High-Variability Training in Second-Language Reading Instruction

by

Kevin Enns

A Thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Linguistics

University of Manitoba

Winnipeg

Abstract

High-variability (HV) training is a technique that uses a wide variety of tokens representative of a specific category's internal variation. Readers should be able to handle material written in both easy-to-read and stylized formats. The conventional approach to reading instruction is to start with low-variable examples before spreading out to less common cases.

This study performs two experiments comparing low-variability (single source) training against HV (multiple source) training of novice readers recognizing Chinese characters. Participants were tested on their ability to recognize characters they had trained with as well as their ability to generalize their training to unfamiliar variations. Results showed that HV training was associated with decreases in response time and increases in overall accuracy, including the ability to generalize training to novel variation not seen during training. This study found evidence showing that HV training is an effective method for English speakers learning to recognize Chinese characters.

Table of Contents

List of Figures

1 Language Variation and Highly Variable Stimulus

Second-language (L2) learners sometimes have a difficult time perceiving language properties that either do not appear in their first language (L1), or directly conflict with knowledge from their L1 (Best, 1995; Flege, 1995). Language teachers usually hope that what they teach in the classroom will be enough to prepare their students to handle real-world language experience. But students who appear to have learned well in a classroom environment often fail to perform in real-world situations. This type of failure can be partially attributed to the idea that students perform better when they are tested in a similar environment to where they were taught (i.e. environmental context-dependency bias) (Grant et al., 1998). Changes in environment and even small changes to surface-level presentation of material can cause disruptions to expectation, making it more difficult for students to recall previously learned material.

A 'language' is not a static entity. It is made up of many different people each speaking in their own individual way. Regional accents, personal mannerisms, and a speaker's unique biology provide a near infinite source of variation. Despite this, language users can still understand each other without difficulty. One of the more challenging aspects of learning any language is being able to handle surface-level variation (i.e. language-internal variation). High-variability (HV) training is a technique that uses a wide variety of tokens representative of a specific category's internal variation.

HV training has been given a lot of attention as an effective technique for teaching students to discriminate between competing L1 and L2 phonemes. Logan, Lively, & Pisoni (1991) used HV

training with Japanese speakers. Japanese speakers often have difficulty perceiving the difference between English [ɹ] and [l].

In English, /ɹ/ and /l/ are discrete phonemes and the tap [ɾ] is an allophone of /t/ and /d/. Japanese 'r' is an independent phoneme.[1] There is no equivalent phoneme for English /ɹ/ or /l/. When a Japanese speaker hears English /ɹ/ or /l/, they often include them as aberrant members of the category of /ɾ/.

English	/ɹ/		/t/		/d/	/l/	
	[ɹ]		[t]	[ɾ]	[d]	[l]	
Japanese	Ø	/ɾ/	/t/	/ɾ/	/d/	/ɾ/	Ø

Figure 1-1: English /ɹ/ & /l/ versus Japanese /ɾ/

Logan et al (1991) created HV training materials from six native English speakers (four male, two female) speaking different instances of the English /ɹ-l/ minimal pair. After participating in a phoneme identification task over multiple sessions, Japanese listeners were better able to identify tokens they had been trained with as well as instances of the /ɹ/ and /l/ phonemes in previously unheard English words. Strange & Dittmann (1984) had previously experimented with similar /ɹ-l/ minimal-pair discrimination tasks. Participants listened to synthesized speech examples generated from a single source (i.e. low-variability (LV)) and found that while discrimination of English /ɹ/ and /l/ could be improved for the synthesized tokens that skill did not generalize to natural speech

[1] Japanese 'r' is a lone non-nasal sonorant consonant which is marked in Figure 1-1 as /ɾ/. It includes English [ɾ], [ɹ], and [l] as uncommon allophones but is not equivalent to the English /ɾ/.

tokens. Logan et al (1991) attributed their results to the use of naturally produced tokens and not the HV nature of the training.

Logan et al (1991)'s findings were replicated and expanded on in Lively, Logan, & Pisoni (1993). Japanese listeners exposed to HV tokens were able to generalize their training to new words spoken by both familiar and unfamiliar speakers. Listeners who were trained by only a single LV source did not. The training used by Logan et al (1991) and Lively, Logan, & Pisoni (1993) produced better results than the training tokens used by Strange & Dittmann (1984) because of their use of HV training.

Continued research revealed that groups trained with HV stimuli also showed an increase in both long-term retention and production skill. Bradlow et al. (1997) demonstrated that adult Japanese listeners exposed to HV training tokens would transfer their perceptual learning of /ɹ/ and /l/ to the production domain. Hardison (2003) tested adult Japanese and Korean speakers with a combination of HV training and visual cues. She was able to replicate the results of Logan et al (1991), Lively, Logan, & Pisoni (1993), and Bradlow et al. (1997) insofar as HV training produced small, but statistically significant better results than equivalent LV training. However, Hardison (2003) qualified these results by claiming that the high variability in training was only partially responsible for increases in perception and the ability to generalize training to novel instances. "Although multiple talkers likely offer benefits in perceptual training because they are a component of a variable stimulus set, it is also likely that performance in generalizing to familiar

and unfamiliar talkers is related to the intelligibility of an individual's articulatory gestures" (p. 517).

Infants acquire information about their first language by collecting various examples spoken by multiple people (Kuhl, 2004; Maye, Werker, & Gerken, 2002; Singh, 2008). It would be a reasonable hypothesis to suggest that infants instinctually use HV stimuli during L1 acquisition. Children, and adolescents, could use HV training in combination with their increased brain plasticity for both L1 and L2 acquisition. A comparison of the ability of young children (6-8 years), older children (8-12 years), adolescents (15-18 years), and adults (25-59 years) to learn the Japanese-English /ɹ-l/ paradigm while using a HV training program revealed that adolescents showed the most significant improvement and adults showed the least improvement[2] (Shinohara, 2014; Shinohara & Iverson, 2013, 2015).

Giannakopoulou, Brown, Clayards, & Wonnacott (2017) tested HV and LV conditions to train the English phonemic contrast /i/ vs /ɪ/ with native Greek speakers. However, experimental results comparing children and adults' ability to perceive this contrast found that both age groups benefited more from LV training tokens over HV. Adults benefited slightly from the HV training but only inconsistently when facing novel speaker and word variants. Children showed benefits

[2] Adolescents > Older Children > Younger Children > Adults (Shinohara, 2014; Shinohara & Iverson, 2013, 2015)

from HV during the initial stages of learning, but in later sessions they received primary benefit from the LV training.

There are several possible explanations for the apparent inconsistency of the efficacy of HV versus LV training and the disparity across the various age groups. Wang & Kuhl (2003) make the point that the inconsistent performance of infants and younger children is due to the lack of pre-existing mental categories. A pre-literate child without explicit phonological awareness, experiencing a novel phonetic category in an L2 would not have a sufficiently complex "mental map" to connect that experience to a pre-existing L1 category. Giannakopoulou, Brown, Clayards, & Wonnacott (2017) suggest that the more important factor to consider is the number of training/testing sessions. Children learn more slowly than adults do, so the more training sessions a child has, the more likely positive results will manifest. Perrachione, Lee, Ha, & Wong (2011) propose that a subject's personal aptitude for perception is the key factor. They claim "high-variability training enhanced learning only for individuals with strong perceptual abilities. Learners with weaker perceptual abilities were actually impaired by high-variability training relative to a low-variability condition." (p. 461).

Despite its mixed successes, explicit HV training is still primarily studied as a method for acquiring phonetic discrimination in an L1 or L2. There have been attempts to use the HV method for other more complex speech patterns. Clopper & Pisoni (2004) have had limited success using HV training to teach people to categorize English dialects. However, to the author's knowledge, there has never been any explicit experimental research testing the effectiveness of HV training on

reading instruction. Written language, like spoken language, contains a high degree of internal variation in the form of contrast, viewing angles and typography.

Brown & Carr (1993) performed a series of experiments testing student's response times for recognizing words and pseudowords presented in hand-written and digital fonts. They found that, despite spending the same amount of time training, students who first learned pseudowords in hand-written form had little difficulty in transferring that skill to typewritten examples, but students who were trained with these easy-to-read digital fonts had trouble with handwritten text. The asymmetric results from Brown & Carr (1993) show that some types of training tokens are more effective at producing generalizing results than others.

Brown & Carr (1993) acknowledge that the weight given to the digital font tokens were probably affected by the participants previous lifetime experiences with printed text. Because of this, Logan (1997) claims that during the initial stages of reading instruction, the best results come if the student is only given simple, easy-to-read text.

> Readers need to be able to decipher all sorts of fonts, from Century to squiggly handwriting. Research on font variation suggests that reading instructions do not have to worry much about this issue. Transfer among professionally printed fonts is excellent. Handwritten fonts sometimes show poor transfer, especially when the handwriting is unique and atypical. However, teachers' handwriting is usually clear and typical, and so should provide no problem. The natural variation in printed and handwritten fonts experienced in the first few years of reading instruction should prepare students reasonably well for the occasional atypical fonts they will experience later in life. (G. D. Logan, 1997, p.139)

Logan (1997) claims that a teacher's clear and typical writing (along with any ad hoc encounters with natural occuring variation) are sufficient for a student to generalize to the majority of the variation they will encounter. It is entirely possible that a student will encounter a sufficient subset of variation without explicit exposure during instruction. I will not make the claim that explicit exposure to variation is required for category creation. However, Logan (1997)'s proposition should be re-evaluated in consideration of the environmental context biases outlined by Grant et al. (1998). Grant et al (1998) describes how small surface-level changes disrupt expectation, making it more difficult for people to recall previously 'learned' material. If a student does not encounter a sufficient subset of variation then their recognition will be disrupted by their first encounter with material that differes from their training regimine. HV training could be an effective method for introducing people to category-internal variation, and in doing so, teaching them to process typographic variation more efficiently and overcome many of these context biases.

To date, the existing body of research discussing HV training in the context of reading instruction has described L1 reading instruction. Learning to read an L2 presents many different

challenges. L1 reading is accompanied by the concurrent development of explicit mental categories for allophones and phonemes (i.e. phonological awareness). The mental categories established for an L1 are often insufficient and/or directly conflict with categories for the phonemic inventory of an L2[3].

Orthographic representation must also be considered. An orthographic symbol might represent a set of phonemes in an L2 that is different from the set in an L1. The ability to match text with its phonetic expression depends in part on how closely the number of phonemes in a language correspond to the number of orthographic tokens available (i.e. orthographic correspondence). For instance, English has more phonemes corresponding to a single orthographic token than does Serbo-Croatian. When written text is partially obscured or difficult to read, a Serbo-Croatian reader has an easier time 'mentally repairing' the text because there are fewer options for what the obscured text could be (Frost & Katz, 1989).

An L2 beginner reader will not only have to organize the new inventory of phonemes, but also remap them to a different set of orthographic symbols. There are further complications if the target L2 uses a completely unfamiliar writing system, like an L1 English speaker learning to read Chinese or Arabic scripts.

---

[3] See §4 for a more detailed discussion of phonemic categories and their creation.

2 Multi-Modal Language Perception

The areas of the brain responsible for language processing are independently linked to both the auditory and visual centers. Bernstein et al. (2002) observed via fMRI that the speech processing centers in the brain[4] can be activated by visual stimuli without directly activating the primary auditory cortex. These types of neural cross-activations support an integrated model of perception. Some non-nativist approaches to language acquisition make the claim that the visual and spoken language processes use many of the same perceptual mechanisms (Bybee, 2013; Campbell, 1995).

Both modalities deal with signals that can be weak, incomplete, or contain noise. Both spoken and written language signals have information that may be missing, redundant, or completely unnecessary for interpretation. The ability of a beginner reader to extract information from a weak or noisy signal depends on how well a person knows which parts of the signal are necessary for understanding and which parts are not.

Despite the existence of cross-modal similarities, visual and auditory perception manifest several differences in the way signals are processed and the way that imperfect signals are dealt with. A key difference can be found in the nature of the signal itself. The information stream in auditory signal is always interpreted in real-time. There is no way to pause and fixate on a

[4] Specifically, the superior temporal gyrus and areas surrounding the primary auditory cortex.

(non-repeating) sound. Real-time auditory perception can be improved by using redundant information or additional context such as discourse cues, situational contexts, or gestures.

Visual signals (insofar as they are used for reading) are limited by the observer's field of view. A person focusses on what is in their foveal (i.e. the fixation point) and parafoveal regions. Information outside of these areas (in the peripheral view) can be detected but not with any degree of detail. Even with explicit training a person cannot detect items in the peripheral range with enough detail to be able to read (Chung, 2007).

A beginner listener will listen to their training sources over and over to identify and extract the features necessary for understanding because there is no (non-technological) way to fixate on an auditory signal for typical spoken language experiences. Increasing fixation (i.e. spending a longer time looking at the material) is a common visual strategy for dealing with complicated or distracting written text (Vlaskamp & Hooge, 2006).

Perrachione et al (2011) criticizes HV training as being beneficial only for persons with a sufficiently high pre-existing perceptual skill. The difficulty that people with lower perceptual skills have, may be the result of the real-time requirement for auditory processing. The use of HV training in a modality where visual fixation is a viable option could be sufficient to overcome perceptual limitations.

3 Deciphering the Signal

When children learn to understand spoken language, they pay careful attention to the cues embedded in the speech signal. By listening to multiple examples of the same token, they learn to recognize which linguistic cues are important and which can be ignored. They learn the intonational differences between questions and commands, the prosodic meanings behind shouting and whispering, and how to recognize different people's voices. Children quickly learn that the actual words spoken are only a part of the whole meaning of an utterance. Full meaning is understood through a combination of multiple cues working together. The same process can be applied to learning to read.

Writing is the interface between spoken language and image that encodes multiple levels of meaning into a series of orthographic characters. Bellantoni & Woolman (1999) describe two levels of meaning in a printed word: the word image, referring to the dictionary definition; and the typographic image, referring to the holistic impression made by the visual presentation. The typographic image can encode information relating to culture, gender, and social status (Stöckl, 2005). Typography and presentation can create meaning, not just with the letterforms, but with color, three-dimensionality, texture, spacing, or orientation (Kress & Van Leeuwen, 2002). Many typefaces incorporate semiotic elements that blur the boundaries between image and letterform (Leeuwen, 2005) as illustrated in the figure below.

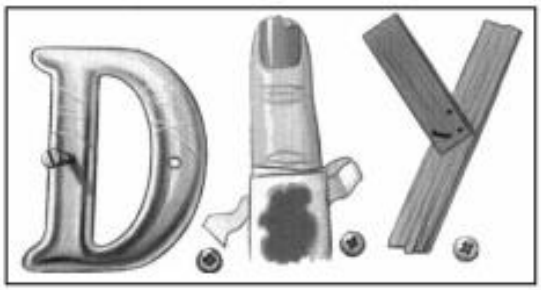


Figure 3-1: Artwork from 'The Guardian, Weekend', 17 July 2004 (Stöckl, 2005)

The typographic image of written language can be manipulated to represent specific features of spoken language such as stress (Brumberger, 2004; Kumpf, 2000). An expert reader can extract meaning from both levels of the text simultaneously.

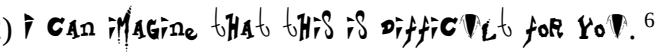
The ideas that typography encodes additional meaning is held as a self-evident truth by artists and graphic designers but has been given surprisingly little study outside of those fields. Linguists have largely ignored the implications of typography and design on reading and its effects on second-language education. Oppenheimer & Frank (2008) go so far as to claim that major theories of language categorization do not sufficiently address issues of presentation, especially for elements that are not explicitly relevant to the feature space. They cite that font size in particular is ignored, despite it having a large impact on perceptual fluency (Mueller, Dunlosky, Tauber, & Rhodes, 2014; Reber, Winkielman, & Schwarz, 1998).

3.1 Letter Identification through Features

Letter identification is a front-end reading task and an essential part of the 'word image' reading process. Consider the following characters:

(1) '', '  ', and ''. [5]

Now consider the following sentence:

(2)  [6]

What makes it so difficult to identify the individual characters in (1)? Why is the sentence in (2) more difficult to read than the rest of this document? The obvious answer is 'Because the letterforms are abnormal'. There is a high degree of irregularity in the letter shapes. Some letters, like the 'u' are represented by a solid geometric shape instead of the typical curved line. The sentence is made up of a mix of lower and uppercase letters, some letters are written as script and other in block type, some are written with thin lines and others with thick.

The examples above were written in a font called Hieronymous Boschian. Readers are sensitive to familiarity and consistency within the typographic image. People will have an easier time parsing information if they've seen it before, or if it is similar to other information nearby (Sanocki

---

[5] 'u', 'f', and 'd'

[6] 'I can imagine that this is difficult for you.'

& Dyson, 2012; Whittlesea & Leboe, 2000). For many readers, this will have been the first time you have ever experienced reading anything written in this font. The wide variety of letterforms is because each of the characters was designed by a different person (Gauthier, Wong, Hayward, & Cheung, 2006). Hieronymous Boschian is difficult to read because the style of each character is independent from the others. Readers are better able to recognize texts when it 'matches', even if the writing system is unfamiliar (Gauthier et al., 2006). Inconsistency in presentation causes slower reading speed and efficiency. This phenomenon is clearly seen in the Stroop Effect, where readers are more prone to recognition errors when a color word is not written in the color of the ink that the name of the color denotes (Stroop, 1935).

Context is also important, seeing strange characters in the context of a word or sentence helps to identify the most likely candidate for each character. It is quite difficult to identify the three single characters in (1), but reading an entire sentence made up of thirty-six characters in (2) can be done with much less difficulty.[7] Using context is not always an option for a beginner. A new learner cannot always figure out characters in the context of words because they do not know the whole words yet. For them, the reading process begins by learning to identify each character

[7] Identifying the example in (2) as 36 individual characters presumes that the reader has sufficient knowledge of alphabetic writing to understand that these words are comprised of separate characters. It is possible that a reader unfamiliar with alphabetic writing systems to perceive each word as its own independent character or glyph. In this case, the example in (2) would be perceived with the same difficulties that example (1).

individually. Character identification is done by breaking them down into component features (LaBerge & Samuels, 1974).

As a person reads, they perform a series of eye movements (called saccades) and fixations. Any characters located in the foveal and parafoveal view during a fixation get broken down into component features to be identified (Pelli, Burns, Farell, & Moore-Page, 2006; Pelli, Palomares, & Majaj, 2004). Character features can include lines, angles, basic geometric shapes, curves, line segment length, rotation, viewing angle, and contrast (Pelli, Farell, & Moore, 2003).

A typical saccade length for an English reader is 15 characters (dependent on viewing distance and text size) or 5 degrees of viewing angle (McConkie & Rayner, 1976). A saccade for a Chinese reader includes 2-3 characters or 3.2 degrees of viewing angle (Chen & Tang, 1998**).** Words and phrases longer than 15 characters in English (or 2-3 in Chinese) would require multiple fixations. (McDonald, 2006).

The amount of information in a single fixation depends on several factors including character complexity, viewing angle, stroke width, and other typographic features. There is an upper limit to how many features can be detected at once (Pelli et al., 2003). If the fixation were restricted to a single character, all focus could be applied to that single character, making it extremely likely that all possible features could be identified and extracted. However, if that character was viewed alongside 15 others, then the upper feature limit would have to be applied across all the characters within the fixation. Each individual character would receive less focus and some essential identification features could be missed.

Pelli et al. (2006) examined the ability of a new reader to learn characters from previously unknown writing systems. They found that simple characters with fewer features are easier to identify, but any character, when presented in isolation, can be perceived at native level proficiency within 3000 exposures. However, when readers looked at a previously recognizable character in the context of others in a character cluster their recognition of the previously known character was reduced to beginner levels. A beginner's recognition fluency breaks down when they are forced to spread their finite feature detection capacity across multiple characters. Their limited experience with the language prevents them from being able to pick out the most salient features in the target characters. Their focus is misspent on features that are not necessary for identification.

A person just learning to read would have to learn which features are critical for identification and which are not. Ideally, the limited perceptual span should be able identify the features that are critical for inter-category identification and ignore any features that represent surface-level category internal variation. A new reader must figure out (or be explicitly taught) which features are most important for their identification and which features can be ignored. Consider the difference in the characters 'O' and 'Q'. The difference between this pair of characters comes down to a single line feature. Experienced readers are most sensitive to features that could turn one letter into another (Gauthier et al., 2006). When making the determination between these two characters, the experienced reader will focus on that single line feature, because that line is what makes the difference between competing characters.

3.2 Words & Context

When new readers focus their limited feature detection abilities on a cluster of characters, they need to be able to identify enough features to successfully identify the entire cluster. But, as described previously, individual feature detection is not the only strategy available for reading. As a person's reading ability increases, they rely on the statistical likelihood that certain clusters of characters are more likely to co-occur (character collocations) and other larger discourse contexts.

Rumelhart & McClelland (1982) proposed an interactive activation model for letter perception (See Figure 3-2 below) that combines bottom-up processing from individual features but also allows for elements of larger top-down contexts.

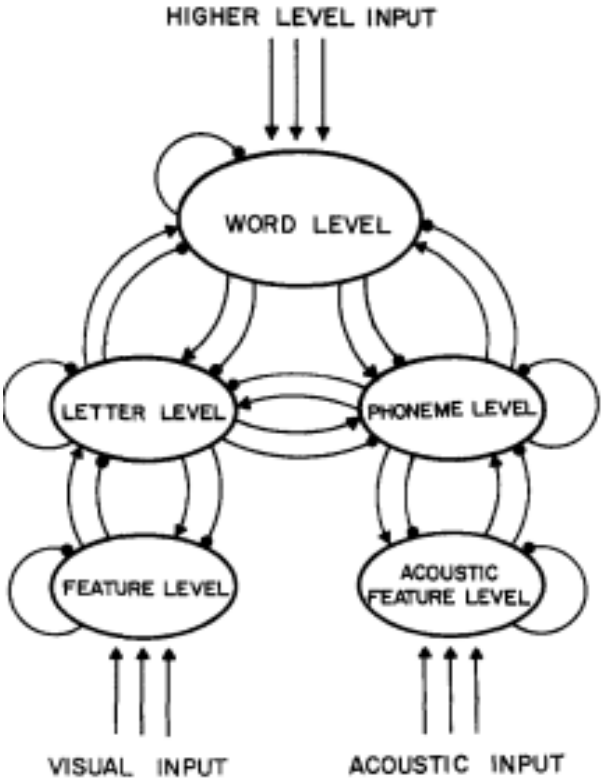


Figure 3-2: Processing levels involved in visual and auditory word perception with interconnection. Each processing stage can facilitate (arrowhead lines) or inhibit (dothead lines) other interconnected stages (McClelland & Rumelhart, 1982)

This model says that contexts larger than individual characters can affect word perception. It also says that the different levels of processing interact back and forth with each other. When presented in isolation it is difficult to determine that the character '▼' written in Hieronymous Boschian represents the letter 'u', but in the context of the entire word, ▼▼▼▼▼▼; the identity of the character becomes much more obvious.

This contextual processing contributes to the 'word superiority effect'. The word superiority effect occurs when a letter shown by itself will take longer to recognize than if that same letter

were presented within the context of a word (McClelland & Johnston, 1977; Wheeler, 1970). There are some limitations to the word superiority effect. Pelli et al (2003) argue that "human word identification never exceeds the accuracy attainable by strictly letter-based models." (p.753). They claim that the supposed benefits gained from the 'word superiority effect' is actually the result of readers being more likely to "incorporate their historical confusion probabilities when reading words than when individual letters" (p. 755) (i.e. statistical learning).

Nearby characters are not always helpful. 'Crowding' is a type of interference where nearby characters inhibit character identification because a person's attention is spread across too many characters in a limited space (Whitney & Levi, 2012). Adults (experienced readers) are better at dealing with crowding than children (inexperienced readers). Jeon et al (2010) suggest that this is due to the developmental differences between adults and children. Adults can deal with crowding because they have a more mature visual cortex. This effect could also be due to adult's having more experience with reading in general. Crowding effects are worse when the adjacent characters are perceived as symbols instead of clusters of letters (Grainger, Tydgat, & Isselé, 2010). It isn't clear to what degree valid but unknown characters are perceived as symbols by beginner readers, but it is likely that these readers experience the effects of crowding to a greater degree than experienced readers do.

The way crowding manifests, and is subsequently managed, is different depending on the writing system. With an alphabetic writing system the easiest way to reduce crowding is to decrease the amount of information in the fixation by either increasing typographic size or adding

additional space between characters (i.e. kerning) or words (Larson, Hazlett, Chaparro, & Picard, 2007; Pelli et al., 2007). These strategies are not as effective in non-alphabetic writing systems.

The typical Chinese writing convention doesn't use spaces between words, so in typical situations the white space strategy is limited to kerning despite research showing that separating Chinese words with white space or even distinguishing the words by alternating color does facilitate reading speed (Perea & Wang, 2017). Crowding effects in Chinese writing is highest when the target and flanking characters have similar stroke counts (Zhang, Zhang, Xue, Liu, & Yu, 2009). For readers, the most effective way of overcoming crowding effects, regardless of the writing system, is practice. Repeated exposure through frequency training increases reading speed and overcomes the effects of crowding (Chung, 2007; Chung & Truong, 2013).

4 Categorization

As people learn, they absorb information and do their best to organize it. Language data are collected and abstracted into categories. Categories can vary significantly in number of items and degree of internal variation. It is also possible for a token to belong to multiple categories simultaneously as shown in the figure below.

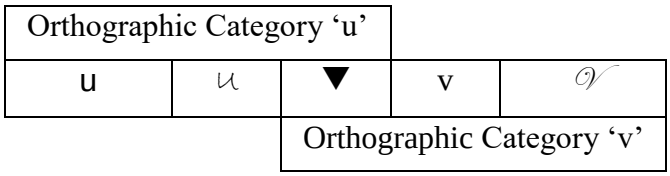


Figure 4-1: Orthographic Category Distribution

Category boundaries develop based on how often specific examples turn up (i.e. frequency) (Aslin, Saffran, & Newport, 1998; Jacoby & Dallas, 1981) as well as how easy it is for recognizable features to be extracted (Elio & Anderson, 1981; Pelli et al., 2006).

There are multiple linguistic frameworks for understanding language categorization. They can be roughly divided into two broad frameworks; the abstractionist and instance types. The abstractionist viewpoint claims that each mental category contains only one absolute representation for every word, morpheme, or phoneme independent of the listener's experience (Chomsky & Halle, 1968; Prince & Smolensky, 2008). Any variations are derived from the core abstraction. In this model people understand speech by connecting what they hear with the abstract representation. Any superfluous information from the signal must be stripped away for recognition to occur.

Instance model frameworks are fundamentally different. The underlying assumption behind instance theories are that new experiences impact the entire system. Tenpenny (1995) identified this as a key weakness of abstractionist theories, arguing that episodic theories are better predictors of word identification. These models support the positions of Pierrehumbert (2002) and Wedel (2006), who showed that people pay attention to fine-grained phonetic details for word identification, and that their lexical representation includes this type of information. Logan (1990) claims that "general instance theory can account for a broad range of cognitive phenomena." (p. 32). This would, of course, include reading and writing.

Several categorization models use the instance framework. Two of the most popular models are the prototype model and the exemplar model. In prototype theory, the incoming stimulus is compared with a 'best' exemplar, typically a high-frequency token, called a prototype (Iverson et al., 2003; Kuhl, 1991; Kuhl & Iverson, 1995). This prototype is an exemplar that is used as canonical representation of the category. A prototype is similar to the abstractionist model in that both use a single representation for their category but differ in that the prototype representation is based on the direct experience of the listener. The representative for categories for both the prototype and exemplar theories rely on the experiences of the listener.

In exemplar theory, a category is not a single token, but an entire collection of memorized tokens, often called exemplars (Gahl & Yu, 2006). Other exemplar models suggest that the category representative is made up of the statistical average of all category members (Bybee, 2013; Pierrehumbert, 2001).

The prototype and exemplar models vary on precisely how new information is initially assigned to or integrated within existing categories. Experienced stimuli are compared to a representative of the group of exemplar tokens. Recognition in both prototype and exemplar theories happens by mapping stimuli onto the exemplars in the mental lexicon.

New stimuli are compared to their category representation to determine what category the stimulus should belong to. A listener would be able to use a perfect category representation to accurately represent all valid examples of a category while rejecting any invalid examples. This implies that every stimulus could be classified as 'correct' or 'incorrect'. An allophone (or its

written counterpart, allograph) is only ‘correct’ insofar as it contains enough identifiable features to be classifiable. If the stimuli are close enough it is incorporated into the appropriate category, and the category representation is modified by the new information. An ideal observer would be able to examine all possible features without distraction or bias. They would also be able to determine to properly account for features that are categorically representative and which features are manifestations of surface-level category internal-variation. The real-world observer is often subjected to distractions and may give undue weight to unnecessary features, requiring them to rely on other contextual clues to determine category membership.

When a person is faced with a token that they have never encountered before or one they are unable to categorize, it will not have any sort of correctness value and the token might become a category of one or ignored entirely. It can be challenging for an L2 learner to create categories because it is often difficult to know whether two tokens should be sorted into two separate categories or be included in just one. Through multiple repeated exposures to information, categories may merge, separate, and have their category representation modified significantly. For many adult L2 learners, the initial process of categorization, comparison, and representation is done with explicit conscious effort. However, once enough information has been gathered, categorization becomes implicit. Implicit categorization is faster and is done without conscious effort. It is one of the key components of perceptual fluency and necessary to become an expert reader (Oppenheimer & Frank, 2008).

5 Reading in a Second-Language

From a linguistic point of view, reading is generally considered to be a secondary language process that is derived from spoken language (Olson, 1996; Perfetti, 1985). Despite this, a substantial body of research indicates that writing fundamentally rewires the parts of the brain responsible for language (Dehaene et al., 2010). Learning to read exploits the brain's natural plasticity and forms connections that allow the integration of multiple cognitive processes.

Simply being able to speak in your first language provides some basic information of that language's structure. At some level, a speaker will implicitly understand their language's inventory of phonemes and words and the relationship between them. They come to understand which features of the surface form are critical for interpretation and which are not. As a person learns to read, their phonemic awareness becomes more explicit and is reinterpreted in the context of their language's writing system (Hatcher, Hulme, & Ellis, 1994; Liberman, 1973; Liberman, Shankweiler, Fischer, & Carter, 1974; Morais, Bertelson, Cary, & Alegria, 1986; Wagner et al., 1987).

Knowledge about your first-language might be helpful or harmful to learning a second language. An L1 learner can create categories inductively, drawing generalizations from the available set of data (Aslin et al., 1998; Elio & Anderson, 1981). L2 learners can approach learning their second language the same way as their first, paying attention to the language single and creating categories inductively. However, they could also create categories deductively, starting

with the meta-knowledge of the existence of categories and their phonemic awareness from their first language, and create and modify existing categories as required.

Recognition can be explicit (consciously effortful) or implicit (automatic). Automatic processing in reading refers to the combined tasks of individual character recognition, word recognition, and the integration of all necessary contextual information from the typographic image. An expert reader recognizes as much as possible automatically. Automatic processing is faster and more accurate, frees up cognitive resources, and ultimately restructures all the processes relating to recognition (LaBerge & Samuels, 1974; Logan, 1988; Segalowitz & Gatbonton, 1995).

Automatic processing comes with repeated practice and learning what to pay attention to. It is important when developing automaticity for reading that the automatic processing be able to cover a wide variety of surface forms. Even the smallest changes, such as a typographical change, can shift the recognition path from automatic to explicit. New readers and readers with low-skill are particularly affected by these sorts of changes (Perfetti, 1992).

6 Current Study

There is a growing body of literature discussing the effectiveness of HV training for long-term retention and recognition in phonological perception and production. Despite the lack of explicit experimental research using HV training for reading instruction, the existence of shared perceptual mechanisms means that is reasonable to assume that the principles that apply to spoken language could be applied to reading as well. The current study explores whether using HV training in L2 reading instruction is an effective way of teaching people to deal with variation in written text.

Variation in written language comes in several forms; contrast, colour, and font are among the most significant. These features make up a written character's typography. I am not claiming that HV training is *required* to deal with typographic variation, but I do propose that, based on experiments with spoken language, training people with HV could be an effective method for teaching people to process typographic variation more efficiently. Samuels (1997) argues that learning from a wide variety of diverse [auditory] tokens allows for better automatic recognition in spite of surface form variation. And when we consider the presence of asymmetric transfer between difficult and easier to read text (Brown & Carr, 1993), an HV approach to training should be better than a training method that focuses on the simpler LV approach advocated by Logan (1997).

This study includes two experiments[8]. The methodology for both was based on Moreland & Zajonc's 1977 experiments investigating the role of exposure effects on affect and recognition. Moreland & Zajonc (1977) showed their participants a series of Japanese kanji[9] at different exposure rates (2 kanji assigned to be shown either 1x, 3x, 9x, and 27x for a total of 80 tokens). After being shown the 80 tokens participants were shown each of the 8 kanji used during the training in succession. One group of participants were asked to evaluate each of the kanji from the original presentation for recognition (yes/no). A second group was asked to determine whether

[8] Both experiments were performed in parallel with each other.

[9] Moreland & Zajonc (1977) refer to kanji as ideographs.

they recognized the character but was also asked to rank their confidence in their response (on a 7-point scale).

The current study uses Chinese characters (which are closely related to Japanese kanji). Single character recognition is one of the earliest skills developed when learning to read. At this stage, character recognition is performed through explicit feature recognition. But as a person's skill grows, they begin to parse larger and larger chunks. In alphabetic writing systems, character combinations are particularly important as there is little practical utility in reading only single characters.

While using character collocations is a necessary and inevitable strategy for effective reading, it is outside the scope of this experiment. Based on Pelli et al. (2006)[10], single characters are more easily identified because all attention is applied to a single character. Upper limits to feature detection within a visual fixation mean that each character in a cluster need to be identified with reduced focus.

The first experiment tested the effectiveness of HV training for the recognition of single characters (unigrams). There is limited practical utility in learning to read single characters but single character identification is more slightly more useful in the Chinese writing system where many characters (and all of the characters used in this study) represent entire words. The first

[10] See §1.1.3 for more details

experiment is intended to provide a baseline for the simplest possible reading case. The second experiment has participants recognize a single character in the context of a pair of characters (bigrams). Although identification characters in bigrams is still a simpler task than reading in a real-world environment, it is a much harder task than single character identification (Pelli et al., 2006) and provides a more robust scenario for testing the efficacy of HV training.

For both experiments typographic variation was created by using multiple fonts.[11] Typography manipulates many of the surface-level features of the character shape making it a useful source of variation for testing whether or not HV training can be effective at overcoming biases associated with the context dependency biases described in Section 1.

Figure 6-1: Unigram Typographic Variation

Both experiments used the same three conditions.

1. Low-variable training and testing (LV/LV).

   This condition uses the same single font (SimSun) for both training and testing.

[11] See Appendix 1 for detailed information on the specific characters and typefaces used.

2. Low-variable training and high-variable testing (LV/HV)

   This condition uses one font (SimSun) during training and a variety of fonts during testing.

3. High-variable training and testing (HV/HV)

   The third condition uses multiple fonts for both training and testing.

The main hypothesis for both experiments is that HV training will make participants faster and more accurate than people in the LV training condition. If a person can develop a stronger category representation for orthographic characters through explicit HV training, then the recognition process would become more efficient, increasing both response speed and accuracy over participants that that only received LV training.

I hypothesize that HV training will prepare participants to recognize characters in fonts that they have never seen before (i.e. training generalization). I further hypothesize that the benefits of HV training will be larger when applied to bigrams (Experiment 2) over unigrams (Experiment 1).

7 Experiment 1: Unigrams

7.1 Participants

Twenty-seven English-speaking students recruited from the University of Manitoba. All participants reported no knowledge of the Chinese writing system or its derivatives[12]. Participants were compensated monetarily or with course credit.

7.2 Stimuli

This experiment made use of the following:

List of 55 Chinese Characters[13]

Chinese characters with concretely imageable definitions (Schwanenflugel, Akin, & Luh, 1992) were extracted from a list of the most frequently used Chinese words. Each character had between 4 and 9 strokes.

10 Truetype Fonts

Ten publicly available fonts capable of displaying simplified Chinese Characters. Fonts were evenly divided between 'print' styles and 'hand-written' styles.

---

[12] Persons with prior knowledge of Chinese (Traditional), Chinese (Simplified), Korean Hanza, and Japanese Kanji were excluded.

[13] See Appendix 1 for detailed information about the Chinese characters and fonts.

7.3 Methodology

The experiment was broken into three parts. An initial training session [1] immediately followed by a testing session [2]. A second testing session [3] was performed two days after the initial testing [2] to fully test the long-term effectiveness and retention of high-variable training[14].

Each participant was randomly assigned to one of three experimental conditions. Participation across each condition was even, with 9 participants per condition. One participant assigned to LV/HV condition did not complete the second testing session. Data collected from this participant's single session was retained.

<table>
<tr><th colspan="2">Training</th><th></th><th colspan="2">Testing</th></tr>
<tr><td rowspan="3">Single Character Training</td><td rowspan="2">Single Font (LV Training)</td><td>LV/LV</td><td>Single Font (LV Testing)</td><td rowspan="3">Unigram Testing</td></tr>
<tr><td>LV/HV</td><td rowspan="2">Multiple fonts (HV Testing)</td></tr>
<tr><td>Multiple fonts (HV Training)</td><td>HV/HV</td></tr>
</table>

Figure 7-1: Experiment 1 – Conditions

All the stimuli were presented on a Samsung SyncMaster T240 color monitor with 5ms refresh rate and 20000:1 dynamic contrast. The monitor brightness was set to factory default settings. The experiment was controlled by a Dell DCNE desktop running E-Prime 2 Professional Software

[14] Dumay & Gaskell (2015) showed that for newly learned words to fully integrate into a person's lexicon requires at least one night's sleep.

Suite (v. 2.0.8.74) developed by Psychology Software Tools. Subjects were seated approximately 50cm from the screen.

Data were analyzed in R (v3.3.2) using mixed-effects models. For reaction time data and mixed-effects linear models were fit using the lmer function of lmerTest (v3.0.1), with participant and item (target character) as random effects. For accuracy data, mixed-effects logistic models were fit using the glmer function, with family="binomial" and with participant and item as random effects. Unless otherwise indicated, all statements in this thesis about a fixed effect being significant are based on (for reaction time or d-prime) the p-value of the t-statistic of the relevant model coefficient using the Satterthwaite approximation for degrees of freedom, or (for accuracy data) on the p-value of the z-value of the relevant model coefficient. The full summaries of these models can be found in Appendix 2. Statements about a fixed effect being non-significant are based on its p-value being greater than 0.05 when that fixed effect is added to the nearest minimal model given in the appendix.

7.3.1 Training

From the 550 typographically unique tokens (55 characters in 10 fonts) each participant was shown one unique set of 55 training tokens (one set per participant).

Five Chinese characters were randomly selected and assigned an exponential exposure frequency [1x, 4x, 9x, 16x, and 25x][15]. All the training tokens for the LV/LV and LV/HV group were shown in the SimSun Truetype font. Tokens used for in the HV/HV group were assigned one of five randomly selected font[16]. Each training token was black on a white background and displayed at a 144pt size. Tokens were shown for 5 seconds before automatically progressing to the next. Each token displayed a Chinese character with its English definition, as seen in the figure below.

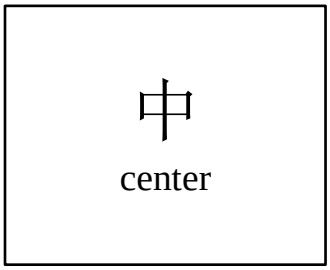


Figure 7-2: Unigram Training Token

Participants were told that they were going to be shown a series of Chinese characters with their corresponding English definitions and that after the training session they would be given a quiz.

[15] An exponential frequency progression was chosen over logarithmic to facilitate a balanced distribution of character and fonts during each of the experimental phases.

[16] 55 tokens / 5 = 11 tokens for each of the 5 fonts

7.3.2 Testing

Each participant was shown a unique set of 10 testing tokens. Five new randomly selected characters were added to the five characters used in the training session to form the testing character tokens. Tokens in the LV testing condition were displayed in the SimSun Truetype font. Tokens in the HV testing conditions were randomly assigned one of ten fonts. Each testing token was black on a white background and displayed at a 144pt size. A sample testing token can be seen in the figure below.

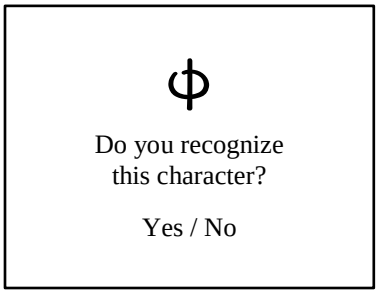


Figure 7-3: Unigram Testing Token

Participants were asked if they recognized the character from the training session. Recognition response (either yes or no) and response time was collected. If they responded ‘yes’ then they were prompted to type in the English definition as seen in the figure below. Both the definition response and speed were collected.

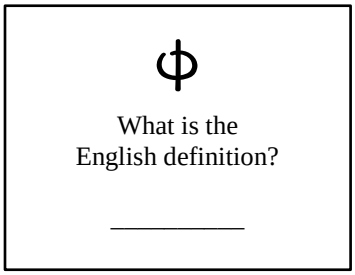


Figure 7-4: Unigram Definition Prompt

The testing in the second session two days later used the same tokens as the first session but were presented in a different order.

7.4 Results

The two key measurements were the response times and accuracy rates[17]. This experiment provided four distinct response types. There were two types of correct answer; correctly identifying a character that was part of their training (i.e. correct trained), or correctly identifying a character that was not (i.e. correct untrained). Similarly, there were two types of errors. Either a participant failed to recognize a character that was part of their training (i.e. a false negative), or they thought they 'recognized' a character that wasn't part of their training (i.e. a false positive).

[17] The following analyses include all recorded data. No outlier data was removed. Additional data and analyses (including analysis of sensitivity using a d-prime analysis using Macmillan & Creelman's (2004) method can be found in Appendix 2).

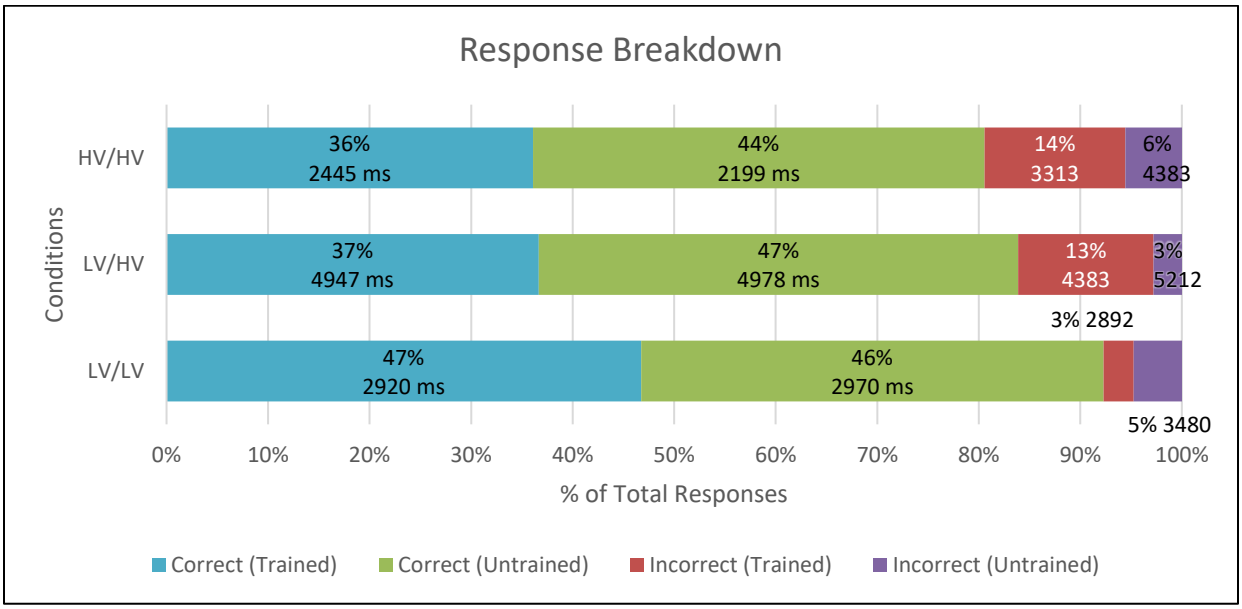


Figure 7-5: A breakdown of responses types as a percentage of total responses for each of the three conditions in first experiment. Includes the mean response time (in milliseconds).

Figure 7-5 above shows that for correct responses the HV/HV condition was very close to the LV/HV condition in terms of accuracy (36% and 44% versus 37% and 47%) and slightly faster compared to both LV trained conditions. Despite a longer response time the LV/HV condition maintained a similar error rate to the HV trained condition. In all three conditions, false negatives (untrained incorrect) were the least likely type of error. False positives, in conditions that used HV testing were the same, despite the longer response times from the LV/HV group.

The following figures show results labeled as either ‘combined’ or ‘generalizing’. The data under the ‘combined’ label includes all the data from all three experimental conditions. The ‘generalizing’ label indicates a subset that limits the data to characters that appeared during training but that used a novel font during the testing (i.e. instances where the training was generalized to

novel forms). The 'generalized' data removes testing tokens that used identical font/character combinations shown to participants during testing. This includes duplicate tokens that occurred due to experimental design (the entire LV/LV condition) and those created as a by-product of the random font/character assignment. The figures below provide an analysis of the response times and accuracy rates as they pertain to the combined and the generalizing subset.

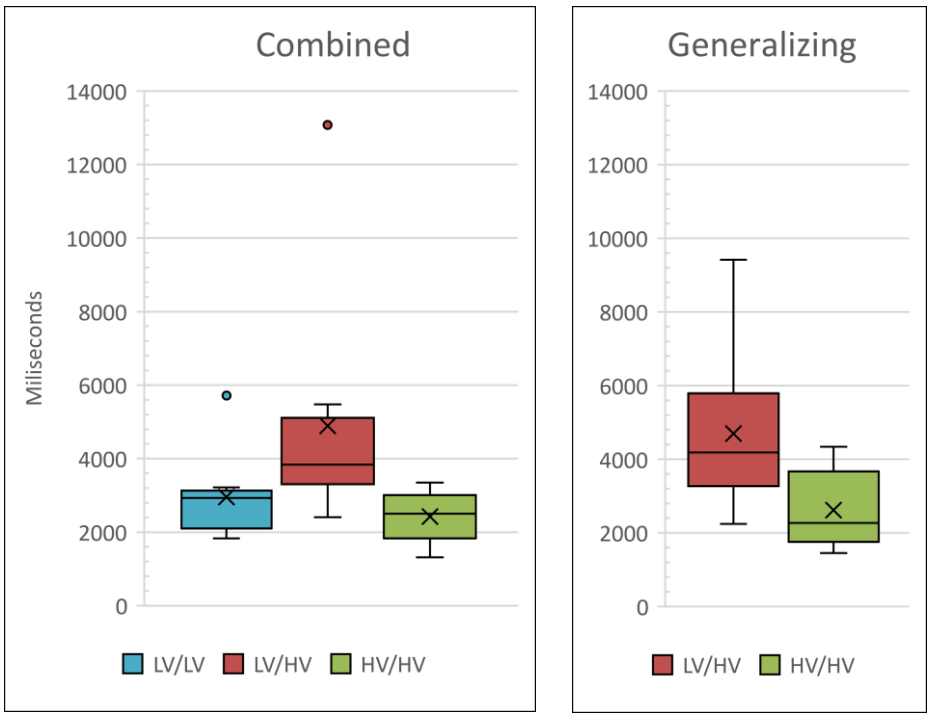


Figure 7-6: Experiment 1 response times (in milliseconds) for all data in each of the three conditions (combined) and the subset of data that excludes identical font/character combinations shown to participants during training and testing (generalizing)[18].

[18] Additional data is available in Appendix 2.

When considering all responses, the control condition LV/LV had a median response time of 2929ms ($\bar{x}$ = 2953ms). The LV/HV and HV/HV conditions had median response times of 3839ms ($\bar{x}$ = 4894ms) and 2505ms ($\bar{x}$ = 2428ms) respectively. A linear model analysis of a log transform showed that the response times for the HV/HV condition were significantly faster than the LV/HV [p = 0.00703]. When considering only familiar characters shown in a new font (i.e. the generalizing subset), LV/HV and HV/HV had median response times of 4181ms ($\bar{x}$ = 4693ms) and 2275ms ($\bar{x}$ =2622ms) respectively. Participants in the HV/HV condition showed a statistically significant difference in response times when generalizing their training to novel fonts [p=0.0118].

The figure below shows the accuracy rates for the ‘combined and ‘generalizing’ datasets. Neither dataset showed a statistically significant difference in accuracy between participants in the LV/HV and HV/HV conditions.

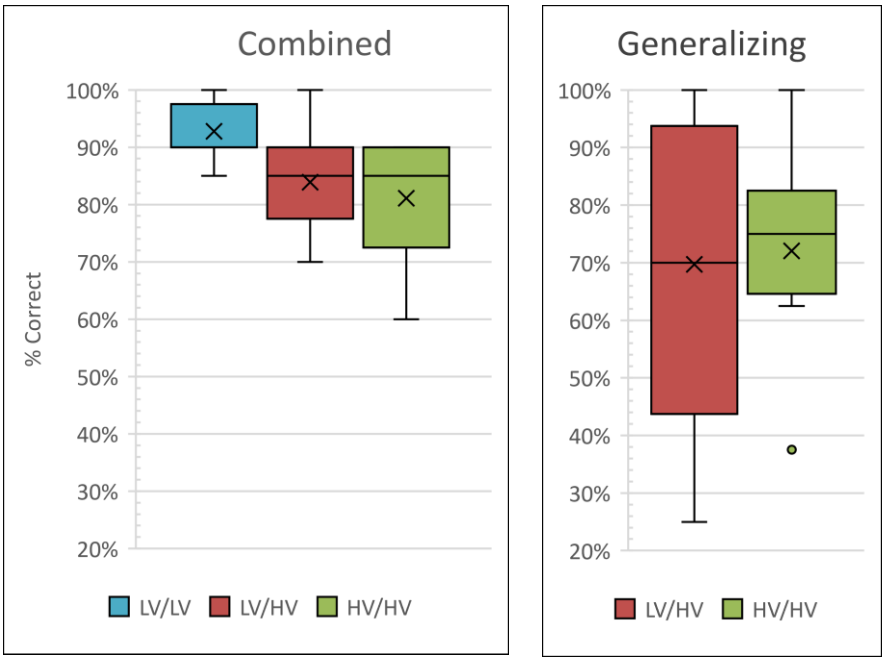


Figure 7-7: Percentage of accurate responses, summarized by participants, for all of the data in each of the three experimental conditions in the first experiment (combined) and the subset of data that excludes identical font/character combinations shown to participants during training and testing (generalizing).

Figure 7-8 below indicate the changes for both the mean response time and accuracy between the first and second sessions.

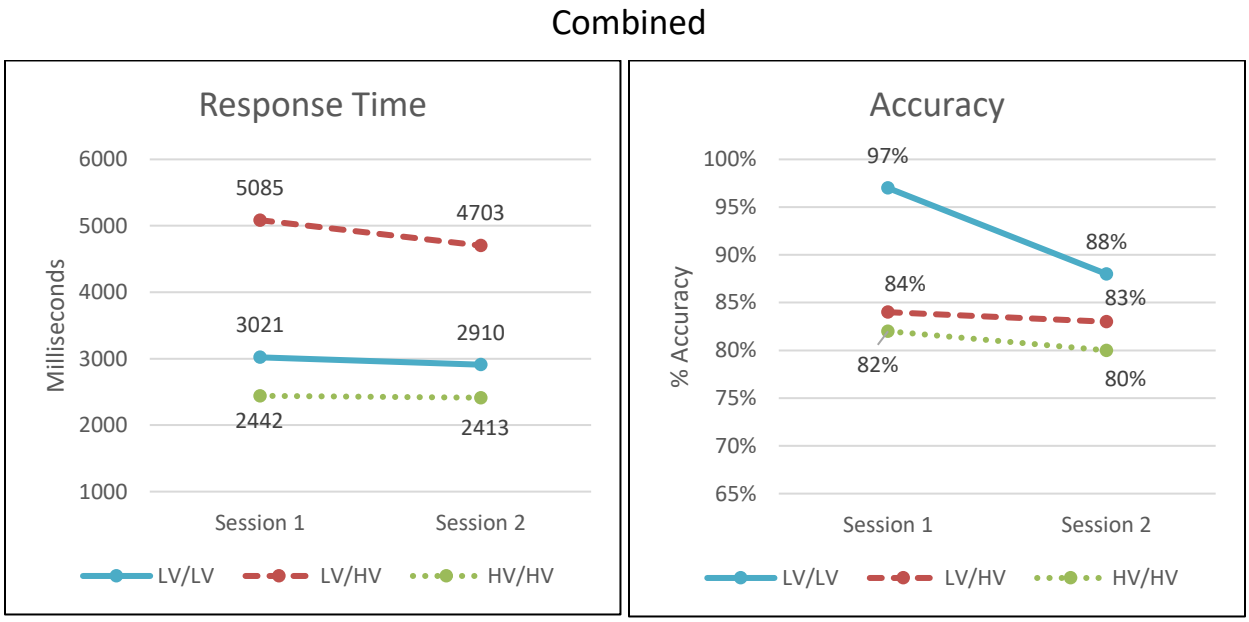


Figure 7-8: The panel on the left shows the changes in response time (in milliseconds) between the first and second session of experiment one. The panel on the right shows the changes (in percent of accurate responses) between the first and second sessions of experiment one.

Participants in every condition were slightly faster and less accurate in the second session. The largest drop in accuracy was the LV/LV control condition which fell from 97% to 88%. The accuracy for both HV testing sessions experienced small drops of only 1-2%. The LV/HV condition had the longest response times in both sessions. The difference in response times for participants in the LV/HV and HV/HV condition were significant for both session one and session

two (S1: [p=0.0114], S2: [p=0.00836])[19]. The differences in accuracy between the LV/HV and HV/HV for both session one and two was not significant.

During training characters were shown 1, 4, 9, 16 or 25 times. Unsurprisingly, the highest accuracy was seen at the higher frequencies and most errors occurred at the lower frequencies.

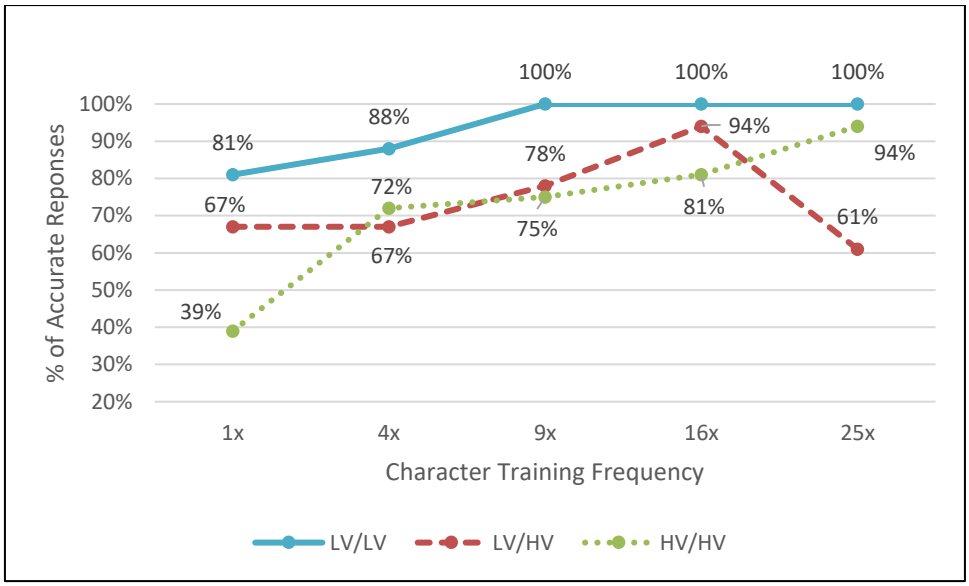


Figure 7-9: Accuracy rates (as percentage of total responses) for each of the characters that appeared during the training session for the three experimental conditions in experiment one. Separated by the training frequency.

[19] The response times for the generalizing subset of data (showing only instances of new fonts combined with previously experienced characters) was statistically significant as well (S1: [p=0.0114], S2: [p=0.00963]). See Appendix 2 for additional details.

As seen in the figure above, each condition showed improvement as the training frequency increased but no experimental condition benefited from frequency more than any other in a statistically significant way.

If the participants responded that they recognized the character from the training phase, they were given an opportunity to provide the English definition for the character. The figure below shows the response time and accuracy rates for the definition responses.

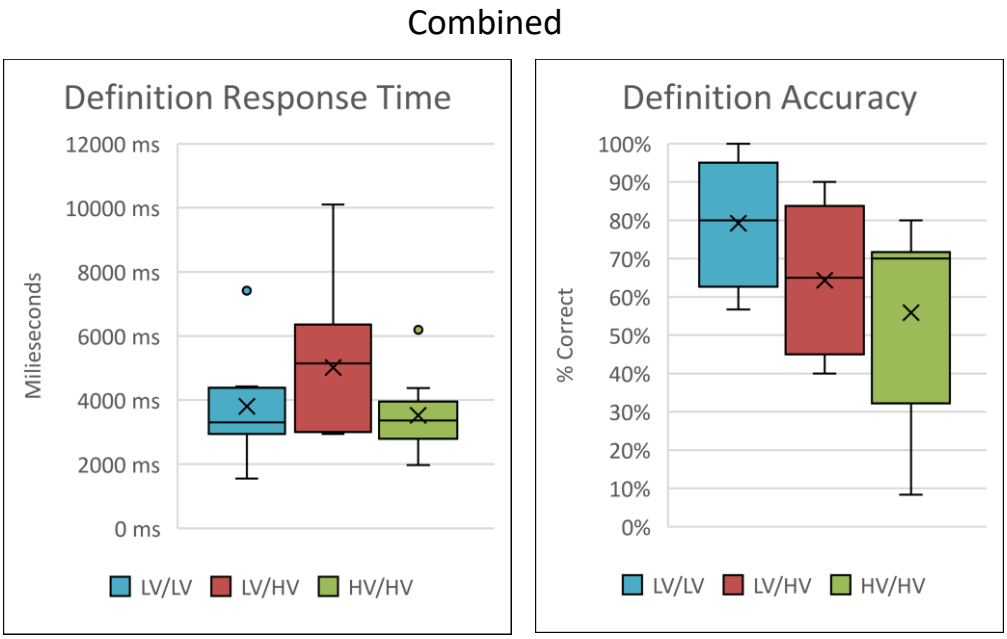


Figure 7-10: The panel on the left shows the definition response times (in milliseconds) for all of the data in each of the three conditions in experiment one. The panel on the right shows the percentage of accurate definition responses summarized by participants.

The LV/HV condition had the slowest response times and lowest accuracy. However, differences in both the definition response time and accuracy rates for LV/HV and HV/HV participants were not statistically significant.

7.5 Summary

The results of this experiments failed to show any significant differences in overall accuracy for people whether they trained with one font or many. Participants in the LV/HV condition were slower than the HV/HV participants but was slightly more accurate, most likely due to the longer average response time. If participants trained with many fonts, participants were faster, with participants in the HV/HV condition having statistically significant faster response time than participants in the LV/HV condition. These results support the hypothesis that HV training would result in a faster response time but fails to provide evidence that HV training could increase accuracy as well.

During the first session, participants in the LV/LV condition were more accurate than both HV testing conditions, but the LV/LV participants showed poorer retention – with only an 8% higher accuracy in the second session. HV training did not provide any benefit to retention.

When considering only the ability to generalize training to new fonts, participants in the HV/HV condition were statistically faster than the LV/HV participants. The median accuracy was higher than the LV/HV condition but not enough to be statistically significant. The hypothesis that HV training would make participants both faster and better at recognizing novel fonts was only partially supported.

The HV/HV participants had the fastest definition responses but had the poorest definition accuracy. Moreland & Zajonc (1977) argued that recognizing a written character is different from

remembering its meaning, stating that recollection requires more cognitive resources. This is supported by the definition response data.

The results of HV training for unigrams was mixed. There was a clear advantage for speed but no corresponding benefit for accuracy. However, the first experiment only deals with unigrams, the simplest possible reading case. Ultimately, the ability to identify unigrams alone is completely unrepresentative of real-world reading. Experiment two examines HV training in a more difficult situation, the identification of unigrams in the context of a bigram.

8 Experiment 2 – Bigrams

Pelli et al. (2006) demonstrated that characters are more easily identified when presented in isolation. It is more difficult to recognize a character when it is placed in the context of an unknown character. Unigram fluency breaks down when a person is forced to spread their attention across multiple characters and is not able to correctly identify the most important features of the character they are looking at. Experiment 2 follows the basic structure of Experiment 1 but asks participants to identify characters located within bigrams. The target character to be identified will be either the left or right of the character pair making up the bigram.

8.1 Participants

Twenty-seven English-speaking students recruited from the University of Manitoba. Requirements for participation were the same as Experiment 1.

8.2 Stimuli

Stimuli used to create the training and testing materials for this experiment were identical to Experiment 1.

8.3 Methodology

The basic design of this experiment was the same as Experiment 1, except that testing used pairs of characters (bigrams) instead of single characters (unigrams).

Each participant was randomly assigned to one of three experimental conditions. Recruitment continued until there was even number of participants that successfully completed all three conditions (8 participants). Two participants assigned to the LV/HV condition and one participant assigned to the HV/HV condition did not complete the second testing session. Data collected from these participants during the first testing session was retained.

<table>
<tr><th colspan="2">Training</th><th></th><th colspan="2">Testing</th></tr>
<tr><td rowspan="3">Single Character Training</td><td rowspan="2">Single Font (LV Training)</td><td>LV/LV</td><td>Single Font (LV Testing)</td><td rowspan="3">Bigram Testing</td></tr>
<tr><td>LV/HV</td><td rowspan="2">Multiple fonts (HV Training)</td></tr>
<tr><td>Multiple fonts (HV Training)</td><td>HV/HV</td></tr>
</table>

Figure 8-1: Experiment 2 Conditions

Equipment and data analysis methodology were the same as Experiment 1.

8.3.1 Training

From the 550 typographically unique tokens (55 characters in 10 fonts) each participant was shown one unique set of 55 training tokens per participant. The training tokens used in this experiment were generated and presented using the same method as Experiment 1. A sample training token can be seen in the figure below.

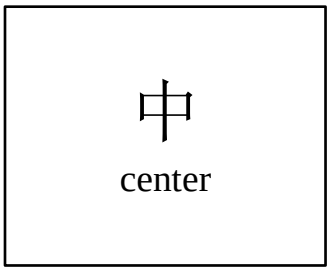


Figure 8-2: Bigram Training Token

8.3.2 Testing

Each participant was shown a unique set of 10 testing tokens. The five characters used to generate the training tokens were combined with fifteen new characters randomly selected from the character set and arranged into bigrams. Tokens in the LV testing condition were displayed in the SimSun font. Tokens in the HV testing conditions were randomly assigned one of the ten fonts in the dataset. Each testing token was black on a white background and displayed at a 144pt size as seen in the figure below.

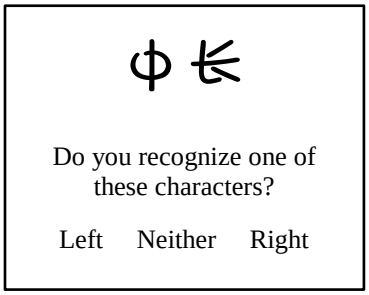


Figure 8-3: Bigram Testing Token

Participants were asked if they recognized the left, right, or neither character in the presented bigram. If they responded ‘left’ or ‘right’ then they were prompted to type in the English definition as seen in the figure below.

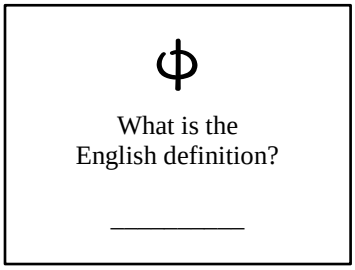


Figure 8-4: Bigram Definition Prompt

Second session testing used the same tokens as the first testing session but presented in a different order.

8.4 Results

Like experiment 1, the most important measurements were the response times and accuracy rates[20]. This experiment provided four distinct response types. There were two types of correct answer; correctly identifying a character that was part of their training (i.e. correct trained), or correctly identifying a character that was not (i.e. correct untrained). Similarly, there were two types of errors. Either a participant failed to recognize a character that was part of their training (i.e. a false negative), or they thought they 'recognized' a character that wasn't part of their training (i.e. a false positive). Figure 8-5 below breaks the data down into these four response types by percentage of total responses and indicates the average response time.

[20] The following analyses include all recorded data. No outlier data was removed. Additional data and analyses (including analysis of sensitivity using a d-prime analysis using Macmillan & Creelman's (2004) method can be found in Appendix 2).

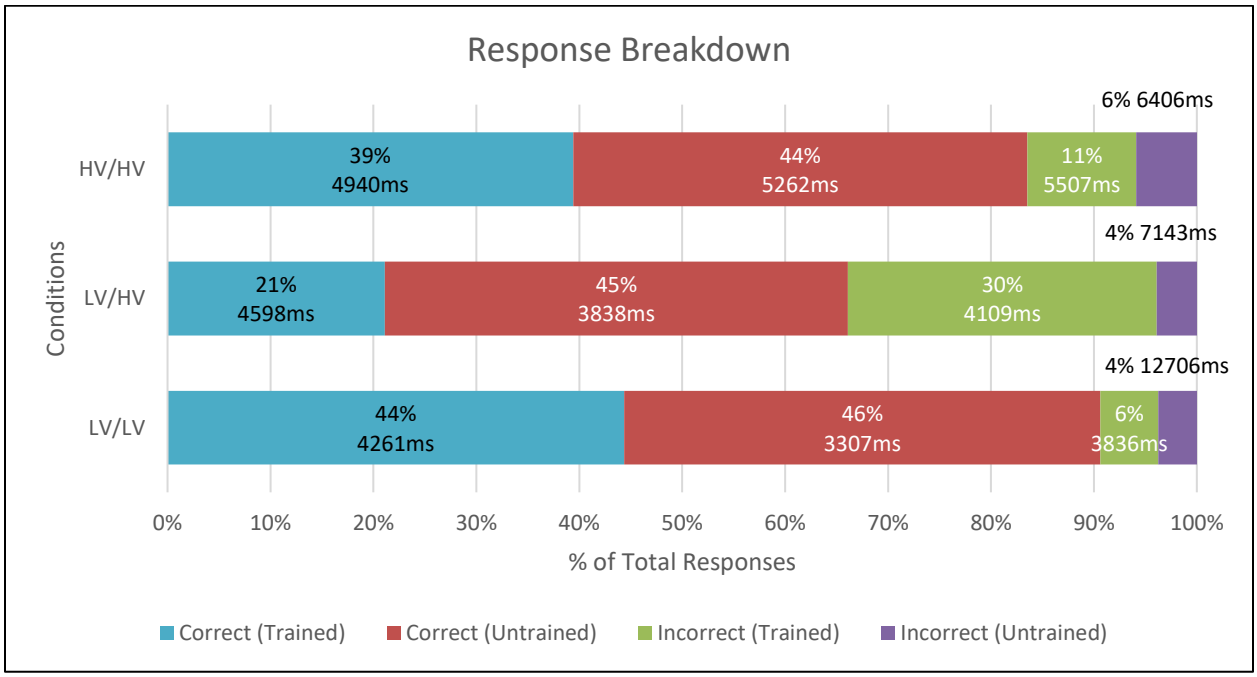


Figure 8-5: A breakdown of response types as a percentage of total responses for each of the three conditions in second experiment. Includes the mean response time (in milliseconds).

Like the first experiment, the response-type breakdown shows that the participants in the LV/LV were the most accurate. Unlike experiment one, participants in the HV/HV condition were much more accurate than LV/HV condition (83% overall accuracy versus 66%). However, the correct responses in HV/HV were all slower compared to both LV trained conditions. False negatives (incorrect trained) responses were the least likely error type but took the most time compared to all other response types. Participants in the LV/HV condition were much more likely to incorrectly identify a never-before-seen font-character token as one they had seen in training with false-positives making up 30% of their total responses.

Similar to the first experiment, the following figures show results labeled as either ‘combined’ or ‘generalizing’. The data under the ‘combined’ label includes all the data from all three

experimental conditions. The ‘generalizing’ label indicates a subset that limits the data to characters that appeared during training but that used a novel font during the testing (and uses instances where the training was generalized to novel forms). The ‘generalized’ data removes testing tokens that used identical font/character combinations shown to participants during testing. This includes duplicate tokens that occurred due to experimental design (the entire LV/LV condition) and those created as a by-product of the random font/character assignment. The figures below provide an analysis of the response times and accuracy rates as they pertain to the combined and the generalizing subset.

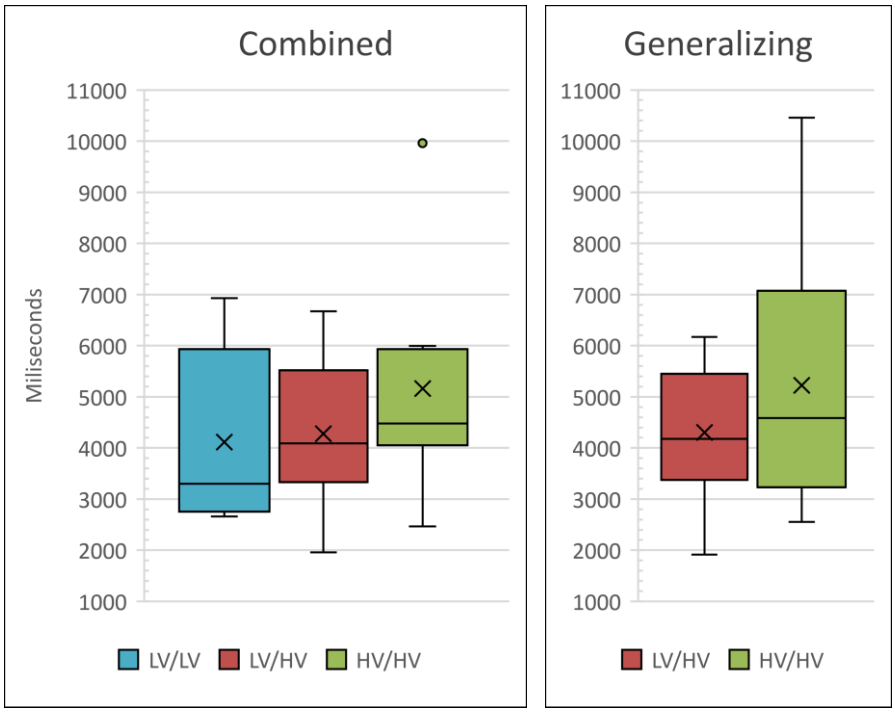


Figure 8-6: Experiment two response times (in milliseconds) for all data in each of the three conditions (combined) and the subset of data that excludes identical font/character combinations shown to participants during training (generalizing)[21].

Unlike the previous experiment, participants in the HV/HV condition were slightly slower than the LV/HV in both the combined and the generalizing data sets. There were no statistically significant difference for either the combined or generalized response times.

[21] Supplemental data is available in Appendix 2.

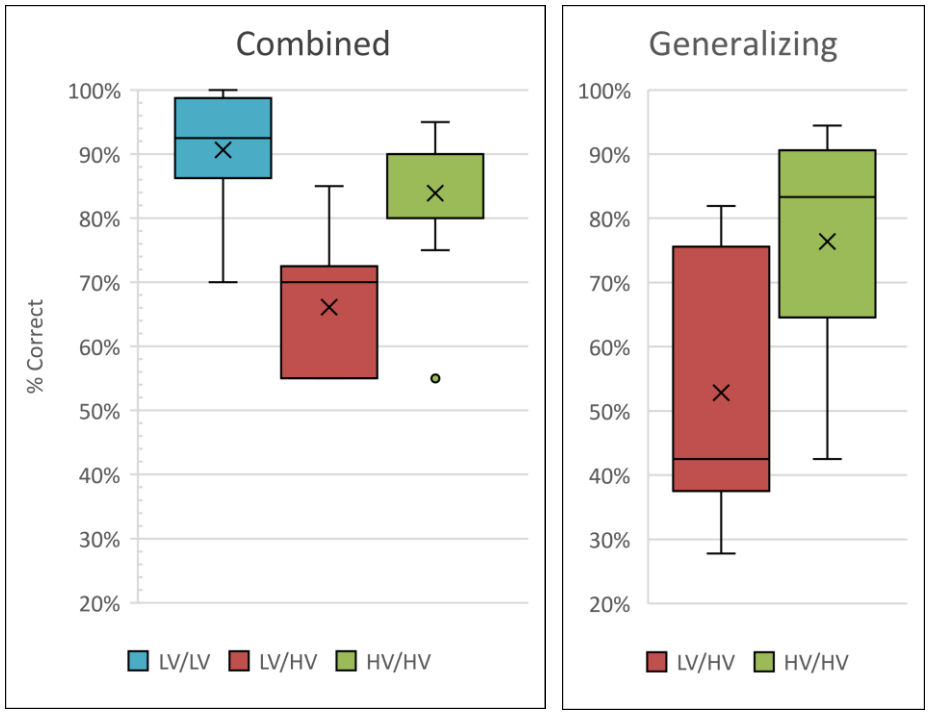


Figure 8-7: Percentage of accurate responses, summarized by participants, for all of the data in each of the three experimental conditions in the first experiment (combined) and the subset of data that excludes identical font/character combinations shown to participants during training and testing (generalizing).

For the combined results, the overall median accuracy for the LV/LV control condition was 93% ($\bar{x}$ = 91%), with two participants achieving a perfect score. The LV/HV condition had a median accuracy of 70% ($\bar{x}$ = 68%) while the HV/HV condition had a median accuracy of 90% ($\bar{x}$ = 84%). The HV/HV condition accuracy was significantly higher than the LV/HV condition [p = 0.000963]. In the generalizing data, the LV/HV condition had a median accuracy of only 43% ($\bar{x}$ = 53%). The HV/HV condition maintained a median accuracy of 83% ($\bar{x}$ = 76%). The difference in accuracy in the generalized data was statistically significant [p = 0.00223].

The figure below illustrates the change in both the mean response time and accuracy between session 1 and 2.

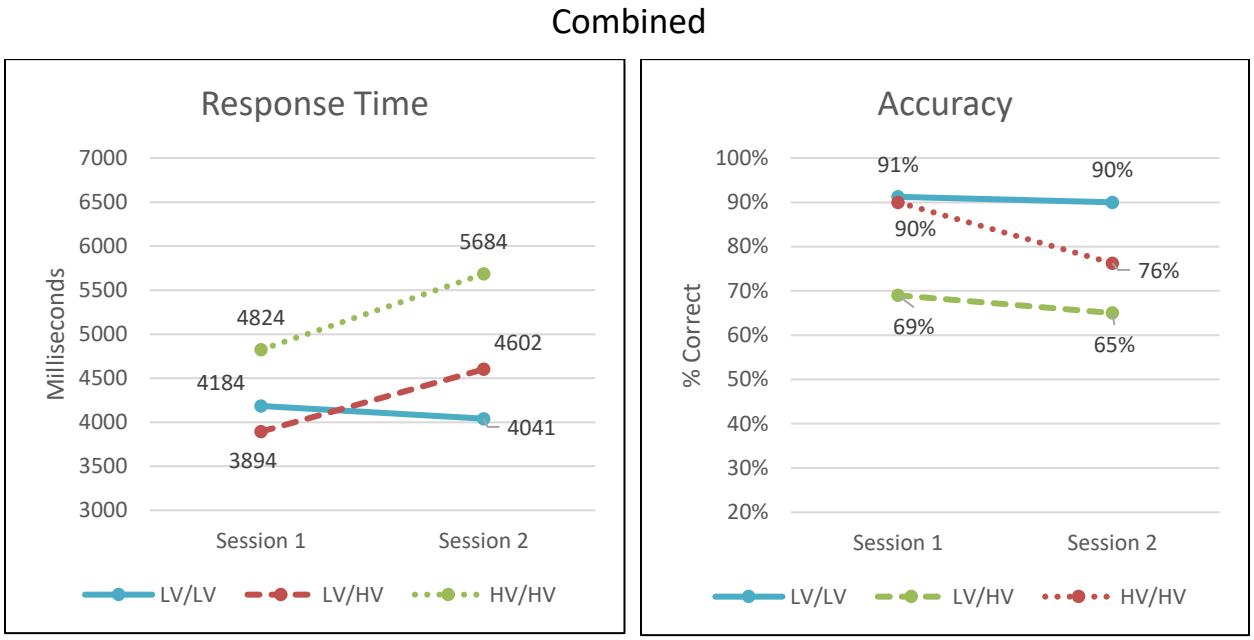


Figure 8-8: The panel on the left shows the changes in response time (in milliseconds) between the first and second session of experiment two. The panel on the right shows the changes (in percent of accurate responses) between the first and second sessions.

The overall response time for both sessions that used HV testing increased between sessions. There was no statistically significant difference in response time in either session one or two. The accuracy of the two conditions that used LV training decreased slightly across both sessions while the HV/HV condition dropped from 90% to 76%. Despite the drop the HV/HV condition remained statistically better than the LV/HV in both conditions (S1: [$p = 0.00214$], S2: [$p = 5.03 \times 10^{-8}$])

Characters were shown 1, 4, 9, 16 or 25 times during training. The expected effect of this would be characters shown at a higher rate would be easier to identify during testing (i.e. the frequency

effect). The figure below shows that this was true for the LV/LV and the HV/HV conditions but not for the LV/HV.

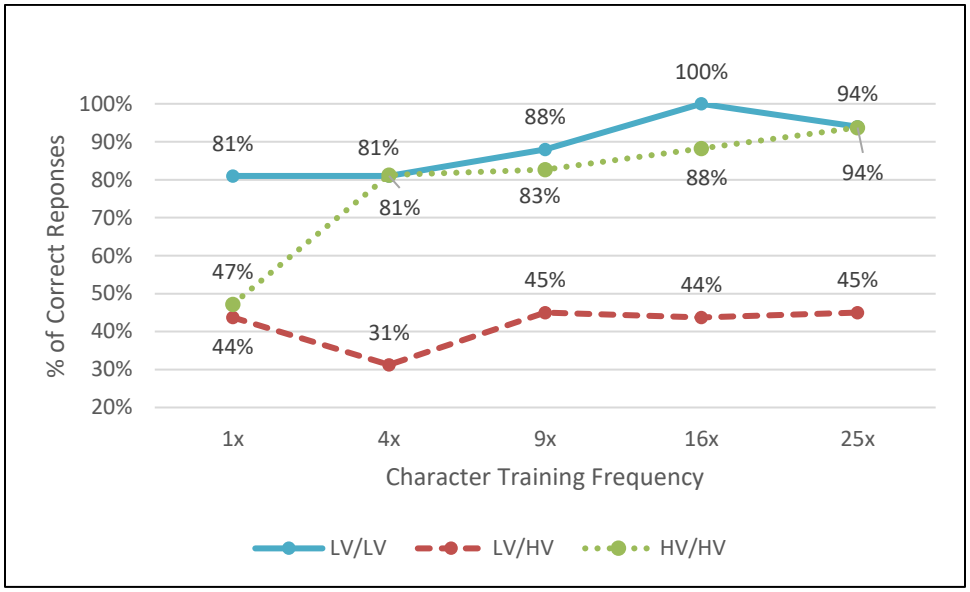


Figure 8-9: Accuracy rates (as percentage of total responses) for each of the characters that appeared during the training session for the three experimental conditions in experiment two. Separated by the training frequency.

The LV/HV accuracy remains stable ($\bar{x}$ is approximately 44 to 45%) for nearly all training frequencies. There was no advantage for HV-training at the very lowest frequencies, but as the frequency increased the HV-trained participants got more accurate while LV/HV participant accuracy remained flat [$p = 0.0014$]. A participant in the LV/HV condition is just as likely to remember a character once as they are after seeing it 25 times. It does not appear that the LV/HV participants are learning to recognize the characters at all. A linear mixed-model analysis of the data showed a three-way interaction between the training condition, participant accuracy and the frequency [$p = 0.0140$].

If the participants responded that they recognized the character from the training phase, they were given an opportunity to provide the English definition for the character. The figure below shows the response time and accuracy rates for the definition responses.

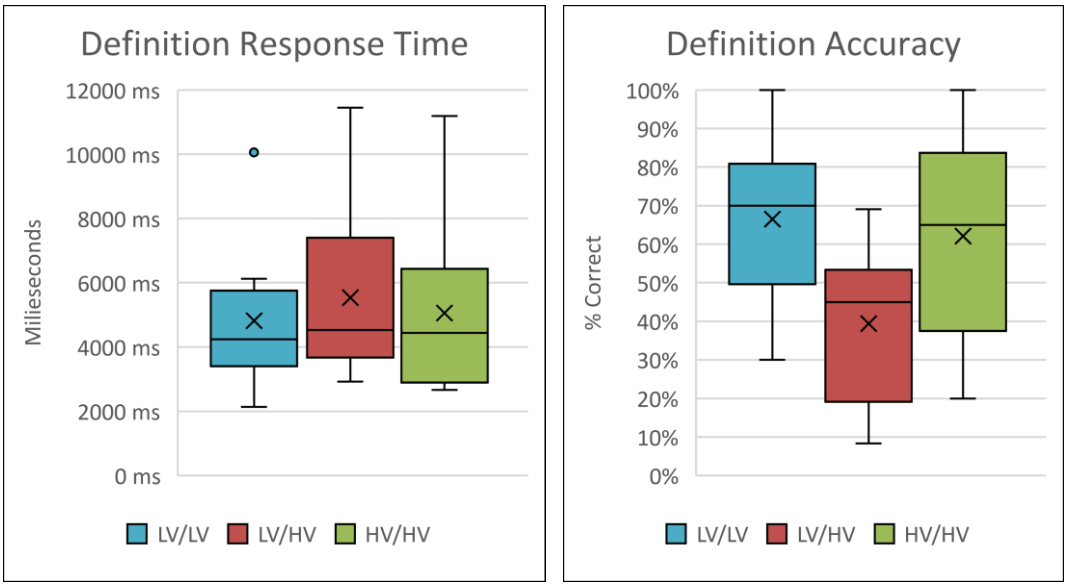


Figure 8-10: The panel on the left shows the definition response times (in milliseconds) for all data in each of the three conditions in experiment one. The panel on the right shows the percentage of accurate definition responses summarized by participants.

There were no statistically significant differences in definition response times or accuracy between the three conditions in this experiment.

8.5 Summary

The results of this experiment showed that participants in the HV/HV condition were significantly more accurate than participants in the LV/HV condition. The accuracy of participants

in the LV/HV condition was only slightly higher than chance[22]. However, the HV/HV condition was slower than the LV/HV condition although the difference was not statistically significant. These results support the hypothesis that HV training would result in higher accuracy but fail to provide evidence that HV training would increase speed as well.

Every condition decreased in accuracy and between the first and second sessions. Participants in the HV/HV had the largest drop at 14%.

When considering only the ability to generalizing training to new fonts, participants in the HV/HV condition were statistically more accurate than the LV/HV participants. The median response time was slightly higher than the LV/HV condition but not enough to be statistically significant. The hypothesis that HV training would make participants both faster and better at recognizing novel fonts was only partially accurate.

As expected, participants in both the LV/LV and HV/HV groups were more accurate the more often they saw a character during training (i.e a typical frequency effect). Surprisingly, the LV/HV condition did not show any increase in accuracy that corresponded with increased frequency during training.

---

[22] Participants in the LV/HV condition for the second experiment were given three options during testing: the character on the right, the character on the left, or neither character. Chance accuracy would be 33%. See §8.3.2 for more details.

The results of HV training for character identification within bigrams was mixed. There was a clear advantage for accuracy but no statistically significant benefit for speed. However, the near-chance accuracy by participants in the LV/HV condition make it clear that LV training was not sufficient to teach people to recognize characters in bigrams.

9 General Discussion

The main purpose of this study was to apply the HV training techniques commonly used with spoken language to L2 reading instruction. The first step of any reading model is to identify if what is being viewed is something that can be read. Visual stimuli that is not recognized as valid orthographic characters will be not be processed as such. This study presented participants with a set of orthographic stimuli that (depending on the assigned training regimen) varied in the degree of category internal variation. The experiments allowed participants to build their orthographic categories based on the training stimuli.

This study asked participants to identify single characters in isolation and in the context of character pairs. Both experimental tasks are much simpler than reading in a real-world environment but still provide useful insight into the effectiveness of HV training as a pedagogical technique. Each experiment used the same three conditions: LV/LV, LV/HV, and HV/HV. The LV/LV condition was the simplest possible version of each of the experimental tasks and is unrepresentative of any real-world scenario. However, it does provide a baseline for the other two conditions.

HV testing with isolated unigrams (experiment 1) showed that the HV training resulted in significantly faster response times. However, there was no statistically significant difference in accuracy between participants in the HV/HV and LV/HV training regimens. In the second experiment where participants were asked to identify single characters located within a pair the results were inverted. The conditions that used HV testing had no statistically significant difference in response time but participants in the HV/HV condition were significantly more accurate than their LV/HV counterparts. HV training produced significant results in either speed or accuracy but not both at the same time. Vlaskamp & Hooge (2006) say that increases in fixation is a common strategy for dealing with complicated written text. It is entirely possible that longer fixations could account for some of the increased time spent on the the more complex bigram stimuli.

A major part of reading fluency is speed and accuracy. Speed and accuracy come from repeated practice (i.e. frequency effects). Figure 8-9 above clearly shows that LV training does not respond typically towards the expected frequency effects. When characters are shown multiple times during training the expected effect is that characters shown at a higher rate would be easier to recognize. During unigram testing, the participants in the LV/HV condition followed this principle. However, LV/HV participants that were tested with bigrams did not gain any benefit from repeated exposures.

The results of the current study run contrary to a popular belief in the educational sphere that successful learning is based on how easily a student can encode the information; that reducing the cognitive load is always beneficial for the learner (Sweller & Chandler, 1994). However, there is

some evidence that typographically based interventions can produce positive educational outcomes.

Diemand-Yauman, Oppenheimer, & Vaughan (2011) tested a cognitive intervention that used typographical disfluency. In their first experiment they taught participants a series of fictional facts, where the 'facts' were written in variety of different typefaces. In their second experiment, they used atypical typographic presentation in real-world classroom instruction. They transcribed high-school classroom materials (for classes including English, History, Physics, and Chemistry) into versions that use non-standard fonts (specifically Haettenschweller, Montotype Corsiva, and Comic Sans Italicized) and various contrast ratios. The new versions were legible but harder to read than normal (i.e. disfluent). They found people that used the 'disfluent' versions of the material showed improved memory performance and test scores. Diemand-Yauman et al claim that the 'disfluent materials' compelled the students to use deeper processing strategies and maintain a higher focus. Diemand-Yauman et al warn of the diminishing returns of disfluency interventions. They hypothesize that a U-shaped curve affects a participant's motivation to learn the material. Both simple-to-consume material and material that is so 'disfluent' as to be illegible will encourage students to not apply sufficient effort to understand the material.

McDaniel, Hines, & Guynn (2002) developed a disfluency intervention by artificially removing letters from words and rearranging words within sentences. Oppenheimer (2006) used word-length and unnecessarily complex synonyms to common terms. Unsurprisingly, these interventions only showed benefits for readers who were particularly skilled to begin with,

manifesting in a steep U-curve of benefit versus disadvantage. Stronger readers can use alternative strategies that allow the reader to figure out the missing characters or obscure definitions by contextual cues derived from the surrounding characters or words. Diemand-Yauman et al do not make any explicit claim that low-perceptual abilities will reduce the effectiveness of the disfluent material, only that the material itself must be sufficiently legible.

A disfluency intervention is not the same as HV training. In Diemand-Yauman et al's experiments each student was presented with only one, albeit atypical, typographic variation. The student used the same typographic variation throughout the entire school term, giving them plenty of time to become comfortable with the initially novel presentation. Disfluency interventions, as described above, would not work in the context of a beginner learner because they are objectively poor readers. A beginner reader is equally unfamiliar with all typographic variations, so manipulation of fonts can only cause disfluency for native readers with an existing base of experience. However, HV training and disfluency interventions are similar in that they are predicated on the hypothesis that an easy-to-consume LV presentation is not the best approach.

An underlying assumption of the current study was that exposure to multiple fonts will positively affect the learner because the learner holds no preference for any specific typeface. Both Logan (1997) and Brown & Carr (1993) pointed out that the transfer between typographic presentation is asymmetric. Students who have learned to read handwritten material first have an easier time transferring their skill to digital fonts, but not the other way around. Handwritten text, by its nature, contains a high degree of internal variability. Words and letters, even when written

by a single person, can vary significantly. This high degree of variability makes the ability to identify the salient features very important. By design, most fonts are consistent in their appearance. Consistent features are more easily identified, extracted and memorized.

Exposure to a variety of fonts allowed the participants to generate a better category representation which in turn lead to better generalization to novel presentation. Participants in this study had no previous experience with written Chinese characters and had, therefore, no pre-existing category representation for those characters.

The mechanisms driving fluent language acquisition are, generally speaking, repeated practice (i.e. frequency training) and learning what features are worth focussing on (i.e attention tuning) The more quickly and effortlessly a person can extract, process and integrate information, regardless of modality, the more fluent they are. Fluency is improved by learning what to pay attention to and by experiencing stimuli over and over. The figure below summarizes the results of the current study in terms of the average length (in seconds) each character identification took and what the overall accuracy for each experimental condition was.

		Response Time	Recognition %
Unigram	LV/LV	3.22 sec	92%
	LV/HV	5.83 sec	84%
	HV/HV	3.11 sec	81%
Bigram	LV/LV	4.54 sec	91%
	LV/HV	6.37 sec	66%
	HV/HV	6.26 sec	84%

Figure 9-1: Combined summary of the current study in terms of the average mean response time (in seconds) and what the overall mean accuracy for each experimental condition.

When response times and accuracy rates are considered together, there is no reason to avoid HV tokens when learning to recognize characters in an L2. Learning from just one typographic style tends to make a person good at recognizing that particular style but doesn't prepare a person for handling variations that they've never seen before. By training with variable fonts from the beginning, a person will be better prepared to handle other variation they might encounter.

HV training has produced mixed results in previous studies (see §1 for more details). In some cases, HV training produces highly beneficial results over LV training, while other times the benefits of HV training were inconsistent and the LV was more effective over the long term[23]. One of the major critiques of the HV training paradigm is that this approach does not work well for people who don't already have a well-developed perceptual ability (Perrachione, Lee, Ha, & Wong, 2011). The figure below (previously part of Figure 8.6) shows an outlier participant in the HV/HV condition. It is entirely possible that this participant was of the type described by Perrachione et al (2011)[24] but the precise relationship between people with low-perceptual ability and their ability to learn to recognize written characters with HV training techniques is a subject for future research.

---

[23] Giannakopoulou, Brown, Clayards, & Wonnacott (2017) claim that Greek speaking children and adults' ability to learn the English /i/ vs /ɪ/ contrast benefited more from LV training tokens over HV (See §1 for a more detailed discussion).

[24] No participants had their pre-existing reading or perceptual skills tested.

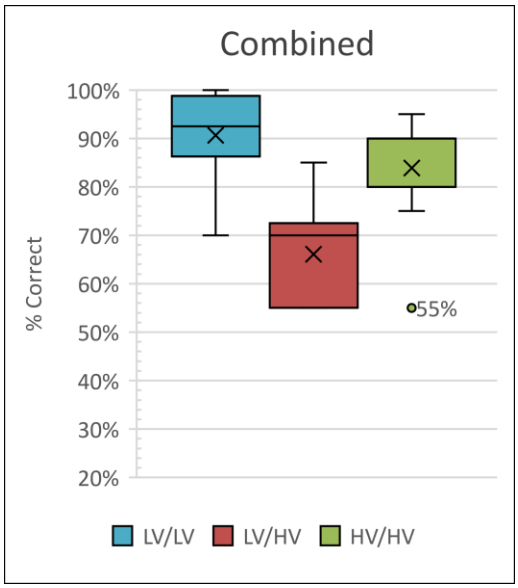


Figure 9-2: Percentage of accurate responses, summarized by participants, for all of the data in each of the three experimental conditions in the first experiment (combined).

The majority of the research investigating the effects of HV training is based on auditory phonological perception. Applying the HV technique to a different modality necessitates new considerations. A basic tenent of the speech in the auditory domain is that stimuli are processed in real-time. A person cannot fixate on spoken language like they could with written language. This begs the question of whether the shortcomings of individual perceptual ability could be mitigated when using the HV paradigm within the visual modality.

Unfortunately, the current study was not designed to independently analyze the effect of fixation time and training method. Participants were permitted to take as much fixation time as they wished during the testing phase. In order to properly determine the degree to which accuracy was derived from fixation time or the training method will have to be left to future experimentation.

This study showed the effectiveness of HV training for single character recognition in isolation and as part of a bigram. This does not mean that HV training is necessarily the best technique for learning to read in a second language. There were no statistically significant benefits for any of the experimental conditions on either definition response speed or accuracy. When it comes to linking written characters to their meaning character recognition is a necessary step (a person can't remember the definition of a word they don't recognize) but HV training does not produce any benefit in teaching people the meanings of the words or characters that they recognize.

The lack of research in this area provides a large number of areas for continuing research. The current study has shown that HV techniques can be effectively applied to written language and found evidence that HV training is effective for English speakers learning to recognize Chinese characters. Further experimentation will have to explore whether this HV training method could also be used to train other writing systems and prove its effectiveness as a useful educational tool in real-world classrooms.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Bellantoni, J., & Woolman, M. (1999). *Type in motion: innovations in digital graphics*. Rizzoli International Publications.

Bernstein, L. E., Auer, E. T., Moore, J. K., Ponton, C. W., Don, M., & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. *Neuroreport*, *13*(3), 311–315.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 171–204). Baltimore, York Press.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y., Yamada, R. A., Tohkura, Y., Akahane-Yamada, R., et al. (1997). Training Japanese listeners to identify English / r / and / l /: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Brown, J. S., & Carr, T. H. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1277–1296.

Brumberger, E. (2004). The Rhetoric of Typography: Effects on Reading Time, Reading Comprehension, and Perceptions of Ethos. *Technical Communication*, *51*(1), 13–24.

Bybee, J. L. (2013). Usage-based Theory and Exemplar Representations of Constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49–69).

Campbell, K. S. (1995). *Coherence, Continuity and Cohesion: Theoretical Foundations for Document Design*. Routledge.

Chen, H.-C., & Tang, C.-K. (1998). The effective visual field in reading Chinese. *Reading and Writing*, *10*(3), 245–254.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. MIT Press.

Chung, S. T. L. (2007). Learning to identify crowded letters: does it improve reading speed? *Vision research*, *47*(25), 3150–3159.

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.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(3), 207–238.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330(6009), 1359-1364.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(6), 397–417.

Flege, J. E. (1995). Second language speech learning theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233–277). York Press.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition*, *17*(3), 302–310.

Gahl, S., & Yu, A. C. L. (2006). Introduction to the special issue on exemplar-based models in linguistics. *The linguistic review*, *23*(3), 213–216.

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

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, 1–35.

Grainger, J., Tydgat, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 673–688.

Grant, H. M., Bredahl, L. C., Clay, J., Ferrie, J., Groves, J. E., McDorman, T. a., & Dark, V. J. (1998). Context-dependent memory for meaningful material: Information for students. *Applied cognitive psychology*, *12*(December), 617–624.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522.

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: the phonological linkage hypothesis. *Child Development*, *65*(1), 41–57.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of experimental psychology*, *110*(3), 306–340.

Kress, G., & Van Leeuwen, T. (2002). Colour as a semiotic mode: notes for a grammar of colour. *Visual Communication*, *1*(3), 343–368.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, *50*(2), 93–107. Springer.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831. Nature Publishing Group.

Kuhl, P. K., & Iverson, P. (1995). Linguistic Experience and the "Perceptual Magnet" Effect. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 121–154).

Kumpf, E. P. (2000). Visual metadiscourse: Designing the considerate text. *Technical Communication Quarterly*, *9*(4), 401–424.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, *6*(2), 293–323.

Larson, K., Hazlett, R. L., Chaparro, S., & Picard, R. W. (2007). Measuring the Aesthetics of Reading. *People and Computers XX—Engage: Proceedings of Human-Computer Interactions 2006* (pp. 41–56).

Van Leeuwen, T. (2005). Typographic Meaning. *Visual Communication*, *4*(2), 137–143.

Liberman, I. Y. (1973). 1. Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society*, *23*(1), 64–77.

Liberman, I. Y., Shankweiler, D. P., Fischer, F. W., & Carter, B. (1974). Explicit Syllable and Phonemic Segmentation in the Young Child. *Journal of Experimental Child Psychology*, *18*(1074), 201–212.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255.

Logan, G. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Logan, G. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*(1), 1–35.

Logan, G. (1997). Automaticity and Reading: Perspectives From the Instance Theory of Automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 123–146.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology Press.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, *22*(3), 249–261.

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*(5), 365–368.

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When Text Difficulty Benefits Less-Skilled Readers. *Journal of Memory and Language*, *46*(3), 544–561.

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition*, *13*(1), 89–98.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. Special Issue: The onset of literacy. *Cognition*, *24*(1–2), 45–64.

Moreland, R. L., & Zajonc, R. B. (1977). Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology*, *35*(4), 191–199.

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*.

Olson, D. R. (1996). Towards a psychology of literacy: On the relations between speech and writing. *Cognition*, *60*(1), 83–104.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. *Applied Cognitive Psychology*, *20*(2), 139–156.

Oppenheimer, D. M., Diemand-Yauman, C., & Vaughan, E. B. (2011). Fortune favors the Bold (and the Italicized): Effects of disfluency. *Cognition*, *118*(1), 111–115.

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, *106*(3), 1178–1194.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*(28), 4646–4674.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, *423*(6941), 752–6.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of vision*, *4*(12), 12.

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), 20.

Perea, M., & Wang, X. (2017). Do alternating-color words facilitate reading aloud text in Chinese? Evidence with developing and adult readers. *Memory & Cognition*, *45*(7), 1160–1170. Retrieved from https://doi.org/10.3758/s13421-017-0717-0

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A. (1992). The Representation Problem in Reading Acquistion. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading Aquisition* (pp. 145–174). Lawrence Erlbaum Associates.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.

Pierrehumbert, J. B. (2001). Exemplar Dynamics: Word Frequency, Lenition and Contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistics Structure* (pp. 137–157). John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussen-hoven & N. Werner (Eds.), *Laboratory Phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter.

Prince, A., & Smolensky, P. (2008). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60–94.

Sanocki, T., & Dyson, M. C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention, Perception, & Psychophysics*, *74*(1), 132–145.

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104.

Segalowitz, N., & Gatbonton, E. (1995). Automaticity and lexical skills in second language fluency: Implications for CALL. *Computer Assisted Language Learning*, *8*(2–3), 129–149.

Shinohara, Y. (2014). Perceptual training of English/r/and/l/for Japanese adults, adolescents and children. UCL (University College London).

Shinohara, Y., & Iverson, P. (2013). Computer-based English /r/-/l/perceptual training for Japanese children. *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19).

Shinohara, Y., & Iverson, P. (2015). Effects of English /r/-/l/perceptual training on Japanese children's production. *The Scottish Consortium for ICPhS*.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833–870.

Stöckl, H. (2005). Typography: body and dress of a text-a signing mode between language and image. *Visual Communication*, *4*(2), 204–214.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131–145.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Sweller, J., & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction*, *12*(3), 185–233.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*(3), 339–363.

Vlaskamp, B. N. S., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision research*, *46*(3), 417–425.

Wagner, R. K., Torgesen, J. K., Bradley, L., Bryant, P., Crowder, B., Mann, V., Stanovich, K., et al. (1987). The Nature of Phonological Processing and Its Causal Role in the Acquisition of Reading Skills. *Psychological Bulletin*, *101*(2), 192–212.

Wang, Y., & Kuhl, P. K. (2003). Evaluating the "critical period" hypothesis: perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age,". *Poster Presented at the 15th International Congress of Phonetic Sciences* (pp. 1537–1540).

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The linguistic review*, *23*(3), 247–274.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85.

Whitney, D., & Levi, D. M. (2012). Visual Crowding: a dundamental limit on conscious perception and object recognition. *Trends Cogn Sci*, *15*(4), 160–168.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of experimental psychology. General*, *129*(1), 84–106.

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.

Appendix 1: Experiment Materials

Font 1: *SimSun.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 2: *Jing Chen Handwritten Pen Chinese Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 3: *Sharp anger Bold Figure Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 4: *The phantom earl Chinese Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 5: *Wandering the Wolf Razor Chinese Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 6: *Hao Li Handwritten pen handwriting Chinese Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 7: *SimHei.ttf*

中	center	水	water	白	white	争	fight	火	fire

长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 8*: Tweezers bread finger Chinese Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 9: *YunHe Sun Ink Brush (Writing Brush) Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat

体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud
东	east	听	hear	带	belt	线	string	房	house

Font 10: *Ye Gen You Sharp Hand Drawing Font-Simplified Chinese.ttf*

中	center	水	water	白	white	争	fight	火	fire
长	long	平	flat	光	light	风	wind	须	beard
明	bright	目	eye	王	king	车	car	早	morning
手	hand	市	market	色	color	林	forest	李	plum
头	head	金	metal	南	south	快	fast	吃	eat
体	body	指	finger	北	north	元	dollar	图	picture
果	fruit	队	team	边	side	字	letter	青	green
西	west	形	shape	觉	sleep	百	hundred	红	red
月	moon	书	book	师	teacher	城	city	响	loud

Appendix 2: Supplementary Data

Supplement to Figure 7-5: Experiment 1 Response Breakdown

	Accuracy			Response Time		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Correct (Trained)	47%	37%	36%	2920 ms	4947 ms	2445 ms
Correct (Untrained)	46%	47%	44%	2970 ms	4978 ms	2199 ms
Incorrect (Trained)	3%	13%	14%	2892 ms	4383 ms	3313 ms
Incorrect (Untrained)	5%	3%	6%	3480 ms	5212 ms	4383 ms

Supplement to Figure 7-6: Experiment 1 Response Times

Unigram Response Time (Combined)			
	Session 1 & 2		
	LV/LV	**LV/HV**	**HV/HV**
Median	2929 ms	3840 ms	2505 ms
Mean	2953 ms	4894 ms	2428 ms
Stdev	1149 ms	3202 ms	708 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 690.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.37853 -0.64793 -0.04699  0.54124  2.81021
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.03275  0.1810
##  Participant (Intercept) 0.10524  0.3244
##  Residual                0.33501  0.5788
## Number of obs: 360, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
```

```
## (Intercept)   8.1701     0.1200 17.5170  68.071  < 2e-16 ***
## ConditionC3  -0.5125     0.1656 15.8230  -3.095  0.00703 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.691
```

Unigram Response Time (Generalizing)			
	Session 1 & 2		
	LV/LV	**LV/HV**	**HV/HV**
Median	NA	4181 ms	2275 ms
Mean	NA	4693 ms	2622 ms
Stdev	NA	2165 ms	1047 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 257.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.41125 -0.54736 -0.05057  0.52099  2.89936
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.04197  0.2049
##  Participant (Intercept) 0.06471  0.2544
##  Residual                0.32215  0.5676
## Number of obs: 133, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.1687     0.1150 14.6690  71.061   <2e-16 ***
## ConditionC3  -0.4704     0.1649 15.5700  -2.853   0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.657
```

Supplement to Figure 7-7: Experiment 1 Accuracy Rates

Unigram Accuracy (Combined)			
	Session 1 & 2		
	LV/LV	**LV/HV**	**HV/HV**
Median	90%	85%	85%
Mean	93%	84%	81%
Stdev	5%	9%	14%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    341.2    356.8   -166.6    333.2      356
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7154  0.3471  0.3957  0.4534  0.6980
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.3311   0.5754
##  Participant (Intercept) 0.1718   0.4144
## Number of obs: 360, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5479     0.2734   5.661  1.5e-08 ***
## TrainingLV    0.2697     0.3566   0.756    0.449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.590
```

Unigram Accuracy (Generalizing)			
	Session 1 & 2		
	LV/LV	LV/HV	HV/HV
Median	NA	70%	75%
Mean	NA	70%	72%
Stdev	NA	27%	18%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    167.2    178.7    -79.6    159.2      129
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9474 -1.1316  0.5135  0.6415  1.0020
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  Target.Def  (Intercept) 1.912e-06 0.001383
##  Participant (Intercept) 5.024e-01 0.708781
## Number of obs: 133, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.8044     0.3923   2.050   0.0403 *
## TrainingLV    0.1923     0.5287   0.364   0.7160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.730
## convergence code: 0
## Model failed to converge with max|grad| = 0.00272208 (tol = 0.001,
component 1)
```

Experiment 1 d-Prime

	Session 1 & 2	Session 1	Session 2
LV/LV	2.87	3.70	2.37
LV/HV	2.22	2.32	2.12

HV/HV	1.81	1.90	1.73

Supplement to Figure 7-6: Experiment 1 Response Times by Session

Unigram Response Time (Combined)						
	Session 1			Session 2		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	2668 ms	3899 ms	2482 ms	3092 ms	3655 ms	2528 ms
Mean	3021 ms	5085 ms	2442 ms	2910 ms	4703 ms	2413 ms
Stdev	2014 ms	3784 ms	756 ms	714 ms	756 ms	741 ms

```
## Session 1
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
##
## REML criterion at convergence: 362.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44398 -0.67276 -0.01491  0.56679  2.53425
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.01049  0.1024
##  Participant (Intercept) 0.12172  0.3489
##  Residual                0.36400  0.6033
## Number of obs: 180, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.1812     0.1337 16.3430  61.204   <2e-16 ***
## ConditionC3  -0.5370     0.1878 15.9310  -2.859   0.0114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.703
```

```
## Session 2
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
##
## REML criterion at convergence: 357.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1713 -0.6649 -0.1072  0.4828  2.9396
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.002256 0.0475
##  Participant (Intercept) 0.080642 0.2840
##  Residual                0.371184 0.6092
## Number of obs: 180, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.1343     0.1147 15.9740  70.933  < 2e-16 ***
## ConditionC3  -0.4868     0.1619 15.9630  -3.007  0.00836 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.706
```

Unigram Response Time (Generalizing)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	NA	4064 ms	2290 ms	NA	4443 ms	2250 ms
Mean	NA	4946 ms	2627 ms	NA	4439 ms	2730 ms
Stdev	NA	2573 ms	913 ms	NA	1909 ms	1536 ms

```
## Session 1
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
##
## REML criterion at convergence: 125.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.89323 -0.54952 -0.05394  0.50448  2.91875
##
```

```
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.0000   0.0000
##  Participant (Intercept) 0.0530   0.2302
##  Residual                0.3452   0.5875
## Number of obs: 65, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.2524     0.1207 13.2920  68.387  < 2e-16 ***
## ConditionC3  -0.5353     0.1860 18.0330  -2.879  0.00999 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.649
```

```
## Session 2
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 138.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5059 -0.4624 -0.1111  0.4378  2.7881
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.1156   0.3400
##  Participant (Intercept) 0.0000   0.0000
##  Residual                0.3335   0.5775
## Number of obs: 68, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.0954     0.1162 47.0900  69.641  < 2e-16 ***
## ConditionC3  -0.4296     0.1610 65.2400  -2.668  0.00963 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC3 -0.604
```

Unigram Response Time Change over Sessions (Combined)			
	Session 1	**Session 2**	**Change**
LV/LV	3021 ms	2910 ms	-111 ms
LV/HV	5085 ms	4703 ms	-382 ms
HV/HV	2442 ms	2413 ms	-29 ms

Supplement to Figure 7-6: Experiment 1 Accuracy by Session

Unigram Accuracy (Combined)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	100%	90%	80%	90%	80%	80%
Mean	97%	84%	82%	88%	83%	80%
Stdev	5%	12%	14%	7%	9%	10%

```
## Session 1
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    172.3    185.1    -82.2    164.3      176
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5227  0.3614  0.3970  0.4733  0.6076
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  Target.Def  (Intercept) 4.832e-10 2.198e-05
##  Participant (Intercept) 2.322e-01 4.818e-01
## Number of obs: 180, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5313     0.3339   4.586 4.52e-06 ***
```

```
## TrainingLV    0.2363     0.4631   0.510     0.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.656
```

```
## Session 2
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    179.2    191.9    -85.6    171.2      176
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2361  0.4472  0.4472  0.5000  0.5000
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0        0
##  Participant (Intercept) 0        0
## Number of obs: 180, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.3863     0.2635   5.261 1.44e-07 ***
## TrainingLV    0.2231     0.3866   0.577    0.564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.682
```

Unigram Accuracy (Generalizing)						
	Session 1			Session 2		
	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV
Median	NA	75%	67%	NA	75%	75%
Mean	NA	68%	68%	NA	71%	76%
Stdev	NA	30%	32%	NA	27%	16%

```
## Session 1
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##     89.4     98.1    -40.7     81.4       61
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4712 -0.9966  0.4814  0.6193  0.9345
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.4814   0.6939
##  Participant (Intercept) 0.3987   0.6314
## Number of obs: 65, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.4445     0.5263   0.845    0.398
## TrainingLV    0.5478     0.7009   0.782    0.434
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.690
```

```
## Session 2
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##     86.5     95.4    -39.3     78.5       64
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7321 -1.6237  0.5774  0.6159  0.6159
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0        0
##  Participant (Intercept) 0        0
## Number of obs: 68, groups:  Target.Def, 37; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0986     0.4364   2.517   0.0118 *
## TrainingLV   -0.1292     0.5620  -0.230   0.8182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.777
```

Unigram Response Accuracy over Sessions (Combined)			
	Session 1	**Session 2**	**Change**
LV/LV	97%	88%	-9%
LV/HV	84%	83%	-1%
HV/HV	82%	80%	-2%

Supplement to Figure 7-9: Experiment 1 Frequency Effects

Unigram Frequency Effects (Accuracy)			
	LV/LV	**LV/HV**	**HV/HV**
1x	81%	67%	39%
4x	88%	67%	72%
9x	100%	78%	75%
16x	100%	94%	81%
25x	100%	61%	94%

Unigram Frequency Effects (Response Times)				
		LV/LV	**LV/HV**	**HV/HV**
1x	**Correct**	5329 ms	3302 ms	2249 ms
	Incorrect	3189 ms	3474 ms	1842 ms
4x	**Correct**	2546 ms	7323 ms	2742 ms
	Incorrect	2446 ms	6456 ms	6795 ms
9x	**Correct**	2322 ms	3914 ms	2009 ms
	Incorrect	NA	1097 ms	4880 ms
25x	**Correct**	2533 ms	5693 ms	2202 ms
	Incorrect	NA	5064 ms	2799 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 694.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.43777 -0.63063 -0.06685  0.53775  2.83014
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.02977  0.1725
##  Participant (Intercept) 0.10597  0.3255
```

```
##  Residual                0.33676  0.5803
```

```
## Number of obs: 360, groups:  Target.Def, 49; Participant, 18
```

```
##
## Fixed effects:
##                          Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)               7.62906    0.12342  19.30000  61.816  < 2e-16 ***
## TrainingLV                0.51369    0.16611  15.80000   3.092  0.00706 **
## log(Training.Freq + 1)    0.02490    0.02605 323.40000   0.956  0.34000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrnnLV
## TrainingLV  -0.674
## lg(Trn.F+1) -0.234  0.007
```

Supplement to Figure 7-10: Experiment 1 Definitions

Unigram Definitions Response Time (All)									
	Session 1 & 2			Session 1			Session 2		
	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV
Mean	3801 ms	5016 ms	3520 ms	4034 ms	4909 ms	3927 ms	3602 ms	5123 ms	3113 ms
Median	3299 ms	5146 ms	3360 ms	3923 ms	3690 ms	3502 ms	3205 ms	3733 ms	3081 ms
Stdev	1617 ms	2407 ms	1197 ms	1647 ms	3070 ms	1817 ms	1800 ms	2995 ms	901 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 251.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.6259 -0.6202 -0.1717  0.4310  3.0827
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.00000  0.0000
##  Participant (Intercept) 0.08016  0.2831
##  Residual                0.27793  0.5272
## Number of obs: 146, groups:  Target.Def, 41; Participant, 18
```

```
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   7.9905     0.1126 14.3630  70.973   <2e-16 ***
## TrainingLV    0.2207     0.1607 14.7460   1.373     0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## TrainingLV -0.701
```

Unigram Definitions Accuracy (All)									
	Session 1 & 2			Session 1			Session 2		
	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV
Mean	79%	64%	56%	89%	68%	57%	65%	61%	55%
Median	80%	65%	70%	83%	80%	67%	69%	60%	60%
Stdev	16%	19%	27%	10%	10%	30%	23%	19%	24%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    327.3    341.9   -159.7    319.3      275
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1526 -0.5811 -0.5111  0.9043  2.4212
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.08119  0.2849
##  Participant (Intercept) 0.37290  0.6107
## Number of obs: 279, groups:  Target.Def, 49; Participant, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8671     0.2919  -2.971  0.00297 **
## TrainingLV   -0.4427     0.4093  -1.082  0.27943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr)
## TrainingLV -0.639
```

Supplement to Figure 8-5: Bigram Response Breakdown

	Accuracy			Response Time		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Correct (Trained)	44%	21%	39%	4261 ms	4598 ms	4940 ms
Correct (Untrained)	46%	45%	44%	3307 ms	3838 ms	5262 ms
Incorrect (Trained)	6%	30%	11%	3836 ms	4109 ms	5507 ms
Incorrect (Untrained)	4%	4%	6%	12706 ms	7143 ms	6406 ms

Supplement to Figure 8-6: Experiment 2 Response Times

Bigram Response Time (All)			
	Session 1 & 2		
	LV/LV	**LV/HV**	**HV/HV**
Median	3298 ms	4089 ms	4481 ms
Mean	4112 ms	4275 ms	5162 ms
Stdev	1694 ms	1395 ms	2087 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 297.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.47043 -0.67653  0.02329  0.65230  2.38405
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.01814  0.1347
##  Participant (Intercept) 0.10700  0.3271
##  Residual                0.25193  0.5019
```

```
## Number of obs: 175, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.2037     0.1193 18.2890  68.746   <2e-16 ***
## ConditionC6   0.1283     0.1705 17.1250   0.753    0.462
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.681
```

Bigram Response Time (Generalizing)			
	Session 1 & 2		
	LV/LV	**LV/HV**	**HV/HV**
Median	NA	4176 ms	4588 ms
Mean	NA	4300 ms	5225 ms
Stdev	NA	1309 ms	2549 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 238.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.37201 -0.57983 -0.00683  0.58888  2.14307
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.05616  0.2370
##  Participant (Intercept) 0.10796  0.3286
##  Residual                0.24109  0.4910
## Number of obs: 134, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.1754     0.1259 18.7090  64.958   <2e-16 ***
## ConditionC6   0.1411     0.1790 17.6480   0.788    0.441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.653
```

Bigram Accuracy (All)			
	Session 1 & 2		
	LV/LV	LV/HV	HV/HV
Median	93%	70%	90%
Mean	91%	68%	84%
Stdev	10%	11%	12%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    211.0    223.7   -101.5    203.0      171
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4248 -0.6452  0.3796  0.5407  1.6720
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.2292   0.4787
##  Participant (Intercept) 0.7154   0.8458
## Number of obs: 175, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5665     0.4565   3.432 0.000600 ***
## TrainingLV   -1.8921     0.5732  -3.301 0.000963 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.762
```

Bigram Accuracy (Generalizing)			
	Session 1 & 2		
	LV/LV	LV/HV	HV/HV
Median	NA	43%	83%
Mean	NA	53%	76%
Stdev	NA	20%	17%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    171.6    183.2    -81.8    163.6      130
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0769 -0.6356  0.4390  0.6975  1.8258
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  Target.Def  (Intercept) 4.297e-09 6.555e-05
##  Participant (Intercept) 6.105e-01 7.813e-01
## Number of obs: 134, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.1532     0.4239   2.721  0.00652 **
## TrainingLV   -1.6998     0.5558  -3.058  0.00223 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.769
```

Experiment 2 d-Prime

Bigram d-Prime			
	Session 1 & 2	**Session 1**	**Session 2**
LV/LV	2.65	2.72	2.59
LV/HV	1.09	1.07	1.12
HV/HV	1.99	2.93	1.43

Supplement to Figure 8-8: Experiment 2 Response Times by Session

Bigram Response Time (Combined)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	3435 ms	3602 ms	4747 ms	3309 ms	4897 ms	4480 ms
Mean	4184 ms	3894 ms	4824 ms	4041 ms	4602 ms	5684 ms
Stdev	1821 ms	1480 ms	1721 ms	1786 ms	1438 ms	2863 ms

```
## Session 1
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
##
## REML criterion at convergence: 167
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9547 -0.4149 -0.1169  0.5331  1.9492
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.03082  0.1756
##  Participant (Intercept) 0.13896  0.3728
##  Residual                0.22697  0.4764
## Number of obs: 95, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.0731     0.1393 17.9040  57.938   <2e-16 ***
## ConditionC6   0.2281     0.1991 16.8490   1.146    0.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.676
```

```
## Session 2
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 129.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.77805 -0.58966 -0.01206  0.55204  1.68948
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.06248  0.2500
##  Participant (Intercept) 0.11100  0.3332
##  Residual                0.17141  0.4140
## Number of obs: 80, groups:  Target.Def, 41; Participant, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  8.36915    0.14305 15.97700  58.505   <2e-16 ***
## ConditionC6  0.02428    0.19578 14.12100   0.124    0.903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.687
```

Bigram Response Time (Generalizing)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	NA	3628 ms	4199 ms	NA	4896 ms	4506 ms
Mean	NA	3738 ms	4489 ms	NA	4582 ms	6031 ms
Stdev	NA	1592 ms	1750 ms	NA	1380 ms	3271 ms

```
## Session 1
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
##
## REML criterion at convergence: 134.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.69493 -0.47373 -0.06337  0.44873  1.61109
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.06332  0.2516
##  Participant (Intercept) 0.15756  0.3969
##  Residual                0.20981  0.4580
## Number of obs: 73, groups:  Target.Def, 40; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.0206     0.1512 16.6350  53.030   <2e-16 ***
## ConditionC6   0.2460     0.2182 16.8960   1.127    0.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.652
```

```
## Session 2
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 99
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4669 -0.5257 -0.1294  0.5454  1.7002
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.1033   0.3214
##  Participant (Intercept) 0.1214   0.3484
##  Residual                0.1285   0.3585
## Number of obs: 61, groups:  Target.Def, 36; Participant, 16
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.3428     0.1514 14.9350  55.104   <2e-16 ***
## ConditionC6   0.1062     0.2106 13.8150   0.504    0.622
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## ConditionC6 -0.662
```

Bigram Response Time Change over Sessions (All)			
	Session 1	**Session 2**	**Change**
LV/LV	4184 ms	4041 ms	-144 ms
LV/HV	3894 ms	4602 ms	708 ms
HV/HV	4824 ms	5684 ms	860 ms

Supplement to Figure 8-8: Experiment 2 Accuracy by Session

Bigram Accuracy (Combined)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	90%	70%	90%	95%	60%	80%
Mean	91%	69%	90%	90%	65%	76%
Stdev	8%	12%	5%	14%	16%	22%

```
## Session 1
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    118.7    128.9    -55.3    110.7       91
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2823 -0.8241  0.4120  0.5392  1.0941
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.3038   0.5512
##  Participant (Intercept) 0.1352   0.3677
```

```
## Number of obs: 95, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.6977     0.5251   3.233  0.00122 **
## TrainingLV   -1.8378     0.5987  -3.069  0.00214 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.826
```

```
## Session 2
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##     78.7     88.2    -35.3     70.7       76
##
## Scaled residuals:
##       Min        1Q    Median        3Q       Max
## -0.122233 -0.001952  0.000000  0.000440  0.126402
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 5868     76.61
##  Participant (Intercept) 2880     53.66
## Number of obs: 80, groups:  Target.Def, 41; Participant, 16
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   27.246      6.101   4.466 7.98e-06 ***
## TrainingLV   -48.579      8.913  -5.450 5.03e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.880
```

Bigram Accuracy (Generalizing)						
	Session 1			**Session 2**		
	LV/LV	**LV/HV**	**HV/HV**	**LV/LV**	**LV/HV**	**HV/HV**
Median	NA	40%	75%	NA	58%	81%
Mean	NA	43%	80%	NA	64%	73%
Stdev	NA	31%	17%	NA	15%	25%

```
## Session 1
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##     Session == "S1")
##
##      AIC      BIC   logLik deviance df.resid
##     97.9    107.0    -44.9     89.9       69
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1595 -0.7958  0.4346  0.7082  1.2531
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.51474  0.7175
##  Participant (Intercept) 0.05037  0.2244
## Number of obs: 73, groups:  Target.Def, 40; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.4643     0.6146   2.382   0.0172 *
## TrainingLV   -1.7130     0.6852  -2.500   0.0124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.843
```

```
## Session 2
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
```

```
##      AIC      BIC   logLik deviance df.resid
##     78.1     86.6    -35.1     70.1       57
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3879 -0.4815 -0.2907  0.5640  1.6183
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.4991   0.7065
##  Participant (Intercept) 2.6263   1.6206
## Number of obs: 61, groups:  Target.Def, 36; Participant, 16
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.176      1.019   1.154    0.248
## TrainingLV    -2.391      1.571  -1.522    0.128
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.833
```

Bigram Accuracy Change over Sessions (All)			
	Session 1	**Session 2**	**Change**
LV/LV	91%	90%	-1%
LV/HV	69%	65%	-4%
HV/HV	90%	76%	-14%

Supplement to Figure 8-9: Experiment 2 Accuracy by Session

Bigram Accuracy by Frequency			
	LV/LV	**LV/HV**	**HV/HV**
1x	81%	44%	47%
4x	81%	33%	82%
9x	88%	44%	82%
16x	100%	44%	88%
25x	94%	44%	94%

Bigram Response Time by Frequency				
		LV/LV	**LV/HV**	**HV/HV**
1x	**Correct**	5628 ms	5652 ms	5735 ms
	Incorrect	3408 ms	3882 ms	5466 ms
4x	**Correct**	4545 ms	4076 ms	3833 ms
	Incorrect	3781 ms	4693 ms	4089 ms
9x	**Correct**	3641 ms	3884 ms	5144 ms
	Incorrect	NA	3181 ms	3672 ms
25x	**Correct**	4253 ms	4588 ms	4000 ms
	Incorrect	1076 ms	3952 ms	6052 ms

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    202.3    221.3    -95.2    190.3      169
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3187 -0.6243  0.2528  0.6505  1.7775
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.1334   0.3652
##  Participant (Intercept) 0.8002   0.8945
## Number of obs: 175, groups:  Target.Def, 43; Participant, 19
##
## Fixed effects:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -0.7327     0.7597  -0.964   0.3348
## TrainingLV                         0.2528     1.0182   0.248   0.8040
## log(Training.Freq + 1)             1.1944     0.3740   3.194   0.0014 **
## TrainingLV:log(Training.Freq + 1)  -1.1289     0.4593  -2.458   0.0140 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) TrnnLV l(T.+1
## TrainingLV  -0.739
## lg(Trn.F+1) -0.791  0.594
```

```
## TLV:(T.F+1)  0.642 -0.807 -0.815
```

Supplement to Figure 8.10: Experiment 2 Definitions

Bigram Definitions Response Time									
	Session 1 & 2			Session 1			Session 2		
	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV
Mean	4817 ms	5541 ms	5054 ms	4917 ms	6191 ms	5367 ms	4716 ms	4731 ms	4742 ms
Median	4241 ms	4531 ms	4443 ms	4975 ms	4975 ms	3874 ms	3409 ms	4360 ms	3298 ms
Stdev	2405 ms	2642 ms	2730 ms	1615 ms	5127 ms	3183 ms	3781 ms	1499 ms	3578 ms

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## REML criterion at convergence: 232.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1458 -0.5807 -0.1328  0.5391  2.9760
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.05315  0.2306
##  Participant (Intercept) 0.08583  0.2930
##  Residual                0.36468  0.6039
## Number of obs: 111, groups:  Target.Def, 42; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   8.2450     0.1311 14.4340  62.912   <2e-16 ***
## TrainingLV    0.1229     0.1885 17.2870   0.652    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.653
```

Bigram Definitions Accuracy									
	Session 1 & 2			Session 1			Session 2		
	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV	LV/LV	LV/HV	HV/HV
Mean	66%	39%	62%	71%	50%	71%	62%	25%	52%
Median	70%	45%	65%	73%	45%	80%	63%	18%	50%
Stdev	22%	21%	28%	18%	30%	23%	31%	27%	39%

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
##
##      AIC      BIC   logLik deviance df.resid
##    107.1    117.9    -49.5     99.1      107
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2615 -0.3546 -0.2432 -0.2079  2.3782
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Target.Def  (Intercept) 0.4963   0.7045
##  Participant (Intercept) 2.4712   1.5720
## Number of obs: 111, groups:  Target.Def, 42; Participant, 19
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.0723     0.8208  -2.525   0.0116 *
## TrainingLV    0.1222     1.0057   0.122   0.9033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## TrainingLV -0.580
```