants finished the training after 8 hr or after attaining 90% accuracy at the presentation duration of 60 ms.

Pretests and posttests

Participants completed three tasks, including perceptual fluency, crowding, and holistic processing, that were identical before and after training.

Measuring perceptual fluency

Perceptual fluency was measured to quantify individual music-reading ability using a sequential matching paradigm (Y. K. Wong & Gauthier, 2010a, 2010b, 2012; Y. K. Wong et al., 2014). A total of 360 four-note sequences were randomly generated using notes between D4 and G5. The contrast for all stimuli was lowered by about 60% from full contrast to avoid a ceiling effect. The images were presented at a visual angle of about 6.80° × 4.35°. The refresh rate of the monitor was 90 Hz. On each trial, a central fixation cross was presented for 200 ms, followed by a 500-ms premask, a target sequence for a varied duration, and a 500-ms postmask. Then the target sequence appeared side by side with another distractor in which one note was shifted by one step (randomly chosen out of the four notes, with the up–down shifts counterbalanced across trials). Participants had to select the target sequence by key press. The duration threshold to keep participants' accuracy at 80% was estimated using QUEST (Watson & Pelli, 1983) four times, each with 40 trials, and perceptual fluency was defined as the average duration thresholds (highly unstable estimates with a QUEST SD of 0.15 or above were excluded).

To control for individual differences not specifically tied to expertise with notes, perceptual fluency for four-digit strings was also measured with an identical procedure. A total of 360 four-digit strings were randomly generated using 0 to 9 (excluding 1). In the distractor string, one digit of the target string was randomly chosen and replaced by another digit randomly drawn from the set. The digit strings were shown with the same lowered contrast and visual angle as the note sequences.

Measuring crowding

The crowding stimuli and measurement followed that reported in our prior work (Y. K. Wong & Gauthier, 2012). Stimuli subtended about 1.3° × 1.3° in visual angle on a cathode ray tube monitor (106.0 cd/m^2) in a dimly lit room. The baseline of the musical stimuli included a line and a dot, sometimes crowded with four additional staff lines and two flanking dots (Figure 2). Participants were asked to judge whether the dot (or the central dot in crowded stimuli) was on a line or on the space. Landolt Cs were used as nonmusical controls, for which participants judged whether the gap was at the top or the bottom. The target–flanker distances ranged from 0.22° to 0.43° for musical stimuli and 0.64° for Landolt Cs—well within the critical spacing for crowding (half of the eccentricity of the stimuli = 1.3°; Bouma, 1970). The target–flanker distances were different for these stimuli because the same target–flanker distance used in the musical stimuli would lead to a floor effect for the Landolt Cs condition (i.e., participants failed to achieve 80% accuracy even with full contrast). Therefore, we increased the target–flanker distance for the Landolt Cs condition.

Each trial included a 500-ms central fixation and a 100-ms stimulus randomly to the left or right of the fixation at 2.6° (Figure 2B). Participants responded by key press without time limit. Trials for each condition were blocked with the order counterbalanced. The log Weber contrast threshold of the original full-contrast image for 75% accuracy was estimated four times, each with 40 trials, using QUEST (Watson & Pelli, 1983). The average log threshold served as the dependent measure, in which zero indicates the original full contrast and a smaller value corresponds to lower contrast. Participants were tested with the musical stimuli followed by the Landolt Cs, each with 24 practice trials with feedback before the testing without feedback.

Measuring holistic processing

The stimuli and design of the holistic processing task were similar to that reported in our prior work (Y. K. Wong & Gauthier, 2010a and 2010b). In brief, 512 four-note sequences were generated for the sequential matching task. The stimuli were black on a white background for about 4.62° × 6.77°. On each trial, a central fixation dot was presented for 1000 ms, followed by a 750-ms sequence, a 500-ms mask, and a second sequence for 2500 ms or until response. Participants performed same–different judgment on the two sequences as fast and as accurately as possible. There were 512 trials in total. Because this task is beyond the scope of this article, the details of the design and the results are not reported below.

Results

Training performance

Performance substantially improved throughout the 8-hr training for all participants. Eight of the 15
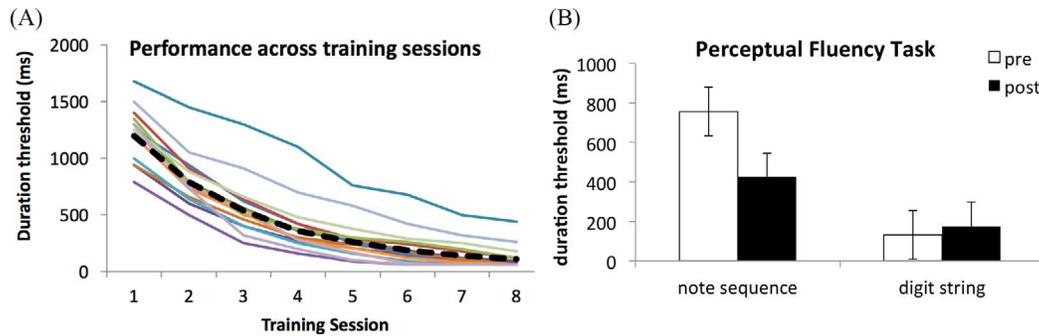
Figure 3. Improvement in music-reading fluency across training. (A) The presentation duration attained by participants in each 1-hr training session. The thin lines show changes of presentation duration in each participant, and the thick dashed line shows the group-averaged changes of presentation duration across training. (B) The duration threshold for maintaining participants at around 80% matching accuracy for music sequences or digit strings pre- and posttraining. Error bars plot the 95% confidence interval for the PrePost × Stimulus Type interaction.

participants finished the training by attaining the presentation duration of 60 ms with 90% accuracy. The mean presentation duration of the final training block was 113.4 ms (range = 60–440 ms, SD = 105.8 ms; Figure 3A).

Training effect on music-reading fluency

The PrePost (pretest, posttest) × Stimulus Type (notes, digits) analysis of variance on average duration threshold indicated a marginally significant main effect of PrePost, $F(1, 14) = 4.54$, $p = 0.051$, with a lower duration threshold at posttest than at pretest. The main effect of stimulus type was significant, $F(1, 14) = 63.2$, $p \leq 0.0001$, with a lower duration threshold for digits than for notes. More important, the PrePost × Stimulus Type interaction was significant, $F(1, 14) = 10.8$, $p = 0.005$, $\eta^2 = 0.44$. Post hoc Scheffé tests ($p < 0.05$) showed that the duration threshold for notes was lower at posttest ($M = 422.9$ ms) than at pretest ($M = 755.9$ ms), whereas that for digits remained similar after training (131.3 and 174.7 ms for pretest and posttest, respectively; Figure 3B), indicating that perceptual fluency for notes was specifically improved. In other words, a few hours of music-reading training effectively improved participants' music-reading fluency.

Training effect on crowding

For musical notes, the PrePost × Crowding (crowded, uncrowded) × Visual Field (left, right) ANOVA on log contrast threshold revealed a main effect of crowding, $F(1, 14) = 220.9$, $p \leq 0.0001$, with a higher contrast threshold for crowded notes than for uncrowded notes. The PrePost × Crowding interaction was significant, $F(1, 14) = 17.1$, $p = 0.001$, $\eta^2 = 0.55$. Scheffé tests ($p < 0.05$) showed that the contrast

threshold decreased after training for crowded stimuli but not for uncrowded stimuli. Interestingly, the three-way interaction between PrePost, crowding, and visual field was marginally significant, $F(1, 14) = 4.46$, $p = 0.053$, $\eta^2 = 0.24$ (Figure 4A). Separate PrePost × Crowding ANOVA were conducted for each visual field to further understand the interaction. In the left visual field, the PrePost × Crowding interaction was significant, $F(1, 14) = 26.6$, $p = 0.0001$, $\eta^2 = 0.66$. Scheffé tests ($p < 0.05$) showed that the contrast threshold for crowded notes decreased after training, whereas that for uncrowded notes increased. In contrast, the PrePost × Crowding interaction did not reach significance in the right visual field ($p = 0.19$). In sum, the music-reading training led to reduced crowding for notes in the left visual field only.

For Landolt Cs, a similar ANOVA on log contrast threshold revealed a main effect of crowding, $F(1, 14) = 197.7$, $p \leq 0.0001$, with a higher contrast threshold for crowded Landolt Cs than for uncrowded Landolt Cs. The main effect of visual field was significant, $F(1, 14) = 8.03$, $p = 0.013$, which interacted with crowding, $F(1, 14) = 7.96$, $p = 0.014$, $\eta^2 = 0.36$. Scheffé tests ($p < 0.05$) showed that the contrast threshold for crowded Landolt C was higher in the left visual field than in the right visual field, whereas that for uncrowded Landolt C was similar across visual fields. Importantly, none of the effects involving PrePost reached significance (all Fs < 1), indicating that the contrast threshold for Landolt Cs remained similar after training (Figure 4B).

We compared the training effects on crowding of musical notes and Landolt Cs with a PrePost × Crowding (crowded, uncrowded) × Object Category (notes, Landolt Cs) × Visual Field (left, right) ANOVA on log contrast threshold. Importantly, the PrePost × Crowding × Object Category interaction was significant, $F(1, 14) = 5.61$, $p = 0.033$, $\eta^2 = 0.29$. Scheffé tests ($p < 0.05$) showed that the contrast threshold for crowded notes was reduced after training, whereas that
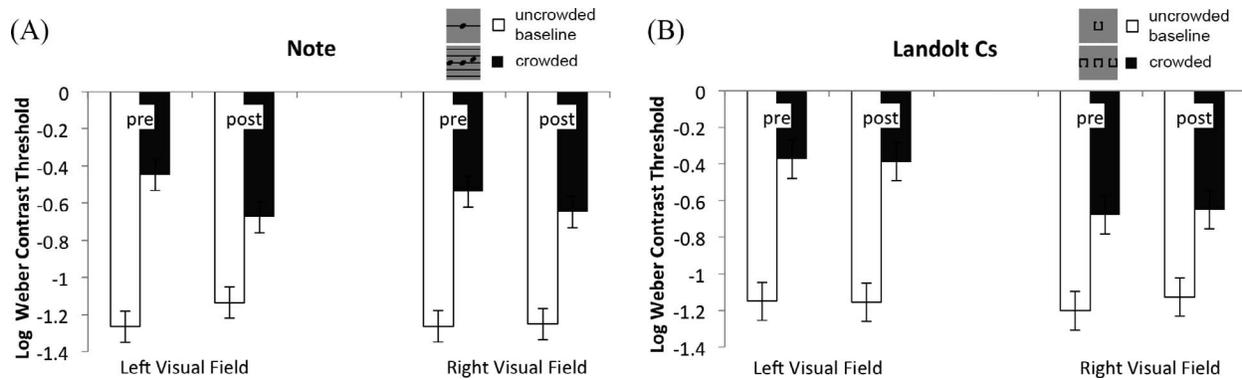
Figure 4. Log Weber contrast threshold for the crowding task pre- and posttraining. Error bars plot the 95% confidence interval for the PrePost $\times$ Crowding $\times$ Visual Field interaction for each stimulus category.

for uncrowded notes, crowded Landolt Cs, and uncrowded Landolt Cs remained similar after training. This was consistent with the interpretation that the music-reading training specifically reduced crowding for notes but not for nonmusical stimuli. This three-way interaction was not modulated by visual field ($p = 0.18$).

To test whether the improvement in music-reading fluency predicted the degree of reduction in crowding, the correlation between the log contrast threshold for crowded notes and the log duration threshold for music-reading fluency was tested separately in the left and right visual fields. Neither of the correlations reached significance ($ps > 0.5$).

Discussion

Following 8 hr of music-reading training in the laboratory, the threshold duration was reduced by 44.1% (from 755.9 ms at pretest to 422.9 ms at posttest) for reading musical notes but not for reading random digit sequences. The perceptual fluency at posttest was comparable with that observed in real-world experts in prior studies (mean expert fluency across studies = 334.3 ms, range = 239.3–465.5 ms; Y. K. Wong & Gauthier, 2010a, 2010b, 2012; Y. K. Wong et al., 2014). Importantly, crowding was reduced with musical stimuli and not with nonmusical stimuli, demonstrating that music-reading training alleviated crowding specifically for musical stimuli. These results suggested that the smaller crowding for musical notes found in the prior study with extant music-reading experts (Y. K. Wong & Gauthier, 2012) was not driven simply by the preselection bias. Instead, visual experience does reduce crowding associated with musical notation, which serves as one of the major perceptual bottlenecks of reading speed.

The current results indicate that crowding limitation can be specific to object category. In the literature, there is consensus that crowding reflects a limitation on object recognition (Levi, 2008; Pelli & Tillman, 2008; Whitney & Levi, 2011). Previous studies have investigated the effect of crowding when the target and flankers were in the same or different object categories (Huckauf, Heller, & Nazir, 1999; Reuther & Chakravarthi, 2014; Yeh, He, & Cavanagh, 2012) or when the target and flankers had the same or different shapes (Kooi, Toet, Tripathy, & Levi, 1994). However, little is known regarding how crowding is modulated by object categories (when the target and flankers are in the same object category) because it is rarely the focus of previous studies (but see Grainger, Tydgat, & Isselé, 2010). Recently, it has been proposed that crowding is independent of object categories (Pelli & Tillman, 2008). In contrast to this proposition, the present study demonstrates that crowding can be reduced specifically for musical stimuli and not for the nonmusical Landolt Cs by lab-induced training. This is consistent with the previous observation of smaller crowding specifically for musical notes in real-world music-reading experts compared with novices (Y. K. Wong & Gauthier, 2012). It is difficult to explain the category specificity of crowding by pure bottom-up accounts of crowding (e.g., general changes in receptive field size or long-range horizontal connections in the visual periphery; Levi & Waugh, 1994) or featural integration (Pelli & Tillman, 2008). Instead, the category specificity of crowding suggests that there is a top-down component of crowding, possibly contributed by higher level object representations, because perceptual expertise similar to that created in the present training typically leads to changes in the representation of the trained objects in the higher visual regions (Gauthier, Skudlarski, Gore, & Anderson, 2010; A. C.-N. Wong, Palmeri, Rogers, Gore, & Gauthier, 2009; A. C.-N. Wong et al., 2012; see also Grainger et al., 2010) and sometimes in both early and higher visual areas (Y. K. Wong, Folstein, &

Gauthier, 2012; Y. K. Wong & Gauthier, 2010a). It is possible that by interacting with higher level representations of musical notation, crowding is alleviated by specifically (for musical notation) reducing inappropriate featural integration (Pelli & Tillman, 2008) or by enhancing the spatial resolution of receptive fields (S. He et al., 1996; Tripathy & Cavanagh, 2002). Future work should clarify the mechanisms underlying the category-specific improvement in crowding, which provides important constraints to the theories of crowding in general.

It is interesting to observe that the music-reading training specifically reduced crowding in the left visual field. Given that the reading direction of music reading is left to right, parafoveal preview of the upcoming notes in the right visual field should help fluent reading as in English text reading (Rayner, 1998). Therefore, one would expect that the right visual field should be more important and hence have larger improvement in music-reading training. Why do we observe reduced crowding in the left instead of the right visual field in the present study? In our prior study (Y. K. Wong & Gauthier, 2012), none of the effects interacted with visual fields. The participants in the present study were intermediate-level readers with a pretest fluency (755.9 ms) somewhere between that of experts (465.5 ms) and novices (1281.0 ms) in our prior study (Y. K. Wong & Gauthier, 2012). One explanation is that the ability to tackle crowding in the right visual field may develop first when a novice learns to read music so as to allow parafoveal preview of the upcoming musical notes on the right. Therefore, intermediates may have already alleviated crowding in the right visual filed to a certain extent, leading to insignificant further reduction after training. In contrast, overcoming crowding in the left visual field may be what it takes to progress from intermediate to expert levels of music-reading fluency. It is possible that the size of the visual span, which limits one's reading speed (Legge et al., 2007; Pelli & Tillman, 2007), is significantly limited by crowding in the left visual field for intermediate music readers. This possibility should be tested in further studies.

The current results complement previous training studies in crowding in several aspects. First, while all prior training studies involved parafoveal and peripheral visual regions in which crowding effects are typically robust (3°–10°; Chung, 2007; Chung & Truong, 2013; Huckauf & Nazir, 2007; Sun et al., 2010), our results demonstrated that crowding can also be alleviated by presenting training stimuli at the central foveal region. Second, the use of music sequences demonstrated that crowding can be reduced by perceptual training with stimulus categories other than letters. Furthermore, results from the current and previous studies lead to an intriguing question concerning the relationship between crowding and reading speed. If crowding serves as one of the major perceptual bottlenecks of reading speed (Legge et al., 2007; Pelli & Tillman, 2007), why did direct and highly effective uncrowding training not lead to improved reading speed (Chung, 2007), whereas training to increase reading speed leads to reduced crowding (as in the current study)?

At least two factors may be involved. First, in previous uncrowding training, participants typically were asked to report the central letter and ignore the flanker letters (e.g., Chung, 2007; Chung & Truong, 2013; Huckauf & Nazir, 2007; Sun et al., 2010). This may have created a perceptual goal that is incompatible with normal reading. For example, normal reading tends to require perception of more than one letter in each word; therefore, learning to ignore flanker letters may be counterproductive in normal reading and hence fail to improve reading speed. Second, recognizing the global shape of the words (Pelli & Tillman, 2007) and spatial configurations within letter or note sequences (A.C.-N. Wong et al., 2011, 2012; A.C.-N. Wong, Wong, Ng, & Ngan, 2016) is important in reading fluency. These can be enhanced in reading training that requires recognition of the whole sequences (as in the current study) but not in uncrowding training (Chung & Truong, 2013). It may thus be unsurprising to observe that uncrowding training led to increased visual span (Chung & Truong, 2013) without improving one's reading speed (Chung, 2007). In contrast, training studies that used more readinglike tasks (e.g., recognizing all letters in letter trigrams; Bernard, Arunkumar, & Chung, 2012; Chung, Legge, & Cheung, 2004; Y. He, Legge, & Yu, 2013) or recognizing all notes in music sequences as in the current study led to enhanced reading speed (Bernard et al., 2012; Chung et al., 2004; Y. He et al., 2013) and reduced crowding (Y. He et al., 2013).

To our best knowledge, this study demonstrates the first music-reading training protocol that uses short-term laboratory training to improve musical note discrimination fluency to a level comparable with extant expert musicians. This method offers a way to study the causal relationship between music-reading training and improvement in constituent visual skills related to musical note recognition (e.g., holistic processing; Y. K. Wong & Gauthier, 2010b) and the causes of the brain activity changes observed in extant music-reading experts (Y. K. Wong & Gauthier, 2010a; Y. K. Wong et al., 2014). Practically, this method may help students improve music-reading fluency, which will in turn allow students to explore and enjoy a wider range of music pieces and to spend more effort on developing other music skills, such as motor planning, emotional expression, sight reading, or the integration of these skills.

Acknowledgments

Corresponding authors: Yetta Kwailing Wong; Alan C.-N. Wong.
Email: yetta.wong@gmail.com; alanwong@cuhk.edu.hk.
Address: Faculty of Education, University of Hong Kong, Pokfulam, Hong Kong; Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong.

References

Bernard, J. B., Arunkumar, A., & Chung, S. T. L. (2012). Can reading-specific training stimuli improve the effect of perceptual learning on peripheral reading speed? *Vision Research, 66,* 17–25.

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

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

Chung, S. T., Legge, G. E., & Cheung, S. H. (2004). Letter-recognition and reading speed in peripheral vision benefit from perceptual learning. *Vision Research, 44,* 695–709.

Chung, S. T. L., & Truong, S. R. (2013). Learning to identify crowded letters: Does the learning depend on the frequency of training? *Vision Research, 77,* 41–50.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2010). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197.

Grainger, J., Tydgat, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performances, 36,* 673–688.

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

Harrison, W. J., Retell, J. D., Remington, R. W., & Mattingley, J. B. (2013). Visual crowding at a distance during predictive remapping. *Current Biology, 23,* 793–798.

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

He, Y., Legge, G. E., & Yu, D. (2013). Sensory and cognitive influences on the training-related improvement of reading speed in peripheral vision. *Journal of Vision, 13*(7):14, 1–14, doi:10.1167/13.7.14. [PubMed] [Article]

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences, 1,* 86–93.

Huckauf, A., Heller, D., & Nazir, T. A. (1999). Lateral masking: Limitations of the feature interaction account. *Perception & Psychophysics, 61,* 177–189, doi:10.3758/BF03211958.

Huckauf, A., & Nazir, T. (2007). How odgcrnwi becomes crowding: Stimulus-specific learning reduces crowding. *Journal of Vision, 7*(2):18, 1–12, doi:10.1167/7.2.18. [PubMed] [Article]

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8,* 255–279, doi:10.1163/156856894X00350.

Legge, G. E., Cheung, S.-H., Yu, D., Chung, S. T. L., Lee, H.-W., & Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *Journal of Vision, 7*(2):9, 1–15, doi:10.1167/7.2.9. [PubMed] [Article]

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

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

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision, 7*(2):24, 1–11, doi:10.1167/7.2.24. [PubMed] [Article]

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of

crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744.

Pelli, D. G. (1997). The Videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation. *PLoS One, 2*(1), e680.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11,* 1129–1135.

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

Reuther, J., & Chakravarthi, R. (2014). Categorical membership modulates crowding: Evidence from characters. *Journal of Vision, 14*(6):5, 1–13, doi:10.1167/14.6.5. [PubMed] [Article]

Sayim, B., & Cavanagh, P. (2013). Grouping and crowding affect target appearance over different spatial scales. *PLoS One, 8,* e71188.

Sun, G. J., Chung, S. T. L., & Tjan, B. S. (2010). Ideal observer analysis of crowding and the reduction of crowding through learning. *Journal of Vision, 10*(5):16, 1–14, doi:10.1167/10.5.16. [PubMed] [Article]

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

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15,* 160–168.

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):15, 1–9, doi:10.1167/9.11.15. [PubMed] [Article]

Wong, A. C.-N., Bukach, C. M., Hsiao, J., Greenspon, E., Ahern, E., Duan, Y., & Lui, K. F. H. (2012). Holistic processing as a hallmark of perceptual expertise for nonface categories including Chinese characters. *Journal of Vision, 12*(13):7, 1–15, doi:10.1167/12.13.7. [PubMed] [Article]

Wong, A. C.-N., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *PLoS One, 6*(6), e20753.

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

Wong, A. C.-N., Wong, Y. K., Ng, T. Y. K., & Ngan, V. S. H. (2016). *Individual differences in sensitivity to configural information predicts word recognition fluency.* Manuscript submitted for publication.

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology: General, 14,* 682–698.

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

Wong, Y. K., & Gauthier, I. (2010b). Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices. *Cognitive, Affective & Behavioral Neuroscience, 10,* 541–551.

Wong, Y. K., & Gauthier, I. (2012). Music-reading expertise alters visual spatial resolution for musical notation. *Psychonomic Bulletin & Review, 19,* 594–600.

Wong, Y. K., Peng, C., Fratus, K. N., Woodman, G. F., & Gauthier, I. (2014). Perceptual expertise and top-down expectation of musical notation engages the primary visual cortex. *Journal of Cognitive Neuroscience, 26,* 1629–1643.

Yeh, S.-L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological Science, 23,* 608–616, doi:10.1177/0956797611434746.