Improving Letter Recognition and Reading in Peripheral Vision: Sensory and Cognitive Constraints

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Yingchen He

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Gordon E Legge
Co-advisor: Sheng He

May 2017

Acknowledgements

If I can only acknowledge one person, that would have to be my adviser, Dr. Gordon E. Legge. He made me truly, deeply understand the word “academic *father*”. Supportive, considerate, patient, responsible, diligent… I am just so fortunate to have a mentor like this. He has helped me grown a lot, and I owe my deepest gratitude to him.

I thank Dr. MiYoung Kwon for her collaborative work on the project in Chapter 2. I also thank Dr. Susana T. L. Chung for sharing her data and providing comments. Thank you to Drs. Sheng He, Steve Engel, and Matthew Johnson, for serving as my committee members, especially Dr. Sheng He for his advice regarding the topic of crowding (Chapters 2 & 3).

Thank you to all lab members, who have accompanied me in my journey as a graduate student. Dr. Aurélie Calabrèse, Muzi Chen, Dr. Hui Wang, and Dr. Quan Lei, it is my great pleasure to know you and be connected with you both in research and in our leisure times. I am also grateful to the help from Rachel Gage, our all-powerful lab manager, for so many things. Also I am thankful to Dr. Tiana Bochsler, Ryan Persons, Justin Page, Mitchell Mastey, Hannah McMunn, Sori Baek, Mark Schatza, Seoyoung Park, who have provided help for different projects in various ways. And I thank the UMN survey service for building the web version of the survey used in Chapter 4. And I would like to thank all of my previous subjects, without whom this dissertation will not be possible.

I also thank my parents for their unconditional love and support. Even when my work is far beyond their scope of understanding, they keep trying to help in their own way.

And I thank my fiancée, Sha Li, whose love and company have made me stronger and unafraid of the unknown.

This thesis research was supported by the U. S. National Institutes of Health grant EY002934 to Prof. Gordon E. Legge and the Doctoral Dissertation Fellowship (University of Minnesota, Twin Cities). Trips to present the results were also supported by David Campbell Fellowship (Department of Psychology, University of Minnesota, Twin Cities) and Graduate Student Research and Travel Award (Department of Psychology, University of Minnesota, Twin Cities).

To my lovely,
keyboard-obsessed cat friend,
Biscuit (伯虎),

who is responsible for all the typurrrrrrrrrros in this dissertation.

Abstract

Reading is an important daily task, but it is very difficult for people who have lost their central vision, because they must use peripheral vision to read. One hypothesis for slow reading speed in peripheral vision is the shrinkage of the visual span, which is the number of identifiable letters within a glimpse. Previous studies have shown that perceptual training tasks of letter recognition can enlarge peripheral visual span, as well as improving peripheral reading speed by 40% or more. This thesis focuses on sensory and cognitive factors that facilitate or limit the training-related improvements, with an ultimate goal of developing rehabilitation protocols for people with central-field loss. Chapter 1 gives an overview of the thesis. Chapter 2 demonstrates that there are common constraints limiting the size of the visual span across languages (Korean and English), and that extensive training of reading Korean characters using peripheral vision enlarges Korean visual span as well as English visual span. This transfer of training suggests a pre-symbolic nature of the visual span, and a strong potential for training benefits to generalize to untrained scripts. Chapter 3 discusses visual crowding, the inability to recognize objects in clutter, which is proposed to be the major sensory factor limiting the size of the visual span and reading. The results lead to the conclusion that reducing the impact of crowding can enlarge the visual span and can potentially facilitate reading, but not when adverse attentional bias is introduced, for example directing attention to one specific, small area in the visual field. By dissociating the influence of sensory and attentional factors, the link between crowding, visual span and reading was clarified.

Finally, Chapter 4 reports on a study where the training was implemented in a word-guessing video game. The game training successfully enlarged the visual span and improved reading speed. Embedding the training in a game enhanced the enjoyment of the training and can temporarily boost letter-recognition performance during the game, but the quality of the training was not altered compared with similar training without the game. Together, the studies presented in this thesis not only speak to the theoretical basis for the training-related changes, but also provide practical guidance for designing potential reading rehabilitation protocols for people with central-field-loss.

Table of Contents

List of Tables

List of Figures

Chapter 1. Overview

Age-related macular degeneration (AMD) is a prevalent condition among elderly people and affects approximately 6.5% of the US population aged 40 years and older (Klein et al., 2011). Macular degeneration is caused by the death of photoreceptors in the macular region of the retina. Two major forms of AMD are the dry form and the wet form. In dry AMD, also termed geographic atrophy, the atrophy of the retinal pigment epithelium cells underlying the retina causes the loss of corresponding photoreceptors, resulting in vision loss. The damage is patchy and does not always influence the fovea, but if it does, the patient will be left with central-field loss. Wet AMD, sometimes called neovascular AMD, is caused by abnormal growth of blood vessels under the retina. These vessels are fragile and may rupture, bleed, and scar, damaging the photoreceptors.

In more advanced cases of AMD, the central visual field is impaired, often bilaterally. Central visual field is the region with the highest visual acuity and is used to fixate and to distinguish fine details. People with central-field loss have to use their peripheral vision for high-acuity daily activities. Reading, being one of the most important daily activities, is highly challenging in peripheral vision. This makes reading extremely hard for some AMD patients, discouraging them from reading on a regular basis. The impairment not only poses difficulties in a person’s daily life, but also impacts the person’s psychological well-being (Mitchell & Bradley, 2006). These consequences motivate researchers to develop rehabilitation methods for people with AMD and provoke the research in the current thesis.

There has been continuous effort to tackle the reading difficulties in people with AMD. Two big obstacles in peripheral reading are the insufficient acuity to distinguish fine details and the poor quality of saccades along the text. To compensate for low acuity, various types of reading aids can be supplied, ranging from simple optical magnifiers to powerful video magnifiers (for a review, see Virgili, Acosta, Grover, Bentley, & Giacomelli, 2013). To improve oculomotor function with a non-foveal reference point, training protocols on eye-movement control have been developed (e.g. Seiple, Szlyk, McMahon, Pulido, & Fishman, 2005). Moreover, to minimize the occlusion of text caused by the scotoma, researchers have tried to optimize the position of the reference point of saccades (Nilsson, Frennesson, & Nilsson, 2003; Vingolo, Salvatore, Domanico, Spadea, & Nebbioso, 2013; Vingolo, Salvatore, & Limoli, 2013) or to modify text presentation (such as 90 degrees rotated text) (Calabrèse, Liu, & Legge, 2017).

In addition to reduced acuity and inaccurate oculomotor control, peripheral vision may have other properties which limit reading speed. With adequate magnification and after eliminating the need for eye movements using rapid serial visual presentation (RSVP) to measure reading speed (Forster, 1970; Rubin & Turano, 1992), the maximum reading speed in peripheral vision is still not comparable to that in central vision: From 0° (fovea) to 20° in the periphery, maximum RSVP reading speed decreased from 807 words-per-minute (wpm) to 135 wpm for normally-sighted subjects (Chung, Mansfield, & Legge, 1998). One proposed factor influencing reading speed is the visual span, which is the number of letters that can be recognized accurately without eye movements (Legge et al., 2007). Figure 1-1 demonstrates the measurement of visual span using a letter-recognition task and the resulting visual span profile. According to the visual span hypothesis, the size of the visual span poses a sensory bottleneck on reading speed. From central to peripheral vision, the decrease in reading speed is correlated with shrinkage in the size of the visual span (Legge et al., 2007).

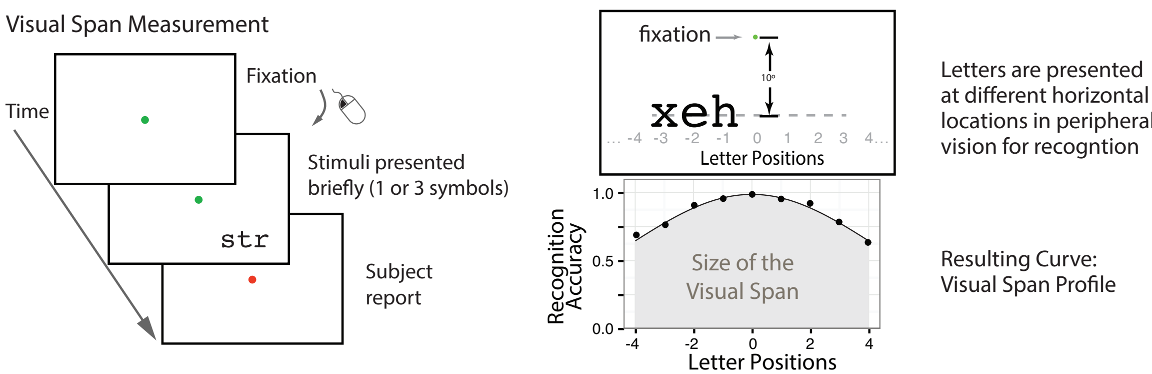

Figure 1-1. Visual span measurement and the visual span profile.

The association between reading speed and the size of the visual span has raised the possibility of improving peripheral reading speed via enlarging the visual span by perceptual learning. Perceptual learning is a process where long-lasting improvements of certain perceptual skills happen after extensive practice on those tasks. For the human visual system, perceptual learning can improve various capabilities such as visual acuity, contrast sensitivity, pattern recognition, motion detection, visual working memory, visual attention, and others (for reviews, see Fine & Jacobs, 2002; Sagi, 2011). Perceptual training tasks, usually reading letters or words displayed in peripheral vision, can increase the size of the peripheral visual span for normally-sighted young adults, accompanied by an improvement of at least 40% in reading speed(Chung, Legge, & Cheung, 2004; He, Legge, & Yu, 2013; Lee, Kwon, Legge, & Gefroh, 2010; Yu, Legge, Park, Gage, & Chung, 2010). Elder adults within the age range of the onset of AMD can also benefit from this type of training, achieving a 60% improvement in reading speed(Yu, Cheung, Legge, & Chung, 2010). Recently, similar training has also been applied to people with central-field loss (Calabrèse et al., 2017; Chung, 2011; Nguyen, Stockum, Hahn, & Trauzettel-Klosinski, 2011) For instance, Chung (2011) trained 6 subjects with an RSVP reading task. Six weekly training sessions yielded a mean improvement of 53% in RSVP reading speed. These successful examples demonstrate the potential clinical value of perceptual training.

This thesis focuses on improving peripheral vision to read, with an ultimate goal of developing rehabilitation protocols for people who have lost their central vision. Specifically, the major interests are the nature of training-related enlargement of the visual span, and factors that facilitate or limit the transfer of the training effect to improved reading performance. Note that in the current thesis all the reported experiments were conducted with normally-sighted subjects. This is to ensure the effectiveness of the training protocols before applying the methods to people with AMD. Chapters 2, 3 and 4 are structured as three research papers, dealing with different aspects of the changes associated with letter-recognition training. The remainder of Chapter 1 summarizes the major findings in these three papers.

Chapter 2. Pre-symbolic Nature of the Visual Span.

Is the training-related enlargement of visual span specific to the symbols used in training? If so, the training will not generalize and will only have a limited benefit. Chapter 2 presents a study examining whether there are common constraints limiting Korean and English recognition, and whether the training benefits will transfer from Korean to English characters. Nine Korean-English bilingual subjects underwent 4 daily training sessions of reading Korean trigrams (three adjacent characters). Training was conducted in the lower visual field. Before and after the training, we measured visual spans for Korean symbols—single Korean components (vowels and consonants), single Korean characters (composed of 2 or 3 components), and Korean trigrams—and also English symbols—single letters and trigrams. The 5 types of testing stimuli are illustrated in Figure 1-2.

Figure 1-2. Five types of testing stimuli used in Chapter 2.

The results showed that the size of both Korean and English visual spans were limited by a physical property of the script, namely pattern complexity, rather than their linguistic property: the size of the visual spans ranked as English letters = Korean components > Korean single characters = English trigrams > Korean trigrams, and was negatively correlated with averaged pattern complexity of the symbol sets. Training enlarged the visual spans for Korean single characters and trigrams, and the benefit also transferred to English trigrams. This successful transfer is likely due to improved recognition of shared features underlying recognition in both scripts. Together, these results provide evidence that the visual span reflects a universal sensory limit for text recognition across languages.

Chapter 3. Linking Crowding, Visual Span, and Reading.

Visual crowding, the inability to recognize objects in clutter, has been found to be the major sensory factor limiting the size of the visual span (He et al., 2013; Pelli et al., 2007). It thus appears that crowding limits reading speed by limiting the size of the visual span. However, this proposed linkage seems to be inconsistent with the finding that after training to recognize tightly-arranged letters in peripheral vision, crowding was reduced but peripheral reading speed was not improved (Chung, 2007). Chapter 3 deals with this discrepancy by disentangling the impact of crowding and spatial attention. 27 normally-sighted college students were trained for 6 days, 1 hours/day, to read letters using peripheral vision. They were randomly assigned to three groups where the training stimuli varied in two key ways: the presence of flankers (whether the target letter for recognition was isolated or crowded) and the spatial distribution of training stimuli (whether the target letters were distributed across different horizontal locations, or always localized at one fixed location). Three training groups were designed (Fig. 1-3): the

Flanked-Local, Flanked-Distributed, and Isolated-Distributed Groups. Before and after the training, the spatial extent of crowding, the size of the visual span, and RSVP reading speed were measured.

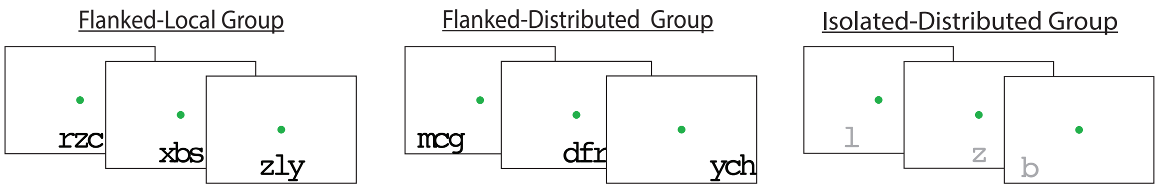

Figure 1-3. Training stimuli for the three training groups in Chapter 3.

The major finding was that spatially distributed training stimuli, as in the two Distributed Groups, appeared to be necessary and sufficient for the training benefit to transfer from letter recognition to increased reading speed. Spatially localized training as in the Flanked-Local Group failed to improve reading, possibly because training facilitated the deployment of attention to one specific, small area in the visual field, which was disadvantageous for reading. We also found, surprisingly, that training without the presence of flankers (the Isolated-Distributed Group) was effective in enlarging the visual span and improving reading speed. This is contrary to the intuition that training with crowded stimuli should be necessary to reduce the impact of crowding. Overall, Chapter 3 suggests that while the visual span represents a sensory bottleneck on reading, there may also be an attentional bottleneck. Reducing the impact of crowding can enlarge the visual span and can potentially facilitate reading, but not when adverse attentional bias is introduced.

Chapter 4. Training With A Game: More Fun But Not More Improvement

Despite the promising improvement related to perceptual training, the training procedure suffers from being laborious and tedious. The nature of perceptual learning is to repeat the same task intensively and for a long time, which is likely to discourage patients from continuing training. Video-game training may help to address this problem. For normally sighted subjects, research has shown that video-game playing can not only elevate subjects' interest in the task, but also yield improvements on various visual skills, from low-level contrast sensitivity to high-level visual attention (for a review, see Bavelier & Green, 2012). Chapter 4 reports on a form of game training we developed, where the trigram training was incorporated into a word game similar to the popular TV show *Wheel of Fortune* (for a .gif demonstration, please go to https://yingchenhe.files.wordpress.com/2017/02/game-movie.gif?w=825). In the game, after completing each trigram-recognition trial, correct letter responses will yield clues to words-to-be-guessed. The subjects were thus motivated to recognize as many letters as possible in order to collect clues for successful guessing.

The benefits of the game training were evaluated in comparison with a no-game trigram training (6 subjects, data from He et al., 2013), including subjective engagement, real-time performance boost, and sustainable learning effects. Six young, normally sighted college students played the game for 4 days, 1.5 hours/day, matching the training time of the no-game training group. Since adding the game reduced the number of training trials to half of that in the no-game group, another six subjects were trained for a total of 12 hours to match the total number of trials to the no-game training. The size of the visual span and RSVP reading speed were measured in the pre- and post-tests.

We found that embedding the trigram training in a word-puzzle game enlarged the visual span, and the improvement largely transferred to an untrained visual field and to improved reading. A survey administered to subjects revealed that the game was more enjoyable compared to the no-game tasks, mostly because of its timely feedback and its competition aspect. We also built a model to separate the effects of training and gaming on performance. Our model showed that incorporating the game in the training protocol boosted real-time letter-recognition performance while playing the game, but sustainable improvement due to training was not enhanced. We conclude that the video-game component of our training protocol enhanced subjective experience but did not enhance the objective improvement in reading speed.

Summary

The studies presented in this thesis demonstrated that the visual span reflects a universal sensory limit for text recognition and reading across languages. But the visual span may not be the only bottleneck for reading; there may also be an attentional bottleneck where narrowly focused spatial attention is disadvantageous for reading. Adding a game component to the training protocol makes the training more enjoyable, although it does not enhance the sustainable training benefit. These results provide insights into the design of a potential reading rehabilitation protocol for people with AMD: it needs to consider both sensory and attentional bottlenecks, and may even include a gaming component to improve user experience. If successful, the training benefits will likely generalize to other untrained scripts, which adds to its clinical value.

Chapter 2. Pre-symbolic Nature of the Visual Span -- Comparing Korean and English Recognition

Introduction

Human pattern recognition is limited by available visual information within a glimpse. The visual span, referring to the number of identifiable letters within one fixation, reflects such sensory limits on reading (Legge et al., 2007). Recently, the concept of visual span has been generalized to faces (He et al., 2015), showing that similar sensory constraints limit both face and letter recognition. This indicates that the visual span reflects presymbolic limits of human vision on pattern recognition.

In peripheral vision, the size of the visual span for English characters can be enlarged through letter recognition training (e.g. Chung, Legge, & Cheung, 2004). The major component of this enlargement is a reduction of between-letter crowding, followed by a minor contribution from reduced errors in encoding the position of the letters (He et al., 2013). The training-related enlargement of the visual span can transfer to an untrained visual-field location (Chung et al., 2004; He et al., 2013; Yu, Legge, et al., 2010) or to an untrained print size (Yu, Legge, et al., 2010). This transfer suggests that the effect is not retinotopically specific and can tolerate some variance in the size of the stimuli.

But is the training-related improvement symbol-specific, or is it transferrable to an untrained set of symbols? One possible symbol-specific mechanism of the improvement is to learn better templates for identification. The templates for identification can be estimated using a classification image technique. Gold and colleagues compared the templates used by an ideal observer and human subjects in a pattern discrimination task (Gold, Sekuler, & Bennett, 2004). While the template for an ideal observer was simply the difference between the two alternatives, human subjects used a sub-optimal template that was only partially correlated with the ideal template. They further confirmed that perceptual training improved the template used by human subjects to be more similar to the ideal's.

If the subjects learned more precise templates of the trained samples, there would be limited transfer of training benefits to an untrained set of symbols. Alternatively, if the underlying mechanism of improvement is not symbol-specific, such as reduced spatial extent over which crowding can happen, training benefits will transfer. Here, to test whether the mechanism of training-related improvement is symbol-specific or not, we will examine whether training to recognize Korean characters transfers to English letter recognition. If transfer occurs, we will conclude that the learning mechanism is not specific to the set of symbols being tested and therefore not primarily due to sharpening of the character templates.

We chose to study transfer from Korean to English mainly for two reasons. First, despite the differences in the symbols, Korean writing and English writing are both alphabetic, making them similar and comparable to each other. Korean language is written in the Korean alphabet (consonant and vowel letters) known as Hangul. The basic letter set contains 24 distinct symbols--14 simple consonants and 10 simple vowels, close to the set size of lowercase English letters (26).

Unlike English writing where letters are arranged horizontally, Korean letters do not stand alone. Instead, they are assembled into blocks of syllables (characters), where each character contains 2-3 letters, or sometimes even more (see Methods for a detailed description). These characters are then arranged horizontally to create a word, similar to how letters are arranged in English words. Given the special arrangement of Korean characters, it is likely that both within-character and between-character crowding are present in Korean reading. This is our second reason to choose Korean as our subject of study, because it allows us to study the nature of the training: If the Korean visual span enlarges after training, is it due to a reduction of within-character or between-character crowding, or both?

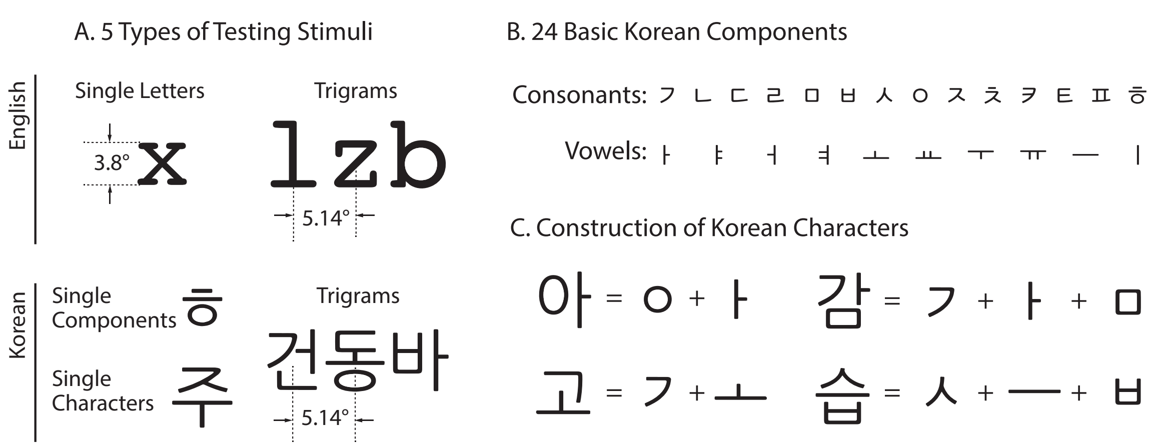

Figure 2-1. Testing stimuli used in Chapter 2.

A. Five types of testing stimuli, including single English letters, English trigrams, single Korean components, single Korean characters, and Korean trigrams. **B.** 24 basic Korean components, including 14 consonants and 10 vowels. **C.** Construction of Korean characters. The top row shows two-component (left) and three-component (right) characters whose vowels contain a major vertical bar. The bottom row shows two-component (left) and three-component (right) characters whose vowels contain a major horizontal bar.

Given the higher complexity of Korean characters compared to English letters, and the negative correlation between perimetric complexity and visual span size tested with Chinese and English characters (Wang, He, & Legge, 2014), we predict a smaller visual span of Korean single characters compared to English letters, and a smaller visual span of Korean trigrams compared to English trigrams. If the training improvement is not symbol-specific, we will observe that our training with Korean characters successfully transfers to English letters.

To summarize, our study has two purposes: 1) to compare Korean and English visual spans, and 2) to examine whether training to read Korean characters enlarges the visual span and whether any such enlargement transfers to the size of the English visual span.

Methods

Participants

4 male and 5 female native Korean speakers were recruited from the University of Minnesota (mean age 21.8, ranging from 19 to 24 years old). All subjects were fluent in English and were able to type in both English and Korean dexterously. Participants all had normal or corrected-to-normal vision, with binocular acuity of -0.04 ± 0.01 logMAR (mean ± SEM, measured by Lighthouse Near Acuity Chart, Lighthouse Low Vision Products, Long Island City, NY). The protocol was approved by the Institutional Review

Board and was in compliance with the Declaration of Helsinki. All subjects gave informed consent prior to the experiment.

Stimuli and Apparatus

The stimuli consisted of black text (English or Korean) on a white background (background luminance 102 cd/m^2; Weber contrast = 98%). We used a NEC MultiSync CRT monitor (model FP2141SB-BK, NEC, Tokyo, Japan; refresh rate = 100 Hz; spatial resolution = 0.04°/pixel) controlled by a Mac Pro Quad-Core computer (model A1186, Apple Inc., Cupertino, CA). Five types of symbols were used in the experiment (Fig. 2-1A): single English letters, English trigrams (strings of three characters), single Korean components (consonants and vowels), single Korean characters, and Korean trigrams. English letters were rendered in Courier, and Korean characters were rendered in Nanum Gothic. The stimuli were generated and presented using MATLAB R2014b with Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). All stimuli were viewed binocularly from 30 cm in a dark room. Viewing distance was maintained using a chin rest and subject's fixation was monitored using a webcam.

Size. All English letters were lowercase and had an x-height of 3.8°. For trigrams, the center-to-center spacing between letters was 1.16 × x-width (standard spacing for Courier, approximately 5.14°). The size of the Korean characters was scaled so that the center-to-center spacing between adjacent characters was also 5.14° (Fig. 2-1A). Both English letter size and Korean character size exceeded their corresponding critical print size for reading at 10 degrees in the lower visual field (Baek, He, & Legge, 2016; Chung et al., 1998), so that the print size was not a limiting factor of reading speed.

Korean character *set.* In a related project we used 900 Korean sentences to compare Korean reading speed in central and peripheral vision (Baek et al., 2016). Here, we used the most frequent 279 characters from those sentences (accounting for 90% occurrences) as our testing stimuli. A full list of the characters is supplied in Appendix 1. We acknowledge that the set size (279) was large when compared to English (only 26 letters), but since reducing the set size would compromise the resemblance of the task to real Korean reading, we kept the set of 279 characters. To minimize the influence of set size, all the characters were printed on a hard copy paper, and we encouraged the subjects to review the list before each testing block. Subjects were not required to memorize the list, but if during the measurement they reported any out-of-set characters, a warning was triggered. The trial was cancelled and we ran replacement trials until the response was within the set. Out of the total trials, 11% contained out-of-set responses.

Korean components *and characters.* There are at least two components in a character: an initial consonant letter and a vowel letter. For a character with two components, they are arranged left-right if the vowel has a major vertical bar, or arranged top-bottom if the vowel has a major horizontal bar. For a character with more than two components, a final consonant letter is also present and will always be placed below the other two letters, no matter how the lead consonant and vowel letters are arranged. Figure

2-1C provides examples for these variations. In our set of 279 characters, 114 of them are two-component characters (40.9%) and the rest are three-component characters.

Generation of testing stimuli. We generated images of English letters, Korean components, and Korean characters before the experiment. English letters and Korean characters were directly "typed" onto the backdrop in Matlab and cut using the default bounding-box for the specified size (282 pixel (W) by 345 pixel (H)). For Korean characters, we first generated images of Korean characters and then cut component images from them (see Appendix 1 for a detailed description). We carefully controlled the shape and size of the components so that they are representative of what appear in common characters.

Experimental Design

Our experiment consisted of 3 parts (Fig. 2-2A): pre-test (day 1), training (day 2-5), and post-test (day 6). Each daily session lasted for approximately 1.5-2 hours.

In the pre- and post-tests, subjects' visual span profiles for both Korean and English were measured (Fig. 2-2A; see later for details). English visual-span profiles were measured in the order of single letters and trigrams. Korean visual-span profiles were measured in the order of single components, single characters, and trigrams. Whether to first measure Korean or English visual span was counterbalanced between subjects in the pre-test, and reversed in the post-test for each subject (for example, Korean-English in the pre-test and English-Korean in the post-test). From day 2 to day 5, subjects underwent training sessions where each day they performed 16 blocks of visual span measurements using Korean trigrams.

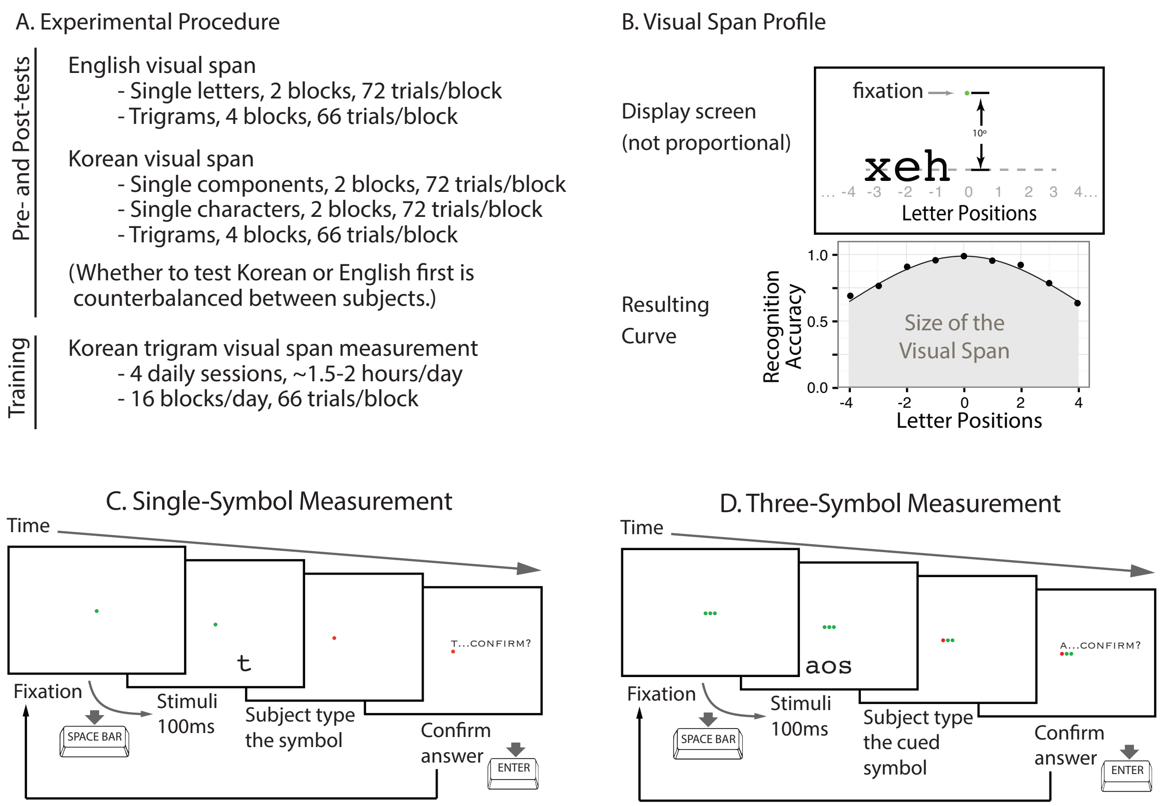

Figure 2-2. Experimental procedure and visual span measurement.

A. Experimental procedure. **B.** Visual span profiles. **C.** Diagram of single-symbol measurement. **D.** Diagram of three-symbol measurement.

Visual Span Measurement

In order to measure the size of the visual span, we performed a letter-recognition task, similar to the task in He et al. (2013). Letters were presented in pre-defined slots, which were horizontally arranged on an imaginary line at 10° in the lower visual field

(Fig. 2-2B). The slot on the fixation midline was labeled 0, and left and right slots were labeled with negative and positive numbers respectively. The center-to-center spacing between adjacent slots was approximately 5.14° (1.16 × x-width, corresponding to standard spacing in the Courier font).

As described previously, there were 5 different types of testing stimuli. In a single-symbol trial (English letters, Korean components, or Korean characters; Fig. 2-2C), the subject first fixated on a dot, then pressed the space bar to initiate a trial. The symbol would appear for 100ms, and the subject was asked to type that symbol using a keyboard. To reduce typing errors, the typed response was shown on the screen, and the subject was asked to confirm the response by hitting the enter key. The visual feedback for typing was rendered in a font different from the testing stimuli (and also rendered in uppercase for English) to minimize the potential benefit on performance caused by seeing the symbol. Within a block, stimuli appeared 8 times on each slot from -4 to 4 in a random order, making a total of 72 trials. In the pre- and post-tests, single English letters, single Korean components, and single Korean characters were each measured in 2 blocks respectively.

For English and Korean trigrams (Fig. 2-2D), we used a partial report method in order to minimize the memory load. In the beginning of a trial, there were 3 horizontally-arranged green dots in the center of the screen. The subject needed to fixate on the middle dot, and press the space bar to initiate a trial. A trigram would appear for 100ms, and 1 of the 3 dots would change to red, indicating which symbol to report. For example, if the left dot changed to red, the subject needed to report the left symbol of the three. The response was recorded in the same way as for single-symbol trials. Within a block, stimuli were centered 6 times on each slot from -5 to 5 in a random order, with the left, middle and right letter each being cued 2 times. Since slots ±5 and ±6 had fewer letters than the other slots, only data from slots -4 to 4 were used in further analysis. In the pre- and post-tests, English and Korean trigrams were each measured in 4 blocks, respectively. To get the visual span profiles, letter-recognition accuracy was plotted against letter positions (Fig. 2-2B). We will report the average accuracy across the 9 slots as a summary of the size of the visual span.

Two-Stage Model for Pattern Recognition

In a previous study comparing letter and face recognition, we modeled pattern recognition as a serial, independent two-stage process to quantify the influence of crowding (He et al., 2015). Briefly, this model assumes that at the first stage, recognition is limited by factors affecting the processing of isolated symbols. The second stage represents the additional interfering effects of nearby symbols on recognition. Each stage is characterized by its reliability, that is, the probability that the correct information is transmitted through this stage. We have built 3 separate models where the first stage corresponds to recognizing single English letters, Korean components, and Korean characters. To distinguish between them, we will use subscripts E, K_component and K_character respectively when describing the corresponding reliability.

According to our model, the reliability of the first stage (R_1) is equal to the probability (corrected for guessing) of correctly identifying isolated symbols. R_1 could be influenced by factors such as visual acuity, contrast sensitivity, and others. For simplicity, we will use "acuity effect" to describe any factor that may influence the recognition of isolated symbols.

The probability of recognizing the same symbol in a crowd (also corrected for guessing) is the product of the reliabilities of the two stages ($R_1 \times R_2$). For example, if the recognition accuracy is 90% for an isolated English letter and 70% for a letter in a trigram, then

$$R_{1,E} = 0.9, \text{ and}$$

$$R_{1,E} \times R_{2,E} = 0.7 .$$

From here, $R_{2,E}$ can be computed as 0.7 / 0.9 = 0.78.

Korean characters are different from English letters, because they are each made up of individually identifiable components. We therefore built two separate models for Korean characters. One model treated the recognition of individual Korean components as the first stage, and represented the influence of other components within the same character in the second stage. If the recognition accuracy is 90% for an isolated component and 80% for a component in an isolated character, then $R_{1,K\_component} = 0.9$ and $R_{2,K\_component} = 0.8 / 0.9 = 0.89$. Note that during single Korean character and Korean trigram measurements, we only ask the subjects to report the identity of the entire character, not the components. However, once we know the identity of the character, for example 감, we can infer that the subject thinks the three components are ㄱ, ㅏ and ㅁ and then score them based on component accuracy. Figure 2-1C gives some examples of this decomposition.

The second model treated the recognition of individual characters as the first stage, and the influence of nearby characters was reflected in the second stage. According to this model, if the recognition accuracy is 70% for an isolated character and 50% for a character in a trigram, then $R_{1,\ K\_character} = 0.7$ and $R_{2,\ K\_character} = 0.7 / 0.9 = 0.78$.

Thus, the values of R_2 can be interpreted as the letter recognition accuracy when stage 1 processing is not a limiting factor, i.e. when the reliability of stage 1 is 100% ($R_1 = 1.0$). In this way we can remove the influence of the acuity effect from the visual span and estimate the influence of within- and between-symbol crowding. The closer to 1 R_2 is, the less crowding there is.

Statistical Analysis

When comparing training effects between types of symbols, we performed a 5×2 repeated-measures ANOVA on the average accuracy of the visual span profiles, with two within-subject factors being symbol type (English letters / English trigrams / Korean components / Korean characters / Korean trigrams) and session type (pre-test / post-test). When evaluating the changes in crowding, we performed a 3×2 repeated-measures ANOVA on the R_2 values, including $R_{2,\ E}$, $R_{2,\ K\_component}$, and $R_{2,\ K\_character}$. Two within-subject factors were symbol type (English letters / Korean components / Korean characters) and session type (pre-test / post-test). If a significant interaction was found, we further analyzed the interaction using *R* with the package *phia* (Post-Hoc Interaction Analysis; Martínez, 2013). The reported p-values were adjusted for multiple comparisons.

Results

In the following sections, we will first compare the visual span profiles for the five types of symbols, and then examine how training changed the size of these visual spans. To preview, we found that the size of visual spans, both before and after training, ranked as: English letters = Korean components > Korean characters = English trigrams > Korean trigrams. After training, the visual spans for Korean characters and Korean trigrams both enlarged, and the enlargement transferred to English trigrams. Using our two-stage model, we found that training reduced the within-character and between-character crowding in Korean recognition, as well as between-letter crowding in English letter recognition.

Comparing Korean and English Visual Spans

We first compared the visual span profiles in the pre- and post-tests for Korean and English symbols. From left to right, Figure 2-3 shows results for English letters, Korean components, Korean characters, English trigrams and Korean trigrams. In both pre- and post-tests, the visual span profiles for English letters and Korean components were very similar and close to perfect recognition, but the profile for Korean characters was narrower and had a lower peak, suggesting a smaller size. The visual span for

English trigrams was larger than that for Korean trigrams, but they both had decreasing recognition accuracy for letter positions farther away from the midline.

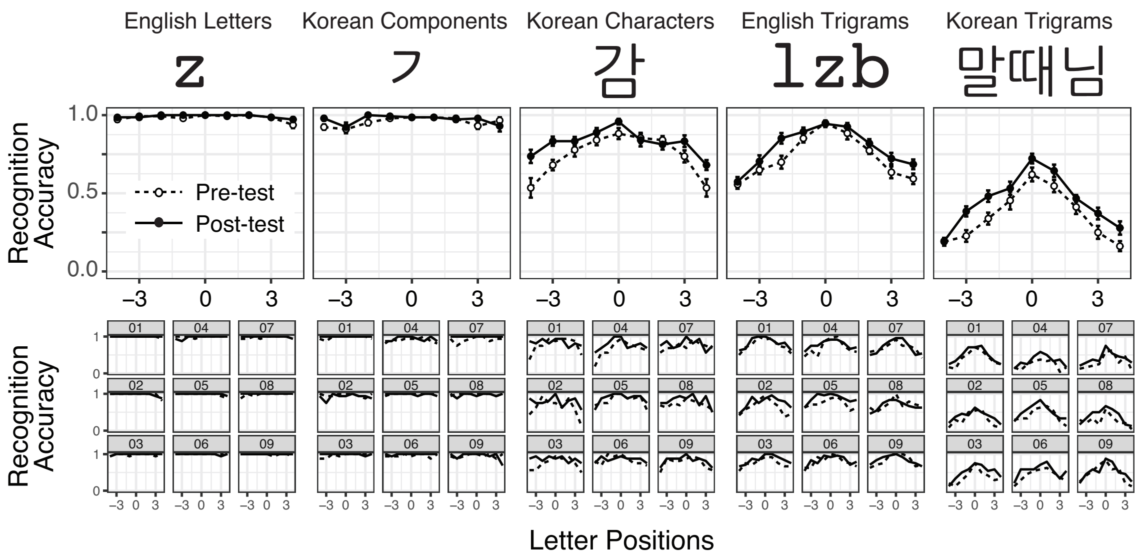

Figure 2-3. Korean and English visual span profiles.

Dashed lines and open symbols: pre-test. Solid lines and filled symbols: post-test. Error bars: ±1 SEM. Small panels on the bottom are individual curves.

We then asked how the differences in the size of the visual span related to the pattern complexity of the different sets of symbols. We used perimetric complexity (Attneave, Arnoult, & Attneave, 1956; Pelli, Burns, Farell, & Moore-Page, 2006) to quantify pattern complexity. For a binary figure, perimetric complexity is well defined as

$$Perimetric\ Complexity\ = \frac{Perimeter^2}{Ink\ Area}.$$

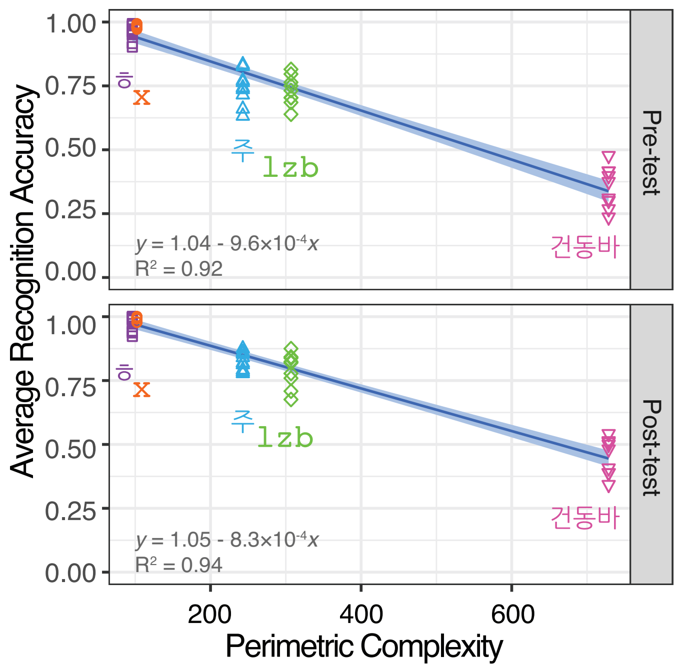

Figure 2-4. Perimetric complexity analysis.

Each data point represents one subject's visual span size in one of the following conditions: Korean components (complexity = 96), English letters (102), Korean characters (243), English trigrams (307), and Korean trigrams (729). Straight line is a linear fit, with shaded area showing 95% confidence interval.

Figure 2-4 shows the correlation between perimetric complexity of the five types of symbols and their visual span size (in terms of average recognition accuracy). From most simple to most complex, the average perimetric complexity of the sets of symbols was ranked as Korean components (96), English letters (102), Korean characters (243), English trigrams (307), and Korean trigrams (729). Their corresponding average accuracy for visual-span profiles was 95.7%, 98.4%, 74.2%, 73.2%, and 35.5% in the pre-test, and 97.2%, 99.2%, 82.4%, 79.1%, and 45.3% in the post-test. The general pattern was that more complex patters corresponded to lower recognition accuracy, but note that there was a small reversal for Korean components and English letters: Korean components had smaller complexity compared to English letters but had slightly poorer recognition performance. This reversal may be caused by the higher similarity between Korean components, which we will further discuss in Appendix 2.

Despite this small reversal, there was a high negative correlation between perimetric complexity and the average accuracy in both the pre-test ($r=-0.96$) and the post-test ($r=-0.97$). The slopes for the linear regressions were -9.6×10^{-4} in the pre-test and -8.3×10^{-4} in the post-test, which means that when perimetric complexity increases by 100, average recognition accuracy will decrease by 9.6% in the pre-test and 8.3% in the post-test. This indicates that pattern complexity has a negative impact on recognition performance, but training can reduce this impact.

Enlargement of Visual Span

Next we return to Figure 2-3 and compare the visual span profiles in the pre- and post-tests to examine the training effect. After training, the visual spans for Korean characters, Korean trigrams, and English trigrams all enlarged, with their profiles moving up and becoming broader. For English letters and Korean components, their visual spans remained unchanged due to a ceiling effect.

A repeated ANOVA of symbol type (English letters / English trigrams / Korean components / Korean characters / Korean trigrams) × session type (pre-test / post-test) was performed on the average recognition accuracy. We found a significant main effect of symbol type ($F(4, 72) = 403.17$, $p<0.001$) and a significant interaction between symbol type and session type ($F(4, 72) = 4.94$, $p=0.0014$): In general, the average accuracy was the highest for English letters (98.4% and 99.2% in the pre- and post-tests), followed by Korean components (95.7% and 97.2%), Korean characters (74.2% and 82.4%), English trigrams (73.2% and 79.1%), and Korean trigrams (35.5% and 45.3%). Post-hoc comparisons showed that the size of visual span ranked as: English letter = Korean letter > Korean character = English trigram > Korean trigram (">" signs mean "significantly larger than", and all of the adjusted $p<0.001$). This rank held true for both the pre- and post-tests. An analysis of the interaction effect showed that after training, visual span only enlarged for Korean single characters (average accuracy +8.2%, adjusted $p<0.001$), Korean trigrams (+9.7%, adjusted $p<0.001$) and English trigrams (+5.9%, adjusted $p=0.0025$), but not for Korean components or English letters. Our training thus successfully enlarged the visual span for the trained symbols, and the training effect also transferred to the untrained English trigrams.

The Reduction of Crowding

We used a two-stage model (see Methods) to estimate the reliability of the two stages, R_1 and R_2, for three types of symbols: Korean components, Korean characters, and English letters. The method of calculation and the results are summarized in Table 2-1. R_1 was equal to the corrected-for-guessing recognition accuracy when these symbols were in isolation, and it reflected the "acuity effect". In the pre- and post-tests, $R_{1,\ K\_component}$ = 95.5% and 97.1%, $R_{1,\ K\_character}$ = 74.1% and 82.3%, and $R_{1,\ E}$ = 98.3% and

99.2%, respectively. In this section we will focus on R_2 (which reflects the influence of crowding) and its change after training.

		Pre-test	Post-test
Korean component	$R_{1, K\_component}$ = Isolated-component accuracy (corrected for guessing)	95.5% ± 1.0%	97.1% ± 1.0%
	$R_{1, K\_component} \times R_{2, K\_component}$ = Component accuracy within isolated characters (corrected for guessing)	88.8% ± 1.2%	92.9% ± 0.6%
	$R_{2, K\_component}$	93.2% ± 1.4%	96.0% ± 0.8%
Korean character	$R_{1, K\_character}$ = Isolated-character accuracy (corrected for guessing)	74.1% ± 2.3%	82.3% ± 1.2%
	$R_{1, K\_character} \times R_{2, K\_character}$ = Character accuracy within trigrams (corrected for guessing)	35.3% ± 2.7%	45.1% ± 2.4%
	$R_{2, K\_character}$	46.6% ± 2.4%	53.9% ± 2.3%
English letter	$R_{1, E}$ = Isolated-letter accuracy (corrected for guessing)	98.3% ± 0.3%	99.2% ± 0.3%
	$R_{1, E} \times R_{2, E}$ = Letter accuracy within trigrams (corrected for guessing)	72.1% ± 1.9%	78.3% ± 2.3%
	$R_{2, E}$	73.3% ± 1.9%	78.9% ± 2.3%

Table 2-1. R_1 and R_2 calculations.

Figure 2-5 shows R_2 profiles for Korean components, Korean characters, and English letters in the pre- (open symbols, dashed lines) and post-tests (filled symbols, solid lines).

Figure 2-5A illustrates the effect of between-component crowding within individual Korean characters. $R_{2,\ K\_component}$ was close to 1 in both the pre-test (mean value 93.2%) and the post-test (mean value 96%). This indicates that between-component crowding was small overall, and a small reduction of within-character crowding occurred after training.

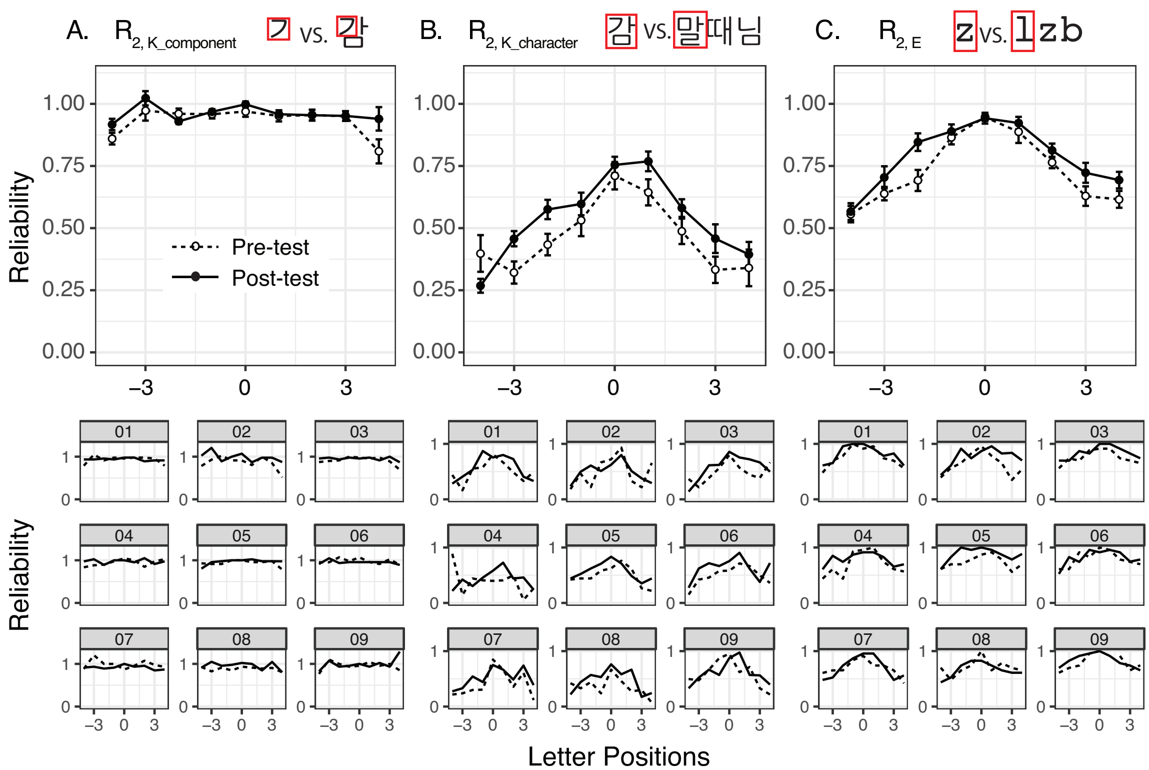

Figure 2-5. Profiles of R_2 reliability in pre- and post-tests.

A. Profiles for $R_{2,\ K\_component}$, reflecting the effect of between-component, within-character crowding on Korean components. **B.** Profiles for $R_{2,\ K\_character}$, reflecting the effect of between-character crowding on Korean characters. **C.** Profiles for $R_{2,\ E}$, reflecting the effect of between-letter crowding on English letters. Small panels on the bottom show individual curves. Dashed lines and open symbols, pre-test; solid lines and filled symbols, post-test; error bars: ±1 SEM.

Figure 2-5B illustrates the effect of between-character crowding on recognizing Korean characters. $R_{2,\,K\_character}$ was the highest on the midline and declined rapidly as the letter position moved farther away from midline, indicating more severe crowding. In the pre-test, $R_{2,\,K\_character}$ ranged from 35.5% to 59.3%, with a mean of 46.6%. In the post-test, it increased notably across almost all letter positions, ranging from 42.7% to 62.8% with a mean of 53.9%. This shows that between-character crowding also reduced as a result of training, and this reduction of between-character crowding appeared to be larger than the reduction of within-character crowding.

Figure 2-5C shows changes of $R_{2,\,E}$, which reflects the effect of between-letter crowding on recognizing English letters. The profiles exhibit a very similar shape as the profiles for Korean characters in panel B, but with larger R_2 values: In the pre-test, $R_{2,\,E}$ ranged from 63.6% to 82.1%, with a mean of 73.3%. In the post-test, it increased notably across almost all letter positions, ranging from 66.3% to 87.6% with a mean of 78.9%.

A two-factor repeated-measures ANOVA was performed on the average R_2 values across letter positions, with two within-subject factors being symbol type (Korean components / Korean characters / English letters) and session type (pre-test / post-test). No significant interaction was found, so we removed the term from the model. In the updated model, we found significant main effects of both symbol type ($F(2, 42) = 475.27$, $p<0.001$) and session type ($F(1, 42) = 19.68$, $p < 0.001$). Further post-hoc comparisons showed that the $R_{2,\,K\_component} > R_{2,\,E} > R_{2,\,K\_character}$, all of the adjusted $p<0.001$. This indicates that crowding is the least severe between Korean components within a character, more severe between English letters within a trigram, and most severe between Korean characters within a trigram. After training, R2 increased for all types of symbols (+2.8% for Korean components, +5.6% for English letters, and +7.3% for Korean characters). Despite the numerical differences in the improvement, the lack of an interaction between symbol type and session type indicates that training reduced crowding for all the symbols similarly, suggesting a successful transfer of training benefit from Korean to English.

Connecting Two Models of Korean Recognition

So far, we have built two separate models for Korean recognition but only one model for English, which makes it hard to compare Korean and English recognition directly. Is it possible to connect the two models for Korean recognition? We will focus our discussion on the data in the pre-test, but the same rules also apply to data in the post-test.

For both English and Korean, words are made up of characters and reading suffers from between-character crowding. But for Korean, characters are made of components, which means that Korean reading also suffers from within-character crowding. In the following discussion we try to unite the influence of between-character and within-character crowding in one model. We chose to treat the recognition of Korean components as the first stage in our united model, because English letters and Korean components had approximately the same pattern complexity and similar recognition accuracy when they are isolated, which provides a good baseline for comparing the two scripts.

First consider recognition of isolated Korean characters. Because of between-component crowding, the reliability of recognizing each *component* was reduced from $R_{1, K\_component}$ = 95.5% to $R_{1, K\_component} \times R_{2, K\_component}$ = 88.8%. Remember that there are two or three components within a character. Assuming that the components are otherwise recognized independently, the average reliability for isolated Korean *character* recognition would be $(R_{1, K\_component} \times R_{2, K\_component})^2$ = 78.8% for 2-component characters and $(R_{1, K\_component} \times R_{2, K\_component})^3$ = 70.0% for 3-component characters. These values were very close to the observed values of 79.5% and 69.9%, suggesting that our assumption of independent component-recognition is reasonable.

Next, consider identifying a Korean character within a trigram. For the same *components*, there is now an additional influence of between-character crowding. How would component recognition be influenced by this between-character crowding? We have quantified the magnitude of between-character crowding at the *character level* using the value of $R_{2, K\_character}$ (46.6%). For example, character recognition accuracy would be reduced from 0.7 when isolated to 0.7 × 0.466 = 0.326 when crowded. But $R_{2, K\_character}$ was not informative regarding the impact of between-character crowding at the *component level*. That is, if the recognition accuracy of a component in an isolated character was 0.7, we do not know what would the recognition accuracy be for a component in a trigram.

Can we use the between-character crowding for English letters, reflected by $R_{2, E}$, to estimate the effect of between-character crowding on Korean components? Since

English letters and Korean components have similar complexity, it is possible that the same between-character crowding factor applies to both English letters and Korean components. If this assumption is true, a Korean component in a trigram would have reliability of recognition of $R_{1,\ K\_component} \times R_{2,\ K\_component} \times R_{2,\ E}$, where $R_{2,\ K\_component}$ represents the influence of within-character crowding and $R_{2,\ E}$ represents the hypothesized between-character crowding at the *component level*. Given that there are 40.9% two-component characters, we can compute the weighted average for the recognition of two-component and three-component characters to estimate the average probability of recognition for a Korean character in a trigram. Using the reliability values from Table 2-1, this weighted average is $0.409 \times (R_{1,\ K\_component} \times R_{2,\ K\_component} \times R_{2,\ E})^2 + 0.591 \times (R_{1,\ K\_component} \times R_{2,\ K\_component} \times R_{2,\ E})^3 = 33.8\%$. This is very close to the observed value of 35.3%, which supports our assumption that English letters and Korean components are limited by the same between-character crowding.

Together, our above analysis united the within- and between-character crowding in Korean recognition. Due to within-character crowding, Korean characters are harder to recognize than English letters. Most importantly, between-character crowding had the same effect on Korean components and English letters, likely because they have similar pattern complexity. These findings point to common constraints underlying Korean and English recognition.

Discussion

In the current study, we sought to test whether similar constraints limit the size of Korean and English visual spans, and whether training-related enlargement of the visual span transfers to untrained symbols. Our first finding is that the size of the visual span is determined by the complexity of the pattern. Below we will discuss how complexity impacts isolated and crowded pattern recognition respectively. Another related influencing factor of pattern recognition is similarity of the symbols within a set, which we will discuss in Appendix 2.

Pattern Complexity Influences Visual Span

Increased pattern complexity was associated with poorer recognition. For isolated symbols, complexity level ranked as English letter = Korean component < Korean character, and the averaged visual-span accuracy ranked in the opposite order, i.e. Korean character < Korean component = English letter. Korean characters are made up of multiple components. Compared to isolated Korean components, multiple components in Korean characters increased pattern complexity and introduced extra within-character crowding, which reduced the size of the visual span. This is consistent with the findings that patterns with higher complexity have higher acuity thresholds (Watson & Ahumada, 2012; J.-Y. Y. Zhang, Zhang, Xue, Liu, & Yu, 2007). Presumably, in more complex patterns the features crowd each other, so that larger between-feature spacing (therefore larger character size) is needed to escape between-feature crowding.

For crowded symbols (English and Korean trigrams), the sizes of their visual spans were smaller than isolated symbols due to between-symbol crowding. Judging from the values of $R_{2\_K,\ character}$ and R_{2_E}, it seems that more complex symbols (Korean characters) have a smaller second-stage reliability and are more influenced by between-symbol crowding. Other studies have also found that more complex patterns are susceptible to more severe crowding (Bernard & Chung, 2011; Wang et al., 2014; J.-Y. Zhang, Zhang, Xue, Liu, & Yu, 2009). But according to our analysis in the section "Connecting Two Models of Korean Recognition", Korean components and English letters suffer similar levels of between-character crowding. Therefore, Korean characters appear to suffer more between-character crowding than English letters only because the effect of crowding is accumulated from all the components.

Our findings here suggest that complexity has a major role of determining the size of the visual span. An increase in pattern complexity is associated with a reduction in the size of the visual span, primarily due to both increased within-symbol crowding and accumulated between-symbol crowding. Whether within- and between-symbol crowding has different neural origins remains to be investigated by future studies.

Visual Span and Its Enlargement

Our second finding is that training to read Korean characters enlarged the Korean visual span and also transferred to English visual span. As we previously discussed, the size of the visual span was directly related to physical properties of the stimuli such as pattern complexity, and independent of the language of the script. These findings provide evidence that the visual span describes pre-symbolic sensory limits on recognition of scripts.

What underlies the training-related enlargement? We rejected the hypothesis that subjects learned the exact templates of the trained samples, because in that case training benefits would not transfer to an untrained set of symbols. But the subjects may actually have learned templates of the features instead of the symbols. Pelli et al. (2006) adopted probability summation theory to quantify the number of features used by human observers in a pattern-recognition task. The theory assumes that features within a pattern have equal energy and are independently detected. A pattern is detected whenever any of its features is detected, but is identified only when all its features are detected. By comparing the contrast threshold for detection and identification, they were able to estimate the number of features within a pattern. The conclusion was that successful identification of a pattern (e.g. a letter/a character/a word) required identifying 7 ± 2 features, independent of pattern complexity.

Pelli and colleagues did not specify the exact list of features, but it is possible that there are shared features between different sets of symbols. Using the “Bubbles” technique (Gosselin & Schyns, 2001), Fiset and colleagues empirically identified critical features for letter recognition(Fiset et al., 2008). They found that for human observers, line terminations were the most important features to identify both uppercase and lowercase letters. Line terminations are shared across English and Korean symbols and may also be important features for Korean recognition. Moreover, it has been shown that perceptual templates can indeed be improved after training to have higher sampling efficiency and to better extract useful information for recognition in the presence of noise (Chung, Levi, & Tjan, 2005; Gold et al., 2004). If the subjects learned better templates of the most important features such as line terminations, which partly overlap across symbol sets, training benefits are likely to be transferred between scripts as demonstrated in the current study.

Connecting Visual Span with Reading

The visual span has been studied in the context of reading. For English reading, the visual span has been proposed to be a sensory bottleneck limiting reading speed (Legge et al., 2007). Is Korean reading similar to English reading, and thus the visual span hypothesis of reading also applies to Korean? We can look at how the size of the visual span and the reading speed for Korean and English depend on eccentricity respectively. For English, Legge and colleagues found that the size of the visual span decreased with eccentricity (Legge, Mansfield, & Chung, 2001), and that log-transformed reading speed had a high correlation with the size of the visual span when eccentricity changed (Legge et al., 2007). It seems that the size of visual span is affected by eccentricity and thereby limiting reading speed.

Although we do not have data for Korean visual span in central vision, it is likely that the size of Korean visual span decreases with eccentricity in a similar manner as English visual span. If the visual span hypothesis holds true for Korean, we will expect that both Korean and English reading speed have similar dependency on eccentricity.

Chung and colleagues measured English reading speed at different retinal eccentricities using Rapid Serial Visual Presentation (RSVP) (Chung et al., 1998). They then used an E_2 factor to describe how maximum reading speed changes with eccentricity, and found an E_2 of 4.13° for the critical duration corresponding to maximum reading speed. This means that when eccentricity is 4.13°, reading speed drops to half its value in central vision. We performed the same analysis on Korean RSVP reading speed data (acquired from the authors of Baek et al., 2016) and obtained a very similar E_2 factor of 4.32°. That is to say, reading speeds for Korean and English have similar dependency on eccentricity, despite the fact that Korean characters are more complex than English letters. This indicates that common visual properties are limiting reading performance for different languages, which is summarized by the visual span.

Conclusion

Our final conclusion is that the visual span describes a presymbolic constraint on pattern recognition, and its size is largely influenced by physical properties of the pattern such as perimetric complexity rather than high-level properties such as language. Training to recognize scripts for one language can enlarge its visual span and the improvement transfers to another language, possibly due to shared critical features between scripts. According to the visual span hypothesis for reading, the enlargement of visual span will likely be associated with an improved reading speed for both languages due to similar visual constraints on reading across languages.

Chapter 3. Linking Crowding, Visual Span, and Reading

Introduction

According to the visual span hypothesis, a sensory bottleneck on reading speed is the size of the visual span, which is the number of letters that can be recognized accurately without eye movements (Legge et al., 2007). By decomposing the errors made in visual span measurements, He et al. (2013) showed that visual crowding is the major factor limiting the size of the visual span. More directly, Pelli et al. (2007) have provided evidence that the visual span refers to the characters that are not crowded, and reading speed is proportional to this uncrowded span. It thus appears that crowding limits reading speed by limiting the size of the visual span.

If this linkage between visual span and reading speed is correct, reduced crowding should result in an enlarged visual span and improved reading speed. In support of this view, it has been found that perceptual training on a trigram letter-recognition task in peripheral vision (Fig. 3-1A) enlarged the visual span and improved reading speed, and the largest component of visual span enlargement was the reduction of crowding (He et al., 2013). But reduced crowding does not always lead to improved reading. For example, reducing crowding by increasing letter spacing beyond standard spacing slows reading down (Chung, 2002; Yu, Cheung, Legge, & Chung, 2007). This, however, does not necessarily argue against the claim that crowding is the major sensory limit on reading speed, because enlarged spacing also resulted in larger eccentricity and degraded word form. Moreover, in this example extra-wide letter spacing resulted in both a shrinkage of the visual span and a decrease in reading speed (Yu et al., 2007), confirming the link between visual span and reading.

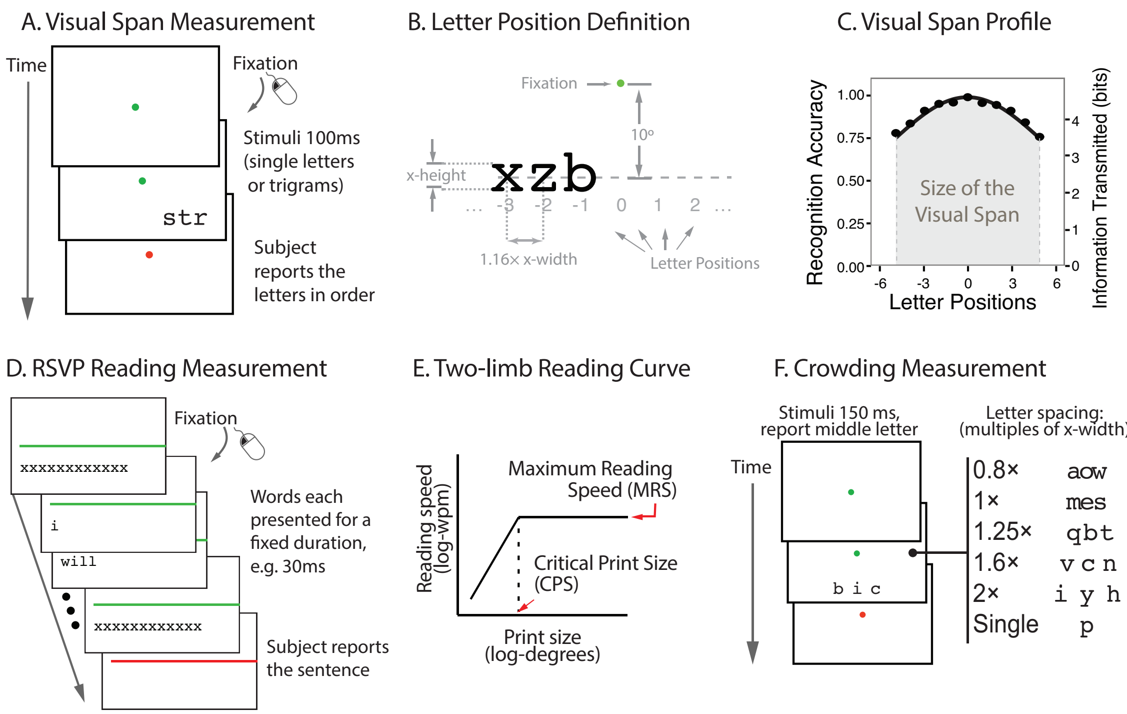

Figure 3-1. Diagrams of measurements.

But a further dissociation between crowding and reading was found by Chung (2007). She found that after six days of training in identifying crowded letters in peripheral vision (hereafter referred to as “uncrowd training”), the spatial extent of crowding (defined as the smallest target-to-flanker spacing yielding 50% recognition accuracy of the target letter) was reduced by 38%, but maximum reading speed barely changed. In a follow-up study, Chung & Truong (2013) showed that a similar training task could both reduce the spatial extent of crowding and enlarge the size of the visual span. These results seem to argue against the proposed linkage between visual span, crowding and reading speed.

Here, we examined two properties of the stimuli used in letter-recognition training that may influence the transfer to reading—the spatial distribution of training stimuli and the presence of flankers.

The first property is the spatial distribution of training stimuli. In a typical trial of trigram visual span measurement (Fig. 3-1A), a trigram appears briefly at a certain vertical eccentricity (say, 10° in the lower field) but an unpredictable horizontal eccentricity. After the trigram disappears, the subject reports the identity of the three letters in order. From trial to trial, the horizontal location of the trigram varies, resembling different letter positions in a word relative to fixation. In a trial of "uncrowd" training as in Chung (2007), the procedure is similar, but the trigram is always centered right below fixation. The trigram has narrower-than-standard letter spacing (0.8× x-width), and only the center (target) letter needs to be reported. As a result, attention is directed to one specific letter position below fixation, and flanking letters can be ignored. For English reading, distributed attention to multiple letters of a word may be more advantageous than focused attention and letter-by-letter processing. In the extreme case of letter-by-letter reading in pure alexia, readers have normal thresholds to recognize single letters but cannot encode several separate visual features simultaneously, resulting in abnormally small attentional span and severely impaired reading (for a review, see Dehaene & Cohen, 2011). Training on a task with focused attention to a single letter located at a fixed location is dissimilar to normal reading. Therefore, spatially localized training may limit the transfer of training benefits to reading compared to training on stimuli distributed at different spatial locations, similar to letters in a line of text. Chung's (2007) "uncrowd" training may still have the potential to improve reading, but limited by its restricted training location. Our prediction is that when the uncrowd training is spatially distributed across locations, reading speed will show a greater improvement than when training occurs at a single spatial location.

A second property, pertinent to training peripheral vision to read, is whether crowding is present or not. That is, whether training trials use isolated letters or letters flanked by other letters. Excessive crowding is often associated with slow reading, not only in peripheral vision for normally-sighted subjects (as we discussed before), but also in central vision for people with amblyopia (Levi, Song, & Pelli, 2007) or dyslexia (Callens, Whitney, Tops, & Brysbaert, 2013; Martelli, Di Filippo, Spinelli, & Zoccolotti, 2009; Moll & Jones, 2013; but see Doron, Manassi, Herzog, & Ahissar, 2015;). Given the association between crowding and reading speed, training without crowded stimuli may fail to reduce the effect of crowding and thus cannot improve reading. But if learning to identify unflanked letters can reduce crowding, reading speed should improve despite the fact that isolated letters only rarely occur in real-life text.

To study these two factors and their influence on training effects, we designed 3 training paradigms for letter recognition (Fig. 3-2): One paradigm is a replication of Chung (2007) where target letters are localized at 10 degrees directly below fixation, with two narrowly-arranged flankers on each side (Flanked-Localized Group). In the second paradigm, target letters were located at 10 degrees vertically below fixation but distributed across different horizontal locations, also with two narrowly-arranged flankers on each side (Flanked-Distributed Group). The third paradigm had isolated, spatially distributed target letters (Isolated-Distributed). By evaluating the training effects on crowding, visual span, and reading speed, our goal is to investigate the necessary and sufficient conditions for transfer of training from letter recognition to the size of the visual span and reading speed.

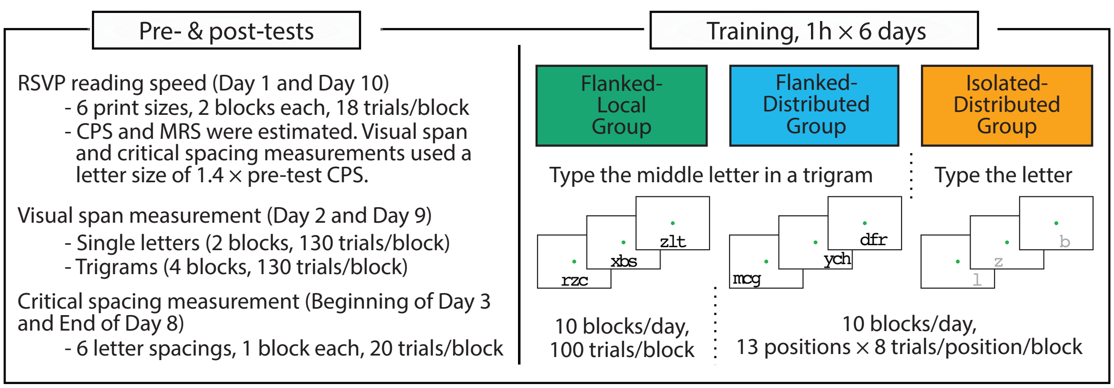

Figure 3-2. Experimental design.

Methods

Participants

12 male and 15 female college students were recruited from the University of Minnesota and randomly assigned to 3 groups (described later). Participants all had normal or corrected-to-normal vision, and their binocular acuity (Lighthouse Near Acuity

Chart, Lighthouse Low Vision Products, Long Island City, NY) and reading performance (MNREAD, Precision Vision, La Salle, IL) were tested before the experiment. Subject information is summarized in Table 3-1. The protocol was approved by the Institutional Review Board and was in compliance with the Declaration of Helsinki. All subjects gave informed consent prior to the experiment.

Group	Gender ratio (M:F)	Age	Visual acuity (LogMAR)	MNREAD Measurements		
				Reading Acuity (LogMAR)	Critical Print Size (LogMAR)	Maximum Reading Speed (wpm)
Flanked-Local	5:4	21.4 ± .7	-.07 ± .01	-.39 ± .03	.06 ± .05	221 ± 6
Flanked-Distributed	3:6	22.7 ± 1.0	-.08 ± .01	-.4 ± .03	-.07 ± .04	234 ± 9
Isolated-Distributed	4:5	21.3 ± .7	-.08 ± .01	-.4 ± .02	-.02 ± .03	213 ± 6
All groups	12:15	21.8 ± .4	-.08 ± .004	-.4 ± .02	.04 ± .02	223 ± 4

Table 3-1. Subject groups' information in Chapter 3 (mean ± SEM).

Reading speed – print size curves from MNREAD measurements were fitted on a log-log scale with the following function using non-linear least squares fitting to determine maximum reading speed and critical print size:

$$RS = \mathrm{MRS} \times (1 - e^{-e^{lrc}(x - x_{int})})$$

RS, reading speed; *MRS*, maximum reading speed; *lrc*, parameter describing the elbow of the curve; *x*, print size; x_{int}, reading acuity. Critical print size was the print size where reading speed reaches 85% of the maximum reading speed.

Stimuli and Apparatus

The stimuli consisted of black lowercase letters on a white background (background luminance 90 cd/m^2; Weber contrast = 99%), except for the training task of the Isolated-Distributed Group where stimuli were gray lowercase letters on a white background (Weber contrast ranges from 7% to 14%). All stimuli were viewed binocularly from 40 cm in a dark room. The letters were rendered in Courier font. Letter spacing in the reading task and visual span measurement was 1.16× x-width (standard spacing) but varied from 0.8× to 2× x-width in the crowding measurement.

The stimuli were generated and presented using MATLAB R2014b with Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). We used a NEC MultiSync CRT monitor (model FP2141SB-BK, NEC, Tokyo, Japan; refresh rate = 100 Hz; spatial resolution = 0.04°/pixel) controlled by a Mac Pro Quad-Core computer (model A1186, Apple Inc., Cupertino, CA). Viewing distance was maintained using a chin rest and subject's fixation was monitored using a webcam.

Experimental Design

The tasks we used are illustrated in Figure 3-1, and the procedure is explained in Figure 3-2. Our experiment consisted of 3 parts (Fig. 3-2): pre-test (2 days), training (6 days), and post-test (2 days). For 18 of the 27 subjects, the experiment took place on 10 consecutive days, whereas for the other subjects scheduling arrangements meant that the experiment spanned 11 (4 subjects), 12 (4 subjects) or 15 (1 subject) days. Since the effectiveness of visual perceptual learning on identifying crowded letters and enlarging the visual span is similar for daily, weekly and biweekly training (Chung & Truong, 2013), the variations in scheduling should have minimal effect in our design.

In the pre- and post-tests, subjects' reading speed, visual span, and spatial extent of crowding were measured. Reading speed was measured on the first and last days. Each individual's critical print size (CPS) for reading was determined after the pre-test (see section "Reading Measurement" below), and a print size of 1.4×CPS was used for subsequent visual span measurement, crowding measurement, and training.

Visual span profiles were measured on Day 2 and Day 9, using both single letters and trigrams (three adjacent letters; see section "Visual Span Measurement"), with some exception for the Flanked-Local Group: Since the measurement of single letters was added later in the data collection process, 5 subjects in this group had no data on single-letter visual span, 3 subjects had the post-test data only, and only 1 subject had both the pre- and post-test data.

The measurement of crowding only took 10 minutes, and therefore we performed the measurement immediately before the first training block on Day 3 and after the last training block on Day 8, similar to Chung (2007).

Training consisted of 6 days of identifying letters, approximately 1 hour/day. We used three training paradigms to separate the effect of 1) the spatial distribution of training stimuli, and 2) the presence of flankers (Fig. 3-2; details see section "Training").

Reading Measurement

Reading speed was measured using Rapid Serial Visual Presentation (RSVP, Forster, 1970; Rubin & Turano, 1992). For each RSVP trial, a sentence was randomly chosen from a pool of 847 sentences for testing. A subject never saw the same sentence twice. The average sentence length was 11 words, ranging from 7 to 15 words. The word length averaged 4 letters, ranging from 1 to 12 letters. In an RSVP trial, a sentence was randomly chosen and presented word-by-word, with the sentence preceded and followed by masks of “xxxxxxxxxx” (Fig. 3-1D). Subjects were asked to fixate on a line without making vertical eye-movements, but horizontal eye-movements along the line were permitted. Words were presented 10° below the fixation line and were left-aligned with the left edge of the fixation line. Subjects read the sentences out loud and the experimenter recorded the number of correctly-read words.

In each block of 18 trials, 6 different word exposure durations were tested in a random order (3 times each). Depending on individual performance during practice, 1 of 3 duration sets could be chosen for a given block: {30, 53, 93, 164, 290, and 511ms}, {53, 93, 164, 290, 511, and 1000ms}, or {93, 164, 290, 511, 1000, and 2000ms}. The choice of set maximally ensures 1) greater than 80% word recognition accuracy for the longest duration in the set, 2) close to chance-level word recognition performance for the shortest duration in the set, and 3) good eye fixation for all durations (fixation becomes poorer as duration becomes longer). The resulting accuracy–duration curve was then fitted with a psychometric function, and the subject’s reading speed (measured in words per minute, wpm) was calculated using the exposure duration yielding 80% accuracy of word recognition.

Reading speed was measured in this way for 6 different print sizes: 0.56°, 0.79°, 1.12°, 1.59°, 2.26°, and 3.2° in x-height. In the pre-test, these 6 print sizes were tested in a random order, one print size in a block, for the first 6 blocks, and then tested in the reverse order for another 6 blocks. The post-test followed the same order as in the pre-test. The resulting reading speed–print size curve was then fitted with a two-limb function on a log-log scale (Fig. 3-1E) to extract the subject's maximum reading speed (MRS) and critical print size (CPS). The slope of the two limbs were constrained to 2.32 and 0 (following Chung, 2007). This curve represents that reading speed remains constant at the maximum reading speed for larger print sizes, but starts to decrease when print size becomes smaller than the critical print size. After extracting the subject's pre-test CPS, a print size of 1.4×CPS was used for subsequent visual span measurement, crowding measurement, and training.

Visual Span Measurement

In order to measure the size of the visual span, we performed a letter-recognition task, as described in He et al. (2013). Figure 3-1A illustrates the basic procedure of a trial: A subject fixated on a dot and clicked the mouse to initiate a trial. Stimuli were randomly-chosen lowercase letters, either in isolation or arranged as triplets (trigrams, placed in 3 adjacent slots). Letter(s) appeared for 100ms in the lower visual field for the subject to identify. For a letter in a trigram to be correct, both its identity and location needed to be correct.

Letters were presented in pre-defined slots, as shown in Figure 3-1B. The slots were horizontally arranged on an imaginary line at 10° in the lower visual field. The slot on the fixation midline was labeled 0, and left and right ones were labeled with negative and positive numbers respectively. The center-to-center spacing between adjacent slots was 1.16× x-width, corresponding to standard spacing in the Courier font. Since each individual used a different print size determined by their critical print size for reading, the actual horizontal eccentricity of the slots varied across subjects. We analyzed the data in the "letter position" space instead of the "visual degree" space, because once print size exceeds the critical print size for reading, the size of the visual span (in terms of number of recognizable letters) remains constant within a reasonable range of print sizes (Legge et al., 2007). By using the "letter position" space, we were able to compare results between different subjects.

In each block of trials, letters or trigrams were centered 10 times on each slot from -6 to 6, including 0 (the midline). This means that a total of 130 letters were presented during a single-letter block, or 390 letters during a trigram block. In a trigram test, since slots ±6 and ±7 had fewer letters than the other slots, only data from slots -5 to 5 were used in further analyses.

To get the visual span profile, letter-recognition accuracy was plotted against letter positions (Fig. 3-1C). We converted letter-recognition accuracy to information transmitted in bits using the formula

$$\textit{Information transmitted in bits} = -0.036996 + 4.6761 \times \textit{letter recognition accuracy},$$

where chance level performance (about 3.8% correct) corresponds to 0 bits of information and 100% accuracy corresponds to about 4.7 bits. The number of bits was added for slots -5 to 5 to estimate the size of the visual span (Fig. 3-1C).

Crowding Measurement

The crowding task measured the spatial extent of crowding. Following Chung (2007), the spatial extent was defined as the letter separation (in multiples of x-width) yielding 50% recognition accuracy of the target letter, corrected for guessing. In one trial (Fig. 3-1F), a target letter appeared briefly (150ms) on the screen, either in isolation or with two flanking letters on its left and right sides. The target letter was always placed at 10° in the lower visual field right below the fixation point. The center-to-center spacing between the target and its flanking letters could be 0.8×, 1×, 1.25×, 1.6×, 2× x-width (Fig. 3-1F). Subjects needed to report the target letter and ignore the flankers.

The 5 conditions with different target-flanker spacing and 1 no-flanker condition were tested in a random order, each in a 20-trial block. The post-test followed the same spacing order as the pre-test. The resulting accuracy–spacing curve was then fitted with a psychometric function to determine the spatial extent of crowding. The reduction of crowding after training was quantified as the percent change in the spatial extent of crowding. For instance, if the spatial extent was 1.5× x-width in the pre-test and 1× in the post-test, then the reduction was (1.5 - 1)/1.5 = 33%.

Training

The training task was to identify target letters, either flanked on both sides by letters or in isolation, at 10° in the lower visual field (Fig. 3-2). One group was trained with crowded, localized letters (Flanked-Local Group): Target letters always appeared in slot 0 with two flanking letters, and the center-to-center spacing was always 0.8× x-width.

One group was trained with crowded, distributed letters (Flanked-Distributed Group): Target letters could appear in any slot from -6 to 6 (defined in the same way as in Figure 3-1B), with two flanking letters at a separation of 0.8× x-width. A third group was trained with isolated, distributed letters (Isolated-Distributed Group): Target letters could appear in any slot from -6 to 6, with no flanking letters, but with reduced contrast and shorter exposure duration to increase task difficulty. Contrast level and the length of exposure duration were chosen for each individual prior to training to achieve roughly 50%-70% correct during practice, so that the task was below ceiling and well above chance.

Each daily training session had 10 blocks. For the Flanked-Local Group, each block had 100 trials. For the Flanked-Distributed and Isolated-Distributed Groups, each block had 104 trials, in which the target letter fell in each one of the 13 slots 8 times.

Data Analysis

When comparing training effects between groups, if not otherwise specified, we performed 3×2 mixed-design ANOVAs, with group (Flanked-Local / Flanked-Distributed / Isolated-Distributed) as the between-subject factor and session type (pre-/post-test) as the within-subject factor. If a significant interaction was found, we further analyzed the interaction using *R* with the package *phia* (Post-Hoc Interaction Analysis; Martínez, 2013). The reported p-values were adjusted for multiple comparisons.

For maximum reading speed, critical print size, and spatial extent of crowding, ANOVA analyses were performed using log-transformed data, because the original curve fittings were performed on a log-scale. When we report group-averaged data in the preand post-tests, we will transform the values back into their original units for easier understanding. For example, “maximum reading speed improved from 183 to 235 wpm” was converted from “averaged log-wpm changed from 2.26 to 2.37”. However, when we report the amount of improvement (such as in Table 3-2), the computation was done in the original unit (say, wpm) without log-transformation, because the changes in log-units are less intuitive to interpret. Take the example of “average improvement in maximum reading speed was 30.3%”. We first compute the percent improvement in wpm for each subject, and then average the percent change across subjects to get the value of 30.3%.

Results

Our main results are summarized in Figure 3-3 and Table 3-2. Left panels of Figure 3-3 show averaged group data with curves fitted to the average values. These curves are for demonstrative purposes only; data analyses were based on fitted values for individual subjects instead of the group-level curves. Right panels show interaction plots for the ANOVAs we performed, and the error bars indicate ± 1 SEM for within-subject ANOVAs (Cousineau, 2005). The error bars can be used as a visual guide for within-group comparisons (between the pre- and post-tests), but not appropriate for between-group comparisons.

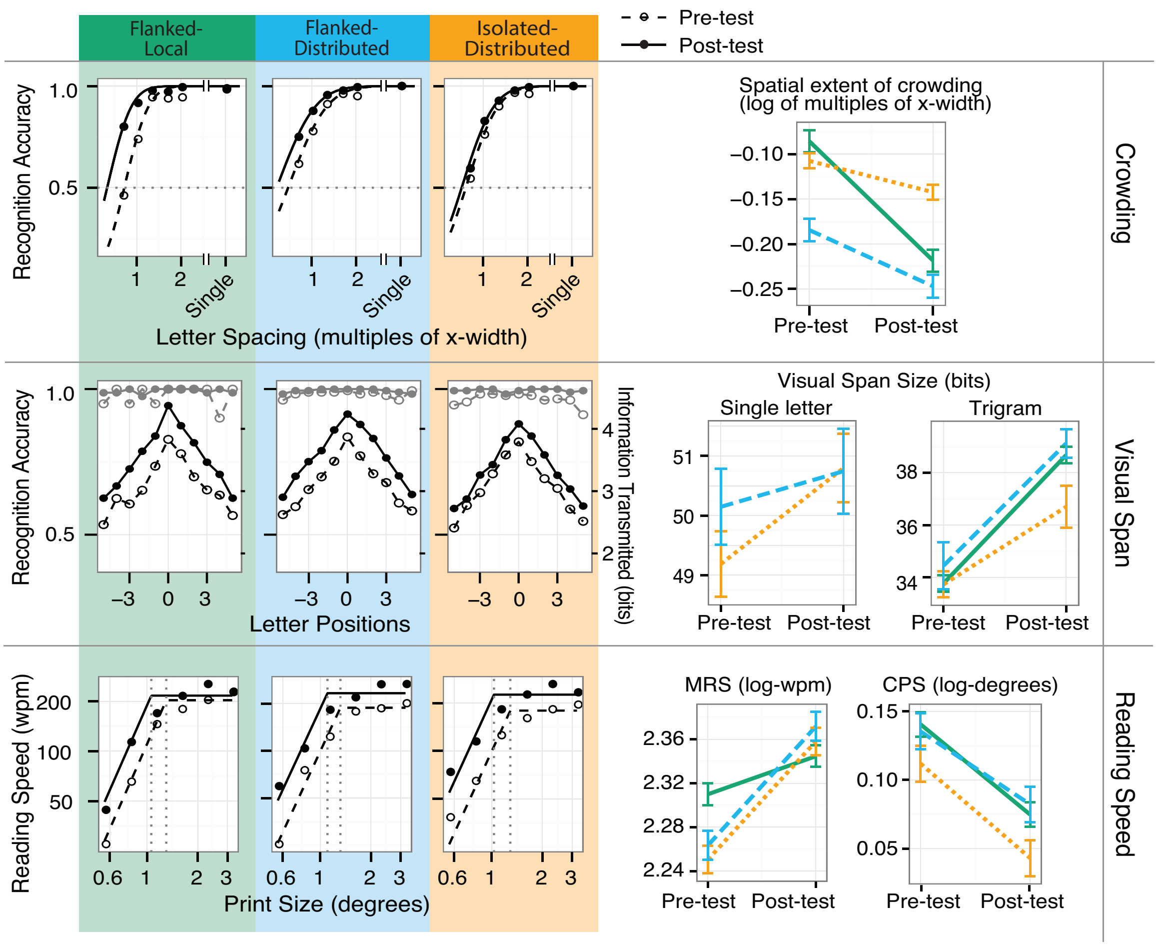

Figure 3-3. Summary of results.

Left panels: group curves fitted on averaged data. Open circles and dashed lines: pre-test; filled circles and solid lines: post-test. **Right panels**: Interaction plots showing group averages before and after training. Green solid lines: Flanked-Local Group; blue dashed lines: Flanked-Distributed Group; orange dotted lines: Isolated-Distributed Group. Error bars indicate ± 1 SEM for within-subject ANOVAs (Cousineau, 2005). **Top row**: Crowding results. Left: crowding curves. Dotted gray lines indicate 52% recognition accuracy (i.e. 50% accuracy after correction for guessing), which is the criterion used to determine the spatial extent of crowding. Right: Interaction plots of the spatial extent of crowding (log of multiples of x-width). **Middle row**: Visual span results. Left: visual-span profiles for single letters (gray) and trigrams (black). Right: Interaction plots of visual span sizes (bits) for single letters and trigrams. **Bottom row**: Reading speed results. Left: reading curves. Dotted vertical lines indicate the positions of fitted critical print sizes before and after training. Right: Interaction plots of maximum reading speed (MRS, log-wpm) and critical print size (CPS, log-degrees).

		Flanked-Local	Flanked-Distributed	Isolated-Distributed
Spatial Extent of Crowding (multiples of x-width)		**↓ 25.4% ± 3.9%**	**↓ 11% ± 5.3%**	↓ 7.2% ± 3.5%
Size of Trigram Visual Span (bits)		**↑ 4.9 ± 0.6**	**↑ 4.7 ± 0.7**	**↑ 3.0 ± 0.8**
Reading Performance	MRS (wpm)	↑ 9.2% ± 5.0%	**↑ 30.3% ± 7.8%**	**↑ 29.9% ± 7.4%**
	CPS (°)	**↓ 13.5% ± 3.5%**	**↓ 10.2% ± 5.9%**	**↓ 13.4% ± 5.3%**

Table 3-2. Average of changes after training (mean ± SEM).

Up- and down-pointing arrows indicate the direction of change (increase or decrease). Significant changes are marked bold and in color (red for increases and green for decreases), as indicated by ANOVA or interaction analyses (see text). MRS, maximum reading speed. CPS, critical print size.

In the following sections, we will first discuss the training effects on the spatial extent of crowding, visual span, and reading separately, and then put them together in a common framework in order to understand the nature of their associations.

Reduction of the Spatial Extent of Crowding

The top rows in Figure 3-3 and Table 3-2 summarize the results for crowding measurements. The crowding curve (accuracy against letter-spacing) for the Flanked-Local Group shifted leftward after training, indicating improved recognition accuracy at smaller letter separations. The Flanked-Distributed Group also showed some improvement, although to a lesser extent. No large improvement was apparent from the crowding curves of the Isolated-Distributed Group.

We then derived the spatial extent of crowding (letter separation yielding 50% recognition accuracy, corrected for guessing) for each subject, and performed an ANOVA to examine the effect of training (see interaction plot). We found a significant main effect of session type ($F(1, 24)=33.25$, $p<0.001$), as well as a significant interaction ($F(2, 24)=4.98$, $p=0.02$). Analysis of interaction showed that only the two Flanked groups had statistically significant reduction of crowding after training (Flanked-Local: spatial extent of crowding decreased from 0.82 to 0.60 times x-width, mean reduction 25.4%, $p<0.001$, adjusted for multiple comparisons; Flanked-Distributed: from 0.69 to 0.61 times x-width, average reduction 11%, adjusted $p=0.03$), but not the Isolated-Distributed Group (from 0.78 to 0.72 times x-width, average reduction 7.2%, adjusted $p=0.13$). The reduction for the Flanked-Local Group was significantly larger than that for the Flanked-Distributed Group (adjusted $p=0.04$) and the Isolated-Distributed Group (adjusted $p=0.008$). No difference was found between the two Distributed groups (adjusted $p=0.50$).

One concern is that the pre-test performance of the Flanked-Distributed Group was poorer than the other two groups, which may leave more room for improvement. We therefore performed a linear regression between the reduction of crowding (difference in spatial extent, log unit) and the pre-test crowding level. The regression line had a negative slope of -0.47 that was significantly different from zero ($p=0.009$), i.e. larger reduction was associated with poorer pre-test level. This indicates that in the previous analysis we had underestimated the reduction of crowding in the Flanked-Distributed Group, because it had a better starting level than the other two groups. A new between-group ANOVA was performed on the reduction of crowding, using the pre-test level as a covariate. After accounting for the pre-test level, there was still a main effect of group ($F(2, 23)=4.62$, $p=0.02$). A post-hoc analysis showed that crowding was only reduced in the two Flanked groups (adjusted $p<0.01$) but not in the Isolated group ($p=0.14$). When comparing between groups, a slightly different pattern emerged: When comparing the two Flanked groups, although the face value of the reduction of crowding was larger for the Flanked-Local group, the difference did not reach significance after pre-test level was taken into account (adjusted $p=0.28$). The reduction for the Flanked-Local Group was significantly larger than for the Isolated-Distributed Group (adjusted $p=0.007$), and the two Distributed groups did not differ (adjusted $p=0.28$).

Taken together, our analyses suggested that the Flanked-Local training was the most effective in reducing the spatial extent of crowding, followed by the Flanked-Distributed training, but the Isolated-Distributed training was not effective.

Enlargement of the Visual Span

The middle rows in Figure 3-3 and Table 3-2 summarize the results for visual span measurement. For single letter visual span profiles (gray lines and symbols), only a very small change was observed. For trigrams (black lines and symbols), all groups exhibited notable training-related enlargement, and the improvement appeared to be smaller for the Isolated-Distributed Group compared to the other two groups.

We then performed ANOVAs on visual span sizes for single letters and trigrams respectively. For single letters, we only used the data from the Flanked-Distributed and Isolated-Distributed Groups, because in the Flanked-Local Group only 1 subject had both the pre- and post-test data in this condition. We found a significant main effect of group ($F(1, 16)=6.88$, $p=0.019$) and a marginal significant main effect of session type ($F(1, 16)=3.75$, $p=0.07$), and a significant interaction between the two factors ($F(1, 16)=5.38$, $p=0.034$). An interaction analysis revealed that only the Isolated-Distributed Group had statistically significant improvement after training (from 49.2 to 50.8 bits, or averaged accuracy from 96.4% to 99.5%; adjusted $p<0.001$), and this improvement was significantly larger ($p=0.02$) than that for the Flanked-Distributed Group (from 50.1 to 50.7 bits, or averaged accuracy from 98.3% to 99.4%). Nevertheless, the absolute value of the improvement was very small due to the ceiling effect, and thus may not reflect large changes in visual functions.

An ANOVA on trigram data revealed a main effect of session type ($F(1, 24)=48.7$, $p<0.001$), indicating enlarged visual spans after training. No main effect of group or interaction was found. Although the groups were not significantly different, the absolute enlargement for the Isolated-Distributed Group (from 33.7 to 35.7, +3 bits, or averaged accuracy from 66.4% to 72.1%) was smaller than that for the other two groups (Flanked-Local Group: from 33.8 to 38.7, +4.9 bits, or averaged accuracy from 66.4% to 76.0%; Flanked-Distributed Group: from 34.4 to 39.1, +4.7 bits, or averaged accuracy from 67.8% to 76.8%). This pattern is also apparent from the interaction plot in Figure 3-3.

This enlargement evaluates changes accumulated over 11 letter slots (from -5 to 5) on the visual span profile. Past studies using a similar trigram training have found enlargement of 5.4 bits accumulated over 9 letter slots (He et al., 2013) or 6.1 bits over 11 slots (Chung et al., 2004), whereas their corresponding no-training control groups had 0.5 bit (He et al., 2013) and about 1.2 bits change (estimated from Fig. 9B in Chung et al., 2004). Our training-related enlargement is smaller compared to previous training paradigms, but seems larger than no-training controls.

Improvement in Maximum Reading Speed

Reading performance is summarized in the bottom rows of Figure 3-3 and Table 3-2. We will mainly focus on the change of maximum reading speed in the following discussion, and report the results on critical print size in Appendix 3.

From the averaged reading curves in Figure 3-3, only the two Distributed groups had larger maximum reading speed after training. For the Flanked-Local Group, maximum reading speed remained almost unchanged. An analysis of individual fitted parameters confirmed the group-level pattern, as shown in the interaction plot. A 3×2 ANOVA on log-MRS revealed no main effect but a marginally significant interaction between group and session type ($F(2, 24)=3.15$, $p=0.06$). Analysis of the interaction showed that only the two Distributed groups showed significant improvement after training (Flanked-Distributed: from 183 to 235 wpm, average improvement 30.3%; Isolated-Distributed: from 178 to 228 wpm, average improvement 29.9%; both of their adjusted $p<0.001$). MRS for the Flanked-Local Group changed from 204 to 221 wpm, which is not statistically significant (a change of 9.2%; adjusted $p=0.15$). Further between-group comparisons showed that the Flanked-Local Group had marginally less improvement in MRS than the other two Distributed groups (both of the adjusted $p=0.087$), but the two Distributed groups were not significantly different (adjusted $p=0.99$). Our design did not include a no-training control group, but control data were available from a similar training study (Chung et al., 2004). In their study, trigram-recognition training improved maximum reading speed by 41%, but without training, maximum reading speed increased by less than 10% (estimated from Fig. 9C in Chung et al., 2004). The improvement for the Flanked-Local Group was similar to their no-training control group.

However, we again have the concern that the difference in the pre-test level may produce misleading group differences. We therefore performed a linear regression of the pre-post difference in log-MRS against the pre-test log-MRS level. This time, the improvement did not depend on the pre-test level, i.e. the slope of the regression line is not significantly different from zero ($p=0.44$). This suggests that higher pre-test performance level cannot account for the lack of improvement in the Flanked-Local Group.

Summary of Results

We used 3 different training paradigms where the training tasks differed in a systematic way, in order to test the influence of 1) spatial distribution of training stimuli, and 2) the presence or absence of flanking letters. As observed in Table 3-2 and summarized from the above analyses, we found that:

1) The spatial extent of crowding only decreased in the two Flanked groups but not in the Isolated group, and the reduction was larger in the Flanked-Local Group than the Flanked-Distributed Group. This suggests that crowded training stimuli are necessary in order to reduce the spatial extent of crowding. Moreover, when all crowded training stimuli are located at one single location, the training effect for that location is larger than when crowded stimuli are distributed across various spatial locations. This reflects some degree of retinotopic specificity in the improvement.
2) The size of the trigram visual span improved for all three groups, and the enlargement was smaller for the Isolated group. This suggests that the visual span enlarges after practicing with letter recognition, no matter flanked or isolated, distributed or localized, although the use of flanked training stimuli may result in a larger improvement.
3) Maximum reading speed improved in the two Distributed groups, and the Local group showed small changes similar to a no-training group. This suggests that the transfer of training to reading speed is limited when training stimuli have fixed spatial locations. For the conditions we have tested, training with distributed targets is both necessary and sufficient for such transfer to occur. Conversely, the

presence of flanking letters is neither a necessary nor sufficient condition for the transfer.

Discussion

Spatial Distribution of Training Stimuli

Our major finding is that spatially distributed training stimuli appear to be necessary and sufficient for the training benefit to transfer from letter recognition to increased reading speed. This result is consistent with previous findings where maximum reading speed improved with distributed training (41% improvement with trigram measurement training, Chung et al., 2004) but not with localized training (7.2% change with "uncrowd" training similar to our Flanked-Local Group, Chung, 2007). In the following paragraphs, we will discuss how spatially distributed training differs from localized training in its influence on reading. We will focus on two aspects: the retinotopic area that receives stimulation, and the deployment of attention during the task.

Maximum reading speed is achieved with large character sizes exceeding the critical print size, as shown in Figure 3-1E. Retinotopically, localized training only has limited spatial overlap with the words in larger print sizes. Retinotopic exposure is important because the effect of perceptual learning is sometimes specific to the trained location. It is possible that in order to better process larger-sized text, the corresponding retinal area has to be stimulated. In localized training, stimuli are constrained to a very small horizontal span, possibly making it less advantageous for reading larger-sized text compared to distributed training. However, insufficient retinotopic exposure cannot fully explain the lack of improvement, because letter-recognition training in the upper (or lower) hemi-field has been shown to improve reading speed in the untrained hemi-field (for example, Chung, Legge, & Cheung, 2004; He et al., 2013; Lee, Kwon, Legge, & Gefroh, 2010; Yu, Legge, Park, Gage, & Chung, 2010). This suggests a higher-level, non-retinotopic mechanism underlying the lack of improved maximum reading speed in the present study.

A more probable explanation for the lack of benefit of localized training is the deployment of visual attention. Localized training may facilitate the deployment of attention to one specific, small area in the visual field. For instance, in a visual search task, if the target more frequently appears in a certain spatial area, attention will be biased towards that area (Geng & Behrmann, 2002). It only takes dozens of trials to learn such bias, but once learned, the bias is not easily unlearned and can last for at least a week (Jiang, Swallow, Rosenbaum, & Herzig, 2012). Similarly, during localized training such as for our Flanked-Local Group, subjects may have acquired a sustained bias of spatial attention towards the letter position right below fixation. When tested with a reading task in the post-test, such an attentional bias may have persisted, limiting the spread of attention to the letters of longer words and negating potential benefit from the training.

Moreover, additional spatial bias of attention may have been introduced by the partial report method we used. In our training paradigm, since only one (middle) letter needed to be reported, attention to the flanking letters was likely diminished. This bias again is a disadvantage for reading where attention needs to be distributed to multiple letters simultaneously. Consistently, the improvement of reading in our study (about 30%) was smaller than in some similar studies where full-report trigram training resulted in 40%-66% improvement in the visual field being trained (Chung et al., 2004; He et al., 2013; Lee et al., 2010; Yu, Legge, et al., 2010).

Impaired visual spatial attention has been associated with reading deficits in people with dyslexia. For example, Schneps and colleagues proposed that in dyslexic reading, the mismatch between sluggish attention and the rightward shift of gaze led to crowded perception and more regressive saccades and thereby slowed down reading (Schneps et al., 2013). Together with our results, these pieces of evidence indicate an important role of attention in reading.

Crowding

We found that the presence of crowding in training was not necessary for the enlargement of visual span or the improvement of reading speed. When training with isolated letters, the spatial extent of crowding did not change, whereas reading and visual span both improved. Given the association between crowding, visual span and reading outlined in the Introduction, this result seems surprising.

But is it true that training with isolated letters did not reduce the impact of crowding? Maybe not. There is indeed evidence that in adults with amblyopia, training with isolated, near-acuity, reduced-contrast letters reduced the spatial extent of foveal crowding (Chung, Li, & Levi, 2012). Amblyopic vision share many similarities with normal peripheral vision, including the way they are influenced by crowding (for example, see a small review section in Levi et al., 2007). Therefore, training with isolated letters may have the potential to reduce the spatial extent of crowding in normal peripheral vision.

The reason for the lack of reduced spatial extent after our training with isolated letters may be the letter size we used. A difference between our training paradigm and the one used by Chung et al. (2012) is that they used near-acuity letters whereas our letter size is well above the acuity limit. Smaller-sized letters contain more higher spatial-frequency components compared to larger letters. Since human contrast sensitivity shifts towards higher-frequency when identifying crowded letters (Chung & Tjan, 2007), smaller-sized letters may be more beneficial for crowded letter-recognition. Therefore, training with near-acuity isolated letters can reduce the spatial extent of crowding whereas training with larger-sized letters may not yield statistically significant improvement, as observed in our study. However, the isolated-letter training in our study may still induce some latent improvement, which forms the foundation of enlarged visual span and improved reading speed.

Linking Crowding, Visual span, and Reading

Now we return to the link between crowding, visual span, and reading.

The link between crowding and visual span seems to break when training with distributed, isolated letters, where the spatial extent of crowding remained unchanged but visual span enlarged. As we previously discussed, in the case of the Isolated-Distributed

Group, training may have reduced the impact of crowding without inducing measurable changes in its spatial extent. Reduced impact of crowding, rather than reduced spatial extent of crowding, was necessary for the enlargement of visual span. The link between crowding and visual span still holds.

The link between visual span and reading seems to break when training with localized, crowded letters, where visual span enlarged but maximum reading speed remained unchanged. We proposed that a disadvantageous attentional bias towards a small area limits the improvement in reading. During rapid sequential presentation of words, such a bias will likely direct attention to a very narrow region within the word, resulting in recognition of only a limited number of letters under time pressure. In contrast, during the measurement of visual span, no matter where the trigram appears, the abrupt and brief appearance of the trigram is a strong exogenous cue. This cue will automatically disengage attention from the previously prioritized location and reorient it to the location of the trigram. The low task demand (only three letters need to be reported rather than a sequence of words) and the unlimited response time further alleviate potential impact of the attentional bias. Therefore, the attentional bias slows down reading but only has minimal impact on visual span. Reading is still limited by the size of the visual span, but the spatial distribution of attention also needs to be taken into account.

We conclude that the visual span represents a sensory bottleneck on reading, but there may also be an attentional bottleneck. Crowding limits reading via limiting the size of the visual span. Reducing the impact of crowding can enlarge the visual span and can potentially facilitate reading, but not when adverse attentional bias is introduced. Our results thus clarify the association between crowding, visual span, and reading.

Chapter 4. Training With A Game: More Fun But Not More Improvement

Introduction

Age-related macular degeneration (AMD) is an important cause of vision loss in the United States, and affects approximately 6.5% of the US population aged 40 years and older (Klein et al., 2011). In advanced cases of AMD, the central visual field is impaired, often bilaterally. The impairment not only poses difficulties in a person's daily life, but also impacts the person's psychological well-being (Mitchell & Bradley, 2006). Reading, being one of the most important daily activities, requires the ability to distinguish fine details and to make appropriate saccadic eye movements along the lines of text. These two skills are highly challenging in peripheral vision, making reading extremely hard for some AMD patients, discouraging them from reading on a regular basis. This motivates the development of reading rehabilitation strategies for people with AMD.

Compared to central vision, the slower reading speed in peripheral vision was hypothesized to result from a smaller size of the visual span, which refers to the number of text letters that can be reliably recognized during reading without eye movements (Legge et al., 2001). The size of the peripheral visual span can be measured by presenting triplets of letters (trigrams) at different horizontal locations in peripheral vision, for example 10° in the lower visual field (Fig. 3-1A). The area under the resulting recognition accuracy-letter position profile is the size of the visual span (Fig. 3-1B).

Perceptual training tasks, usually reading letters or words displayed in peripheral vision, can increase the size of the peripheral visual span for normally-sighted young adults, accompanied by an improvement of at least 40% in reading speed(Chung et al., 2004; He et al., 2013; Lee et al., 2010; Yu, Legge, et al., 2010). Elder adults within the age range of the onset of AMD can also benefit from this type of training, achieving a 60% improvement in reading speed (Yu, Cheung, et al., 2010). Recently, similar training has also been applied to people with central-field loss and successfully improved their reading (Calabrèse et al., 2017; Chung, 2011; Nguyen et al., 2011).

Although perceptual learning appears to be an effective training method in the laboratory, it still faces the problem of unsatisfactory compliance. The nature of perceptual learning is to repeat the same task intensively for a long time, which can discourage patients from continuing training. There are instances where subjects withdraw from the training due to boredom, as reported by e.g. Chen, Chen, Fu, Chien, & Lu, 2008 and noted in our lab's subject database. In the domain of public health, non-compliance is a widely-recognized problem for many long-term therapeutic procedures, such as physical exercise (Middleton, 2004), prescription of drug or supplement (Hubbard, Elia, Holdoway, & Stratton, 2012; Maningat, Gordon, & Breslow, 2013), or cognitive behavior therapy (Matthews, Arnedt, McCarthy, Cuddihy, & Aloia, 2013), even when the benefits from those treatments are known and well-established. Improving compliance is thus of great interest in order to maximize therapeutic effects and to minimize unnecessary healthcare costs.

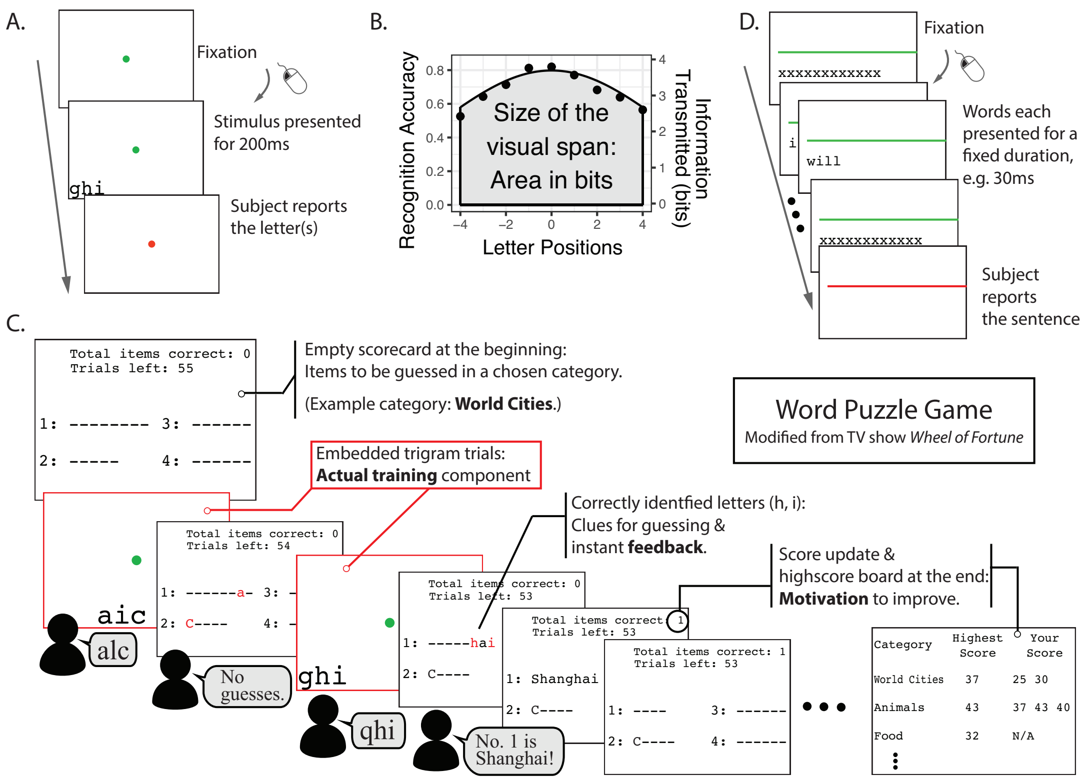

Figure 4-1. Three tasks in the experiment.

A. Letter-recognition task. A letter or a trigram (three adjacent letters) appeared briefly in the upper or lower visual field (shown here is a trigram in the lower visual field). The task is to identify the letters in order. **B.** Calculation of the size of the visual span. **C.** Word-puzzle game based on the trigram task (see Methods). For demonstration clarity, only a few items are shown here in the scorecard and the highscore board. In the real game the subjects saw a scoreboard with 10 items to be guessed, and all 29 categories in the highscore board. The letter recognition trial was the same as in the trigram task, but the fixation screen (a single green dot) and the post-trial screen (a single red dot) are omitted in this demonstration for clarity. **D.** RSVP reading task. A sentence was displayed word-by-word at 10° in the upper or lower visual field (shown here is lower visual field), preceded and followed by masks. The task was to read the sentence out loud, and the number of correctly-read words is recorded.

Video-game training may help to address the problem of compliance in procedures involving perceptual learning. For normally sighted subjects, research has shown that video-game playing can not only elevate subjects' interest in the task, but also yield improvements on various visual skills, from low-level contrast sensitivity to high-level visual attention (for a review, see Bavelier & Green, 2012). Compared to perceptual training on basic tasks, the advantages of using video games are the engaging experience and the non-specific training effects. Researchers have attempted to use video games as a new visual rehabilitation method for visual impairment, such as for amblyopia (Bayliss, Vedamurthy, Nahum, Levi, & Bavelier, 2013; Holmes et al., 2016; Kelly et al., 2016; Li, Ngo, Nguyen, & Levi, 2011; To et al., 2011).

To the best of the authors' knowledge, there is no video-game training tested with AMD subjects yet. So far, training with video games has been typically tested with young adults, and the games have often included first-person shooting games or non-action puzzles. It is doubtful whether elder subjects enjoy these types of video games as much as younger adults, especially the fast-paced action games. But there is evidence that for elder subjects, playing video games can elevate their engagement, no matter whether the game was an action game or a puzzle game (Belchior, Marsiske, Sisco, Yam, & Mann, 2012). They retain adequate brain plasticity to benefit from the game training (Belchior et al., 2013).

Recently, a game-like training interface was designed for people with central-field-loss to induce an eccentric retinal location (Preferred Retinal Locus, PRL) for fixation (Liu & Kwon, 2016). Normally-sighted subjects were tested, and the training successfully improved oculomotor control as well as the adoption of PRL-like nonfoveal behavior in letter and sentence reading.

Inspired and encouraged by past gaming studies, our research goal is to improve reading in peripheral vision by embedding an existing training procedure in a game context. The basic procedure of the game is illustrated in Figure 4-1C. Briefly speaking, we designed a word-guessing game similar to the TV show *Wheel of Fortune*. In our game, the goal is to guess as many words in a given category (for example, world cities) as possible with a limited number of trials. A trial begins with a trigram-recognition task similar to the visual span measurement. Correctly recognized letters would then become clues for the word-guessing game. In order to guess more words, subjects are motivated to recognize more letters to collect as many clues as possible. Presumably, the guessing part will make the training more enjoyable and can engage the subjects for longer periods of training.

We tested the game-training with young, normally-sighted subjects to see whether their peripheral visual span and reading speed would improve, and how the improvement compares to a non-game training. We hypothesize that the enjoyment associated with the game will enhance the motivation to learn, which may result in greater training benefits.

METHODS

Subjects

5 male and 7 female college students were recruited from the University of Minnesota and randomly assigned to 2 groups-- a Short-game or a Long-game group (described later in "Procedure"). Our analyses will also include data collected by He et al. (2013), where 7 male and 5 female college students were recruited and randomly assigned to a No-game group or a Control group (originally called the "training group" and the "control group" by He et al. (2013)). All participants were native English speakers and had normal or corrected-to-normal vision. Their binocular acuity (Lighthouse Near Acuity Chart, Lighthouse Low Vision Products, Long Island City, NY), reading performance (MNREAD, Precision Vision, La Salle, IL) and other information are given in Table 4-1. The protocol was approved by the Institutional Review Board and was in compliance with the Declaration of Helsinki. All subjects gave informed consent prior to the experiment.

Stimuli and Apparatus

Stimuli consisted of black lowercase letters on a white background (background luminance 102 cd/m^2; Weber contrast = 98%). All stimuli were viewed binocularly from 40 cm in a dark room. The letters were rendered in Courier font with an x-height of 3°, and letter spacing was 1.16× x-width (standard spacing, 4.06°). This letter size was chosen because it exceeds the critical print size at 10° in peripheral vision (Chung et al., 1998), and thus letter size did not limit reading speed .

The stimuli were generated and presented using MATLAB R2010a with Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). We used a NEC MultiSync CRT monitor (model FP2141SB-BK, NEC, Tokyo, Japan; refresh rate = 100 Hz; spatial resolution = 0.04°/pixel) controlled by a Mac Pro Quad-Core computer (model A1186, Apple Inc., Cupertino, CA). Viewing distance was maintained using a chin rest and the subject's fixation was monitored using a webcam.

Group	Gender ratio (M:F)	Age	Visual acuity (LogMAR)	MNREAD Measurements Reading Acuity (LogMAR)	Critical Print Size (LogMAR)	Maximum Reading Speed (wpm)
Short-game	1:5	20.8 ± 1.1	-.09 ± .007	-.40 ± .02	.02 ± .02	205 ± 5
Long-game	4:2	19.2 ± .5	-.08 ± .008	-.35 ± .02	.07 ± .04	214 ± 16
No-game	4:2	19.6 ± 2.1	-.07 ± .02	-.38 ± .04	.09 ± .03	205 ± 11
Control	3:3	22.4 ± 2.1	-.07 ± .01	-.25 ± .04	.17 ± .03	191 ± 6
All groups	12:12	20.5 ± .5	-.08 ± .007	-.34 ± .02	.09 ± .02	204 ± 5

Table 4-1. Subject groups' information in Chapter 4 (mean ± SEM).

Reading speed – print size curves from MNREAD measurements were fitted on a log-log scale with the following function using non-linear least squares fitting to determine maximum reading speed and critical print size:

$$RS = \mathrm{MRS} \times (1 - e^{-e^{lrc}(x - x_{int})})$$

RS, reading speed; *MRS*, maximum reading speed; *lrc*, parameter describing the elbow of the curve; *x*, print size; x_{int}, reading acuity. Critical print size was the print size where reading speed reaches 85% of the maximum reading speed.

Procedure

The experiment consisted of three parts: pre-test, training, and post-test (Fig. 4-2). In the pre- and post-tests, subjects' visual span profiles (using single letters and trigrams) and reading speeds (using ordered and unordered text) were measured. The sequence of these tests was single-letter visual span measurement (4 blocks), trigram visual span measurement (8 blocks), RSVP reading with ordered text (4 blocks) and with unordered text (4 blocks). Half of the blocks were tested at 10° in the upper visual field and half at 10° in the lower visual field, interleaved. The order of the two visual fields, i.e. upper-lower or lower-upper, was kept the same in the pre- and post-tests for each subject but counterbalanced between subjects.

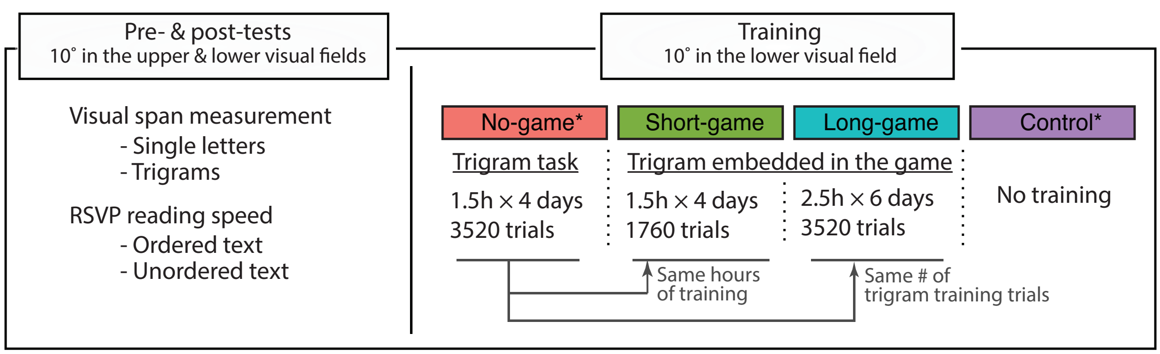

Figure 4-2. Experimental procedure.

* Data from He et al., 2013

Training always happened at 10° in the lower visual field. The training task was either the trigram letter-recognition task by itself (hereafter referred to as "trigram training"), or the trigram task modified to include the game (hereafter referred to as "game training"). There were three different training groups: No-game, Short-game, and Long-game groups. Subjects in the No-game group received 4 daily sessions of trigram training, each session consisting of 16 blocks and lasting about 1.5 hours (data obtained from He et al., 2013). The Short-game group also underwent 4 daily sessions of game training, matching the total training time of the No-game group (6 hours). However, since the game contains both the training component (the presentation of the trigram) and the gaming component (word-guessing, described later), the time to complete one game trial was twice as much as to complete one non-game trial. Therefore, the Short-game group only had half the number of training trials as the No-game group (8 blocks/day). We then included a Long-game group to match the total number of training trials of the No-game group (3520 trials) with the training time doubled. Their training lasted for 6 days and had a total of 64 training blocks. Finally, we included data from a control group where no training occurred during the corresponding training time (data obtained from He et al., 2013).

In the following paragraphs, we will first describe the trigram letter-recognition method used in measuring the visual span and our modification of it to include the "Wheel-of-Fortune" game. We will then describe the method for measuring reading speed.

Visual Span Measurement

In order to measure the size of the visual span, we performed a letter-recognition task, as described in He et al. (2013). Figure 4-1A illustrates the basic procedure of a trial:

A subject fixated on a dot and clicked the mouse to initiate a trial. Stimuli were randomly-chosen lowercase letters, either in isolation or arranged as triplets (trigrams). Letter(s) appeared for 200ms for the subject to identify. For a letter in a trigram to be correct, both its identity and location need to be correct. The letters were positioned in slots that were horizontally arranged on an imaginary line at 10° in the lower (or upper) visual field. The slot centering on the fixation midline was labeled 0, and left and right slots were labeled with negative and positive numbers respectively. The center-to-center spacing between adjacent slots was 1.16 × x-width, corresponding to standard spacing in the Courier font. In each block, trigrams were centered 5 times on each slot from -5 to 5. This means that a total of 165 letters were presented during a single block. Since the letters in slots ±5 (10 letters/slot) and ±6 (5 letters/slot) were less than the other slots (15 letters/slot), only data from slots -4 to 4 were used in further analysis.

To get the visual span profile, letter-recognition accuracy was plotted against letter positions (Fig. 4-1B). We converted letter-recognition accuracy to information transmitted in bits using the formula

Information transmitted in bits = -0.036996 + 4.6761×letter recognition accuracy,

where chance level performance corresponds to 0 bits of information and 100% accuracy corresponds to about 4.7 bits. The number of bits was added for slots -4 to +4 to estimate the size of the visual span.

Training

The trigram training used for the No-game group was exactly the same as the trigram measurement. The game training used for the Short- and Long-game groups was modified from the trigram training. We modeled our game on the popular American TV show *Wheel of Fortune*, where the goal is to guess items from a given category using letter clues. We developed word lists in 29 different categories for the subjects to choose from, including Male actors, Female actors, American cities, World cities, Animals, Sports, Plants, Instruments, Movies, Jobs, Household items, Food, Cars, Academic fields, Politics, Summer related, Adjectives for people, Business, Characters from children's story and cartoons, Characters from comics and superheroes, Computer software, Dance, Drugstore items, Makeup and skin care, Psychology, Rocks and minerals (easy), Rocks and minerals (expert), Sport teams (MLB, NBA, NFL, NHL), and TV series. For a complete list of words, see Appendix 4.

The procedure of the game is demonstrated in Figure 4-1C. The subject first picked a favorite category of words to guess, say, world cities. The game block started with a scorecard with 10 empty entries, randomly chosen from the unguessed words in the specified category. In the program, these 10 words were then concatenated to form a string, with the ending of the last word connected back to the first word. Three adjacent letters were picked randomly from this string, converted to lowercase, and presented as a trigram to the subject for 200ms. In the given example in Figure 4-1C, the first two words were actually Shanghai and Cairo, but at the beginning the subject only saw empty entries “--------“ and “-----“. The concatenated hidden string was “…ShanghaiCairo…”. Three adjacent letters “aic” was chosen from the string and presented to the subject for recognition. The subject reported “alc”, and therefore only got 2 of 3 letters correct. The correct letters a and c were then shown on the scorecard, resulting in “------a-“ and “C----“. At this point, the subject did not have enough clues to guess any words and decided to move on to the next trial. The correct letters a and c were removed from the hidden string to form a new string: “…Shanghiairo…”. Again, three adjacent letters “ghi” were chosen and presented, and the subject reported “qhi”. This time, only the last two letters h and i were correct. Now the first entry became “-----hai”, and the subject was confident to guess “Shanghai” for this entry. This was a correct guess, so a cheerful tone was played, one point was added to the “Total items correct” count, the entry was fully revealed on the screen, and a new empty entry (“Rome”) replaced the old one. In the hidden string, all the letters corresponding to the old entry were removed, and the letters from the new entry were inserted, becoming “…Romeairo…”. In the case of a wrong guess, a punishing tone was played and the correct entry was also revealed and replaced. In a block of 55 trials, the more letters the subject correctly recognized, the more clues were available. At the end of each block, a score board appeared, showing the highest score (most items guessed) in each category achieved by any subject, as well as the current subject’s personal records.

Reading Measurement

Reading speed was measured using Rapid Serial Visual Presentation (RSVP, Forster, 1970; Rubin & Turano, 1992). For each RSVP trial, a sentence was randomly chosen from a pool of 847 sentences for testing. A subject never saw the same sentence twice. The average sentence length was 11 words, ranging from 7 to 15 words. The word length averaged 4 letters, ranging from 1 to 12 letters. In an RSVP trial, the sentence was presented word-by-word, preceded and followed by masks of "xxxxxxxxxx" (Fig. 4-1D). Subjects were asked to fixate on a line without making vertical eye-movements, but horizontal eye-movements along the line were permitted. During a trial, words were presented 10° above or below the fixation line and were left-aligned with the left edge of the fixation line. Subjects read the words out loud and the experimenter counted the number of correctly-read words. The sentence was either ordered (words were presented in their original, meaningful order) or unordered (scrambled word order). Ordered text provides sentence context information and benefits reading. By comparing reading speed of ordered and unordered text, we were able to evaluate how subjects utilize context information during reading.

In each block of 36 trials, 6 different word exposure durations were tested randomly (6 times each). Depending on individual subject's performance during practice, one of two duration sets could be chosen: {30, 53, 93, 164, 290, and 511ms}, or {53, 93, 164, 290, 511, and 900ms}. The choice of set maximally ensured 1) greater than 80% word recognition accuracy for the longest duration in the set, 2) close to chance-level word recognition performance for the shortest duration in the set, and 3) good eye fixation for all durations (fixation becomes poorer as duration becomes longer). The resulting accuracy – duration curve was then fitted with a psychometric function, and the subject's reading speed was calculated using the exposure duration yielding 80% accuracy of word recognition.

Statistical Analysis

When comparing training effects between groups, we first computed individual changes after training in the appropriate unit (i.e. % change in wpm for reading speed, # of bits change for visual span). If not otherwise specified, we performed 2×2×4 mixed-design ANOVAs, with session type (pre-/post-test) and visual field (trained lower/untrained upper) being the within-subject factors, and group (No-game/Short-game/Long-game/Control) as the between-subject factor. In our model, we included the main effects of session type, visual field, and group, as well as an interaction term between group and session type. Significant interaction would indicate that training has differential effects among groups, i.e. some groups improve more than other groups. To keep our model easy to interpret and powerful, we did not include other interaction terms. If a significant interaction between group and session type was found, we further analyzed the interaction using *R* with the package *phia* (Post-Hoc Interaction Analysis; Martínez, 2013). The reported p-values were adjusted for multiple comparisons.

Reading speed data (wpm) were log-transformed into log-wpm before the analyses in order to satisfy the normality assumption, and were transformed back to wpm when reporting results for easier understanding.

Supplementary Survey Experiment

A few months after collecting behavior data, we conducted a supplementary survey experiment. The purpose was to measure the subject's enjoyment when undergoing the training, their motivation to improve on the trained task, and their expected improvement after the training. A complete list of questions can be found in Appendix 5. We sent invitations to the 18 training subjects to participate in this follow-up experiment, but only 8 of them came back, including 2 from the Short-game group, 5 from the Long-game group, and 1 from the No-game group.

To measure subjective enjoyment, we chose 21 questions from the *eGameFlow* scale (Fu, Su, & Yu, 2009) and modified them to suit our task. These questions cover eight key factors of enjoyment regarding e-learning games, including concentration, goal clarity, feedback, challenge, autonomy, immersion, social interaction, and knowledge improvement. To measure subject's motivation to improve, we chose 9 questions from the Achievement Motivation Inventory (AMI, Schuler, Thornton III, Frintrup, & Mueller-Hanson, 2004), each question representing one facet of motivation (adopted from the method of Shaw, 2011) : compensatory effort, persistence, confidence in success, eagerness to learn, internality, goal setting, preference for difficult tasks, pride in productivity, and status orientation. The enjoyment and motivation questions were then combined into one survey of 30 questions. All survey items were scored on a 7-point Likert scale with 1 representing “strongly disagree” and 7 representing “strongly agree”.

The returning subjects first performed 2 blocks of a certain task, for example the trigram letter recognition task. After that, they completed the combined survey to evaluate their level of pleasure and motivation during that task. The same procedure was repeated for three tasks: trigram letter recognition, the game, and RSVP reading. The sequence of these three tasks was balanced between subjects using a Latin square.

RESULTS

This section is structured in three parts. In Part I, we will report the behavior results of visual span enlargement and reading speed improvement after training. In Part II, we will construct a model to decompose the magnitude of improvement into contributions from trigram training, gaming, and the influence of starting level. In Part III, we will briefly report our survey results.

Part I. Behavioral Results

Visual Span

Figure 4-3A shows the visual span profiles for single letters (dark gray) and trigrams (black). From the profiles, the absolute values of single letter visual spans did not change much from the pre-test to the post-test for any group. In contrast, trigram visual span profiles increased in height and became broader in the post-test for all three training groups, in both the trained (lower) and untrained (upper) visual fields. No improvement occurred for the control group.

The size of the visual span was computed as the area under each visual span profile in bits of information transmitted (Figure 4-3B; see Methods). A 2×2×4 mixed-design ANOVA was performed, with session type (pre-/post-test) and visual field (trained lower/untrained upper) as the within-subject factors, and group (No-game/Short-game/Long-game/Control) as the between-subject factor.

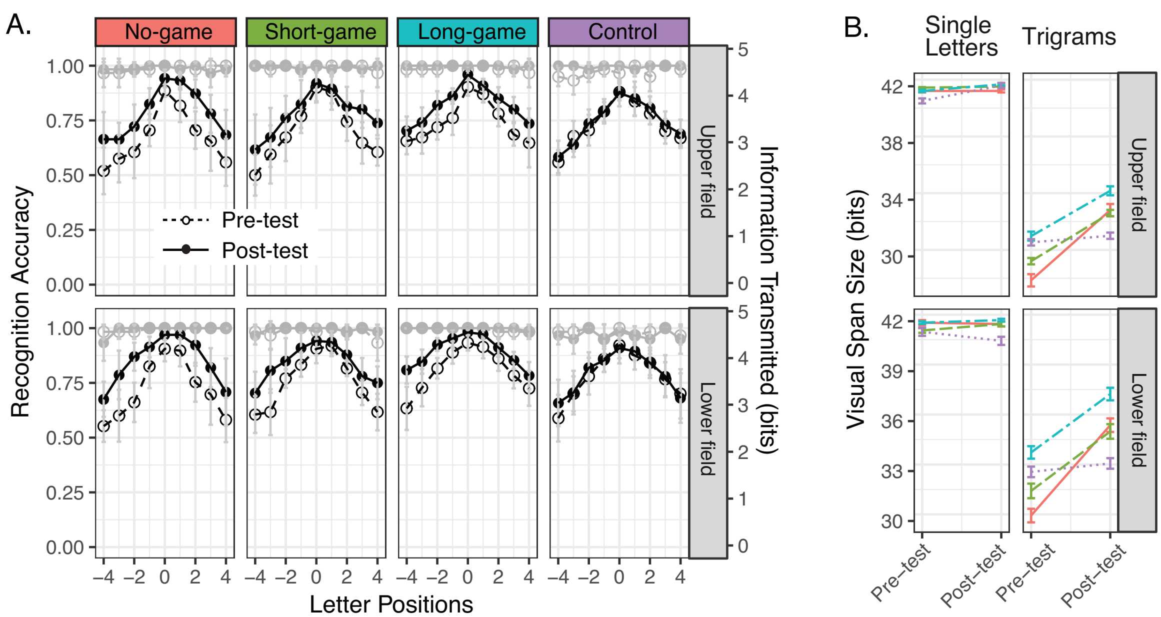

Figure 4-3. Visual span enlargement after training.

A. Visual span profiles. Dark gray: single letters. Black: Trigrams. Dashed lines: pre-test profiles. Solid lines: post-test profiles. Error bars: ±1 SEM. **B.** Interaction plot of visual span size (bits). Red solid lines, No-game group. Green dashed lines, Short-game group. Blue dash-dot lines, Long-game group. Purple dotted lines, Control group.

For single letters, there were no main effects of session type, visual field, group, or the interaction between session type and group. Taken upper and lower visual fields together, visual span size increased (non-significantly) from 41.2 bits in the pre-test to 41.4 bits in the post-test, corresponding to a small change of average accuracy from 98.6% to 99.1%.

For trigrams, we found significant main effects of session type (pre-test: 30.8 bits; post-test: 33.9 bits; $F(1, 67)=92.15$, $p<0.001$) and visual field (lower field: 33.5 bits; upper field: 31.1 bits; $F(1, 67)=78.40$, $p<0.001$). No main effect of group was found, but there was a significant interaction between session type and group ($F(3, 67) = 13.00$, $p<0.001$). Further analysis of the interaction revealed that for all the training groups, visual span size increased significantly from the pre-test to the post-test (No-game: 5.2 bits; Short-game: 3.5 bits; Long-game: 3.4 bits; all of their adjusted $p<0.001$), but not the Control group (0.5 bit, adjusted $p=0.37$). Pairwise comparisons indicate that the comparison of enlargement is statistically significant when comparing any training group to the control group (all of these adjusted $p<0.001$) and when comparing the No-game group to the Long-game group (adjusted $p=0.049$), marginally significant when comparing the No-game group to the Short-game group (adjusted $p=0.05$), but not significant when comparing the Short-game to the Long-game group.

Taken together, these analyses indicate that:

1) In general, visual span in the upper visual field is smaller than that in the lower visual field.

2) All three types of training enlarged trigram visual span in the trained (lower) field, and the training effect transferred to the untrained (upper) field, with a mean transfer ratio of 92%.

3) The improvement with training is significantly larger than no training, and the no-game training produced larger improvement compared to the game training.

Reading Speed

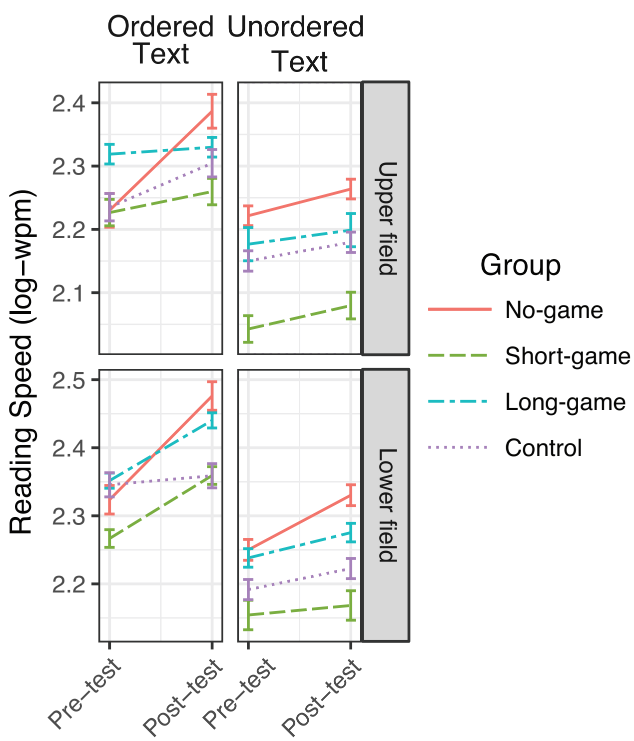

Figure 4-4: Interaction plot of reading speed.

Figure 4-4 shows the reading speed before and after training for the 4 groups, tested using both ordered and unordered text. A 2×2×4 mixed-design ANOVA was performed (session type × visual field × group) on reading speed of ordered and unordered text respectively. Additionally, we compared the context gain (the ratio between reading speeds for ordered and unordered text) in the pre- and post-tests. Compared to unordered text, ordered text provides extra context information and thus increases reading speed. By such a comparison, we evaluated whether the subjects were able to utilize context information better after training.

For ordered text, main effects of session type (pre-test: 194 wpm; post-test: 231 wpm; $F(1, 67)=24.44$, $p<0.001$) and visual field (upper field: 193 wpm; lower field: 232 wpm; $F(1, 67)=25.43$, $p<0.001$) were found, as well as a significant interaction between session type and group ($F(3, 67)=2.80$, $p=0.046$). Further analysis of the interaction revealed that among the 4 groups, only the No-game training produced significant improvement after training. But in spite of the p-values, the three training groups all exhibited some improvement in the trained field (No-game: 45%; Short-game: 25%; Long-game: 23%), compared to 5% improvement in the control group. The average transfer rate of training effect to the untrained field was 90%.

For unordered text, main effects of visual field and session type were found, but there was no main effect of group or any interaction between session type and group. Reading speed was faster in the lower field (169 wpm) compared to the upper field (146 wpm, $F(1, 67) = 48.50$, $p < 0.001$), and faster in the post-test (164 wpm) compared to the pre-test (151 wpm, $F(1, 67) = 4.93$, $p = 0.030$). Despite being statistically significant, the improvement was rather small compared to ordered text.

We then computed the context gain (the ratio of reading speeds for ordered and unordered text) for all the subjects. Across all subjects, visual fields, and testing sessions, 95% of the time the context gain was greater than 1, confirming the benefit of sentence context on reading. A similar 2×2×4 ANOVA on context gain revealed only a main effect of session type ($F(1, 67) = 5.82$, $p = 0.019$): Context gain increased from 1.31 in the pre-test to 1.44 in the post-test, and there was no statistically-significant difference between groups or between visual fields. This suggests that all subjects were better at utilizing context information in the post-test, regardless of the training they received. One thing to note is that the change in context gain for the Control group (+0.04) appears to be smaller than that for the training groups (average +0.16), but we have insufficient statistical power to conclude that there is a significant difference between groups.

Taken together, these results indicate that:

1) Reading speed in general is slower in the upper visual field compared to the lower visual field, and slower when measured with unordered text compared to ordered text.
2) Only the No-game training significantly improved reading ordered text, but the two game-training groups also yielded some level of improvement.
3) In the post-test, subjects were better at utilizing sentence context information during reading.

Connecting visual span and reading

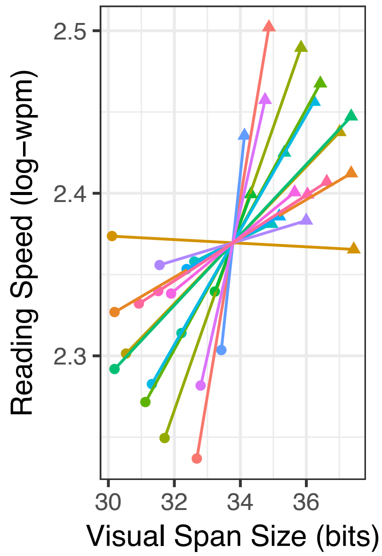

Figure 4-5. Correlation of reading speed and visual span size associated with training.

Each line indicates the change from pre-test (circles) to post-test (triangles) for an individual subject. The centers of all the lines were aligned to the mean value of all the subjects for clearer visual demonstration. The average slope of all the lines is 0.041 ± 0.08 log-wpm/bit (or 10% improvement in wpm per bit).

The size of the visual span has been found to tightly correlate with reading speed, and 1 bit enlargement of visual span on average corresponds to an increase in reading speed of 0.03 log-wpm (or 7% improvement in wpm; Legge et al., 2007). We plotted log-transformed reading speed (log-wpm) against the size of the visual span (in bits) in the trained lower visual field for all the subjects from the training groups (Fig. 4-5). Each line segment indicates one subject and it has two data points corresponding to the pre-test (circles) and the post-test (triangles). To better illustrate the distribution of the slopes of the lines, the centers of all the segments were aligned to the mean value of all the subjects.

These lines have similar slopes, and the average slope was 0.041 ± 0.008 log-wpm/bit (or 10% improvement in wpm per bit). Our finding here further supports an invariant relationship between the size of the visual span and reading speed, and thus supports the visual span hypothesis that the visual span is a sensory bottleneck on reading.

Part II. Modeling

Rationale

We were surprised that the game seemed to produce less improvement than the no-game training. To better understand what happened during the process of training, we computed the average recognition accuracy for all the trigram blocks in the lower field during the pre-test, training, and post-test for the three training groups.

Figure 4-6A plots the average trigram-recognition accuracy against block number for each training group. The first and last four blocks are the pre- and post-tests, respectively. Three patterns can be observed here. First, the overall performance was the highest for the Long-game group, followed by the Short-game group, and then by the No-game group. Lower starting level in the No-game group may have left larger room for improvement compared to the two gaming groups. Second, the learning curve was the steepest at the beginning of training but became less steep as training progressed. Third, for the two gaming groups, there was a sudden change in performance at the transition from the pre-test to training, and from the end of training to the post-test. Recall that the game was played only during training, and not during the pre- or post-tests. From the end of the pre-test to training, the addition of the game suddenly boosted performance, whereas from the end of training to the post-test, the removal of the game yielded a sharp drop in performance. It is possible that the game enhances performance only when the game is being played but is not sustained when the game is absent.

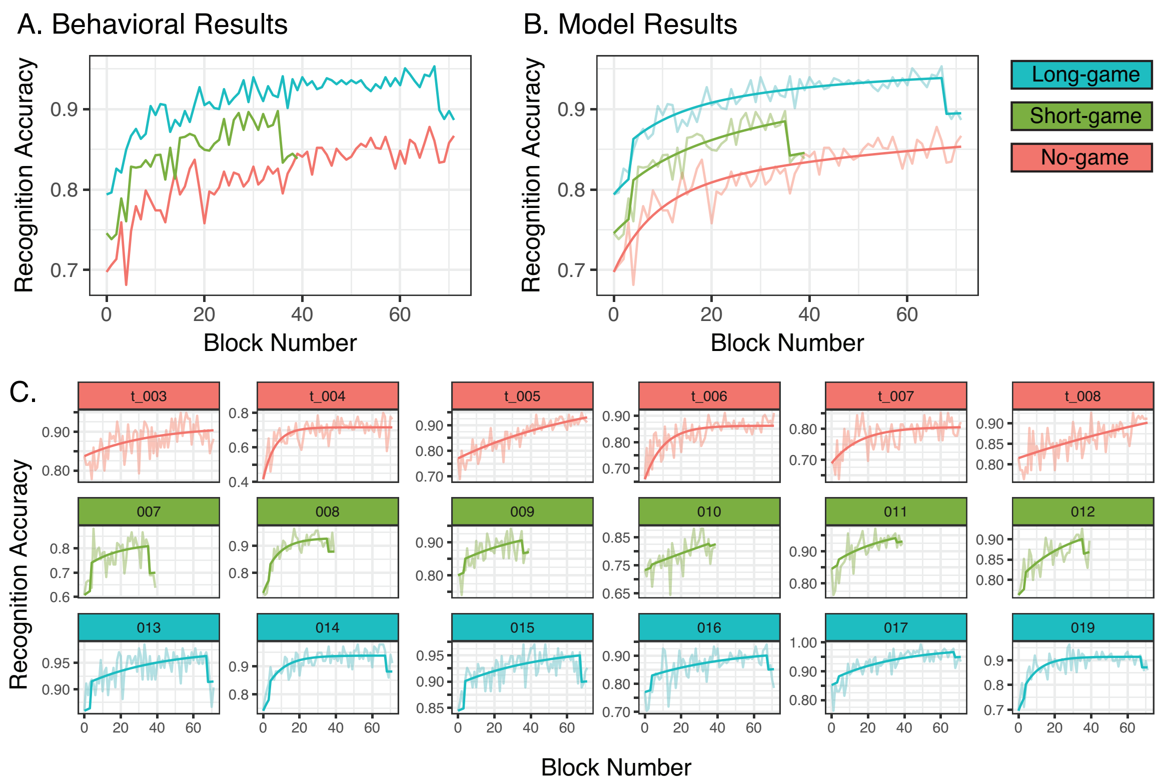

Figure 4-6. Training progress and model prediction.

A. Group-averaged recognition accuracy against block number. The first and last four blocks are the pre- and post-tests, respectively. **B.** Group-averaged training progress overlaid with averaged model fitting. **C.** Individual fitting of training progress. Red curves, No-game group. Green curves, Short-game group. Blue curves, Long-game group. In panels B and C, fitted curves have full contrast and raw data have reduced contrast.

Model description

Based on these three observations, we built a model to describe the training progress:

$$x_n = x_1 + k * (1 - x_1) * \frac{1 - \lambda^{n-1}}{1 - \lambda} + Game$$

Briefly speaking, the performance for the nth block, x_n, is determined by three factors: Performance (proportion correct in letter recognition) on the first block (x_1), cumulative improvement from all previous blocks, ($k * (1 - x_1) * \frac{1-\lambda^{n-1}}{1-\lambda}$), and a "Gaming boost" that is only present for the gaming blocks (*Game*). The improvement was always limited by the ceiling value of 1, that is, 100% letter recognition. We assume that the initial improvement from block 1 to block 2 is proportional to the difference ($1 - x_1$) with k as a factor describing the proportionality. The block-to-block improvement decreases by a factor of λ (1 > λ > 0) as training progresses. For example, for a no-game block, assume the starting level x_1 is 0.6, k is 0.02, and λ is 0.95. From block 1 to block 2, the improvement is $k * (1 - x_1) = 0.008$, and $x_2 = 0.608$; from block 2 to block 3, the improvement decreases to $k * (1 - x_1) * \lambda = 0.0076$, and $x_3 = 0.6156$; and so on.

We used non-linear mixed-effect (NLME) modeling to estimate the values of k, λ and *Game*. Initial examination of the data revealed that there was no need to include a fixed effect of group in the model, i.e. there was no systematic difference between groups. Therefore, our model only included a random effect to describe the between-subject variance. We fitted individual curves of training progress (Fig. 4-6C) and then averaged them to get the group curves (Fig. 4-6B). Overall, the estimated values (mean ± standard error) were $k = 0.0278 \pm 0.0055$, $\lambda = 0.9498 \pm 0.0106$, and $Game = 0.0455 \pm 0.0077$. This means that 1) the average initial improvement from block 1 to block 2 was about 2.8% of the room left for improvement (i.e. the difference between the starting accuracy level and the ceiling value of 1.0, $(1 - x_1)$), 2) the improvement on average decreased by a factor of 0.95 every block, and 3) the game context boosted the recognition accuracy up by an extra 4.55%, regardless of the block number.

The estimated between-subject standard deviation (i.e. the random effect in the NLME model) was 0.02 for k, 0.04 for λ, and 0.03 for the Gaming Boost. The variability was relatively large, which is also apparent from the individual fitted curves (Fig. 4-6B): for some subjects the training curves rose fast and reached a plateau in an early stage of the training, whereas for some other subjects the training curves had a shallower initial slope but kept increasing throughout the training period. The Gaming Boost also differed widely across subjects from the gaming groups, ranging from 1% to 11%.

Despite individual differences, there were no systematic differences between groups. Our modeling results suggest that the core component of the game training, which is the recognition of trigrams, produces the same effects as the no-game trigram training. The game can enhance performance only in the context of the game, but the quality of the training is not altered.

Part III. Survey

Since we only have data from 8 subjects for the survey experiment, we will only report descriptive results. Here, we focus on the difference between the trigram training and the game training.

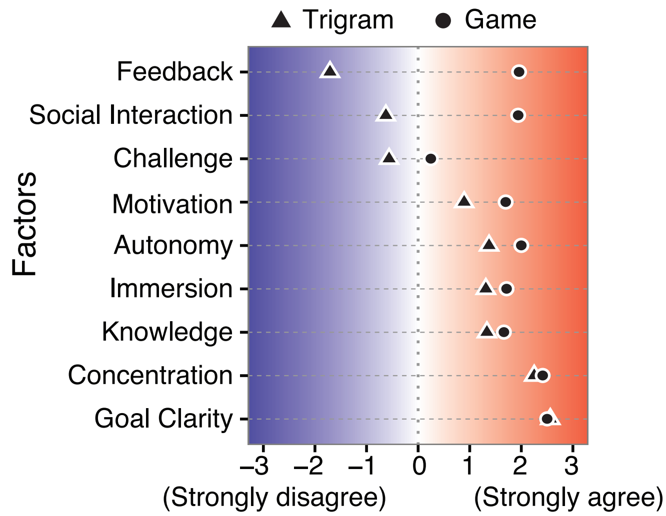

Figure 4-7: Survey result.

Triangles: evaluation of the trigram training. Circles: evaluation of the game training. From -3 to 3, the scale corresponds to the levels from "strongly disagree" to "strongly agree" in the survey.

Figure 4-7 summarizes subjects' enjoyment and motivation when they are performing these two tasks (see Appendix 5 for a detailed graph and a list of individual questions). The enjoyment evaluation was summarized by the 8 key factors (see Methods), and the 9 questions in the motivation survey were averaged as one "Motivation" factor. Comparing between trigram and game training, the largest difference occurred mainly in two factors: Feedback and Social Interaction. In the game training, subjects received clearer feedback and experienced competition with other subjects, which elevated the enjoyment of the task. The two types of training do not have notable differences in other factors of enjoyment or motivation. It is surprising to find that there is no difference in the Motivation factor. Retrospectively, we realized that since the "Motivation" survey was modified from the Achievement Motivation Inventory, it may only reflect a stable personal trait instead of a real-time motivation to complete the task. Overall, the survey revealed that the advantages of the game training are its immediate feedback on the subject's responses and its inclusion of a motivating competition between players. It might be helpful to include these key features in a clinical rehabilitation procedure to enhance compliance.

DISCUSSION

Game vs. Non-game Training

Our study shows that embedding the trigram training in a word-puzzle game can enlarge visual span, and the improvement largely transfers to the untrained visual field. The amount of enlargement for the gaming groups appeared to be smaller than that for the No-game group. The fact that game-training did not have greater training benefits than non-game training may seem surprising at first, but it is not inconsistent with previous literature. Using video games to improve visual skills has received a lot of attention recently, but most experiments focus mainly on 1) comparing gamers vs. non-gamers, 2) comparing action-video games vs. non-action video games, or 3) comparing game-training with a no-training control. To the best of the authors' knowledge, very few studies have compared the same training with and without a game. In one related study, Belchior et al. (2013) examined the useful field of view (UFOV) of elderly subjects and compared the training effect of action video-game training, non-action video-game training, and training with the UFOV test itself, i.e. a non-game training. They did not find superior training benefits induced by the game training, and if anything, the improvement on selective attention was largest for the UFOV training group (the difference was not statistically significant). This pattern is the same as the one we found, which seems to suggest that despite being more enjoyable, game training may not provide superior performance benefits compared with no-game training.

But the comparison between game training and non-game training is never straightforward. Most video games used for training were based on commercial software. Therefore, the dynamic visual presentation in the games is hard to quantify and to control, making it difficult to match any aspect of the game training with non-game training except for the training time. In contrast to previous studies, our word-puzzle game has a dissociable "training component" (the letter-recognition task) and "gaming component" (word guessing), enabling a direct match of either stimuli exposure time or total training time between the two types of training. As illustrated by our modeling results, the training component embedded in the game was as effective as the no-game training. The apparent difference between groups can largely be accounted for by variation in individual starting level.

Perks of the Game

Despite its inability to amplify training effects on visual span and reading speed, the gaming component had an additive boost on performance while the game was present (the Gaming Boost). What is the origin of this temporary effect that elevates online visual perception?

According to our survey results, the game was more enjoyable compared to the no-game task mostly because of its timely feedback and its competition aspect. Feedback is an important factor determining the result of learning. In some cases, perceptual learning cannot occur without feedback (Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006). In some other cases, while feedback is not needed for perceptual learning to happen, faked feedback, given randomly while assuming a steeper learning curve, can boost the learning effect (Shibata, Yamagishi, Ishii, & Kawato, 2009). However, since in our study the dynamics of learning is not altered in the gaming groups, feedback is unlikely the underlying mechanism for the Gaming Boost.

The role of competition has been studied widely in the context of educational games. Take one example of a computer-programming course (Burguillo, 2010). The purpose of the course was to teach students to learn programming skills, and the students were divided into groups to program a virtual player to compete in a game with other groups. Over 5 years of the course, end-of-semester surveys showed that students were interested in the game and found that the competition approach motivated them to learn.

But elevated motivation is still not enough to account for the Gaming Boost observed here.

It may be hard to pinpoint a specific feature in the game, for example feedback or competition, that causes the Gaming Boost. Instead, it might be the integrated experience of game playing that is crucial for the boost. A PET study provided evidence that dopamine was released in the striatum during the game play, and the release level increases with increased game performance, especially in the ventral striatum (Koepp et al., 1998). Dopamine is an important neuromodulator in the visual system, and it has been proposed to modulate human contrast sensitivity (for a review, see Masson et al., 1993). Levodopa, the precursor to dopamine, can be used to treat contrast sensitivity deficits in people with Parkinson's Disease (for example, Bulens, Meerwaldt, Van der Wildt, & Van Deursen, 1987). Together, it seems possible that the engaging experience of the game flow induced an elevation of dopamine level, which then resulted in a temporary boost of visual function and therefore enhanced performance. This benefit, however, did not affect learning itself and thus the effect was not sustained after the removal of the game.

Visual Span and Reading

The improvement in reading speed following training shows a similar pattern to the enlargement of the visual span, where non-game training yielded more improvement than the game-training. Averaged across subjects, one bit enlargement of visual span was associated with an increase of 0.041 ± 0.08 log-wpm, equivalent to a 10% increase in reading speed . Compared to previously reported values (0.024~0.036 log-wpm/bit, average 0.03, Legge et al., 2007), our value is slightly larger. This is possibly due to the difference in the number of slots included to calculate the size of the visual span. The current study measured the size of the visual span using 11 letter slots, whereas some studies analyzed in Legge et al. (2007) used 13-15 slots. More slots result in larger measured visual-span sizes and larger changes in the size, which then results in smaller slope values. Despite the small difference in slope values, our finding confirmed the link between visual span and reading, and that enlarging the visual span in peripheral vision is associated with an improvement in reading speed.

Application

While we support the benefit of using a game, we acknowledge that our "Wheel of Fortune" game has only been tested with normally-sighted subjects and cannot be used for any clinical purposes yet. There are many issues to be addressed before any video-game training can be applied to macular degeneration. A big obstacle in the game design comes from the lack of reliable fixation for people with central-field loss. To make appropriate eye movements, a new fixation reference point needs to be established in peripheral vision, called a preferred retinal locus (PRL). Without a well-defined PRL, it would be hard to display the training stimuli in the targeted visual field. Another concern is that, as sometimes games can be hard to learn (Bayliss et al., 2013), complex rules and physical operation of the game may be discouraging. The Wheel-of-Fortune game has very simple rules, but the current design requires extensive eye movements to examine the scorecard. While it may be good to include eye-movement practice, too much demand on eye movements will make the game difficult and frustrating, especially when the patient does not have good fixation. Possible modifications of the design include reducing the number of words displayed on the scorecard, or providing auditory feedback instead of the visual display. A third issue to address is the equipment required for playing the game. In eye-movement or PRL training studies, special-purpose equipment such as an eye-tracker or a scanning laser ophthalmoscope (SLO) is often required. If game training involves using special equipment, its usage will be limited. Considering the effort and expense to arrange clinic visits, reliable home-based training would be preferable. The Wheel-of-Fortune game does not involve the use of any clinical device, but in the current stage an experimenter is needed to enter responses and to score the guesses. One future direction is to utilize voice recognition technology to automate the game.

Another issue is the extent to which the training transfers to useful everyday tasks. There are already commercial "brain training" programs (e.g. the *Lumosity* platform) aimed to enhance people's cognitive ability, which are batches of small games targeting specific cognitive skills. A review of 374 published works, either listed on a website called *Cognitive Training Data* (http://www.cognitivetrainingdata.org) or cited by leading brain training companies, found that despite improvements on the trained tasks, there was little evidence of strong transfer of training benefit to untrained skills (Simons et al., 2016).

As for the application of video game training in low-vision rehabilitation, two randomized clinical trials were conducted recently to compare video game training against traditional patching therapy for children with amblyopia(Holmes et al., 2016; Kelly et al., 2016). Although they both found improved visual acuity in the amblyopic eye associated with game training, neither of them found any improvement in stereoacuity. In the current study, we found that the gaming component only temporarily boosted performance but did not result in any sustained benefits. It is possible that for many games, the major benefit of playing (compared to a no-game training) is to enhance user experience and to elevate persistence of training. Therefore, while being optimistic about using games for perceptual training, we have to be realistic about the expected outcome. Playing video games is an alternative training method, but it may not necessarily yield long-lasting performance benefits in everyday function.

References

Attneave, F., Arnoult, M. D., & Attneave, F. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452–471. http://doi.org/10.1037/h0044049

Baek, S. R., He, Y., & Legge, G. E. (2016). Reading speed for Korean text in central and peripheral vision. *Investigative Ophtalmology and Visual Science*, *57*(12), 1948–1948.

Bavelier, D., & Green, C. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, *35*, 391–416. http://doi.org/10.1146/annurev-neuro-060909

Bayliss, J., Vedamurthy, I., Nahum, M., Levi, D. M., & Bavelier, D. (2013). Lazy eye shooter: making a game therapy for visual recovery in adult amblyopia usable, 352–360. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39241-2_39

Belchior, P., Marsiske, M., Sisco, S. M., Yam, A., Bavelier, D., Ball, K., & Mann, W. C. (2013). Video game training to improve selective visual attention in older adults. *Computers in Human Behavior*, *29*(4), 1318–1324. http://doi.org/10.1016/j.chb.2013.01.034

Belchior, P., Marsiske, M., Sisco, S., Yam, A., & Mann, W. (2012). Older adults' engagement with a video game training program. *Activities, Adaptation & Aging*, *36*(4), 269–279. http://doi.org/10.1080/01924788.2012.702307

Bernard, J., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, *11*(8), 1–16. http://doi.org/10.1167/11.8.1

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. Retrieved from http://bbs.bioguider.com/images/upfile/2006-4/200641014348.pdf

Bulens, C., Meerwaldt, J. D., Van der Wildt, G. J., & Van Deursen, J. B. P. (1987). Effect of levodopa treatment on contrast sensitivity in Parkinson's disease. *Annals of Neurology*, *22*(3), 365–369. http://doi.org/10.1002/ana.410220313

Burguillo, J. C. (2010). Using game theory and Competition-based Learning to stimulate student motivation and performance. *Computers & Education*, *55*(2), 566–575. http://doi.org/10.1016/j.compedu.2010.02.018

Calabrèse, A., Liu, T., & Legge, G. E. (2017). Does Vertical Reading Help People with Macular Degeneration: An Exploratory Study. *PLOS ONE*, *12*(1), e0170743. http://doi.org/10.1371/journal.pone.0170743

Callens, M., Whitney, C., Tops, W., & Brysbaert, M. (2013). No deficiency in left-to-right processing of words in dyslexia but evidence for enhanced visual crowding. *Quarterly Journal of Experimental Psychology (2006)*, 1–33. http://doi.org/10.1080/17470218.2013.766898

Chen, P.-L., Chen, J.-T., Fu, J.-J., Chien, K.-H., & Lu, D.-W. (2008). A pilot study of anisometropic amblyopia improved in adults and children by perceptual learning: an alternative treatment to patching. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *28*(5), 422–8. http://doi.org/10.1111/j.1475-1313.2008.00588.x

Chung, S. T. L. (2002). The Effect of Letter Spacing on Reading Speed in Central and Peripheral Vision. *Optometry and Vision Science*, *43*(4). http://doi.org/10.1097/00006324-200012001-00034

Chung, S. T. L. (2007). Learning to identify crowded letters: Does it improve reading speed? *Vision Research*, *47*(25), 3150–3159. http://doi.org/10.1016/j.visres.2007.08.017

Chung, S. T. L. (2011). Improving reading speed for people with central vision loss through perceptual learning. *Investigative Ophthalmology & Visual Science*, *52*(2), 1164–70. http://doi.org/10.1167/iovs.10-6034

Chung, S. T. L., Legge, G. E., & Cheung, S.-H. (2004). Letter-recognition and reading speed in peripheral vision benefit from perceptual learning. *Vision Research*, *44*(7), 695–709. http://doi.org/10.1016/j.visres.2003.09.028

Chung, S. T. L., Levi, D. M., & Tjan, B. S. (2005). Learning letter identification in peripheral vision. *Vision Research*, *45*(11), 1399–412. http://doi.org/10.1016/j.visres.2004.11.021

Chung, S. T. L., Li, R. W., & Levi, D. M. (2012). Learning to Identify Near-Acuity Letters, either with or without Flankers, Results in Improved Letter Size and Spacing Limits in Adults with Amblyopia. *PLoS ONE*, *7*(4), e35829. http://doi.org/10.1371/journal.pone.0035829

Chung, S. T. L., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, *38*(19), 2949–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9797990

Chung, S. T. L., & Tjan, B. S. (2007). Shift in spatial scale in identifying crowded letters. *Vision Research*, *47*(4), 437–51. http://doi.org/10.1016/j.visres.2006.11.012

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. http://doi.org/10.1016/j.visres.2012.11.009

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–62. http://doi.org/10.1016/j.tics.2011.04.003

Doron, A., Manassi, M., Herzog, M. H., & Ahissar, M. (2015). Intact crowding and temporal masking in dyslexia. *Journal of Vision*, *15*(14), 1–17. http://doi.org/10.1167/15.14.13

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: a review. *Journal of Vision*, *2*(2), 190–203. http://doi.org/10:1167/2.2.5

Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for uppercase and lowercase letter identification. *Recherche*, 1–30.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, *8*(4), 215–221. http://doi.org/10.3758/BF03210208

Fu, F.-L. L., Su, R.-C. C., & Yu, S.-C. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101–112. http://doi.org/10.1016/j.compedu.2008.07.004

Geng, J. J., & Behrmann, M. (2002). Probability Cuing of Target Location Facilitates Visual Search Implicitly in Normal Participants and Patients with Hemispatial Neglect. *Psychological Science*, *13*(6), 520–525. http://doi.org/10.1111/1467-9280.00491

Gervais, M. J., Harvey, L. O., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(5), 655–66. http://doi.org/10.1037//0096-1523.10.5.655

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*(2), 167–207. http://doi.org/10.1016/j.cogsci.2003.10.005

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. http://doi.org/10.1016/S0042-6989(01)00097-9

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), 1–14. http://doi.org/10.1167/13.7.14

He, Y., Scholz, J. M., Gage, R., Kallie, C. S., Liu, T., & Legge, G. E. (2015). Comparing the visual spans for faces and letters. *Journal of Vision*, *15*(8), 7. http://doi.org/10.1167/15.8.7

Holmes, J. M., Manh, V. M., Lazar, E. L., Beck, R. W., Birch, E. E., Kraker, R. T., … Wallace, D. K. (2016). Effect of a Binocular iPad Game vs Part-time Patching in Children Aged 5 to 12 Years With Amblyopia. *JAMA Ophthalmology*, *33647*, 1–9. http://doi.org/10.1001/jamaophthalmol.2016.4262

Hubbard, G. P., Elia, M., Holdoway, A., & Stratton, R. J. (2012). A systematic review of compliance to oral nutritional supplements. *Clinical Nutrition*, *31*(3), 293–312. http://doi.org/10.1016/j.clnu.2011.11.020

Jiang, Y. V., Swallow, K. M., Rosenbaum, G. M., & Herzig, C. (2012). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), No Pagination Specified. http://doi.org/10.1037/a0027611

Kelly, K. R., Jost, R. M., Dao, L., Beauchamp, C. L., Leffler, J. N., & Birch, E. E. (2016). Binocular iPad Game vs Patching for Treatment of Amblyopia in Children. *JAMA Ophthalmology*, *75231*(5), 67–84. http://doi.org/10.1001/jamaophthalmol.2016.4224

Klein, R., Chou, C., Klein, B. E. K., Zhang, X., Meuer, S. M., & Saaddine, J. B. (2011). Prevalence of Age-Related Macular Degeneration in the US Population, *129*(1), 75–80.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., … Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*(6682), 266–8. http://doi.org/10.1038/30498

Lee, H.-W., Kwon, M., Legge, G. E., & Gefroh, J. J. (2010). Training improves reading speed in peripheral vision: Is it due to attention? *Journal of Vision*, *10*(6), 1–15. http://doi.org/10.1167/10.6.18.Introduction

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

Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, *41*(6), 725–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11248262

Levi, D. M., Song, S., & Pelli, D. G. (2007). Amblyopic reading is crowded. *Journal of Vision*, *7*(2), 1–17. http://doi.org/10.1167/7.2.21.Introduction

Li, R. W., Ngo, C., Nguyen, J., & Levi, D. M. (2011). Video-game play induces plasticity in the visual system of adults with amblyopia. *PLoS Biology*, *9*(8), e1001135. http://doi.org/10.1371/journal.pbio.1001135

Liu, R., & Kwon, M. (2016). Integrating oculomotor and perceptual training to induce a pseudofovea : A model system for studying central vision loss. *Journal of Vision*, *16(6)*(10), 1–21. http://doi.org/10.1167/16.6.10.doi

Maningat, P., Gordon, B. R., & Breslow, J. L. (2013). How Do We Improve Patient Compliance and Adherence to Long-Term Statin Therapy? *Current Atherosclerosis Reports*, *15*(1), 291. http://doi.org/10.1007/s11883-012-0291-7

Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, *9*(4), 1–18. http://doi.org/10.1167/9.4.14

Martinez, H. D. R. (2015). *Analysing interactions of fitted models*. Retrieved from http://cran.wustl.edu/web/packages/phia/vignettes/phia.pdf

Masson, G., Mestre, D., Blin, O., G., M., D., M., & O., B. (1993). Dopaminergic modulation of visual sensitivity in man. *Fundamental & Clinical Pharmacology*, *7*(8), 449–463. Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L23313596%5Cnhttp://sfx.bibl.ulaval.ca:9003/sfx_local?sid=EMBASE&issn=07673981&id=doi:&atitle=Dopaminergic+modulation+of+visual+sensitivity+in+man&stitle=FUNDAM.+CLIN.+PHARMACOL.&tit

Matthews, E. E., Arnedt, J. T., McCarthy, M. S., Cuddihy, L. J., & Aloia, M. S. (2013). Adherence to cognitive behavioral therapy for insomnia: A systematic review. *Sleep Medicine Reviews*, *17*(6), 453–464. http://doi.org/10.1016/j.smrv.2013.01.001

Middleton, A. (2004). Chronic Low Back Pain: Patient Compliance With Physiotherapy

Advice and Exercise, Perceived Barriers and Motivation. *Physical Therapy Reviews*, *9*(3), 153–160. http://doi.org/10.1179/108331904225006312

Mitchell, J., & Bradley, C. (2006). Quality of life in age-related macular degeneration: a review of the literature. *Health and Quality of Life Outcomes*, *4*, 97. http://doi.org/10.1186/1477-7525-4-97

Moll, K., & Jones, M. (2013). Naming fluency in dyslexic and nondyslexic readers: Differential effects of visual crowding in foveal, parafoveal, and peripheral vision. *Quarterly Journal of Experimental Psychology (2006)*, *0*(0), 1–7. http://doi.org/10.1080/17470218.2013.840852

Nguyen, N. X., Stockum, A., Hahn, G. a, & Trauzettel-Klosinski, S. (2011). Training to improve reading speed in patients with juvenile macular dystrophy: a randomized study comparing two training methods. *Acta Ophthalmologica*, *89*(1), e82-8. http://doi.org/10.1111/j.1755-3768.2010.02081.x

Nilsson, U. L., Frennesson, C., & Nilsson, S. E. G. (2003). Patients with AMD and a large absolute central scotoma can be trained successfully to use eccentric viewing, as demonstrated in a scanning laser ophthalmoscope. *Vision Research*, *43*(16), 1777–1787. http://doi.org/10.1016/S0042-6989(03)00219-0

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*(28), 4646–4674. http://doi.org/10.1016/j.visres.2006.04.023

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2), 1–36. http://doi.org/10.1167/7.2.20.Modeling

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research*, *32*(5), 895–902. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16005930

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, *51*(13), 1552–1566. http://doi.org/10.1016/j.visres.2010.10.019

Schneps, M. H., Thomson, J. M., Sonnert, G., Pomplun, M., Chen, C., & Heffner-Wong, A. (2013). Shorter lines facilitate reading in those who struggle. *PloS One*, *8*(8), e71161. http://doi.org/10.1371/journal.pone.0071161

Schuler, H., Thornton III, G. C., Frintrup, A., & Mueller-Hanson, R. (2004). *Achievement motivation inventory test manual.* Goettingen, Germany: Hogrefe.

Seiple, W., Szlyk, J. P., McMahon, T., Pulido, J., & Fishman, G. a. (2005). Eye-movement training for reading in patients with age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *46*(8), 2886–96. http://doi.org/10.1167/iovs.04-1296

Seitz, A. R., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*(9), 966–73. http://doi.org/10.1167/6.9.9

Shaw, S. (2011). Cognitive and motivational contributors to aptitude: A study of spoken and signed language interpreting students. *Interpreting*, *13*(1), 70–84. http://doi.org/10.1075/intp.13.1.05sha

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research*, *49*(21), 2574–85. http://doi.org/10.1016/j.visres.2009.06.009

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest*, *17*(3), 103–186. http://doi.org/10.1177/1529100616661983

To, L., Thompson, B., Blum, J. R., Maehara, G., Hess, R. F., & Cooperstock, J. R. (2011). A game platform for treatment of amblyopia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering : A Publication of the IEEE Engineering in Medicine and Biology Society*, *19*(3), 280–9. http://doi.org/10.1109/TNSRE.2011.2115255

Vingolo, E. M., Salvatore, S., Domanico, D., Spadea, L., & Nebbioso, M. (2013). Visual rehabilitation in patients with myopic maculopathy: our experience. *Canadian Journal of Ophthalmology. Journal Canadien D'ophtalmologie*, *48*(5), 438–42. http://doi.org/10.1016/j.jcjo.2013.08.004

Vingolo, E. M., Salvatore, S., & Limoli, P. G. (2013). MP-1 biofeedback: luminous pattern stimulus versus acoustic biofeedback in age related macular degeneration (AMD). *Applied Psychophysiology and Biofeedback*, *38*(1), 11–6. http://doi.org/10.1007/s10484-012-9203-4

Virgili, G., Acosta, R., Grover, L. L., Bentley, S. A., & Giacomelli, G. (2013). Reading aids for adults with low vision. *Cochrane Database Syst Rev*, *18*(10), CD003303. http://doi.org/10.1002/14651858.CD003303.pub2

Wang, H., He, X., & Legge, G. E. (2014). Effect of pattern complexity on the visual span for Chinese and alphabet characters. *Journal of Vision*, *14*(8), 1–17. http://doi.org/10.1167/14.8.6.doi

Watson, A. B., & Ahumada, A. J. (2012). Modeling Acuity for Optotypes Varying in Complexity. *Investigative Ophtalmology and Visual Science*, *12*(1996), 1–19. http://doi.org/10.1167/12.10.19.Introduction

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2), 1–10. http://doi.org/10.1167/7.2.2

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2010). Reading speed in the peripheral visual field of older adults: Does it benefit from perceptual learning? *Vision Research*, *50*(9), 860–9. http://doi.org/10.1016/j.visres.2010.02.006

Yu, D., Legge, G. E., Park, H., Gage, E., & Chung, S. T. L. (2010). Development of a training protocol to improve reading performance in peripheral vision. *Vision Research*, *50*(1), 36–45. http://doi.org/10.1016/j.visres.2009.10.005

Zhang, J.-Y. Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2007). Legibility variations of Chinese characters and implications for visual acuity measurement in Chinese reading population. *Investigative Ophthalmology & Visual Science*, *48*(5), 2383–90. http://doi.org/10.1167/iovs.06-1195

Zhang, J.-Y., 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. http://doi.org/10.1016/j.visres.2008.09.021

Appendix 1. Korean Stimuli Used for Chapter 2

가	각	간	갈	감	갑	갔	강	같	개	거	건
걸	겁	것	게	겠	겨	경	계	고	곳	공	과
관	교	구	귀	그	금	기	긴	길	까	깨	께
꼭	끄	끼	나	난	날	남	내	너	넘	넣	네
노	놀	놓	누	눈	느	는	늘	니	님	다	단
달	답	당	대	더	던	데	도	돈	돌	동	돼
되	된	두	드	든	들	디	따	때	떨	라	락
란	람	랑	래	러	럽	렇	레	려	렸	로	록
루	르	른	를	름	리	린	림	마	만	많	말
맛	망	맞	매	머	먹	며	면	명	모	못	무
문	물	미	바	반	받	발	밥	방	배	버	번
보	복	부	분	불	비	빠	빨	쁜	사	산	살
상	새	색	생	서	선	성	세	셨	소	속	손
수	스	습	시	식	신	실	싫	심	싶	쓰	씨
아	안	앉	않	알	았	야	약	양	어	언	얼
엄	없	었	에	여	연	열	였	영	예	오	온
올	옷	와	왔	요	용	우	운	울	웃	워	원
위	유	으	은	을	음	의	이	인	일	입	있
자	작	잘	잠	잡	장	재	저	적	전	절	점
접	정	제	져	졌	조	종	좋	주	줄	중	지
직	진	질	집	차	참	찾	책	처	추	치	친
침	커	크	키	타	터	트	틀	파	편	프	피
하	학	한	할	함	합	항	해	했	행	혼	화
회	후	히									

Figure A1-1. Full list of 279 Korean characters.

Although Korean components can be directly "typed" onto the backdrop in Matlab and cut using the default bounding box for the specified size (282 pixel (W) by 345 pixel (H)), we did not adopt this method. Our concern was that components typed in this way were generally larger than they appear in characters (for example ㅐ versus the upper-right component in the character 냄) and thus would not allow us to test the real acuity limit for Korean recognition. We therefore first generated images of Korean characters and then cut component images from them. Due to the variations in the structure of the characters (two or three components, vertical or horizontal), the components within a character could vary in size and shape. A component in a two-component character may appear bigger than the same component in a three-component character (for example the component ㄴ in 내 versus in 냄). To make sure that we could test the acuity limit in recognizing Korean characters, we chose three-letter characters to cut our testing images of Korean letters (see Fig. A1-2). When generating the characters for cutting consonants, we kept the vowel and the tail components unchanged while varying only the lead consonant. Similarly, when generating the characters for cutting vowels, we kept the lead and the tail consonants unchanged while varying only the vowel. In this way we intended to minimize the influence of letter configuration on the shape of the components we cut.

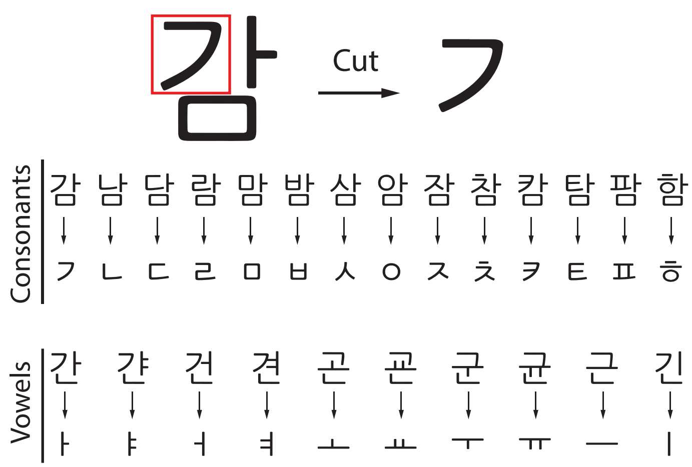

Figure A1-2. Generation of Korean Component Stimuli.

Appendix 2. Pattern Similarity of Korean and English Symbols

In addition to complexity, the similarity between stimuli in the symbol sets also influences recognition. Similarity within a given symbol set is sometimes defined functionally by a confusion matrix between the items in the set. By definition, higher similarity means more chances to mistake one item for another. Empirically, when identifying isolated patterns, the similarity between alternatives is a good predictor of the contrast threshold for a template-matching ideal observer (Pelli et al., 2006). In crowded conditions, higher target-flanker similarity results in more identification errors and mislocation (reporting the correct identity of a flanker instead of the target) errors (Bernard & Chung, 2011).

In parallel, here we found that although Korean components had slightly smaller perimetric complexity compared to English letters (96 vs. 102), their recognition performance was slightly worse than that for English letters (averaged accuracy 95.7% versus 98.4% before training; not significantly different). We quantified the overlap between stimuli by computing pairwise Euclidean distance (Gervais, Harvey, & Roberts, 1984) and found a higher similarity (smaller distance) between Korean components than between English letters (Fig. A2-1). For Korean component-recognition, when compared to English letter-recognition, the disadvantage of higher similarity seemed to offset the potential benefit of smaller perimetric complexity. But although here we discuss the effect of complexity and similarity separately, they often interact in real-world situations. For example, natural scripts with higher complexity are often found to have less similarity (Pelli et al., 2006). To separate their effects, carefully controlled paradigms are needed.

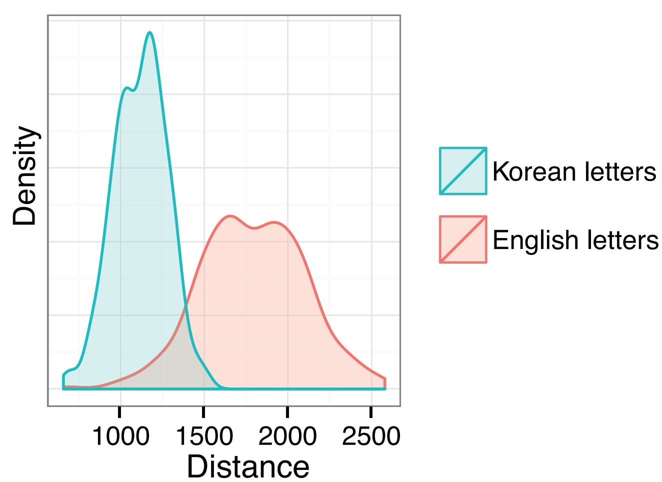

Figure A2-1. Similarity density plot: Korean components vs. English letters. The mean distance between stimuli (off-diagonal) is 1118 for Korean components (blue), and 1791 for English letters (red).

Appendix 3. Changes of Critical Print Size in Chapter 3

This is a continuation of the discussion in Chapter 3. From the averaged reading curves in Figure 3-3, all three groups had reduced critical print size (vertical dashed lines) after training. A 3×2 ANOVA was performed for critical print size after log-transformation, with group (Flanked-Local / Flanked-Distributed / Isolated-Distributed) as the between-subject factor and session type (pre-/post-test) as the within-subject factor. We found a significant main effect of session type ($F(1, 24)=7.67$, $p=0.01$): overall, critical print size decreased from 1.35 to 1.17 degrees after training (average reduction 12.3%). No main effect of group or any interaction effect was found.

Our results indicate that critical print size decreased no matter whether the training was localized or distributed. If the attentional bias limits the improvement of maximum reading speed, why does it not limit the reduction of critical print size? Note that unlike maximum reading speed, the reduction of critical print size is more closely related to the improvement in reading speed for smaller-sized text. When text size is smaller, even a localized training will have large retinotopic overlap with most words. For example, in our study the mean training size for the localized training group (1.4 times the subjects' average critical print size) was 1.85° in x-height, whereas the two smallest sizes in the reading test were 0.56° and 0.79° in x-height. On average, training stimuli (three letters spaced 0.8 times x-width) spanned 5.6° horizontally. A standard-length English word has 6 characters (including space, punctuation, ect.), and a 6-character word would span 4.4° when x-height is 0.56° or 6.2° when x-height is 0.79°. Therefore, our training has enough coverage to stimulate the letters in most of the words for the two smallest print sizes in our reading task. In this case, even if a spatial bias was introduced during training, the preferred location largely overlaps with the words and thus the bias does not have a detrimental effect.

One issue may evoke doubt on the above discussion. We found seemingly inconsistent results with a previous study: using similar distributed training, we found reduced critical print size after training whereas Chung et al. (2004) didn't. But the results are not conflicting if we focus on the improvement in reading smaller-sized text instead of on critical print size per se. It seems that in Chung et al. (2004), reading speed (in log-wpm) improved uniformly across all print sizes, resulting in increased maximum reading speed but no change in critical print size. As can be inferred from the reading curve in Fig. 3-1E, if reading speed improved uniformly across all print sizes, critical print size will remain unchanged despite the improvement in reading small-sized text. This is the case in Chung et al. (2004) where improvement in reading smaller-sized text does not lead to a reduced critical print size. Our results are consistent with Chung et al. (2004) and Chung (2007) in that both distributed and localized training can improve reading speed for small-sized text.

In summary, critical print size decreased in all three training groups. This indicates that neither crowded training stimuli nor spatially distributed training was necessary for such decrease.

Appendix 4. Complete Word List for the Game in Chapter 4

Academic fields

Accounting
Advertising
Agriculture
Algebra
American History
Archaeology
Architecture and Design
Astronomy
Astrophysics
Biochemistry
Biology
Biomedical Engineering
Business
Business Law
Cell Biology
Chemical Engineering
Chemistry
Civil Engineering
Communication Studies
Computer Science
Cultural Studies
Dance
Dentistry
Earth Science
Economics
Education
Electrical Engineering
Entrepreneurship
Environmental Studies
Ethnic Studies
Finance
Genetics
Geography
Geology
Geometry
Health Science
History
Journalism
Kinesiology
Language
Leisure Studies
Life Science
Linguistics
Literature
Marketing
Management
Mathematics
Mechanical Engineering
Media Studies
Meteorology
Microbiology
Music Theory
Neuroscience
Nursing
Nutrition
Occupational Therapy
Optometry
Organic Chemistry
Pediatrics
Performing Arts
Pharmacy
Philosophy
Physical Education
Physics
Physiotherapy
Planetary Science
Political Science
Psychology
Religion
Science
Sexuality
Social Studies
Social Work
Sociology
Sport Management
Sport Medicine
Sports Nutrition
Sports Science
Statistics
Systems Science
Theater
Toxicology
Transportation
Visual Arts
World History
Zoology

Adjectives for people

Abrasive
Abrupt
Active
Adventurous
Angry
Annoying
Anxious
Apathetic
Acquisitive
Bald
Beautiful
Blue
Blunt
Bilious
Bitter
Boisterous
Bold
Bored
Boring
Brave
Bright
Broadminded
Callous
Careful
Careless
Caring
Cheerful
Choosy
Classy
Clever
Cold
Compassionate
Complicated
Compulsive
Condescending
Confident
Conformist
Conscientious
Controlled
Controlling

Cranky
Crass
Crazy
Creative
Crude
Curious
Curt
Cute
Dangerous
Dark
Defiant
Delicate
Depressed
Diseased
Distracted
Domineering
Down
Earnest
Easy
Empathetic
Enthusiastic
Excitable
Exciting
Extroverted
Fake
Faint
Famous
Fanatical
Fashionable
Feeble
Flaky
Flexible
Flighty
Fine
Finicky
Foolish
Foppish
Frenetic
Friendly
Garrulous
Geeky
Genius
Gloomy
Haggard
Happy
Healthy
Heroic
Humble
Hysterical
Ignorant
Impressionable
Impassive
Impatient
Impertinent
Impetuous
Impulsive
Indisposed
Inflexible
Infirm
Inquisitive
Insensitive
Inspired
Insulting
Interesting
Intolerant
Introverted
Irascible
Irrational
Irresponsible
Irritable
Jaded
Jolly
Joyful
Languid
Lazy
Loquacious
Loud
Lucky
Lugubrious
Lusty
Malicious
Mean
Mercurial
Mischievous
Morbid
Moody
Morose
Mournful
Narcissistic
Nerdy
Nervous
Neurotic
Obsessive
Open
Overzealous
Pallid
Passionate
Passive
Peaked
Peevish
Perky
Picky
Pining
Pleasant
Poor
Poorly
Popular
Priggish
Private
Pushy
Quiet
Racist
Responsive
Rich
Rude
Sanguine
Seedy
Sensitive
Short
Skittish
Sickly
Silly
Smelly
Spirited
Spiritless
Spunky
Sleepy
Stiff
Stony
Stout
Stubborn
Stuffy
Surly
Sympathetic
Temperamental
Testy
Thin
Torpid
Touchy
Tough
Troubled
Truthful
Ugly
Unbalanced
Uncaring
Unconcerned
Uncontrolled
Unfeeling
Unlucky
Unreasonable
Unresponsive
Unique
Vehement
Vigorous
Violent
Volatile
Vulgar
Wacky
Warm
Weak
Well
Young
Zealous

American cities

Albany
Albuquerque
Anaheim
Anchorage
Arlington
Atlanta
Austin
Baltimore
Baton Rouge
Birmingham
Boston
Boulder
Charlotte
Chicago
Cincinnati
Cleveland
Colorado Springs
Columbus
Dallas
Denver
Des Moines
Detroit
El Paso
Fargo
Fort Lauderdale
Fort Worth
Fresno
Green Bay
Hollywood
Honolulu
Houston
Indianapolis
Jacksonville
Kansas City
Key West
Knoxville
Las Vegas
Lincoln

Little Rock
Long Beach
Los Angeles
Louisville
Madison
Memphis
Mesa
Miami
Milwaukee
Minneapolis
Nashville
New Orleans
New York
Oakland
Oklahoma City
Omaha
Orlando
Philadelphia
Phoenix
Pittsburgh
Portland
Providence
Raleigh
Rochester
Sacramento
Saint Paul
Salt Lake City
San Antonio
San Diego
San Francisco
San Jose
Santa Fe
Savannah
Scottsdale
Seattle
St Louis
Tacoma
Tallahassee
Tampa
Topeka
Tucson
Tulsa
Virginia Beach
Washington DC
Wichita

Animals

Alligator
Alpaca
Anaconda
Anteater
Antelope
Armadillo
Baboon
Bear
Beaver
Blobfish
Blue Jay
Boa Constrictor
Buffalo
Bumblebee
Butterfly
Camel
Cardinal
Caterpillar
Cheetah
Chicken
Chimpanzee
Chinchilla
Chipmunk
Cobra
Cockroach
Condor
Cougar
Crab
Crane
Crayfish
Cricket
Crocodile
Crow
Deer
Dolphin
Donkey
Dove
Dragonfly
Duck
Eagle
Earthworm
Egret
Elephant
Ferret
Flamingo
Frog
Gazelle
Gecko
Giraffe
Goat
Goldfish
Goose
Gopher
Gorilla
Grasshopper
Guinea Pig
Hagfish
Hamster
Heron
Hedgehog
Hermit Crab
Hippopotamus
Horse
Hummingbird
Hyena
Ibex
Iguana
Impala
Jackal
Jaguar
Jellyfish
Kangaroo
Kitten
King Cobra
Koala
Ladybug
Leopard
Lion
Lizard
Lobster
Loon
Marmoset
Minnow
Mole
Monkey
Moose
Mosquito
Naked Mole
Rat
Narwhal
Newt
Octopus
Orangutang
Ostrich
Otter
Oyster
Panda Bear
Panther
Parrot
Peacock
Pelican
Penguin
Platypus
Polar Bear
Porcupine
Puppy
Python
Rabbit
Rattlesnake
Red Panda
Rhinoceros
Robin
Salamander
Salmon
Scorpion
Seagull
Seal
Shark
Skunk
Sloth
Slug
Snail
Sparrow
Spider
Squid
Squirrel
Starfish
Sun Bear
Swordfish
Tadpole
Tapir
Tiger
Tortoise
Toucan
Turkey
Turtle
Vulture
Wallaby
Walrus
Warthog
Weasel
Whale
Wildebeest
Wolf
Woodpecker
Zebra

Business

Advantage	Debt	Future	Mistake	Rich
Advertisement	Debtor	Goal	Model	Rise
Advice	Decision	Goods	Money	Risk
Agenda	Decrease	Growth	Objective	Royalties
Analyst	Deficit	Guarantee	Offer	Safety
Apology	Delivery	Hours	Operation	Salary
Authorization	Department	Human	Opinion	Sales
Bankrupt	Description	Resources	Option	Schedule
Bidder	Design	Import	Order	Sector
Bill	Difference	Improvement	Output	Service
Billionaire	Disadvantage	Increase	Partnership	Securities
Bond	Discount	Industry	Patent	Share
Boss	Distribution	Innovation	Payment	Shareholders
Brand	Economy	Instructions	Penalty	Signature
Briefcase	Employee	Insurance	Permission	Strategy
Broker	Employer	Interest	Possibility	Stock
Bubble	Enquiry	Inventory	Poverty	Success
Budget	Entrepreneur	Invention	Power	Suggestion
Buyer	Environment	Investment	Premium	Suit
Capital	Equipment	Invoice	Product	Supply
Capitalism	Estimate	Journal	Production	Support
Career	Executive	Knowledge	Profit	Target
Commerce	Expand	Labor	Promotion	Taxes
Commission	Experience	Leadership	Property	Technology
Comparison	Explanation	Liability	Purchase	Trade
Competition	Export	License	Ranking	Transport
Competitor	Facilities	Limit	Reduction	Trip
Confirmation	Factory	Loss	Refund	Turnover
Consumer	Failure	Manufacture	Reminder	User
Cooperative	Feedback	Margin	Rent	Value
Corporation	Finance	Market	Repairs	Wages
Costs	Firm	Mediator	Report	Waste
Creditor	Founder	Media	Responsibility	Wholesaler
Currency	Franchise	Message	Result	
Customer	Free	Millionaire	Retailer	
Deadline	Fund	Mission	Revenue	

Cars

Accord	Altima	Bentley	Buick	Camry
Acura	Aston Martin	Bristol	Cadillac	Cherokee
Alfa Romeo	Audi	Bugatti	Camaro	Chevrolet

Chrysler
Citroen
Civic
Corvette
Crown Victoria
Daihatsu
Delorean
Dodge
Ferrari
Fiat
Ford
Honda
Hummer
Humvee
Hyundai
Infiniti
Isuzu
Jaguar
Jeep
Jetta
Kia Motors
Lamborghini
Land Rover
Lexus
Lincoln
Lotus
Maserati
Mazda
Mercedes Benz
Mercury
Mini
Minivan
Mitsubishi
Mustang
Nissan
Oldsmobile
Opel
Plymouth
Pontiac
Porsche
Prius
Range Rover
Rolls Royce
Rover
Saab
Saturn
Scion
Shelby Super Cars
Smart
Studebaker
Subaru
Suzuki
Thunderbird
Toyota
Trans Am
Triumph
Viper
Volkswagen
Volvo
Winnebago

Characters from children's story and cartoons

Aladdin
Alice in Wonderland
Babar
Bambi
Batman
Bert
Black Beauty
Bugs Bunny
Cat in the Hat
Cat Woman
Chesire Cat
Christopher Robin
Cinderella
Clifford
Cookie Monster
Cowardly Lion
Donald Duck
Dorothy
Eeyore
Elmo
Ernie
Garfield
Goofy
Gretel
Grinch
Hansel
Harry Potter
Hello Kitty
Henny Penny
Horton
Jemima Puddle Duck
Kanga
Kermit
Lassie
Little Mermaid
Little Miss Muffet
Little Red Ridinghood
Mad Hatter
Mary Poppins
Mickey Mouse
Minnie Mouse
Miss Piggy
Mother Goose
Mr Toad
Nancy Drew
Old Mother Hubbard
Old Yeller
Oscar the Grouch
Peter Pan
Peter Rabbit
Piglet
Pippy Longstocking
Pinocchio
Pluto
Puss in Boots
Queen of Hearts
Ramona Crumby
Rapunsel
Rin Tin Tin
Robinhood
Robinson Crusoe
Rumplestiltskin
Scarecrow
Sleeping Beauty
Snow White
Spiderman
Superman
Tarzan
The Count
Tinkerbell
Tin Woodman
Ugly Duckling
Wendy
Willy Wonka
Winnie the Pooh
Wizard of Oz

Characters from comics and superheroes

Aquaman
Asteri
Banshee
Batgirl
Batman
Beast
Black Panther
Black Widow
Bumblebee
Captain America
Captain Planet
Captain Universe
Catwoman

Cyclops
Daredevil
Deadpool
Doctor Octopus
Dr Manhattan
Elektra
Emma Frost
Flash gordon
Gambit
Ghost Rider
Green Arrow
Green Lantern
Green Hornet
Hawkeye
Hellboy
Hulk
Human Torch
Iceman
Invisible
Woman
Ironman
Jean Grey
Lex Luthor
Lois Lane
Lucky Luke
Magneto
Mighty Mouse
Mr Fantastic
Nick Fury
Optimus Prime
Professor X
Punisher
Robin
Rorschach
Silver surfer
Spawn
Spiderman
Storm
Supergirl
Superman
The Flash
The Spirit
The Thing
The Tick
Thor
Underdog
Wolverine
Wonder Woman
Wonder Twins

Computer software

ACDSee
ActiveX
Address Book
Adobe
Dreamweaver
Adobe Flash Player
Adobe Illustrator
Adobe Acrobat
Adobe Photoshop
Audio Hijack Pro
AutoCAD
BASIC
C Programming
Language
Dropbox
EndNote
Evernote
FaceTime
Final Cut Pro
Finder
Foxit Reader
GNU Octave
Goldwave
Google Chrome
Google Earth
Google Talk
Graphic
Internet Explorer
iChat
iMovie
iPhoto
iTunes
JavaScript
Macromedia Flash
Maple
Mathematica
MATLAB
Mendeley
Microsoft Excel
Microsoft Outlook
Microsoft
PowerPoint
Microsoft Word
Mindjet
MindManager
Mozilla Firefox
Mozilla
Thunderbird
Notepad
Norton AntiVirus
Numbers
OpenGL
Opera
Origin
Outlook Express
Pages
Perl
Photo Booth
Picasa
Preview
Prezi
Python
QuickTime Player
R Statistics
Program
RealPlayer
RStudio
Safari
SketchUp
Skim
Skitch
Skype
Software
Spotify
SPSS
Terminal
TeXShop
Text Editor
VLC Media Player
VoiceOver
Windows Media
Player
Windows Movie
Maker
Winrar
WinZip
Xcode

Dance

Acrobatic
Allemande
Artistic
Audience
Ballet
Ballroom
Band
Cakewalk
Choreography
Chorus
Dancers
Classes
Classical
Clockwise
Clog

Collaborative
Composition
Costumes
Craft
Curtsy
Dancing Master
Decor
Disco
Elegance
Emote
Entertainment
Exercise
Flamenco
Folk Dancing
Fox Trot
Garment
Gesture
Glide
Grace
Groups
Gymnasium
Hora
Jazz
Jitterbug
Jump
Leap
Leotard
Lift
Minuet
Modern Dance
Movement
Music
Partner
Passion
Patterns
Piece
Popular
Dancing
Position
Posture
Practice
Rehearsal
Rhythm
Role
Rotation
Rumba
Samba
Square Dancing
Steps
Studio
Style
Sway
Swirl
Tango
Tap Dancing
Tarantella
Technique
Tempos
Theater
Troupe
Twirl
Unitards
Waltz

Drugstore items

Allergy Relief
Medicine
Aloe Vera
Aspirin
Bandaids
Bath Salts
Beach Ball
Book
Brush
Candy
Catsup
Cereal
Chapstick
Chips
Chocolate Bar
Clay Face Mask
Clear Eyes
Coffee
Cologne
Comb
Conditioner
Condoms
Cotton Balls
Cotton Swabs
Cough Syrup
Deodorant
Disinfectant
Ear Plugs
Eye Shadow
Eyelash Curlers
Eyeliner
Face Wash
False Eyelashes
False Nails
Floss
Foundation
Gift Card
Granola Bars
Greeting Card
Hair Dye
Hair Spray
Ice Cream
Insect Repellent
Instant Oatmeal
Ipecac
Lipstick
Lottery Tickets
Macaroni and
Cheese
Magazines
Makeup Remover
Mascara
Mouthwash
Mustard
Nail Clippers
Nail File
Nail Polish
Nasal Decongestant
NoDoz
Nuts
Peanut Butter
Pepper
Perfume
Popcorn
Razor
Rouge
Salt
Scrub
Shampoo
Shaving Cream
Soap
Soda
Sun Hat
Sunscreen
Suntan Oil
Tampons
Toilet Paper
Tooth Brush
Toothpaste
Tums
Tweezers
Vasoline
Vitamins

Female actors

Angelica Huston
Angelina Jolie
Anne Bancroft

Anne Hathaway
Audrey Hepburn
Barbra Streisand
Bette Davis
Brittany Murphy
Brooke Shields
Cameron Diaz
Cate Blanchett
Charlize Theron
Cher
Claudette Colbert
Courteney Cox
Debra Winger
Diane Keaton
Drew Barrymore
Elizabeth Taylor
Ellen Pompeo
Emma Stone
Emma Thompson
Emma Watson
Eva Longoria
Eva Mendes
Farrah Fawcett
Faye Dunaway
Geena Davis
Gena Rowlands
Ginger Rogers
Glenn Close
Goldie Hawn
Grace Kelly
Gwyneth Paltrow
Halle Berry
Heather Locklear
Helen Hayes
Helen Hunt
Helen Mirren
Helena Bonham Carter
Hilary Swank
Ingrid Bergman
Jamie Lee Curtis
Jane Fonda
Jennifer Aniston
Jennifer Connelly
Jennifer Garner
Jennifer Hudson
Jennifer Lawrence
Jessica Alba
Jessica Chastain
Jessica Lange
Joan Crawford
Joan Fontaine
Jodie Foster
Judi Dench
Judy Garland
Julia Roberts
Julie Andrews
Kate Beckinsale
Kate Bosworth
Kate Winslet
Katharine Hepburn
Kathy Bates
Kim Basinger
Kirsten Dunst
Lindsay Lohan
Liv Tyler
Liv Ullmann
Liza Minnelli
Lucy Liu
Marilyn Monroe
Marisa Tomei
Marlene Dietrich
Mary Tyler Moore
Meg Ryan
Megan Fox
Meryl Streep
Michelle Pfeiffer
Mila Kunis
Mira Sorvino
Natalie Portman
Nicole Kidman
Patricia Arquette
Penelope Cruz
Rachel McAdams
Reese Witherspoon
Renee Zellweger
Rosanna Arquette
Sally Field
Salma Hayek
Sandra Bullock
Sarah Jessica Parker
Scarlett Johansson
Sharon Stone
Shirley MacLaine
Sigourney Weaver
Sissy Spacek
Sophia Loren
Susan Sarandon
Uma Thurman
Vanessa Redgrave
Vivien Leigh
Whoopi Goldberg

Food

Almond
Apple
Apricot
Avocado
Bacon
Bagel
Banana
Beans
Beef
Beets
Blackberry
Boysenberry
Blueberry
Bread
Broccoli
Brownie
Brussels
Sprouts
Burrito
Butter
Cabbage
Candy
Cake
Carrot
Cashew
Casserole
Celery
Cereal
Cheesecake
Cherry
Chips
Chicken
Chocolate
Cookie
Cottage Cheese

Crackers
Croissant
Cupcake
Doughnut
Eggplant
English Muffin
Fish
French Fries
Fruit Salad
Grape
Grapefruit
Granola
Guacamole
Gummy Bears
Hamburger
Herbs
Hot Dog
Ice Cream
Jalapeno

Jelly
Jelly Beans
Kiwi
Lasagna
Lettuce
Licorice
Lobster
Macaroni
Mango
Marshmallow
Meatballs
Meatloaf
Milk
Milkshake
Muffin
Mushroom
Nachos
Noodles
Oatmeal

Olives
Omelet
Onion
Onion Rings
Orange
Pancake
Peach
Peanut
Peanut Butter
Pecan Pie
Pepper
Pepperoni
Pickle
Pineapple
Pistachio
Pizza
Plum
Popcorn
Popsicle

Pork
Pork Chop
Potato
Pretzel
Pudding
Pumpkin
Raspberry
Ribs
Salad
Sandwich
Sauce
Sausage
Shish Kabob
Shrimp
Soda
Soup
Spaghetti
Spices
Spinach

Squash
Steak
Strawberry
Sushi
Sweet Potatoes
Taco
Tangerine
Tiramisu
Toast
Tofu
Tomato
Tuna
Turkey
Waffle
Walnut
Watermelon
Yogurt

Household items

Air Conditioner
Alarm Clock
Batteries
Blanket
Blender
Blinds
Blowdryer
Board Games
Books
Bottle Opener
Bowl
Broom
Calculator
Calendar
Camera
Candle
Can Opener
Cell Phone
Chair
Clock
Clothes

Coffee Maker
Coffee Pot
Coffee Table
Computer
Couch
Cups
Curling Iron
Curtains
Desk
Dishwasher
Doorbell
Dryer
DVDs
Fire Alarm
Fireplace
Flashlight
Flatscreen TV
Fork
Freezer
Garage Door
Garbage Can

Glasses
Grill
Headphones
Iron
Knives
Lamp
Lawn Mower
Mattress
Medicine
Microwave
Oven
Mirror
Modem
Napkins
Oven
Paintings
Pans
Paper Towel
Pencil
Plates
Pictures

Pillow
Pitcher
Plant
Plate
Poster
Pots
Printer
Radio
Refrigerator
Remote
Router
Rugs
Scale
Scanner
Scissors
Sewing Machine
Shoes
Smoke Detector
Soap

Spoon
Sponge
Stapler
Stereo
Stove
Stool
Stuffed Animal
Table
Tape
Telephone
Toaster
Towel
Trash Can
Vacuum Cleaner
Video Games
Washing Machine

Jobs

Accountant
Actor
Actress
Ambassador
Artist
Astronaut
Astronomer
Athlete
Attorney
Author
Babysitter
Baker
Ballerina
Banker
Barber
Beautician
Blacksmith
Bookkeeper
Broker
Burglar
Butcher
Butler
Captain
Carpenter
Cartographer
Cashier
Chef
Clerk
Coach
Composer
Contractor
Cook
Craftsman
Criminal
Crook
Custodian
Dancer
Dentist
Designer
Detective
Dictator
Director
Disc Jockey
Doctor
Doorman
Editor
Electrician
Engineer
Entertainer
Entrepreneur
Executive
Farmer
Firefighter
Fisherman
Foreman
Gardener
Governor
Guide
Hairdresser
Hobo
Hunter
Instructor
Intern
Interpreter
Inventor
Janitor
Jockey
Journalist
Judge
Landlord
Lawyer
Lecturer
Librarian
Lifeguard
Locksmith
Magician
Maid
Manager
Mayor
Mechanic
Model
Musician
Navigator
Novelist
Nurse
Painter
Park Ranger
Pastor
Pharmacist
Photographer
Physician
Pianist
Pilot
Plumber
Poet
Police Officer
Politician
President
Priest
Principal
Producer
Professor
Publisher
Referee
Reporter
Researcher
Sailor
Salesperson
Scholar
Scientist
Security Guard
Senator
Sheriff
Shoemaker
Singer
Soldier
Stockbroker
Student
Surgeon
Tailor
Taxi Driver
Teacher
Translator
Travel Agent
Truck Driver
Tutor
Umpire
Undertaker
Usher
Veterinarian
Waiter
Waitress
Watchmaker
Welder
Zookeeper

Makeup and skin care

Absorbent
Acne
Age Spot
Aging
Allergy
Balm
Bath Gel
Bath Salt
Beautiful
Blackhead
Blemish
Blending Brush
Blush Brush
Body Butter
Body Lotion
Body Mist
Body Scrub
Brow Brush
Bronzing
Powder
Brush
Cheeks
Cheekbones
Clarifying
Cleanser
Combination
Skin
Complexion
Concealer
Contour
Cosmetic
Cotton Ball
Day Cream
Deep Cleansing
Dermatologist
Dryness
Essential Oil
Exfoliation
Extract
Eye Cream
Eye Pencil
Eye Shadow
Eye Smudger
Eyebrow Pencil
Eyebrow
Tweezers
Eyelash Curler
Eyeliner
False Eyelashes
Firming
Foundation
Fragrance
Hand Cream
Highlighter
Illuminating
Imperfection
Lip Balm
Lip Gloss
Lip Pencil
Lipstick

Liquid Eyeliner
Liquid
Foundation
Makeup
Remover
Mascara
Massage Oil
Matte
Mirror
Moisturize
Oily Skin
Pores
Powder
Powder Puff
Pressed
Powder
Primer
Radiant
Redness
Sensitive Skin
Shade
Sheer
Shimmering
Skin Tone
Smoky Eyes
Smooth
Soap
Soften
Sponge
Stylish
T Zone
Toner
Waterproof
Waxing
Wipe
Wrinkle

Male actors

Adam Sandler
Adrien Brody
Al Pacino
Alan Arkin
Alec Baldwin
Andy Garcia
Anthony Hopkins
Arnold Schwarzenegger
Ben Affleck
Ben Kingsley
Ben Stiller
Burt Reynolds
Bill Cosby
Bill Murray
Brad Pitt
Bradley Cooper
Bruce Willis
Charles Chaplin
Charlie Sheen
Chris Rock
Chris Tucker
Christian Bale
Christian Slater
Christopher Reeve
Christopher Walken
Clint Eastwood
Colin Farrell
Colin Firth
Daniel Craig
Danny Devito
David Arquette
Dennis Quaid
Denzel Washington
Dustin Hoffman
Ed Harris
Eddie Murphy
Edward Norton
Elijah Wood
Gene Hackman
George Clooney
Gerard Butler
Harrison Ford
Heath Ledger
Hugh Jackman
Jack Black
Jack Nicholson
Jackie Chan
Jake Gyllenhaal
James Franco
Jason Statham
Jim Carrey
Joaquin Phoenix
John Cusack
John Malkovich
John Travolta
John Wayne
Johnny Depp
Jude Law
Keanu Reeves
Kevin Costner
Kevin Spacey
Laurence Fishburne
Leonardo DiCaprio
Liam Neeson
Leslie Nielsen
Mark Wahlberg
Marlon Brando
Martin Lawrence
Martin Sheen
Matt Damon
Matthew Broderick
Matthew Perry
Mel Gibson
Michael Douglas
Michael Caine
Michael J Fox
Mickey Rourke
Mike Myers
Morgan Freeman
Nicolas Cage
Orlando Bloom
Owen Wilson
Patrick Swayze
Pierce Brosnan
Ryan Gosling
Richard Gere
Robert De Niro
Robert Downey Jr
Robert Duvall
Robert Redford
Robin Williams
Roger Moore
Russell Crowe
Ryan Reynolds
Samuel L Jackson
Sean Connery
Sean Penn
Shia LaBeouf
Steve Carell

Steve Martin
Sylvester Stellone
Timothy Dalton
Tom Cruise
Tom Hanks
Tommy Lee Jones
Val Kilmer
Vin Diesel
Will Ferrell
Will Smith
Woody Allen

Movies

A Beautiful Mind
A Clockwork Orange
Alien
American Beauty
American History X
American psycho
Apocalypse Now
Armageddon
Avatar
Back to the Future
Bad boys
Batman
Be kind rewind
Big Daddy
Black Swan
Blood Diamond
Braveheart
Cadillac Records
Casablanca
Casino Royale
Catch Me If You Can
Citizen Kane
Dead Poets Society
Deep Impact
Die Another Day
Die Hard
Django Unchained
Fargo
Fight Club
Finding Nemo
Forrest Gump
Gangs of New York
Ghostbusters
Gladiator
GoldenEye
Gone With the Wind
Good Will Hunting
Goodfellas
Gremlins
Groundhog Day
Grumpy Old Men
Harry Potter
Home Alone
Hook
I am legend
Inception
Independence Day
Indiana Jones
Inglourious Basterds
Its a Wonderful Life
James Bond
Jaws
Jumanji
Juno
Jurassic Park
Kill Bill
King Kong
Limitless
Little Miss Sunshine
Lost in Translation
Mad Max
Marie Antoinette
Meet the Fockers
Memoirs of a Geisha
Men in black
Midnight in Paris
Milk
Mission Impossible
Monsters Inc
No Country for Old Men
Pearl Harbor
Philadelphia
Pirates of the Caribbean
Pitch perfect
Psycho
Pulp Fiction
Raging bull
Rain Man
Requiem for a Dream
Rocky
Rudy
Saving Private Ryan
Scarface
Schindlers List
Scream
Seven
Shutter Island
Silver Linings Playbook
Sleepless in Seattle
Sleepy Hollow
Slumdog Millionaire
Spiderman
Star Trek
Star Wars
Taxi Driver
The Addams Family
The Aviator
The Dark Knight
The Darjeeling Limited
The Da Vinci Code
The Departed
The Godfather
The goonies
The Great Gatsby
The Green Mile
The Grinch
The Hangover
The Ladykillers
The Lion King
The Lord of the Rings
The Man in the Iron Mask

The Matrix
The Notebook
The Pianist
The Shawshank Redemption
The Shining
The Silence of the Lambs
The Sixth Sense
The Social Network
The Sound of Music
The Terminal
The Terminator
The Thomas Crown Affair
The Truman show
The Wizard of Oz
The World Is Not Enough
Titanic
Tomorrow Never Dies
Top Gun
Toy Story
Training Day
Trainspotting
Transformers
Twister
WALL E
Wedding Crashers
Zoolander

Instruments

Accordion
Acoustic Guitar
Bagpipe
Banjo
Bass
Bass Clarinet
Bass Drum
Bass Guitar
Bassoon
Beat
Bell
Bongo Drum
Cello
Chimes
Clarinet
Clef
Conga Drum
Cow Bell
Cymbals
Didgeridoo
Drum
Drumsticks
Electric Guitar
Electric Organ
English Horn
Fiddle
Flute
French Horn
Glockenspiel
Gong
Grand Piano
Harmonica
Harmony
Harp
Headphones
Horn
Kazoo
Kettledrum
Keyboard
Lyre
Maracas
Marimba
Melody
Microphone
Music
Notes
Oboe
Octave
Organ
Percussion
Piano
Piccolo
Pipe Organ
Pitch
Rattle
Recorder
Reed
Rhythm
Saxophone
Snare Drum
Steel Drum
Steel Guitar
String Bass
Symphony
Tambourine
Triangle
Trombone
Trumpet
Tuba
Ukulele
Viola
Violin
Whistle
Wooden Blocks
Xylophone

Plants

Alfalfa
Algae
Annual
Anemone
Apple Blossom
Aster
Azalea
Basil
Bamboo
Bark
Bachelor Button
Bellflower
Bluebonnet
Bergamot
Bitterroot
Blossom
Brush
Bulb
Bush
Buttercup
Butterflyweed
Cactus
Calla Lily
Camellia
Carnation
Cherry Blossom
Cherokee Rose
Chrysanthemum
Clover
Columbine
Cosmos
Cow Parsnip
Daffodil
Dahlia
Daisy
Dandelion
Dogbane
Dogwood
Echinacea
Evening Primrose
Fern
Flower
Forget Me Not
Fuschia
Geranium
Goldenrod
Grass
Hawthorn

Herb
Hibiscus
Honeysuckle
Horticulture
Hydrangea
Indian Blanket
Indian Paintbrush
Indian Pipe
Iris
Jasmine
Jessamine
Lady Slipper
Lavender
Leaf
Lemon Grass
Lilac
Lily
Lobelia
Magnolia
Marigold
Mayflower
Milkweed
Mint
Mistletoe
Morning Glory
Moss
Mountain Laurel
Mushroom
Myrtle
Nicotiana
Orange Blossom
Oregon Grape
Orchid
Pansy
Pasque Flower
Passionflower
Peach Blossom
Peony
Perennial
Petal
Plant
Poinsettia
Poison Ivy
Pollen
Poppy
Primrose
Rhododendron
Root
Rose
Safflower
Sagebrush
Sapling
Seaweed
Shamrock
Snap Dragon
Soil
Spore
Sprout
Starflower
Stem
Sunflower
Sweet William
Syringa
Tarragon
Tickseed
Thistle
Thorn
Thyme
Tiger Lily
Tree Fern
Trillium
Tulip
Tumbleweed
Venus Flytrap
Vine
Violet
Water
Weed
Wild Flower
Wild Prairie Rose
Wintergreen
Wolfsbane
Wormwood
Yucca

Politics

Absentee
Activist
Advertising
Advise
Aggressive
Amendment
Ballot
Beliefs
Bill
Budget
Bureaucracy
Cabinet
Campaign
Candidate
Centrist
Checks and Balances
Congress
Conservative
Constituent
Constitution
Controversy
Debate
Decision
Delegate
Democracy
Democrat
Diplomat
Electoral College
Ethics
Exit Poll
Federal
Filibuster
Front Runner
Funding
Government
Incumbency
Independent
Issues
Judge
Landslide
Legislature
Liberal
Lobbyist
Majority
Media
Mentor
Minority
National
Nominate
Nominee
Party
Platform
Policy
Politician
Poll
Popular
Press
Primary
Propaganda
Race
Ratify
Reform
Regulate
Republican
Senate
Slogan
Spending
Statute
Straw Poll
Subcommittee
Taxes
Term
Ticket
Veto
Vote

Psychology

ABA Design
Abnormal Psychology
Abraham Maslow
Adaptation
Addiction
Adjustment
Aging
Albert Bandura
Altruism
Amnesia
Anxiety Disorder
Attachment
Attention
Attitudes
Auditory
Autism
Aversion Therapy
B F Skinner
Behavior
Behavioral Therapy
Behaviorism
Beliefs
Bias
Bipolar Disorder
Brain
Butterfly Effect
Carl Rogers
Case Study
Catharsis
Cerebral Cortex
Clinical Psychology
Cognition
Cognitive Psychology
Conditioning
Contingency
Contrast Sensitivity
Control Group
Counseling
Culture
Decision Making
Dementia
Depression
Determinism
Development
Disorder
Dream Analysis
Eating Disorder
Edward Thorndike
Egocentrism
Emotion
Empathy
Encoding
Eric Erikson
Evidence
Expectancy
Experiment
Explicit Memory
Extinction
Extraversion
False Alarm
Feeling
Fluid Intelligence
fMRI
Frontal Lobe
Gender Identity
Gestalt
Group Therapy
Humanism
Hypothesis
Illusion
Independent Variable
Intelligence
Ivan Pavlov
Jean Piaget
Laboratory
Learned Helplessness
Learning
Libido
Longitudinal Design
Memory
Mental Age
Mind
Morality
Motivation
Motor Cortex
Neuron
Narcissism
Narrative Theory
Observational Learning
Observer Bias
Occipital Lobe
Operational Definition
Panic Disorder
Parietal Lobe
Perception
Perceptual Consistancy
Personality
Placebo Effect
Plasticity
PTSD
Prejudice
Problem Solving
Projective Test
Psychiatrist
Psychoanalysis
Psychometric Function
Psychotherapy
Punishment
Recall
Recognition
Reinforcement
Relationships
Research
Reward System
Sample
Schizophrenia
Sensation
Sigmund Freud
Signal Detection Theory
Significance
Social Phobia
Statistics
Stereotypes
Study
Survey
Temporal Lobe
Theory
Thinking
Threshold
Traits
Unconscious
Visual Cortex

White Matter
Wilhelm Wundt
William James
Working Memory

Rocks and minerals (easy)

Agate
Amber
Amethyst
Apatite
Aquamarine
Aventurine
Azurite
Barite
Basalt
Beryl
Bloodstone
Blue Lace
Agate
Botswana
Agate
Calcite
Carnelian
Chalcedony
Chalk
Charoite
Citrine
Copper
Coral
Corumdum
Crysocolla
Crystal
Cubic
Zirconium
Desert Rose
Diamond
Emerald
Feldspar
Fire Agate
Fire Opal
Fluorite
Fordite
Garnet
Geode
Gneiss
Gold
Goldstone
Granite
Gypsum
Hematite
Indian Paint
Stone
Iolite
Iron
Jade
Jasper
Labradorite
Lapis Lazuli
Limestone
Malachite
Marble
Meteorite
Mexican Crazy
Lace Agate
Mica
Moldavite
Mookaite
Moonstone
Moss Agate
Obsidian
Onyx
Opal
Pearl
Peridot
Petrified Wood
Picture Jasper
Pietersite
Platinum
Pumice
Pyrite
Quartz
Rhodochrosite
Rhodonite
Rose Quartz
Ruby
Rutilated
Quartz
Salt
Sandstone
Sapphire
Serpentine
Shale
Silver
Slate
Smoky Quartz
Snowflake
Obsidian
Soapstone
Sodalite
Spinel
Sulfur
Talc
Tanzanite
Tiger Iron
Topaz
Tourmaline
Turquoise
White Howlite
Zircon

Rocks and minerals (expert)

Alabaster
Alexandrite
Amazonite
Ammolite
Andesite
Angelite
Anthracite Coal
Apophyllite
Aragonite
Arkose
Bauxite
Bituminous
Coal
Bornite
Breccia
Brecciated
Jasper
Bumblebee
Jasper
Celestite
Chalcopyrite
Chert
Chondrite
Coprolite
Coquina
Cuprite
Dendritic
Jasper
Diatomite
Diorite
Dunite
Eilat Stone
Epidote
Erythrite
Fuchsite
Gabbro
Galena
Halite
Heulandite
Hickoryite
Hornblende
Kaleidascope
Agate
Kambaba
Jasper
Kaolin
Kyanite
Larsonite
Lenticular
Jasper
Lepidolite
Lignite Coal
Lodestone
Magneprase
Mariposite
Mohogany
Obsidian
Noreena Jasper
Okenite
Olivine
Orpiment
Orthoclase
Pegmatite
Phyllite
Picasso Jasper

Potato Stone
Prehnite
Print Stone
Pyromorphite
Quartzite
Rain Forest
Jasper
Rainbow
Hematite
Rainbow
Moonstone
Rainbow
Obsidian
Realgar
Red Poppy
Jasper
Rhyolite
Ruby Zoisite
Scapolite
Schist
Scoria
Selenite
Septarian
Nodule
Seraphinite
Serenity Jasper
Sheelite
Sphalerite
Sphene
Staurolite
Stilbite
Sugilite
Tektite
Thunderegg
Tiffany Stone
Travertine
Tremolite
Triphylite
Turgite
Turritella
Agate
Ulexite
Uralite
Vanadium
Varicite
Variolite
Wollastonite
Youngite
Zebra Stone

Sport teams (MLB, NBA, NFL, NHL)

Anaheim Ducks
Arizona Cardinals
Arizona Diamondbacks
Atlanta Braves
Atlanta Falcons
Atlanta Falcons
Atlanta Hawks
Baltimore Orioles
Baltimore Ravens
Boston Bruins
Boston Celtics
Boston Red Sox
Brooklyn Nets
Buffalo Bills
Buffalo Sabres
Calgary Flames
Carolina Hurricanes
Carolina Panthers
Charlotte Bobcats
Chicago Bears
Chicago Blackhawks
Chicago Bulls
Chicago Cubs
Chicago White Sox
Cincinnati Bengals
Cincinnati Reds
Cleveland Browns
Cleveland Cavaliers
Cleveland Indians
Colorado Avalanche
Colorado Rockies
Columbus Blue Jackets
Dallas Cowboys
Dallas Mavericks
Dallas Stars
Denver Broncos
Denver Nuggets
Detroit Lions
Detroit Pistons
Detroit Red Wings
Detroit Tigers
Edmonton Oilers
Florida Panthers
Golden State Warriors
Green Bay Packers
Houston Astros
Houston Rockets
Houston Texans
Indiana Pacers
Indianapolis Colts
Jacksonville Jaguars
Kansas City Chiefs
Kansas City Royals
Los Angeles Clippers
Los Angeles Dodgers
Los Angeles Kings
Los Angeles Lakers
Memphis Grizzlies
Miami Dolphins
Miami Heat
Miami Marlins
Milwaukee Brewers
Milwaukee Bucks
Minnesota Timberwolves
Minnesota Twins
Minnesota Vikings
Minnesota Wild
Montreal Canadiens
Nashville Predators
New England Patriots
New Jersey Devils
New Orleans Pelicans
New Orleans Saints
New York Giants
New York Islanders
New York Jets
New York Knicks
New York Mets
New York Rangers
New York Yankees
Oakland Athletics
Oakland Raiders
Oklahoma City Thunder
Orlando Magic
Ottawa Senators
Philadelphia Eagles
Philadelphia Flyers
Philadelphia Phillies
Phoenix Coyotes
Phoenix Suns
Pittsburgh Penguins
Pittsburgh Pirates
Pittsburgh Steelers

Portland Trail Blazers
Sacramento Kings
San Antonio Spurs
San Diego Chargers
San Diego Padres
San Francisco Giants
San Jose Sharks
Seattle Mariners
Seattle Seahawks
St Louis Blues
St Louis Cardinals
St Louis Rams
Tampa Bay Buccaneers
Tampa Bay Lightning
Tampa Bay Rays
Tennessee Titans
Texas Rangers
Toronto Blue Jays
Toronto Maple Leafs
Toronto Raptors
Utah Jazz
Vancouver Canucks
Washington Capitals
Washington Nationals
Washington Redskins
Washington Wizards
Winnipeg Jets

Sports

Aerobics
Archery
Athlete
Badminton
Ball
Baseball
Basketball
Bicycle
Billiards
Bobsledding
Bowling
Boxing
Broomball
Coach
Cricket
Cross Country
Curling
Cycling
Darts
Diving
Dodgeball
Dunk
Fencing
Field Hockey
Figure Skating
Football
Game
Goal
Golf
Gymnastics
Handball
High Jump
Hockey
Home Run
Horse Racing
Ice Skates
Javelin
Karate
Kickball
Lacrosse
Long Jump
Luge
Marathon
Nordic Skiing
Olympics
Paintball
Pilates
Ping Pong
Playoffs
Pole Vault
Polo
Race
Racing
Racquetball
Referee
Relay
Rock Climbing
Rowing
Rugby
Running
Scoreboard
Skateboarding
Skiing
Slalom
Snowboarding
Soccer
Speed Skating
Squash
Sumo
Wrestling
Surfing
Swimming
Tennis
Touchdown
Volleyball
Water Polo
Weightlifting
Wrestling

Summer related

Air Conditioner
August
Backpacking
Baseball
Bathing Suit
Beach
Beach Ball
Biking
Bikini
Boating
Camping
Canoeing
Diving
Festival
Fireworks
Fishing
Flip Flops
Floaties
Flower
Frisbee
Gardening
Golfing
Grilling
Heat
Hiking
Hot Dogs
Humidity
Ice Cream
July
June
Lake
Lemonade
Lightning
Ocean
Outdoors
Outside
Parade
Park
Parties
Picnic
Pool
Popsicle
River
Road Trip
Rollerblading
Running
Sailing
Sandals
Sandcastle
Season
Shorts
Showers
Sightseeing
Sprinkler
Sunburn

Sundress
Sunflower
Sunglasses
Sunny
Sunscreen
Sunshine
Sweating
Swimming
Swimsuit
Tanning
Theme Park
Thunderstorm
Travel
Trip
Tropical
Vacation
Warm
Watermelon
Waterpark
Waterski

TV series

Alias
Ally McBeal
American Dad
American Horror Story
American Idol
Arrested Development
Battlestar Galactica
Bones
Boy Meets World
Breaking Bad
Bridezillas
Buffy the Vampire Slayer
Charmed
Cheers
Criminal Minds
Dancing with the Stars
Deadwood
Deadliest Catch
Desperate Housewives
Dexter
Doctor Who
Downton Abbey
Entourage
Everybody Loves Raymond
Family Guy
Frasier
Freaks and Geeks
Friday Night Lights
Friends
Fringe
Full House
Futurama
Game of Thrones
Glee
Gossip Girl
Hannibal
Heroes
Home Improvement
Homeland
House
House of Cards
How I Met Your Mother
I Love Lucy
King of the Hill
Law and Order
Lie to Me
Looney Tunes
Lost
MacGyver
Madmen
Malcolm in the Middle
Married with Children
Melrose Place
Miami Vice
Modern Family
Monk
MythBusters
NCIS
New Girl
Nip Tuck
One Tree Hill
Parks and Recreation
Pawn Stars
Pretty Little Liars
Prison Break
Psych
Rescue Me
Revenge
Revolution
Saturday Night Live
Saved by the Bell
Scrubs
Seinfeld
Sherlock
Smallville
Sons of Anarchy
South Park
Sponge Bob Squarepants
Star Trek
Storage Wars
Supernatural
The Big Bang Theory
The Cosby Show
The Golden Girls
The Good Wife
The L Word
The Magic School Bus
The Mentalist
The Office
The Simpsons
The Sopranos
The Twilight Zone
The Vampire Diaries
The Walking Dead
The West Wing
The Wire
The Wonder Years
The X Files
Tom and Jerry
Top Gear
True Blood
Two and a Half Men
Ugly Betty
Weeds
White Collar
Xena Warrior Princess

World cities

Amsterdam	Cologne	Lagos	Salvador
Athens	Copenhagen	Lima	Santiago
Baghdad	Damascus	London	Sao Paulo
Bali	Delhi	Madrid	Seoul
Bangalore	Dublin	Mecca	Shanghai
Bangkok	Florence	Melbourne	Singapore
Barcelona	Frankfurt	Mexico City	St Petersburg
Beijing	Hamburg	Milan	Stockholm
Berlin	Hanoi	Montreal	Sydney
Bogota	Havana	Moscow	Taipei
Bombay	Hong Kong	Mumbai	Tehran
Brussels	Istanbul	Munich	Tokyo
Budapest	Jakarta	Nairobi	Toronto
Buenos Aires	Jerusalem	Oslo	Vancouver
Cairo	Johannesburg	Paris	Venice
Calcutta	Karachi	Prague	Vienna
Cancun	Kiev	Rio de Janeiro	Warsaw
Cape Town	Kinshasa	Riyadh	Yokohama
Casablanca	Kuala Lumpur	Rome	

Appendix 5. Detailed Survey Results in Chapter 4

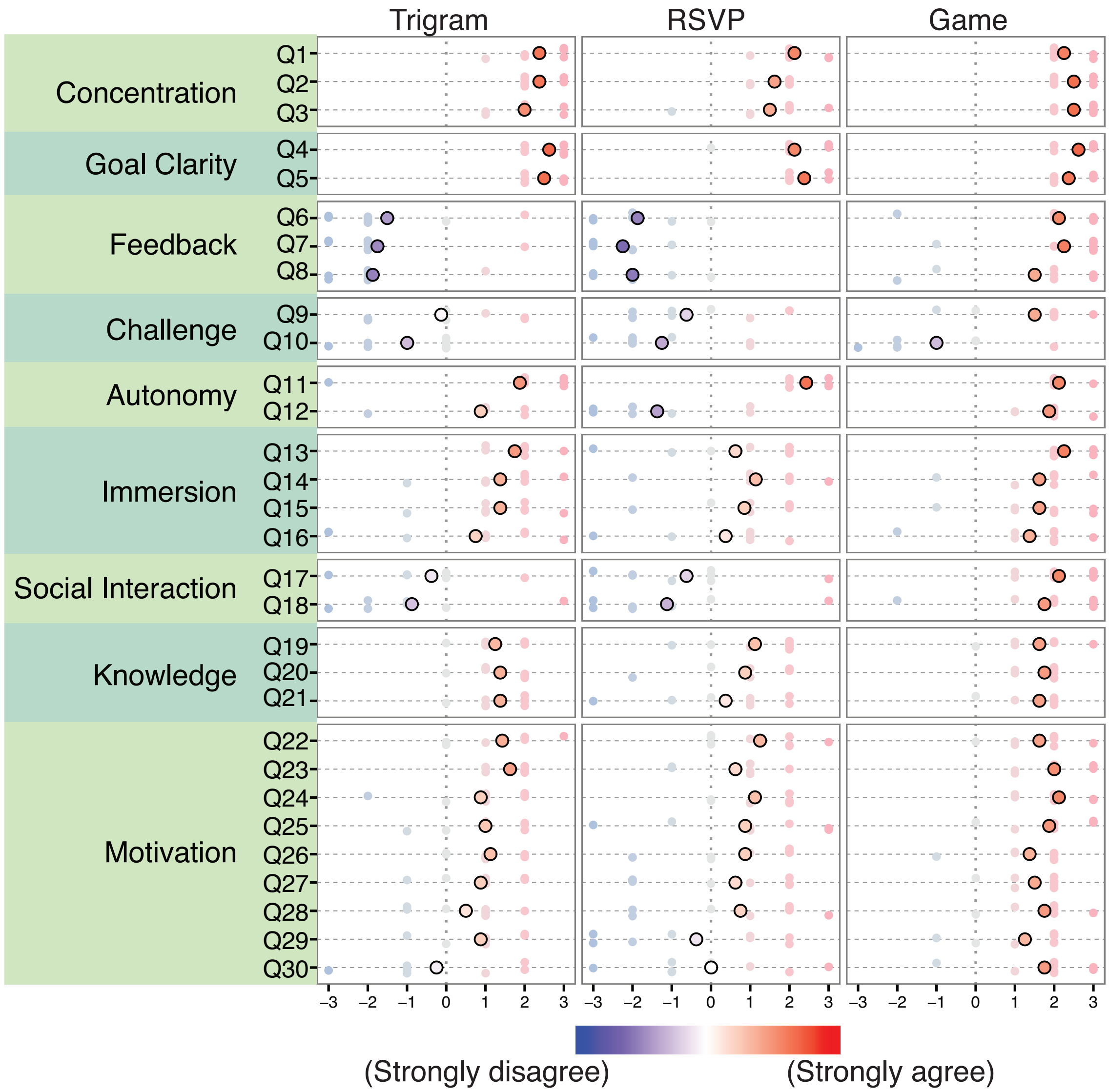

Figure A5-1: Detailed survey results with questions.

Circles with black outline: mean score of the 8 subjects. Translucent dots: individual responses. From -3 to 3, the scale corresponds to the levels from "strongly disagree" (blue) to "strongly agree" (red) in the survey.

Below is a list of all 30 questions used in the survey:

Concentration

Q1 – Most of the activities are related to the learning goal.

Q2 – The workload of the tasks is adequate.

Q3 – I am not distracted from the learning tasks I should concentrate on.

Goal Clarity

Q4 – The procedure is presented at the beginning of the task.

Q5 – The procedure is presented clearly.

Feedback

Q6 – I receive feedback on my progress throughout the task.

Q7 – I receive immediate feedback on each of my responses.

Q8 – I receive information on my overall performance immediately.

Challenge

Q9 – The task provides new challenges at an appropriate pacing.

Q10 – The difficulty of the tasks increases as my skills improves.

Autonomy

Q11 – I know what I am supposed to do at all times during the task.

Q12 – I feel a sense of control over the task.

Immersion

Q13 – I can become involved in the task.

Q14 – I become unaware of my surroundings during the task.

Q15 – I forget about time passing during the task.

Q16 – I temporarily forget worries about everyday life during the task.

Social Interaction

Q17 – The competition in the task is helpful to learning.

Q18 – I feel competitive toward other participants in the task.

Knowledge

Q19 – I want to improve the peripheral vision skills taught by the task.

Q20 – I catch the basic idea of the peripheral vision skills to be improved by the task.

Q21 – I feel the task increases my peripheral vision.

Motivation

Q22 – I want to get positive comments on my performance on this task.

Q23 – I feel I can control my performance on this task through my own effort.

Q24 – I derive a sense of satisfaction when I feel my performance on this task has improved.

Q25 – I prefer this task to be challenging as opposed to being easy.

Q26 – I regard every trial on this task as a new opportunity to improve.

Q27 – I set and maintain high standards of performance for myself in this task.

Q28 – I feel a need to try harder on this task if I suspect that my performance was poor.

Q29 – Even when this task is difficult, I expect to do well.

Q30 – I find it easy to concentrate on this task for long periods of time without becoming tired.