Temporal Integration of English Words: Evidence for a Processing Hierarchy in Visual Word Recognition

by

Ronald Chu

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Psychology
University of Toronto

© Copyright by Ronald Chu 2012
Temporal Integration of English Words: Evidence for a Processing 
Hierarchy in Visual Word Recognition

Ronald Chu

Master of Arts

Psychology
University of Toronto

2012

Abstract

Several models of visual word recognition suggest a processing hierarchy; basic orthographic features are processed early and whole-word representations are processed late in the hierarchy. Unfortunately, given the extreme efficiency of the visual word recognition system, studies typically focus on one specific level of the processing hierarchy (e.g., orthographic, phonological and/or semantic processing). Furthermore, different paradigms are used to study different levels of the hierarchy. Fortunately, data across different studies in the literature do converge to two distinct temporal thresholds for letter perception and whole-word integration. The current experiments assessed the temporal thresholds for both letter perception and whole-word integration using a single novel paradigm. The results demonstrated distinct temporal thresholds for letter perception and whole-word integration which agree with those reported in the literature. Thus, the current experiments provide further behavioral evidence that the visual word recognition is a hierarchical process.
Acknowledgments

I’d like to acknowledge my supervisor Steve Joordens and my subsidiary advisor Eyal Reingold for their feedback and contributions to this manuscript.
# Table of Contents

List of Tables .................................................................................................................. v

List of Figures .................................................................................................................. vi

Chapter 1 Temporal Integration of Visual Word Recognition ........................................... 1

1 Temporal Integration- High and Low Resolution .......................................................... 3
  1.1 High Resolution Processes ..................................................................................... 5
  1.2 Low Resolution Processes ................................................................................... 8
  1.3 Current Experiments .............................................................................................. 13
    1.3.1 Experiment 1a ............................................................................................... 15
    1.3.2 Experiment 1b ............................................................................................... 21
    1.3.3 Experiment 1c ............................................................................................... 24
    1.3.4 Experiment 2 ............................................................................................... 27
  1.4 General Discussion ................................................................................................. 39
  1.5 Conclusion .............................................................................................................. 47

References ..................................................................................................................... 49
List of Tables

Table 1 - Mean Accuracy and Standard Deviations for Experiments 1a-c

Table 2 - Mean Percentage of Correct, Neighborhood and Noise Responses for Experiment 2
List of Figures

Figure 1- Sample presentation sequence for a typical trial of the experimental paradigm

Figure 2- Accuracy data for Experiment 1a

Figure 3- Accuracy data for Experiment 1b

Figure 4- Accuracy data for Experiment 1c

Figure 5- Accuracy data for Experiment 2

Figure 6- Percentage of Neighborhood Responses for Experiment 2

Figure 7- Percentage of Noise Responses for Experiment 2
Chapter 1
Temporal Integration of Visual Word Recognition
Our visual systems are constantly bombarded with information, yet miraculously this mass of information is processed very rapidly, allowing for a coherent, continuous percept of the world. Though we perceive the visual world as continuous, like a film, research has revealed that our brains actually process visual information across chunks of time, like a slideshow. Furthermore, the duration of these chunks will vary depending on the complexity of the stimulus; basic stimuli are processed quickly, while complex stimuli are processed over longer periods.

Research in psychophysics and visual perception has uncovered very basic perceptual processes that operate at very high temporal resolutions. For example, research on the critical flicker frequency (CFF) has shown that individual flashes of light can be discriminated at very high presentation frequencies (~50Hz); at faster frequencies, the individual flashes are perceived as a continuous light source. Recently, there has been emerging interest in the literature regarding more complex processes with slower temporal resolutions. The processing of faces, words and orientation and color across space for example has been demonstrated to occur at a time scale of approximately 5-20 Hz (Holcombe, 2009; Holcombe, 2001; Bodelon, Fallah & Reynolds, 2007). Thus, there seems to be a clear disparity in the temporal resolution for the processing of basic visual features (e.g., flashes of light and line orientations) and more complex visual stimuli made up of multiple features (e.g., faces and words).

The current research suggests a temporal processing hierarchy, wherein increasingly complex stimuli are integrated over longer periods of time. That is, our visual systems have naturally developed perceptual processes that can identify basic elements at a very high temporal resolution. However, our visual systems can also combine these basic elements to perceive more
complex stimuli, though this occurs at a much slower resolution. The goal of the current manuscript is to assess the specific temporal hierarchy of visual word recognition.

While several word recognition models depict visual word recognition as a hierarchical process (e.g., IAM, McClelland & Rumelhart, 1989; DRC, Coltheart et al., 2001; SERIOL, Whitney, 2001; SOLAR, Davis, 1999) behavioral studies typically focus on one specific level of the hierarchy. For example, letter perception (Petit & Grainger, 2002; Ziegler et al., 2000), whole-word integration (Holcome & Judson, 2007; Forget, Buiatti & Dehaene, 2009; Mewhort & Beal, 1977) and semantic processing (Lucas, 2000; Ortells, Vellido, Daza & Noguera, 2006) are usually assessed across separate experiments and usually separate studies. As such, current behavioral evidence for such a processing hierarchy come from distinct, though converging, lines of research. The goal of the current experiments was to provide further behavioral evidence for such a hierarchy using a single novel paradigm designed to assess the temporal characteristics of both letter perception and whole-word integration.

This manuscript will first outline the research regarding fast and slow perceptual processes. This will be followed by a focus on the research demonstrating a temporal processing hierarchy for visual word recognition. Finally, I will describe a novel paradigm that has successfully captured two distinct thresholds for letter perception and whole-word integration. Importantly, the reported thresholds are consistent with those reported in the literature.

1 Temporal Integration - High and Low Resolution

The visual system has often been described as a hierarchical system; basic location-mediated features (e.g., color, orientation and luminance) are assessed early in the processing stream and higher-order location invariant representations (e.g., faces and words) processed later in the processing stream. Several studies of object recognition have revealed anatomical
evidence to support this general hierarchical structure in the ventral stream of the visual system.

Following stimulation of the retina, the first cortical region to receive visual information is the primary visual cortex (V1) (Hendry & Reid, 2000; Dacey, 2000). Electrophysiological readings of primates have revealed that V1 neurons respond to very basic features like bar orientations (Hubel & Wiesel, 1968). Further downstream the visual information proceeds through areas V2, V4 and IT where information regarding edges and illusory contours (Peterhans & von der Heydt, 1991), colour and form (Gallant, Braun & Van Essen, 1993; Schein & Desimone, 1990), and specific shapes (Booth & Rolls, 1998; Desimone, Albright, Gross & Bruce, 1984) are processed, respectively.

Perhaps not surprisingly, the hierarchical nature of visual processing is also reflected in the temporal integration thresholds for specific features and shapes. That is, the integration thresholds increase in duration as you venture further down the visual processing stream, especially when the neuronal receptive fields respond to objects made up of spatially separated features (e.g., faces and words) (Holcombe, 2009). The following section will briefly describe the research regarding the integration thresholds for individual and conjunctive features. This manuscript is currently supporting the stance initially proposed by Holcombe (2009), that the processing of individual features (e.g., colour, luminance, orientation) and spatially overlapped conjunctive features (e.g., a green, rightward bar) occurs at a fast timescale, while the processing of objects made up of spatially separated features (e.g., faces and words) occurs at a slow timescale.
1.1 High Resolution Processes

The research outlined in the following section is aimed to demonstrate the temporal integration thresholds for various basic features. The temporal integration threshold is determined as the fastest presentation rate at which the alternating stimuli can be perceived as separate; at rates quicker than the threshold the stimuli are perceived as being “fused” together (e.g., alternating sinusoidal gratings of opposing directions are perceived as a static plaid pattern at very fast alternations). Typically, such studies will present alternating stimuli at variable alternation rates to determine the frequency at which stimuli discrimination is not possible.

Flicker Frequency - Luminance and Colour

The integration threshold of individual flashes of light has been studied for nearly a century (Bartley, 1936; Bartley, 1939). The critical flicker frequency, defined as the minimum frequency at which individually presented flashes are perceived as a continuous light source, has been demonstrated to be approximately 50 Hz, though factors such as stimuli intensity and luminance can modulate the CFF (Bartley, 1939). Since then, CFFs as high as 70Hz have been reported (Brettl, Shi, & Strasburger, 2006). Perhaps not surprisingly, given the further downstream processing of color (V4), Bodelon, Fallah & Reynolds (2007) reported a slightly slower threshold of 45Hz for color.

Orientation Alone

Neurophysiological studies have found that V1 neurons are tuned specifically to individual bar orientations (Hubel & Wiesel, 1968); incidentally the reported temporal resolutions for bar orientations are fairly rapid. However, the temporal resolution for bar
orientations alone has not been thoroughly studied and there is seemingly conflicting results regarding the specific temporal integration threshold for line gratings. Using an orientation discrimination task, Bodelon, Fallah & Reynolds (2007) reported that alternating presentations of sinusoidal gratings of opposing directions led to stimulus integration at 125Hz. However, subjective ratings of stimulus integrations revealed a threshold of 38Hz (Brettel, Shi & Strasburger, 2006).

The discrepancy in integration thresholds observed for discrimination versus subjective ratings tasks may be due to the differing methodology across the studies. More specifically, Bodelon, Fallah & Reynolds (2007) may have overestimated the threshold frequency due to the nature of their task. Participants were presented with a stream of alternating sinusoidal gratings of opposing orientation and color (e.g., a red rightward facing grating and a green leftward facing grating). Immediately following the presentation stream, the participants were presented with four orientation choices. One alternative was a grating presented during the presentation stream. Another alternative was a misconjunction of a color and an orientation that was presented in the sequence. The final two alternatives included orientations not included in the presentation stream. Thus, half of the multiple choice options included orientations that were presented within the stream. The authors determined orientation discrimination by measuring the percentage of responses that corresponded to the presented orientations. However, since the other two orientations were never presented in the stream, the ‘orientation discriminations’ could be made even if alternating stimuli were integrated into a single percept.

So what does their reported temporal threshold of 125Hz actually correspond to? The authors took extra care to modify the luminance so that at very high temporal frequencies, all orientation information disappeared (the fused stimuli looks like a blank circle of uniform color
and luminance). Thus, their proposed threshold for orientation discrimination actually corresponds to a luminance (flicker) threshold. Given that the integration thresholds vary as a function of luminance (Shwartz, 1992; Anderson & Vingrys, 2000), the unusually high frequency of 125Hz might correspond to the CFF of their specific, luminance-tuned, stimuli.

Thus, it is most likely that the temporal threshold for line orientations corresponds closer to the 38Hz reported by Brettel, Shi & Strasburger (2006). This is further supported by the fact that the critical flicker frequencies have been reported to be 50-70Hz (Brettl, Shi, & Strasburger, 2006; Holcombe, 2006, Bartley, 1939) and even standard fluorescent lighting is capped at a frequency of 100Hz. It is hard to imagine that the temporal resolution for line orientations would be greater than that for basic flashes of light, which is processed upstream of the primary visual cortex (Mize & Murphy, 1976).

Orientation and Luminance Binding

Sinusoidal gratings of differing luminance and color have been used to study the integration threshold for spatially overlapped, conjunctive features (i.e., orientation and luminance or orientation and color). Holcombe (2001) was the first to demonstrate the integration threshold for orientation and luminance binding. In Experiment 2, Holcombe (2001) presented two alternating sinusoidal gratings of equal amplitude but differing luminance and opposing orientations; in half of the trials, the rightward facing gratings were of higher luminance and in the other half they were of lower luminance. Furthermore, the alternating gratings were presented at the same spatial location and at various alternation rates.

Participants were asked to determine the orientation of the higher luminance gratings. At alternation rates of 21Hz subjects successfully identified the orientations for the brighter
gratings. Interestingly, performance dropped to chance levels at the 42 Hz condition. The findings suggest that the integration threshold for line orientations and luminance lies somewhere between 21-42Hz. Subsequent studies have also replicated this threshold (Holcombe & Cavanagh, 2001; Bodelon, Fallah & Reynolds, 2007). Thus, the research has demonstrated that the integration threshold for the conjunction of luminance and orientation is slower than that for orientation and luminance alone.

**Orientation and Color Binding**

Researchers have also measured the temporal threshold for the conjunctive processing of orientation and color (Bodelon, Fallah & Reynolds, 2007; Holcombe & Cavanagh, 2001). Using presentation streams of alternating sinusoidal gratings of opposing orientation and color the authors reported temporal integration thresholds for colour and orientation conjunctions of 30Hz (Bodelon, Fallah & Reynolds, 2007) and 18.8Hz (Holcombe and Cavanagh, 2001). The discrepancy across experiments was attributed to difference in luminance and contrast, which has been shown to influence integration thresholds (Shwartz, 1992, Anderson & Vingrys, 2000). Importantly, the findings support the claim that the temporal resolution for conjunctive features is slower than that of the constituent features. Again, this is in accordance with the processing hierarchy of the visual system, wherein individual visual features are processed early in the stream and more complex multi-feature objects are processed later in the stream (Hubel & Wiesel, 1968; Booth & Rolls, 1998; Desimone, Albright, Gross & Bruce, 1984).

### 1.2 Low Resolution Processes

The preceding section demonstrated variety of stimuli that are integrated across very rapid thresholds. Interestingly, the research suggests that processing of conjunctive features occurs at a slower resolution than the processing the individual features. Recent research has also
uncovered that processing of conjunctive features across space occurs on an even slower (~10Hz) time scale. These include studies demonstrating binding of spatially separated orientation and color, verbal stimuli and faces.

**Spatially Separated Orientation and Color**

Holcombe and Cavanagh (2001) demonstrated that the integration threshold for conjunctive features increased when the feature were separated spatially. The authors presented participants with alternating presentations of circular stimuli; the top of the stimuli contains either rightward or leftward facing grey sinusoidal gratings, while the bottom half of the stimuli was either red or green. Following each stimuli stream the participants were asked to determine which orientation was paired with the red half circle. Thus, to make correct responses the participants must successfully bind the color and orientation information across space.

The authors found a critical threshold for accurately identifying the color-orientation pairing to be less than 3Hz. Conversely, when the color and orientation were superimposed, the critical threshold increased to 18.8Hz. The authors suggest that attention might have been the mediating processes that allow these spatially distinct features to be combined, and that this attentive binding required a longer integration period. Importantly, these findings demonstrate that the temporal threshold for spatially separated features is significantly higher than that for spatially superimposed features.

One concern regarding the Holcombe and Cavanagh (2001) study is the lack of ecological validity; people rarely have to integrate spatially disparate features, like color and orientation, into a single percept. Thus, the longer threshold may simply reflect the novelty of
their specific task. However, recent research in face perception and word recognition suggests that these long thresholds may be a part of everyday perceptual processing.

Faces

Face perception has been thoroughly studied in cognitive psychology (see Tsao & Livingstone, 2008 for a comprehensive review). However, studies of face perception rarely look at how facial information is integrated across time. Thus, there is still a question regarding whether all facial information is analyzed simultaneously or separately and integrated over time. Anaki, Boyd & Moscovitch (2007), recently addressed this issue using a partial presentation paradigm and demonstrated that face perception has an integration threshold of approximately 5Hz.

They presented participants with famous faces separated into thirds, with each third containing one critical piece of facial information (e.g., the eyes, nose, and mouth). These face parts were then presented in sequential manner at different presentation frequencies. The participants were tasked to identify the famous face following each trial. Given that working memory could aid in face perception, identification is possible even at the lowest presentation frequencies. Thus, the critical factor was not determined by accuracy or reaction time alone. Instead, the authors determined the integration threshold by looking at the frequency where the inversion effect disappeared. The inversion effect is the finding that inverted faces are much harder to identify than upright faces (Diamond & Carey, 1986; Valentine, 1988; Yin, 1969). More importantly, this effect is unique to face perception and as such has been used as a marker for ‘normal’ face perception.
The results from the experiments demonstrated that the inversion effect was present up until 5Hz presentation frequencies. At presentation frequencies longer than 5Hz, inverted and upright faces were identified with comparable speed and accuracy. This suggests that, at longer presentation frequencies, subjects were adopting strategies not typically used in normal face perception. More importantly this study demonstrates that even with a highly specialized task, such as face perception, the binding of spatially disparate features require long integration times.

**Words**

Models of word recognition usually assume parallel processing of a word’s constituent letters (e.g., McClelland & Rumelhart, 1981; Coltheart et al., 2001, Whitney, 2001). While these models briefly mention a temporal reset mechanism, there has been very little empirical attention regarding the temporal integration in word recognition. However, recent studies using partial target presentation methodologies have shed some light on the temporal characteristics of letter integration, demonstrating a temporal threshold, wherein successively presented word fragments are processed as a singular verbal percept.

Forget, Buiatti and Dehaene (2009) presented participants with alternating presentations of a word’s even and odd positioned letters (e.g., C_A_P_G_E and _H_M_A_N_ for the word CHAMPAGNE) at various frequencies. The authors demonstrated a 12.5Hz threshold for letter integration; as long as these even/odd letter pairs were alternated at frequencies faster than 12.5Hz they fused into a single percept of the whole word. Similarly, Holcombe and Judson (2007) presented participants with two alternating words at various frequencies and demonstrated that at frequencies faster than 5Hz, the alternating words appeared as if they were transparently overlaid.
Interestingly, Mewhort and Beal (1977), demonstrated a similar integration window, even when letters were presented individually. Their task was a simple word identification task, wherein participants were asked to identify words presented one letter at a time. The letter presentations occurred in rapid succession (5-85Hz) and each letter maintained their original position from the target word. For example, the presentation of CAT would consist of the following stimuli: C _ _ , _ A _ , _ _ T. The results demonstrated that word identification was most accurate with 60Hz presentations. Furthermore, accuracy dropped at slower frequencies, reaching a minimum at 8Hz. These results are consistent with the ~10Hz integration window suggested by Holcombe and Judson (2007) and by Forget, Buiatti and Dehaene (2009). Together, the studies suggest that there is indeed a temporal threshold for visual word recognition and that all verbal stimuli presented within that threshold are integrated into a singular percept.

**Plausible Neural Mechanism for a Processing Hierarchy in Visual Word Recognition**

Dehaene, Cohen, Sigman and Vinckier (2005) proposed a plausible neural model for visual word recognition wherein basic perceptual features of words and letters are processed early in the visual stream; contrast, orientation and local contrast are processed in the LGN, V1 and V2, respectively. Further downstream, case-specific letter shapes are processed in V4 while case-invariant abstract detectors are processed in V8. Finally, local bigrams and reoccurring substrings (e.g., morphemes and short words) are processed in occipitotemporal sulcus. Such a processing hierarchy does explain why whole-word recognition is associated with a much longer integration threshold than more basic visual features like line orientation (Brettl, Shi, & Strasburger, 2006; Holcombe, 2006, Bartley, 1939), luminance (Bartley, 1936; Bartley, 1939), and colour (Bodelon, Fallah & Reynolds, 2007), which are processed further upstream.
Dahaene et al., (2004) demonstrated neurophysiological evidence for such a processing hierarchy used a masked priming paradigm under fMRI. Specifically, they studied activation at visual word form area (VWFA), an extended strip of the posterior occipitotemporal cortex, which has been demonstrated to activate in response to visual word presentation (Cohen et al., 2000, 2002, Nobre, Allison & McCarthy, 1994). The authors observed fMRI adaptation (reduced activation following repetition priming) even when the repetition prime and target varied in case (e.g., home-HOME), suggesting that case-invariant processing of letter is associated with activation of the VWFA. Critically, the authors also revealed hierarchical processing of this specific strip of occipitotemporal cortex.

Analysis of fMRI adaptation revealed three functionally distinct subdivisions of VWFA. Adaptation at the posterior VWFA was dependent on prime-target overlap at specific letter positions; fMRI adaptation disappeared when the prime and target was offset by one letter. More anteriorly, adaptation persisted even after small shifts in letter positions. Furthermore, adaptation in this region was comparable for both repetition and anagram primes. Finally, in the most anterior region, adaptation was stronger for repetition than anagram primes. Thus, the evidence suggests a hierarchical structure for orthographic processing in the VFMA; location-specific letter information is processed posteriorly, location-invariant chunks are processed more anteriorly and finally whole-words are processed at the most anterior VWFA. However, the question remains, is this structural hierarchy also reflected in a temporal processing hierarchy?

1.3 Current Experiments

The preceding section outlined a series of studies demonstrating both slow and fast integration thresholds for individual and conjunctive features. The evidence currently suggests that the processing of basic features (e.g., colour, light flashes & line orientations) occur at a very
high temporal resolution. The temporal resolution slows down when perception necessitates the binding of multiple features (e.g., spatially overlapped luminance and orientation) and is even slower when the features are spatially disparate (e.g., faces and words). Importantly, the slow threshold for spatially disparate feature exists even with highly practiced tasks such as face perception and word recognition.

Thus, the evidence suggests a temporal hierarchy for visual processing wherein basic features are processed very rapidly and more complex stimuli integrate over a longer time period. The goal of the current experiments was to assess the temporal hierarchy of one specific visual process, visual word recognition. Several word recognition models suggest a processing hierarchy in visual word recognition wherein individual letters are processed first, while the integrated whole-word is processed later downstream. Indeed, several studies have demonstrated distinct temporal thresholds for letter perception (Petit & Grainger, 2002; Ziegler et al., 2000) and whole-word integration (Holcome & Judson, 2007; Forget, Buiatti & Dehaene, 2009; Mewhort & Beal, 1977).

However, these thresholds have only been reported across different studies, usually using different paradigms. Specifically, the thresholds for letter perception and whole-word integration have typically been assessed with the masked priming (Petit & Grainger, 2002; Ziegler et al., 2000) and partial-presentation paradigms (Holcome & Judson, 2007; Forget, Buiatti & Dehaene, 2009; Mewhort & Beal, 1977), respectively. Given the extreme efficiency of visual word recognition, no single paradigm has successfully captured the thresholds of both of these processes. The results from the current experiments have successfully demonstrated distinct thresholds for letter perception and whole-word integration using a single novel paradigm. Specifically, the data demonstrated that the integration of individual features for letter perception
occurs over approximately 60ms while the integration of individual letters for whole-word perception happens over approximately 100ms.

In short, I used a modified version of the partial target presentation paradigm employed by Mewhort and Beal (1977) wherein a target word was broken down to parts and presented successively, along with distractor (noise) characters. Critically, I manipulated the presentation frequency of the presentation stream across three SOA conditions. The results demonstrated that word identification was maximally successful at the shortest SOA. Furthermore, participants were significantly more susceptible to interference from noise at the longer SOAs. Critically, this significant decrease in accuracy and increase in interference from noise may be a consequence of the longer SOAs exceeding the temporal thresholds for letter perception and word integration.

In the proceeding section, I will first outline three experiments to better demonstrate the parameters that influence target identification in this paradigm. This will be followed by a fourth experiment demonstrating distinct temporal thresholds for letter perception and word integration. Importantly, the reported temporal thresholds correspond to those found in the previous literature (Petit & Grainger, 2002; Ziegler et al., 2000; Holcome & Judson, 2007; Forget, Buiatti & Dehaene, 2009; Mewhort & Beal, 1977).

1.3.1 Experiment 1a

All of the following experiments used a modified version of the partial target presentation paradigm employed by Mewhort and Beal (1977). Specifically, target words were broken down into its constituent letters and presented as a stream of strings composed of target letters and distractor (noise) (e.g., C##, #A#, ##T would be a typical presentation stream for the target word ‘CAT’). Participants were then tasked with identifying the target word. Critically, successful
target identification necessitated integration of the target letters across successive string presentations.

The goal of Experiment 1a was to assess the ability for different types of distractor (noise) characters to interfere with target identification. Specifically, the current experiment used three types of noise characters: letters, letter-like symbols (e.g., @ and $) and non-letter-like symbols (e.g., % and &). Furthermore, blanks spaces were used as ‘noise’ in a fourth control condition.

Methods

Participants. Thirteen students from the University of Toronto participated in Experiment 1a for course credit. All subjects had normal or corrected to normal vision with proficient English.

Design. Experiment 1a constituted a 4 (noise type, blank spaces vs. non-letter-like symbol vs. letter-like symbol vs. random letter) x 2 (word length, three vs. four letters) x 5 (presentation count, 12 vs. 24 vs. 36 vs. 48 vs. 60) factorial design, with all factors manipulated within subjects. The dependent measure was identification accuracy.

Materials. The experiment stimuli consisted of 231 three letter and 751 four letter words selected from the MCWord Orthographic Wordform Database (Medler and Binder, 2005). All words had a minimum frequency of 10 occurrences per million. The frequency of the three letter words ranged from 10 to 61,445 occurrences per million, with a mean of 920.9, while the range for four letter words was 10 to 11,437 occurrences per million, with a mean of 214.7. The number of orthographic neighbours (N, as defined by Coltheart, Davelaar, Jonasson, & Besner, 1977) ranged from 0 to 25 with a mean of 13.9 for three letter words and 0 to 24 with a mean of
9.9 for four letter words. Two hundred three and four letter words were randomly selected from the stimulus pool to use as targets for each participant. Noise characters were selected from the available E-Prime character pool; a total of 14 non-letter-like symbols and 19 letter-like symbols were selected.

**String Generation.** All strings were generated at the beginning of each trial. The presentation count condition determined the number of strings generated and presented for each trial. All strings maintained the original target word length and contained one target letter. Each string was generated by first generating the target letters. Target letter positions were selected sequentially and maintained their original target word positions (e.g., position one (C_ _) for the first presentation, position two (_ A_) for the second presentation, position three (_ _ T) for the third presentation, then position one again (C_ _) for the fourth presentation etc.). This cycle continued until the required number of strings, as denoted by the presentation count condition, was generated.

Once the target letters were generated, a single randomly selected character, consistent with the noise type condition, was used to fill in all remaining letter positions. For example, if the selected noise character was a ‘#’, then the following series of stimuli would be generated: C##, #A#, ##T (Figure 1). In the single random letter condition, a restriction was made to ensure that the noise letter was not a constituent letter of the target word. Thus, the letters C, A and T were restricted from being used as noise for the word CAT.
**Apparatus.** Participants were tested individually, seated approximately 40cm from a 16in Dell CRT monitor. Each participant completed 400 trials. Stimuli presentation and data collection were achieved through E-Prime software. The experiments were conducted on a Dell Inspiron 1545 computer that ran on an Intel Core 2 Duo processor. Stimuli were synchronized to a screen refresh rate of 85Hz (11.7ms).

Participants were instructed at the onset of the experiment that they will be tasked with identifying words hidden in a flash of seemingly random letter strings. Participants were instructed to type the hidden word at the end of each trial.

**Procedure.** Each trial began with the presentation of a forward mask of five hashmarks (#####) for 2000ms. This mask was immediately followed by the presentation stream. Each string was presented for 11.7ms (one screen refresh at 85Hz) and immediately replaced by the subsequent string. The strings were presented in the order they were generated, thus ensuring that the target letters were presented in sequential order (see above). A backwards mask, identical to the forward mask, immediately followed presentation of the final string. Following the backwards mask, participants were prompted to type the word that was ‘hidden’ in each trial.

*Figure 1*- A typical presentation sequence of strings for the target word ‘CAT’. Notice that only one target letter is presented per presentation; the #’s denote the noise character.
Participants were encouraged to input real words, but nonword inputs were also accepted. Answers were correct only if all letters matched with the target word (case insensitive). All strings were presented in black Courier 18-point font against a white background.

Results and Discussion

The mean accuracies of word identification across the conditions of Experiment 1a are depicted in Figure 2. The mean accuracies and standard errors for all conditions in this and all subsequent experiments are presented in Table 1. The accuracy data demonstrated that all three factors of word length, presentation count and noise type significantly influenced target identification. When blanks were presented with the target letters, accuracy increased quickly with repetition and asymptote at a high level of accuracy, replicating the results from Mewhort & Beal (1977). However, when noise characters were used instead, overall accuracy and asymptotic performance was much lower. Moreover, the reduction in accuracy was most dramatic when letters were used, a little less dramatic for letter-like characters, and least dramatic for non-letter-like characters.

![Figure 2](image.png)

**Figure 2**- Accuracy data for all noise type conditions, across all presentation count conditions. The data is collapsed across both three and four letter words. Error bars represent the standard errors for each noise type condition.
Data Analysis. The data were analysed using a three-way analysis of variance (ANOVA) with noise type (blank spaces vs. non-letter-like symbol vs. level letter-like symbol vs. single random letter) word length (three vs. four) and presentation count (12 vs. 24 vs. 36 vs. 48 vs. 60) as within subject factors. Effects were considered significant at the p=.05 level.

There were significant main effects of word length, $F(1, 12)=12.637, p<0.01$, noise type, $F(3,36)=975.549, p<0.01$ and presentation count, $F(4,48)=36.859, p<0.01$. Three letter words ($M=0.369, SEM=0.017$) were identified with greater accuracy than four letter words ($M=0.325, SEM=0.013$). Furthermore, there was a general trend of increasing accuracy associated with higher presentation count. Pairwise comparisons between the noise type conditions revealed that the blank-space condition ($M=0.938, SEM=0.014$) was associated with the highest identification accuracy, while the single random letter condition was associated with the lowest accuracy ($M=0.068, SEM=0.014$). Furthermore, accuracy for the letter-like symbol condition ($M=0.112, SEM=0.012$) was significantly lower than that for the non-letter-like symbol condition ($M=0.271, SEM=0.027$). All pairwise comparisons were significant.

Significant interactions between noise type and presentation count $F(12,144)=4.021, p<0.01$, and noise type and word length $F(3, 36)=10.856, P<0.01$, were present. The interactions demonstrate that the effect of noise type was stronger with higher presentation counts and that the word length difference disappeared with the blank noise type condition, due to the performance ceiling associated with the blank noise type condition. No other significant interactions were observed.

Discussion. Similar to Mewhort and Bael (1977), Experiment 1a demonstrated a performance ceiling when the target letters were presented without interference. More importantly, the results also demonstrated that word identification was significantly impaired by
the inclusion of noise characters. Furthermore, the level of impairment was dependent on the type of noise characters used; target identification was most severely impaired by the inclusion of letters. Together, these results suggest that target identification is especially vulnerable to interference from letters and characters resembling letters.

The extremely low accuracy (M= 0.068, SEM= 0.014) for the single random letter condition prompted me to explore the subject responses for an explanation. I hypothesized that the persistent presentation of a single noise letter was too pervasive for the subjects to ignore. Indeed, an informal assessment of the response data revealed that participants frequently responded with words in which the noise character was a constituent letter. To test this hypothesis, I decided to run a simple modification of Experiment 1a, wherein the noise positions were filled in from a pool of multiple noise characters sampled randomly, without replacement.

1.3.2 Experiment 1b

**Method**

Experiment 1b was the same as Experiment 1a with the exception that the noise positions were filled with different noise characters, which were sampled randomly, without replacement.

**Participants.** Thirteen students from the University of Toronto participated in Experiment 2 for course credit.

**Results and Discussion**

The mean accuracies of word identification across the conditions of Experiment 1b are depicted in Figure 3. The current experiment demonstrated similar results to Experiment 1a; different classes of noise characters interrupted integration to varying degrees. As with
Experiment 1a, the results from the current experiment revealed that integration was most impaired by letters and least impaired by non-letter-like symbols. A comparison between Figure 2 and Figure 3 revealed that sampling the noise characters randomly, without replacement facilitated target identification; the mean accuracies were essentially shifted upwards. As expected, there was no difference in mean accuracy of the blank conditions between the two experiments.

Data Analysis. The data were analysed via a three-way analysis of variance (ANOVA) with noise type (blank spaces vs. non-letter-like symbol vs. level letter-like symbol vs. single random letter), word length (three vs. four) and presentation count (12 vs. 24 vs. 36 vs. 48 vs. 60) as within subject factors. Effects were considered significant at the $p= 0.05$ level.

![Figure 3](image-url)  
**Figure 3** - Accuracy data for all noise type conditions (sampled randomly, without replacement), across all presentation count conditions. The data is collapsed across three and four letter words. Error bars represent the standard errors for each noise type condition.
There were significant main effects of word length, $F(1,12)= 211.371$, $p<0.01$, noise type, $F(3,36)= 348.677$, $p<0.001$ and presentation count, $F(4,48)= 85.162$, $P<0.01$. Three letter words ($M=0.590$, SEM=0.036) were identified with higher accuracy than four letter words ($M=0.336$, SEM= 0.024). There was a general trend of higher accuracy associated with increasing presentation counts.

Pairwise comparisons of the noise type conditions demonstrated that the blank space condition ($M=0.921$, SEM=0.025) was associated with the highest accuracy, while the letter condition ($M=0.198$, SEM= 0.029) was associated with the lowest accuracy. Accuracy for the letter-like symbol condition ($M=0.332$, SEM= 0.037) was lower than that for the non-letter-like symbol condition ($M=0.401$, SEM= 0.039). All pairwise comparisons were statistically significant.

Significant two-way interactions were observed between word length and presentation count, $F(4,48)=6.536$, $p<0.01$ and word length and noise type, $F(3, 36)= 40.199$, $p< 0.01$, demonstrating that the word length difference increased with presentation count, and disappeared for the blank noise type condition. Furthermore, there was also a significant two-way interaction between noise type and presentation count $F(12, 144)= 3.429$, $p<0.01$; the effect of noise type was more prominent at higher presentation counts. A significant three-way interaction was observed between word length, noise type and presentation count, $F(12,144)=2.060$, $p<0.05$, indicating that the noise type x presentation count interaction was present only with three letter words, due to the performance floor associated with four letter words.
Discussion. The results from Experiment 1b demonstrated that sampling noise characters without replacement improved target identification. Indeed, comparisons between the two experiments demonstrate that the means for Experiment 1b were higher than Experiment 1a. Furthermore, the main effect of noise type found in Experiment 1a persisted in Experiment 1b, validating the hypothesis that sampling noise letters randomly without replacement would cause an upward shift in the accuracy data from Experiment 1a. Given that the only difference between experiments was the random sampling of noise characters in Experiment 1b, the data suggest that the number of noise characters available for sampling was a critical factor influencing target identification.

1.3.3 Experiment 1c

Experiment 1c was designed to directly examine the role of the signal-to-noise ratio in target identification. More specifically, the number of noise letters available for sampling was manipulated. Only letters were used as noise because there were more letters available than symbols, and using letters as noise provided a better representation of what the word recognition system typically encounters in non-experimental situations.

Methods

Participants. Thirteen students from the University of Toronto participated in Experiment 1c for course credit.

Design. Experiment 1c consisted of a 2(word length, 3 vs. 4) x 4(noise sample size, 1 vs. 5 vs. 15 vs. 26) x 5(presentation count, 12 vs. 24 vs. 36 vs. 48 vs. 60) factorial design, with all factors manipulated within subjects. The dependent measure was identification accuracy.
Stimuli. All stimuli were the same as Experiment 1b with the following exceptions: only letters were used as noise characters and the number of noise letters available for sampling was dependent on the noise sampling size condition. Prior to each trial, a set number of noise letters were randomly selected from a pool of 26 available letters, with the exception the target word’s constituent letters could not be selected as noise characters. This set of noise letters was then randomly sampled, without replacement, to fill in the noise positions in the same manner as the previous experiments. Thus, in contrast to the previous experiments the noise sampling size in the current experiment varied across trials, some trials will had one noise character available for sampling, others with five, some with fifteen etc.

Apparatus & Procedure. The apparatus and procedure were the same as in Experiment 2

Results. The mean accuracies for word detection across the conditions are depicted in Figure 4. The results demonstrated that the factors of word length, noise sampling size and presentation count significantly influenced word identification. As with the other experiments, the mean accuracy increased with increasing presentation counts; the effect was more prominent with larger noise sampling sizes. A first glance of the data suggests a general increase of detection accuracy associated with larger sampling sizes, which peaks at a sampling size of 15 characters. Interestingly, the mean accuracy for the 15-character sampling size condition was marginally greater than that for the 26-character sampling size condition at most presentation
count conditions.

Figure 4- Accuracy data associated with all noise character sampling sizes, across all presentation count conditions. The data is collapsed across three and four letter words. Error bars represent the standard errors for each noise sampling size condition.

Data Analysis. The data were analysed via a three-way analysis of variance (ANOVA) with noise sample size (1 vs. 5 vs. 15 vs. 26) word length (three vs. four) and presentation count (12 vs. 24 vs. 36 vs. 48 vs. 60) as within subject factors. Effects were considered significant at the p= .05 level.

There were significant main effects of word length, F(1,12)= 1.777, p<0.01, noise sampling size, F(3,36)= 18.852, p<0.01, and presentation count F(4,48)= 15.124, p<0.01. Three letter words (M= 0.147, SEM= 0.019) were identified with greater accuracy than four letter words (M= 0.030, SEM= 0.004). There was also a general trend of increasing accuracy associated with higher presentation count. Similarly, there was also a general trend of increasing accuracy associated with increasing the noise size, reaching asymptotic performance at a noise sampling size of 15 noise characters.
Significant two-way interactions were found between word length and noise sampling size, F(3, 26)= 18.601, p< 0.01, and word length and presentation count F(4,48)= 8.393, p<0.01, demonstrating that the word length difference increased with larger noise sampling size and larger presentation count. Both interactions were likely due to the performance floor associated with four letter words.

Discussion. The results from Experiment 1c confirmed the hypothesis that increasing the noise sampling size facilitated target identification. Interestingly, the facilitation of target identification reached a ceiling after 15 noise characters; accuracy actually dropped slightly with 26 noise characters available for sampling (though the difference was not statistically significant). These results support the hypothesis that the signal-to-noise ratio plays a significant role in target identification.

The findings from the current experiments demonstrate that the inclusion of noise characters significantly impaired target integration. Furthermore, the level of impairment was dependent on the nature and consistency of the presented noise characters. Experiment 1a demonstrated that using letters as noise most severely impaired target identification. Experiment 1b demonstrated that sampling the noise characters randomly without replacement improved target identification relative to Experiment 1a. Finally, Experiment 1c corroborated findings from Experiment 1b, demonstrating that increasing the sampling size of the noise characters facilitated target identification.

1.3.4 Experiment 2

Experiments 1a, 1b and 1c demonstrated a multitude of factors that can influence target identification. The goal of the following experiment was to test the temporal threshold for word integration using a set of presentation parameters to maximize noise interference, while avoiding
performance floors. The primary manipulation in Experiment 2 was the presentation duration of each string in the presentation stream (SOA). Specifically, each string was presented for 23ms, 46ms or 94ms.

However, as evidenced by the general high performance associated with the Mewhort and Bael (1977) experiments, participants can easily identify targets even at very long SOAs, when the targets were presented without noise. Presumably, the participants are able to hold the target letters in working memory. To preclude such a strategy, I maximized interference from the noise characters; each presentation consisted of a 1:1 signal-to-noise ratio. Four letter target words were used to maintain this ratio (i.e., two noise letters and two target letters were presented at each presentation). Furthermore, the noise sampling size of four letters was selected to further increase the interference from the noise characters. However, as demonstrated in Experiment 1c, a small noise sampling size and a low signal-to-noise ratio leads to extremely low accuracy. Thus, to improve performance, each trial consisted of 60 string presentations. This set of parameters ensured maximal noise interference, while mitigating the possibility for performance floors.

Given the 1:1 signal-to-noise ratio, I propose that targets must be integrated across at least two presentations (e.g., HOXY, WVME for HOME). Thus, if verbal stimuli do indeed integrate across 100ms, then the 23ms and 46ms SOA conditions should fall within this threshold, while the 94ms SOA condition will be too long for verbal integration. This should be reflected in significantly lower performance associated with the 94ms condition.

Methods

Participants. Seventeen students from the University of Toronto participated in Experiment 1 for course credit.
**Design.** Experiment 3 consisted of a 2(Frequency, High vs. Low) x 3(SOA, 23ms vs. 46ms vs. 94ms) factorial design with all factors manipulated within subjects. The dependent measure was identification accuracy and % incorrect trials with neighborhood responses.

**Stimuli.** A new set of stimuli were selected to control for neighborhood size and bigram frequency. Only four letter words were used to ensure a 1:1 signal-to-noise ratio. The total stimuli set consisted of 120 high and 120 low frequency words CELEX word data base with mean frequencies of 4.34 and 805.391 occurrences per million, respectively. For low frequency words, the mean neighborhood size and bigram frequency were 3.29 and 13.17, respectively. For high frequency words, the mean neighborhood size and bigram frequency were 4.85 and 16.163, respectively.

Prior to each trial, four noise letters were randomly selected from a pool of 26 available letters, with the exception the target word’s constituent letters could not be selected as noise characters. Furthermore, all constituent letters of the target’s orthographic neighbors were also excluded (e.g., D, P and U were excluded as noise for the target word HOME because DOME, HOPE and HUME are neighbors). This set of four noise letters was then randomly sampled, without replacement, to fill in the noise positions in the same manner as the previous experiments.

**Apparatus & Procedure.** The apparatus and procedure were the same as in Experiment 1 with the exception that the participants were given explicit instructions to respond only with real English words.

**Results**
Three dependent measures were assessed in Experiment 2. I reasoned that there would be three general types of responses in the current experiment: correct responses, neighborhood responses and noise responses. Neighborhood responses consisted of responses with the target’s orthographic neighbors while noise responses consisted of a least one noise letter in the response. Since, neighborhood characters were excluded from being used as noise, the noise response, correct response and neighborhood responses are three mutually exclusive responses. Together, these three responses consisted of 87, 85 and 90 percent of the trials for the 23ms, 46ms and 94ms SOA conditions, respectively.

Three analyses were conducted, first with accuracy as the dependent measure. This was followed by a subsequent analysis of the percentage of trials with neighborhood responses (responses with a target’s orthographic neighbor) as the dependent measure. A final analysis was conducted to assess the percentage of trials with at least one noise character in the response.

Accuracy

The analysis of accuracy revealed significant main effects of Frequency and SOA. As depicted in Figure 5, accuracy decreased as SOA increased. Furthermore, high frequency words were easier to identify than low frequency words. The Frequency x SOA interaction was not significant.
Data Analysis. The data were analysed via a two-way analysis of variance (ANOVA) with word frequency (high vs. low) and SOA (23ms vs. 46ms vs. 94ms) as within subject factors. Effects were considered significant at the p= .05 level.

There were significant main effects of word frequency, F(1,16)= 54.658, p<0.01 and SOA, F(2,32)= 30.889, p<0.01. High frequency words (M= 0.406, SEM= 0.029) were significantly easier to identify than low frequency words (M=0.243, 0.027). Furthermore Bonferonni corrected, pairwise comparisons revealed that the 23ms SOA condition (M= 0.445, SEM= 0.034) was associated with significantly higher accuracy than both the 46ms (M=0.286, SEM= 0.024) and 94ms (M= 0.243, SEM= 0.32) conditions. The difference between the 46ms and 94ms conditions was not significant.

The two-way interaction between word frequency and SOA duration was not statistically significant F(2,32)= 1.908, p > 0.05.

**Figure 5:** Accuracy for high and low frequency words across all three SOA conditions. Error bars reflect the standard error of the mean.
Neighborhood Responses

The analysis of neighborhood responses revealed a significant main effect of SOA as well as a significant Frequency x SOA interaction; the main effect of frequency was not significant. As depicted in Figure 6, the probability for a neighborhood responses decreased as SOA increased. However, the effect of SOA was only significant for low frequency words.

![Figure 6](image.png)

**Figure 6**: Percentage of trials with neighborhood responses for high and low frequency words across all three SOA conditions. Error bars reflect the standard error of the mean.

**Data Analysis.** There was a significant main effect of SOA, F(2,32)= 12.316, p<0.01. Bonferonni correct pairwise comparisons revealed that the 94ms ISI condition (M= 0.018, SEM= 0.004) was associated with a significantly lower probability for neighborhood responses than both the 23ms (M= 0.051, SEM= 0.008) and 46ms (M= 0.041, SEM = 0.003) conditions. The difference between the 23ms and 46ms condition was not significant. The main effect of word frequency was not significant, F(1,16)=0.763, p>0.05.

The two-way interaction between word frequency and SOA duration was also statistically significant F(2,32)= 5.229, p< 0.05. As such, separate repeated measures analyses were
conducted for high and low frequency conditions. The analyses revealed a significant main effect of SOA for low frequency words $F(2,32)=21.363$, $p<0.01$ but not for high frequency words $F(2,32)=1.974$, $p>0.05$. Subsequent Bonferroni-corrected pairwise comparisons of the low frequency condition revealed that the 94ms SOA condition ($M=0.011$, $SEM=0.004$) was associated with a significantly lower probability for neighborhood responses than both the 23ms ($M=0.058$, $SEM=0.009$) and 46ms ($M=0.049$, $SEM=0.005$) conditions. The difference between the 23ms and 46ms condition was not significant.

Responses with Noise Characters

The analysis revealed significant main effects of SOA and Frequency. As depicted in Figure 7, there was a higher probability for noise responses at longer SOA s and with low frequency targets. The SOA x Frequency interaction was not significant.

![Figure 7: Percentage of trials with noise responses for high and low frequency words across all three SOA conditions. Error bars reflect the standard error of the mean.](image-url)
**Data Analysis** There was a significant main effect of SOA, $F(2.32)= 70.618$, $p<0.01$. Bonferonni-corrected pairwise comparisons revealed that the 94ms SOA condition ($M= 0.627$, $SEM= 0.032$) was associated with the highest probability for responses with noise characters than both the 46ms ($M= 0.519$, $SEM= 0.027$) and 23ms ($M= 0.344$, $SEM = 0.030$) conditions; the 23ms ISI condition had the lowest probability for noise character response. All pairwise comparisons were statistically significant. The main effect of word frequency was also significant, $F(1,16)=28.283$, $p<0.01$. Low frequency words were associated with a higher probability for noise character response than high frequency words. The SOA x Frequency interaction was not significant.
Figure 8: Relative frequencies (in percentages) of trials with noise responses, correct responses and neighborhood responses across the three SOA conditions. Note that the ratio of correct responses to noise responses decrease as SOA increases. Data are collapsed across high and low frequency words.
Discussion

The results from Experiment 2 have clearly demonstrated that while participants successfully identified targets across all SOAs, target identification was significantly easier at very rapid presentation speeds (Table 2). Furthermore, the participants were more susceptible to interference from noise characters as SOA increased, as evidenced by the significant increase in percentage of responses with noise characters and the significant decrease in accuracy at the longer SOAs (Figure 8). The data suggests that participants were using two distinct identification strategies at short and long SOAs. Specifically, target identification at the shortest SOA (23ms) was accomplished with the ‘normal’ ballistic word recognition process. Contrarily, at the longest SOAs (94ms), the participants were using more conscious (not automatic) effort to ‘build the string in their head’.

As support for this hypothesis, the 94ms condition is completely incompatible with the previously reported threshold of approximately 10Hz (100ms) for verbal stimuli (Mewhort & Bael, 1972; Holcombe & Judson, 2007; Forget, Buiatti & Dehaene, 2009). Specifically, given the current experimental parameters, complete integration of all four target letters requires at least two successive presentations of the stimuli stream (e.g., HOXY, WVME for HOME). Two presentations at 23ms and 46ms fall within the 100ms threshold while the two presentations at 94ms exceed the threshold.

It is also interesting to note that during the incorrect trials, participants were significantly more likely to respond with the target’s orthographic neighbor at the shortest SOA. This further supports the notion that the shortest SOA elicited the ballistic word recognition system. That is, this seemingly paradoxical finding of higher probabilities for both target identification and neighborhood response suggests that the shortest SOA condition ‘activated’ the lexical system.
On the contrary, at the longest SOA, the lexical system is no longer automatically ‘activated’, as evidenced by the lower percentage for target identification and neighborhood response. Together, the evidence suggests that the shortest SOA condition corresponded with ‘normal’ lexical processing.

There were also some peculiar findings regarding the 43ms SOA condition. Specifically, this condition was associated with an accuracy that was comparable to the 94ms SOA condition, both of which were significantly lower than the 23ms SOA condition. The significantly lower accuracy associated with the 43ms condition is especially puzzling since two presentations at 43ms should still fall within the 100ms integration threshold.

The explanation for these peculiar findings may come from the masked priming literature. Specifically, several masked priming studies have demonstrated threshold for conscious awareness of masked primes is approximately 60ms (Forster & Davis, 1984; Forster, Davis, Schoknect & Carter, 1987; Petit & Grainger, 2002; Forster, Hector & Mohan, 2003). At this prime duration, participants are aware that a verbal stimulus has been presented but cannot identify the prime. Thus, it is plausible that while the target letters at the 46ms are still integrating within the 100ms threshold, some of the individual letters in each string are also reaching conscious awareness. Critically, this means that some noise characters may also reach conscious awareness, creating a response competition. At 23ms however, the participants are no longer aware of the individual letters in the string, and only the integrated whole reaches conscious awareness. This is evidenced by the fact that participants are significantly more likely to respond with noise characters at the 43ms SOA when compared to the 23ms SOA condition.

Evidence for integration at the 43ms SOA condition comes from the finding that probability for neighborhood response was comparable between the 23ms and 46ms conditions,
both of which were significantly higher than the longest SOA condition. Critically, the neighborhood responses, as an outcome measure, signal partial integration of the target letters. Recall that the target word’s neighborhood letters were never included in the presentation stream. Importantly, a neighborhood response means that the participant responded with a letter that was never even presented. As such, successfully integration of most of the target letters is necessary for the participants to even consider an orthographic neighbor as a response. Thus, the higher probability for neighborhood responses suggests that both the 23ms and 43ms SOA were short enough to allow for at least partial integration of target letters across the presentation stream. The 94ms SOA however is too long and as such, the target letters did not integrate across the presentation stream.

In summary, the pattern of results for Experiment 2 (e.g., lower accuracy and higher noise interference at longer SOAs) appears to be a consequence of the longer SOAs exceeding the temporal thresholds for letter perception and word integration. Firstly, the 43ms and 94ms SOA conditions either approach or exceed the 60ms threshold for letter perception. This is evidenced by the significantly lower accuracy and higher probability for noise response associated with these two conditions. Secondly, only the 94ms SOA condition exceeds the 100ms threshold for word integration, as evidenced by the significantly lower chance for neighborhood response. Critically, these thresholds are in line with those reported in the literature for letter perception (Forster & Davis, 1984; Forster, Davis, Schoknect & Carter, 1987; Petit & Grainger, 2002; Forster, Hector & Mohan, 2003) and word integration (Mewhort & Bael, 1972; Holcombe & Judson, 2007; Forget, Buiatti & Dehaene, 2009). Together, the results provide evidence for a temporal processing hierarchy wherein individual letters are processed within approximately 60ms while integration of the entire word occurs at approximately 100ms.
1.4 General Discussion

The preceding experiments demonstrated a novel paradigm to assess temporal integration of English words. Specifically, by presenting a target’s constituent letters along with noise characters, target identification was systematically influenced by several experimental factors. Specifically, Experiments 1a, 1b and 1c demonstrated that the factors of noise type, signal-to-noise ratio and presentation count all influenced the probability for successful target identification. These factors were then controlled in Experiment 2 to assess the role of stimulus onset asynchrony on target identification. The results revealed that while target identification was successful at all SOAs, accuracy was significantly higher at the shortest SOA. Furthermore, interference from noise characters increased at longer SOAs, as evidenced by the significantly larger proportion of trials with noise characters in the response.

Together, these results suggest that participants adopted qualitatively different approaches to target identification at short and long SOAs. Specifically, the evidence suggests that at the shortest SOA (23ms), participants adopted a ‘normal’ ballistic word recognition approach, whereas at the longest SOA (94ms), participants were identifying individual target letters and consciously binding these letters, possible in working memory. The following section will further discuss the possible mechanisms for target identification at the short and long SOAs.

Target Identification at Short SOAs- Ballistic Word Recognition

Behavioral Evidence

There are several lines of evidence suggesting that normal word recognition is indeed a ballistic process. The following section will first outline research that supports the ballistic nature
of word recognition. This will be followed by an argument that target identification at the shortest SOA condition in Experiment 2 involved the same ballistic process.

One of the longest standing pieces of evidence for the ballistic nature of word recognition is the Stroop effect. In a typical Stroop experiment, participants are presented with words (usually names of colours) in various coloured fonts. In some trials the word and the colour of the font are congruent (e.g., RED), while in others they are incongruent (e.g., BLUE, GREEN). The participants are then tasked to respond with the colour of the font, while ignoring the actual word. The consistently replicated finding is that participants are much slower to respond in incongruent trials (see MacCleod, 1991 for extensive review). Presumably, the incongruency between the font colour and word creates a response competition. Critically, this incongruency effect is mitigated when participants are tasked to read the name of the colour while ignoring the colour of the font. As such, this suggests that participants are incapable of inhibiting the process of reading; word recognition is ballistic.

Another line of support for the ballistic nature of word recognition comes from the masked priming literature. In a typical critical trial of a masked priming experiment, a related letter string (e.g., related in orthography) is briefly presented and immediately masked, prior to target word presentation. Performance on such trials is then compared to trials wherein the target words were preceded by an orthographically unrelated prime, thus giving a measure of the priming effect. One striking finding is that the priming effects disappear when participants are aware of the primes (Forster, 1998; Forster, Mohan & Hector, 2003). Thus, it has been argued that priming effects in masked priming paradigms reflect unconscious processing of the prime (Forster, 1998; Forster, Mohan & Hector, 2003); if priming effects are only present when participants are unaware of the primes, then it is unlikely that such effects are within conscious
control. Furthermore, this also suggests that the effects of masked primes on lexical processing are transient, resetting after a certain amount of lexical activity (e.g., if either the prime or target reaches conscious awareness).

The initial use of the masked priming technique was for the study of subliminal processing (Marcel, 1983). However, in a seminal study, Forster and Davis (1984) combined the masking priming methodology with the lexical decision task and revealed a facilitative effect of a masked identity prime (a prime that is identical to the target), which was larger for low frequency compared to high frequency words (the frequency attenuation effect). Since then, several studies have used the masked priming technique to further assess the role of various orthographic manipulations on the priming effects. Most relevant to the current discussions are the studies employing the substitution priming technique.

Substitution primes are typically formed by substituting one letter of the target word with another letter, essentially the Coltheart et al. (1977) definition of an orthographic neighbour. Furthermore, substitution primes can be either words or nonwords, though a majority of the earlier research focused on nonword primes (Hinton, Liversedge, Underwood, 1997; Forster, Davis, Schoknecht, & Carter, 1987; Perea & Rosa, 2000). The critical factor is that the direction of the priming effect changes depending on the lexicality of the primes. More specifically, nonword primes tend to facilitate (Ferrand & Grainger, 1994, 1992; Forster, 1987; Perfetti & Bell, 1991), while word primes can facilitate or inhibit target response, depending on the relative frequency of the prime; higher frequency primes inhibit while lower frequency prime facilitate (Segui & Grainger, 1990; Bijeljac-Babic, Biardeau & Grainger, 1997; DeMoor & Brysbaert, 2000). These studies suggest that priming effects of word primes reflect an influence on lexical, as opposed to prelexical perceptual, processes.
Together, the evidence demonstrates that masked substitution primes can directly affect orthographic processing and that such effects lie outside of conscious awareness. Relevant to the current experiments, a typical trial in the current experiments can be conceptualized as successive masked presentations of the target’s substitution primes. The following section will discuss three distinct models proposed to explain the masked priming effects and how each model can account for the current findings.

Models of Masked Priming

There are currently three major models proposed to explain the masked priming effects: the activation-based account (Forster, Mohan & Hector, 2003; Davis, 2003), the Bayesian Reader account (Norris & Kinoshita, 2008) and the entry-open account (Forster & Davis 1984). Proponents of activation-based models (e.g. Coltheart et al., 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981) have suggested that masked priming effects reflect persistent activation of the lexical system (Forster, Mohan and Hector, 2003; Davis, 2003); the unconscious primes provide sub threshold activation to the relevant lexical nodes, facilitating the processing of the subsequent target. However, at longer prime durations, activation reaches above threshold levels causing a reset of the lexical system, thus abolishing the persistent activation.

Forster, Mohan and Hector (2003) conceptualized masked priming effects in terms of ‘savings’ effects. They argued that, if primes were considered as persistent activation of the relevant nodes in the lexical system, then subsequent target processing will be facilitated based on the amount of persistent activation induced by the prime. Importantly, the magnitude of the priming effect will be directly related to the duration of the prime. Indeed, the authors found that the magnitude of priming on lexical decision latencies was very closely tied the duration of the prime-target SOA. In the context of persistent activation, the presentation stream at the shortest
SOA condition of Experiment 2 may be conceptualized as a series of masked substitution primes, with each prime progressively activating the relevant target nodes.

Norris and Kinoshita (2008) provided a Bayesian Reader account for masked priming wherein the prime and target were essentially perceived as a single perceptual unit. Specifically, they suggest that the masked primes reflect partial processing of perceptual evidence for the prime, which in turn, influences the prior probabilities for the subsequent target. Critically, this account for masked priming suggests that successive stimuli presented within a short timeframe are processed together. Thus, in this context, the successive presentation stream at the shortest SOAs of Experiment 2 can be conceptualized as being fused into a unitary percept. At longer SOAs however, the successive strings are no longer processes as a single perceptual unit, but instead as individual units. Once again the current findings of greater accuracy for the shortest SOA (individual strings are fused into a unitary percept of the target) and greater chance for noise interference at the longer SOAs (strings are processed as individual units) are compatible with model proposed by Norris and Kinoshita (2008).

Forster and Davis (1984) proposed and entry-opening account based on a serial-scan model for lexical access (Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982). In the serial-scan models the lexical system conducts a serial verification process to determine the match between the activated candidates and the verbal stimulus. In such models, word frequency denotes the serial search priority in the orthographic lexicon, thus higher frequency words are searched first. Furthermore, the serial-scan model proposes a series of two nested scans. Upon target presentation the lexical system first conducts a rapid ‘sloppy’ scan wherein several closely matching lexical entries (e.g., orthographic neighbors) are tagged for lexical access. This rapid
scan is followed by a slower more methodical scan which looks for a specific match between the target and the opened lexical entries.

According the entry-opening account, masked priming effects are conceptualized as “post-access” effects. Specifically, the model suggests that a serial scan upon prime presentation simply opens access to the target. It is then assumed that this opening persists until target presentation, thus facilitating target processing since the target’s lexical entry has already been opened. In this account, the prime does not directly influence processing of the subsequent target. Instead, it simply opens the door for target processing.

However, the entry-open model cannot account for the findings of the current experiments. Specifically, in all reported experiments, the targets were never presented in their entirety. Instead, targets were presented as a stream of substitution primes. Critically, the participants successfully identified the targets in many of the trials, with accuracy depending on the presentation parameters. According to the entry-open account however, primes can only open access to the target, actual target processing however, necessitates an actual presentation of the target. Thus, according to this account, the participants should never be able to identify the target. Furthermore, for this same reason, the fact that target identification was successful in any of these experiments is troublesome for serial-scan models in general, which also necessitates target presentation for successful lexical access.

The preceding section has provided evidence that target identification at the shortest (23ms) SOA works on mechanisms similar to that for masked priming. Critically, only the 23ms SOA condition falls confidently within the 60ms threshold for letter perception. Furthermore, the 23ms SOA condition also falls comfortably within the previously reported integration threshold of 100ms for verbal stimuli (Mewhort & Bael, 1972; Holcombe & Judson, 2007; Forget, Buiatti
Together, the evidence suggests that participants were using the normal ballistic word recognition system for target identification at the 23ms SOAs. However, there still remains the question of how do participants identify targets at the longer SOAs? The following section will describe a proposed mechanism for target identification at above-threshold SOAs.

**Target Identification at Long SOAs- Letter by Letter Reading**

I propose that in the longest SOA condition (94ms) of the Experiment 2, participants may be binding their responses in working memory. There are two main reasons for this hypothesis. Firstly, this SOA precludes the integration of all four target letters within the previously reported 100ms threshold (Mewhort & Beal, 1977; Forget, Buiatti & Dehaene, 2009; Holcombe & Judson, 2007). Secondly, the 94ms SOA condition is long enough to allow conscious awareness of the individual letters of the string, as evidenced by the higher probability for participants to respond with noise characters. For these reasons, I propose that the participants were individually picking out potential ‘target’ letters and integrating them in working memory.

Though normal ‘ballistic’ word recognition likely does not include working memory, there is a specific case of dyslexia that has been proposed to necessitate a working memory component. Letter-by-letter (LBL) reading is a severe case of dyslexia wherein patients read words in a serial, letter-by-letter manner, taking up to 1000ms to process each letter. Researchers have proposed that LBL readers suffer from a perceptual deficit for individual letters (Berhmann et al., 1998). As such, they must focus an obscene amount to time to identify each individual letters and adopt a serial letter-by-letter reading method to accommodate for the perceptual deficit. Binder and Mohr (1992) proposed a mechanism for LBL reading where words are read serially in the right hemisphere, combined by verbal working memory processes in the left hemisphere and then processed in the tradition lexical processing system.
During the long SOA trials, participants may be identifying target words in a manner similar to LBL readers. That is, they were building the string in working memory before being processed in the lexical system. However, given this proposed mechanism, it is curious to find word frequency effects across all SOAs; several researchers and word recognition models suggest that the word frequency effects are a consequence of automatic lexical access (Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982; McClelland and Rumelhart 1982; Whitney, 2001; Coltheart et al., 2001; Seidenberg & McClelland, 1989; Sears, Hino & Lupker, 1999).

However, Berhmann et al., (1998) also reported robust word frequency effects in LBL. The authors suggested that the frequency effect in LBL might be due to excitatory feedback from orthographic representations supporting the letter identification process. That is, letters from higher frequency words will receive the greatest processing benefit because higher frequency words provide more excitatory feedback to the letter level. Thus, it is still plausible that the participants were more inclined to select target letters from high frequency words given the higher processing benefit for letters from high frequency words. On the contrary, when the targets were low frequency words, then the participants were just as likely to select noise letters when building the string. Indeed, low frequency words were associated with a greater percentage of responses with noise characters.

Of course, such claims are still highly speculative. Thus, I propose a future study to better assess the role of working memory in the current paradigm. More specifically, I propose to replicate the current experiments mentioned above, along with a concurrent working memory task (e.g., digit span, letter span or reading span). If working memory does contribute to target identification at the long SOA conditions then I expect performance to drop significantly in the
presence of a concurrent working memory task. Furthermore, the WM task should have no influence on performance at the short SOAs, which should only employ the ballistic word recognition system and have no WM component.

1.5 Conclusion

This manuscript outlined a series of experiments that demonstrated a novel paradigm to assess the temporal characteristics of word recognition. Specifically, target words were presented in parts along with noise characters at varying presentation frequencies. The results suggest that while target identification was successful at all presentation frequencies, only the fastest presentation frequency was associated with the normal ballistic word recognition processes. At slower frequencies, participants adopted a strategy that presumably required working memory.

I propose that the qualitatively different approaches to target identification at long and short SOAs are a consequence of the hierarchical structure of the visual word recognition system. Specifically, the longer SOA conditions necessitated a conscious strategy for target identification because the longer SOAs surpassed the thresholds for letter perception and whole-word integration. The shortest SOA however, was well within the thresholds for both these processes and as such, target identification was possible via normal ballistic word recognition.

Together, the results from the current experiment provide evidence for a temporal processing hierarchy for visual word recognition. Specifically, individual features integrate to support letter perception within 60ms and the individual letters are then integrated into a unitary percept over 100ms to support whole-word recognition. This temporal hierarchy supports the structural hierarchy proposed by Dehaene, Cohen, Sigman and Vinckier (2005). Critically, this hierarchy reflects a general processing hierarchy for all visual stimuli wherein progressively complex visual stimuli are processed further downstream and over longer periods of time.
Interestingly, the data have also demonstrated that target identification was successful even at above threshold presentations durations. I propose that such target identification necessitates the role of working memory though further research necessary to identify the role of working memory in binding processes that are not automatic and require conscious effort.
References


