Skilled Readers Process Words Letter by Letter in a Nearly Optimal Sequence

Caroline Blais,

Université de Montréal

Daniel Fiset,

University of Victoria

Martin Arguin, Pierre Jolicoeur, and Frédéric Gosselin

Université de Montréal

Abstract

Skilled adult readers, in contrast to beginners, show no increase in reading latencies as a function of the number of letters in words up to seven letters. This absence of a word-length effect is typically interpreted as evidence for parallel letter processing in visual word recognition (Coltheart, Curtis, Atkins & Haller, 1993; Fiset, Arguin, Bub, Humphreys & Riddoch, 2005; Grainger & Jacobs, 1996; Weekes, 1997). However, alternative views have been proposed (e.g., left-to-right serial strategy; Kwantes & Mewhort, 1999; Whitney, 2001) and the strategy of letter extraction is still under debate. The present study directly examined the space-time use of letter information while reading using the *Bubbles* technique (Gosselin & Schyns, 2001a; Vinette, Gosselin & Schyns, 2004). Ten participants each read 5,000 five-letter words sampled in space-time within a 200 ms window. We found that, on average, the third and fourth letters were used effectively from 42 to 133 ms after word onset, and that the first letter was used between 42 and 75 ms and again, in conjunction with the second letter, between 142 and 175 ms. To benchmark human performance, we introduced a family of ideal observer models that could optimally use all the available information in order to read words. Of all the models that were examined, the ideal reader that processed one letter at a time was the one that offered the best fit to human performance. Indeed, this serial reader is compatible with the absence of a word-length effect for words from four- to seven letters in length.

Skilled Readers Process Words Letter by Letter in a Nearly Optimal Sequence

Gutenberg's invention has democratized the written word to such an extent that it is estimated that nowadays the average English reader has been exposed to more than 100 million printed words before he reaches the age of 25 (Geisler & Murray, 2003). As a result of this intense exposure skilled adult readers, in contrast to beginners, show no increase in reading latencies as a function of the number of letters in a word (i.e., word length effect), at least for relatively short words (Cohen et al., 2003; Fiset, Arguin & McCabe, 2006; Weekes, 1997).

This finding is interpreted by the vast majority of researchers as evidence that letters are processed in a parallel manner when they form words no longer than six or seven letters (Coltheart, Curtis, Atkins & Haller, 1993; Fiset, Arguin, Bub, Humphreys & Riddoch, 2005; Rayner & Johnson, 2005; McCandliss, Cohen & Dehaene, 2003; Weekes, 1997). However, alternative views have been proposed (e.g., left-to-right serial strategy; Kwantes & Mewhort, 1999; Whitney, 2001) and the parallel letter extraction strategy remains under debate. Indeed, proponents of this strategy can only indirectly infer parallel processing since studies addressing this question have relied on methods that did not allow direct tracking of the extraction of letter information through time.

In the present investigation, we directly track the spatio-temporal progression of attentional resource allocation in skilled readers while they are performing a word identification task. Our approach directly assesses how proficient readers use visual information (the various letters) across time in order to achieve word recognition and unambiguously reveals the processing strategy used by these proficient readers to recognize letter strings.

Attentional deployment during word reading

There is uncertainty concerning the information extraction strategy used by proficient readers in order to recognize written words or non-words. This debate largely focuses on two competing hypotheses: parallel processing versus sequential processing.

Parallel processing implies that several letters of a word are processed simultaneously. In its most extreme form, parallel processing supposes that all the letters of a word are processed simultaneously. In the present paper, this form of parallel processing will be referred to as *fully parallel processing*. Fully parallel processing may be either exhaustive, in which case the beginning and the end of processing are synchronized across all letters, or self-terminating, in which case the end of the processing may occur at different moments for each letter (Lamberts, 2005; Pashler, 1998). An alternative version of parallel processing assumes that more than one letter of a word, but not necessarily all of them, are processed at the same time. In the present paper, this form of parallel processing will be referred to as *partially parallel processing*. This hypothesis is conditional on the simultaneous processing of two or more letters at least once (regardless of duration) during the recognition of a word. It also implies that the beginning and/or the end of the processing of each letter may occur at different times. For example, “ends-in” processing may be classified as partially parallel if the processing of the outer (or inner) letters overlap in time.

Most theories of word recognition postulate that letters are processed in a parallel manner (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Coltheart & Rastle, 1994; Harm & Seidenberg, 1999; McClelland & Rumelhart, 1981). The hypothesis of parallel processing is based, at least in part, on the absence of a length effect with this stimulus type (Cohen et al., 2003; Fiset, Arguin & McCabe, 2006, Weekes, 1997). A length effect is observed with pseudo-words, however, and several researchers propose that it reflects a letter extraction strategy that is sequential (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Coltheart & Rastle, 1994; Weekes, 1997).

Theoretically, however, the absence of a length effect with words could be obtained even with sequential processing, and a few researchers postulate that words are indeed processed sequentially (Kwantes & Mehwort, 1999; Whitney, 2001). Sequential processing (under a strict definition) means that the letters are processed one at a time. This hypothesis is conditional on the absence of temporal overlap in the processing of individual letters in a word and it implies that the beginning and the end of processing occur at different moments for each letter. Most theories that involve sequential processing with words and/or pseudo-words postulate that it is performed from left-to-right (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Coltheart & Rastle, 1994; Kwantes & Mewhort, 1999, Whitney, 2001). However, this is not the only possibility. Indeed, in both partially parallel and sequential processing the various letters could theoretically be examined in many different orders (Lamberts, 2005), including left-to-right, outside-in or random information extraction.

Methodological options for tracking attention during word reading

Eye tracking (Rayner, 1978, 1998) has been used successfully in numerous studies investigating the extraction of visual and lexical information in reading. This method can be advantageous as it leaves the stimulus unaltered and does not interfere with the normal reading process. However, its temporal resolution is insufficient to reveal the extraction of visual information in the recognition of individual words. In fact, the average duration of ocular fixation when reading English text (200-250 ms; Rayner, 1998; Sereno & Rayner, 2000; Sereno, Rayner & Posner, 1998) is about the same as the time needed to recognize an isolated word (less than 250 ms; McCandliss et al., 2003). In addition, using eye tracking for our present purpose would require assuming that eye fixation at any given time unequivocally reveals what is being processed at that time. Such an assumption is hardly tenable since the locus of attention can be dissociated from gaze (Jonides, 1981; Posner, 1980; see however Rayner, 1998; Deubel & Schneider, 2003; Godijn & Theeuwes, 2003). Finally, eye-tracking studies of reading have shown that short words are usually apprehended in a single eye fixation (i.e., mean saccade size is 7-9 letters; mean fixation time while reading is about 200-250 ms; Rayner, 1998). Attentional saccades, however, can occur several times within a single eye fixation (i.e., estimates of the time needed to plan and execute an attentional saccade typically range between 50 ms and 85 ms; e.g., Wolfe, 1998; Wolfe, Alvarez & Horowitz, 2000).

A radical methodological shift is therefore required in order to precisely study the deployment of attention during reading. We have opted for a classification image technique. *Classification image techniques* have already been used to uncover the features involved in letter discrimination regardless of time (Watson & Rosenholtz, 1997; Watson, 1998; Gosselin & Schyns, 2003; Fiset et al., submitted). They have also been used successfully in the study of features involved at different moments in basic image discrimination (Neri & Heeger, 2002), face identification (Vinette et al., 2004), and illusory shape discrimination (Gold & Shubel, 2006). They have however never been used to examine the reading of letter strings, nor the identification of single letters through time. Classification image methods are ideal for our purposes because they can assess the use of information directly and because their spatio-temporal resolution is theoretically unlimited—the only limits stemming from the properties of the visual system and those of the equipment used to display the stimuli. Among the different classification image approaches available, we have chosen *Bubbles* because only *Bubbles* directly identifies the specific visual information that allows an observer to perform a particular task effectively (Gosselin & Schyns, 2002, 2004a, 2004b; Murray & Gold, 2004).

The following analogy illustrates how we applied the *Bubbles* technique to word reading. In the late stages of Emmenthal cheese production, a bacteria releases carbon dioxide gas and this process generates bubbles that become the famous holes. Imagine a word revealed by an animated sequence of masks very similar to a succession of thin, opaque slices of cheese cut from a brick of Emmenthal. This is in essence what we did in the present experiment: On each trial, a target word was randomly sampled across space and time by a collection of tridimensional (i.e., height, width, and time) Gaussian windows (or bubbles; see Figure 1).

Insert Figure 1 about here

Thus, on any given trial, several masks (or slices of cheese, to pursue with our analogy) were successively placed over a target letter string so as to modulate the availability of visual information (partial and/or complete letters) across time. One fundamental feature of *Bubbles* is that the sampling of the stimulus on any given trial is entirely random. This involves a significant cost in terms of the number of trials required to uncover information use (e.g., we ran a total of 50,000 trials in the present experiment). This cost however, is offset by a major advantage: the minimization of experimental bias. In particular, it is impossible for the reader to adjust on-line to the position of the bubbles because they appear briefly and at random locations. The bubbles might interact with the information typically used by the observer but this possible interaction is unlikely to be problematic.

The random location of the bubbles over time and the large number of trials allows for the identification, in space-time, of the information that is significantly correlated with reading accuracy. In other words, this method allows to directly assess which letters are used at what time when reading a word.

Bubbles experiment

Method

Participants. Ten students from the Université de Montréal took part in the experiment. All had normal or corrected-to-normal visual acuity.

Materials and stimuli. Stimuli were displayed on a high-resolution Sony monitor with a refresh rate of 120 Hz. The experiment ran on a Macintosh G4 computer. The experimental program was written in Matlab, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The viewing distance was maintained constant at 91 cm by using a chinrest. Stimuli were lowercase words printed in Courier 40 points. They appeared in dark grey over a light grey background (fixed luminance of 64.8 cd/m^2; minimum and maximum luminances were 0.48 and 130 cd/m^2, respectively). The luminance contrast of the stimulus was adjusted according to the criteria and procedure described below.

Stimuli were constructed from a list of 1,000 five-letter French words. We chose five-letter words because they are relatively taxing for the visual system while they clearly remain within the alleged parallel-processing boundary (i.e., a small but systematic word-length effect is observed for words made of seven letters or more). These five-letter words subtended a vertical x horizontal spatial extent of 9.8 x 46.1 mm (0.6 x 2.9 deg of visual angle or 25 x 118 pixels). The list of words was constructed using BRULEX, a lexical database for French (Content, Mousty & Radeau, 1990), and it excluded all words with diacritic marks (e.g., é, ê, à, and so on). We also discarded extremely low frequency words with which the participants may have been unfamiliar.

The stimuli consisted of movies made from a sequence of 24 frames shown for 8.33 ms each (total duration of 200 ms; this is less than the time required to plan and execute an eye saccade). On any given trial, a randomly selected word was sampled using Gaussian apertures (bubbles) randomly positioned in space-time. Each bubble had a standard deviation of 0.195 degree of visual angle (or 8 pixels) in the spatial domain and a full-width half maximum of 48.06 ms (5.77 frames), in the temporal domain (see Figure 1). This is less than the time required to plan and execute an attentional saccade. The number of bubbles was adjusted twice during the experiment, as will be explained shortly.

Procedure. Each participant completed a total of 5,000 trials divided in 20 experimental sessions (250 trials each) spread over 10 weeks on average. Reading accuracy was maintained at 51% correct. The usual procedure to maintain accuracy at a fixed level in a *Bubbles* experiment is to adjust the number of bubbles on a trial-by-trial basis so as to make the task more difficult (or easier) when the accuracy of the subject is higher (or lower) than the targeted accuracy (e.g., Gosselin & Schyns, 2001a). A trial-by-trial modification of the number of bubbles requires the online computation of each dynamic stimulus. Since this computation took on average 8 seconds, we opted for a dual strategy in order to maintain performance at a fixed level while avoiding excessively long inter-trial intervals. Specifically, we estimated the number of bubbles necessary to obtain approximately 51% accuracy twice during the experiment—at the beginning and halfway through—and we also adjusted the contrast of the target word revealed by the bubbles on a trial-by-trial basis throughout the experimental sessions.

This procedure allowed us to generate the bubbles' masks for each trial before the experimental sessions, thus reducing the time required to prepare the upcoming trial to approximately 4 s, which was relatively comfortable for participants. We estimated the number of bubbles required by each participant to read accurately on 51% of trials in two 50-trial *setting sessions*, one before the first *experimental session* and the other before the eleventh *experimental session*. In these setting sessions, the luminance contrast of the stimuli was kept constant at 0.1, and the number of bubbles was adjusted on a trial-by-trial basis using a gradient descent procedure (Hertz, Krogh & Palmer, 1991). On each trial a homogenous grey screen was first displayed for 250 ms and a 1300 Hz pure tone accompanied this stimulus for the first 122 ms. The grey screen was immediately followed by a dynamically bubbled word displayed for 200 ms at the center of the screen. This was in turn immediately followed by a homogenous grey screen that remained visible until the participant responded. The participant's task was to read the target word aloud as quickly and as accurately as possible. This triggered, via a response key, a dialog box into which the subject typed his/her response using the appropriate computer keyboard keys followed by the «ENTER» key. No feedback was provided to participants during the setting sessions.

The sequence of events for each trial in the *experimental sessions* was identical to that of the *setting sessions*. The number of bubbles per trial for the first ten experimental sessions was determined by the first setting session and remained constant. It was then readjusted after the second setting session and remained constant for the remaining ten experimental sessions. Performance was further fine-tuned at 51% correct by adjusting the luminance contrast of the stimulus on a trial-by-trial basis using QUEST, with the initial contrast being set at 0.1 for all participants (Watson & Pelli, 1983). No feedback was provided to participants during the experimental sessions.

Results and Discussion

In the first half of the experiment, a mean of 332.4 bubbles (between 222 and 530 across participants) with an average Weber contrast of 0.094 (the average contrast was of 0.1 and 0.088 at the beginning and the end, respectively) were necessary to maintain performance at 51% correct. In the second half of the experiment, the corresponding numbers were 252.3 bubbles (between 172 and 455 across participants) with an average Weber contrast of 0.099 (the average contrast was of 0.1 and 0.098 at the beginning and the end, respectively). The word "bulle" displayed in Figure 1 is revealed by 332 apertures, which approximately corresponds to the number of bubbles participants required, on average, in the first half of the experiment.

The efficient use of the spatio-temporal information in the stimulus was determined using an analysis procedure that amounts to a multiple linear regression on the bubbles masks (explanatory variables) and the participant's response accuracy (predictor variable). First, we constructed one regression coefficient volume per session and per subject by subtracting the sum of the bubbles masks that led to an incorrect answer from the sum of the bubbles masks that led to a correct answer. These volumes of regression coefficients will be referred to as *classification movies*, which is a natural extension of *classification image*, a term widely used to refer to planes of regression coefficients (e.g., Eckstein & Ahumada, 2002). The elements of these movies will be referred to as *voxels*. Second, we constructed a group classification movie that combined the classification movies of all subjects, weighted by the number of bubbles used. If all parts of the stimulus (the various letters) were of equal importance for success in the task (word identification), the voxels would be uniform. Any local divergence from uniformity indicates that this particular part of the stimulus (in space-time) was particularly important for the task at hand. The statistical analysis was restricted to the central horizontal strip one third the height of the group classification movie (43 x 128 pixels), which approximately represents the area occupied by the word. The remainder of the group classification movie was used to estimate the mean and the standard deviation of the null hypothesis, and to Z-score the group classification movie. Finally, we conducted a one-tailed Pixel test (Chauvin, Worsley, Schyns, Arguin & Gosselin, 2005) on the group classification movie (S_r = 132,096 voxels; full-width half maximum = 12.696; Z_{crit} = 4.168; $p < .025$).

Figure 2 shows the thresholded classification movie in a rich tridimensional graphic. The space-time voxels reaching statistical significance are depicted in bright red in the center of the figure and are overlaid on the word “javel”. Four relatively small blobs (< 66 voxels) and, more importantly, three larger blobs can clearly be seen. The largest blob (1890 voxels) is shaped like a croissant, the second largest (797 voxels) resembles a bottle, and the third largest is almost round, looking more or less like a “dragée” (234 voxels). Note that the raw significant voxels were convolved with a small Gaussian kernel (spatial std = 1.5 pixels; temporal std = 0.141 frames or 1.18 ms) to reduce aliasing.

Insert Figure 2 about here

This tridimensional representation ultimately comes short of fully conveying the exact shape and location of the blobs in space-time. To remedy this problem, the significant voxels were projected onto three bidimensional planes: the back wall—to isolate the spatial left-right and up-down dimensions; the floor—to isolate the time and the left-right dimensions; and the right wall—to isolate the time and the up-down dimensions. The number of significant voxels projected onto a single pixel on the planes is represented by red saturation (see legend). To further facilitate space-time localization, we have added dim grey lines delimiting the 24 frames on the time dimension and the three different regions of the letters on the space dimensions (i.e., body, ascenders, descenders).

A glance at the right wall reveals two moments especially correlated with accurate reading: one between frames 5 and 10 and the other between frames 19 and 21. It also shows that most of the voxels correlated with reading accuracy are located in the top half of the body region and, to a lesser extent, in the ascender region. Interestingly, there are no significant voxels in the descender regions. Looking at the back wall, it can be seen that the letter positions most correlated with reading accuracy are 4 (letter positions are numbered from 1—leftmost—to 5—rightmost), followed by 3, then 1, then 2, and, finally, 5. The floor helps to visualize the interactions between the importance of letter positions and time in reading accuracy.

To summarize, the space-time use of letter positions 3 and 4 forms the large croissant-shaped blob. It begins around frame 4 (50 - 58.33 ms) or 5 (41.67 - 50 ms) and ends around frame 14 (116.67 - 125 ms) or 15 (125-133.33 ms). The early space-time use of letter position 1 forms the dragée-shaped blob. It approximately extends from frame 5 (41.67 - 50 ms) to frame 8 (66.67 - 75 ms). Together, the croissant- and the dragée-shaped blobs are responsible for the first burst of activity that can be distinctively seen on the right wall. The second burst of activity on the right wall is caused by the bottle-shaped blob. It corresponds to the effective use of letter positions 1 and 2 between frames 17 (141.67 - 150 ms) and 20 (166.67 - 175 ms).

Comparison to electroencephalographic reading data.

Our results might seem at odd with the electroencephalographic (EEG) literature on reading. Indeed, distinct electrical responses to different stimulus categories occur considerably later than the beginning of the activity correlated with correct word identification as observed herein. In fact, the earliest EEG response that differentiates between orthographic and non-orthographic stimuli has been shown to occur approximately 200 ms after stimulus onset (Bentin, Mouchetant-Rostaing, Giard, Echallier & Pernier, 1999; Nobre, Allison & McCarthy, 1994; Rossion, Joyce, Cottrell & Tarr, 2003). Specifically, this response consists of a negative component showing a larger amplitude, in the left hemisphere only, for words as compared to other complex objects. The amplitude of this component was found to be identical for words and for non-words, and it has been suggested that it reflects visual mechanisms tuned to orthographic processing, a stage of processing that occurs earlier than the lexical, phonological and semantic processing stages (Nobre, Allison & McCarthy, 1994). The present results are in fact not inconsistent with the time course of this component. Indeed, they show that the correct identification of words *depends* on the presentation of letters as early as 50 ms after stimulus onset; what the EEG results imply is that the visual information presented as early as 50 ms after stimulus onset is processed for approximately 150 ms before it is reflected in the orthographic specific electrical component of the brain.

Qualitative evaluation of the reading models proposed in the literature.

As mentioned previously, the hypotheses concerned with possible strategies of information extraction in reading that have been formulated so far usually focus on the parallel vs. sequential nature of the processing, and the majority of reading models postulate parallel letter processing when the stimulus is a word.

The fully parallel hypothesis implies that all the letters that form a word are processed simultaneously, and that the end of the processing of each letter occurs either at the same moment (exhaustive processing) or at different moments (self-terminating processing). In exhaustive processing, no specific time slot is thought to be more important, or special, throughout the processing. This is true also, at least at the beginning of processing, for self-terminating processing. A fully parallel model would therefore be associated with a classification volume in which the relative importance of the five letter positions is equal on each frame (McCabe, Blais, & Gosselin, 2005).

To illustrate, consider the following toy problem: a fully parallel reader is exposed for the duration of two frames to a pseudo-word composed of two randomly chosen letters. Each bubble mask could be represented as a 2 x 2 matrix: $\begin{pmatrix} l_1t_2 & l_2t_2 \\ l_1t_1 & l_2t_1 \end{pmatrix}$, where l_xt_y is equal to 1 when there is a bubble on the xth letter position of the word for the yth frame and is equal to 0 otherwise. For this reader, all the bubble masks revealing *at least* letter position one and two during either of the two frames would lead to a correct word identification. Nine bubble masks (out of fifteen) respond to those criteria: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, and $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$; and the six that remain do not: $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, and $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ (we are purposefully neglecting correct guessing here since it has no impact on the main argument we are making). The classification volume computations described in the results section—in the ongoing example, it is a classification plane—would consist in summing up all bubble masks weighted by +1/*number of bubbles*, if the mask led to a correct word identification, and by -1/*number of bubbles*, if it did not: $1/2\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} +$ $1/2\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} + 1/2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + 1/2\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + 1/3\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + 1/3\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + 1/3\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} + 1/3\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

$$+ 1/4\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} - \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} - \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} - \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} - 1/2\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} - 1/2\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix} =$$

$$\begin{pmatrix}0.75 & 0.75\\ 0.75 & 0.75\end{pmatrix}.$$

Our results do not support such a fully parallel model. Indeed, there are no time intervals during which all the letters of the stimulus are processed simultaneously and there is a clear modulation of the relative importance of the five letter positions across time. Both of these facts are incompatible with a fully parallel model.

Partially parallel models propose that only subsets of the constituent letters of a word are being processed simultaneously. In that case, the order in which these subsets of letters are extracted may be systematic or not. If it is systematic, each time slot becomes special and the moment at which the letters are revealed in a bubbles task matters: if a letter is not revealed at that very moment when the reading system would usually process it, the system will not have all the information it needs to correctly recognize the target and performance will suffer. Therefore, a partially parallel model in which letter subsets are extracted in a systematic order predicts a modulation of the attentional resources allocated to each letter position across time. In contrast, any model—partially parallel or serial—in which the order of extraction is random would make predictions equivalent to those of the fully parallel hypothesis. Indeed, readers that deploy their attention randomly, regardless of the number of letters they can process simultaneously, would also generate a homogenous classification volume. As we saw, those predictions are infirmed by the present results. The data presented herein could therefore be explained by a partially parallel model in which the order of extraction is systematic since we observe a clear modulation of the importance, or diagnosticity, of each letter position over time.

One hypothesis that has been proposed about the order of letter extraction in a partially parallel model is that the outside letters are processed before the inside letters (Jordan, Patching & Thomas, 2003; Jordan, Patching & Milner, 2000). The present data appear incongruent with this hypothesis, however, as the inside letters 3 and 4 become useful before outside letter 1.

The last type of reading model encountered in the reading literature is sequential processing, which proposes that the various letters that form a word are being processed one at a time. The idea of the "special" quality of each time slot that was discussed in the context of the partially parallel hypothesis is also relevant to the sequential processing hypothesis: if the order of extraction is systematic, each time slot becomes special; if it is completely random across trials, the results should not show any modulations of the importance, or diagnosticity, of each letter position over time. Most of the authors who propose that a sequential strategy is being used, at least with a certain type of letter strings, suggest that the extraction of the letters occurs in a systematic left-to-right order (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Coltheart & Rastle, 1994; Grainger & Jacobs, 1995; Kwantes & Mewhort, 1999; Whitney, 2001; but see Lamberts, 2005). The results reported in the present study are clearly incongruent with a simple left-to-right strategy of information extraction. Indeed, the relevant processing focuses mainly on the middle and middle-right portions of the word and attention does not appear to smoothly move across adjacent letter positions. Indeed, relevant processing skips intermediate positions and even at time proceeds from right-to-left (e.g., from letter 4 to 2 or 3 to 1). Nevertheless, our results could still be accounted for by a kind of sequential processing in which the order of letter extraction is systematic but not from left-to-right.

In summary, the results of the experiment are consistent with a sequential or partially parallel processing strategy in which the order of extraction is systematic, but not from left-to-right nor from outside-in. The analyses reported *so far* do not permit to discriminate between these two general hypotheses. In fact, even though two or more letters may appear to have been processed simultaneously in the thresholded classification movie (Figure 2), it should be kept in mind that the group classification movie was elaborated from the weighted sum of all bubble masks. This implies that two letters that are revealed simultaneously in the results may have actually been used independently on different trials. For example, in a hypothetical case where participants would use, during frame 5, letter position 3 on half the trials and letter position 4 on the other half of the trials, the weighted sum of all bubbles mask would reveal both letter 3 and 4 on frame 5. We will come back to this issue in a subsequent section, where human data and reading models are compared quantitatively.

A Family of Ideal Readers

Ideal observers optimally use all the information available to perform the task at hand (Giesler, 1989; Legge, Klitz & Tjan, 1997; Legge, Hooven, Klitz, Mansfield, & Tjan, 2002; Mamassian, Landy & Maloney, 2002; Pelli, Farell & Moore, 2003; Anderson, 1991; Gosselin & Schyns, 2001b). Their purpose is not so much to fit human data but to understand how human data diverges from an ideal implementation. In other words, ideal observers can be seen as benchmarks for human performance. Before turning to a new family of ideal readers, we will present Mr. Chips, an ideal-observer model of eye saccades, so as to provide a firm grasp of what ideal-observer models actually are.

Mr. Chips

Legge, Klitz, and Tjan (1997) have proposed an ideal-observer model for a simple reading situation: Mr. Chips. The purpose of the model is to examine how basic visual, motor, and cognitive constraints influence reading behavior in people with normal and low vision. This model provides novel accounts of characteristic patterns of normal eye movements while reading, including the presence of regressive saccades, word skipping, and a preferred gaze landing position in words. The researchers also describe how the model provides insights into the reading difficulties encountered by people with visual-field loss resulting from eye disease.

Visual data are obtained through a simplified "retina". The retina can have any sequence of high-resolution slots in which letters can be recognized, and low-resolution slots or scotomas where only spacing information is available. The authors refer to the set of high-resolution slots as Mr. Chips's "visual span." Mr. Chips has access to lexical information that consists of a dictionary containing all allowable words and their frequencies. Oculomotor information is available in the form of statistical knowledge of eye-movement accuracy. Mr. Chips' eye movements may be "noisy," that is, of somewhat uncertain length, but when Mr. Chips plans a saccade of length *L*, he knows the probability distribution of the landing sites.

Mr. Chips reads texts made of words drawn at random from the dictionary. The model makes no use of syntax or semantics. Mr. Chips' task is to read through the text in as few saccades as possible and to identify all the words sequentially without error. Mathematically, Mr. Chips uses an entropy minimization principle. At any point, he may have partial information about the current word (some of the letters or word length information). He then makes the saccade that minimizes uncertainty (bits of information) about the current word. He cannot depart from the current word until he has unambiguously identified it (i.e., its uncertainty is zero).

Optimal attentional deployment within reading saccades

In this section, we will describe a family of *ideal readers*: models that read words by using all the available information in an optimal manner. In a sense, the realm of this family of ideal readers begins where Mr. Chips' ends: they do their work in between eye saccades.

We have made several simplifying assumptions as we elaborated these ideal readers. First, we assumed that the letters are the atoms of information for word reading (e.g., Pelli, Farell & Moore, 2003). We believe this to be an oversimplification since letters can be divided into features (e.g., Pelli, Burns, Farrell & Moore-Page, 2006; Fiset et al., submitted). Second, we assumed that time ticks in a discrete manner (e.g., VanRullen & Koch, 2003). Third, we assumed two processing stages: the first corresponding to bottom-up, encapsulated, low-level visual processes (e.g., Pylyshyn, 1999), and the second corresponding to a bidirectional attentional mechanism (i.e., bottom-up and top-down). In other words, the deployment of this second processing stage is informed by lexical knowledge. Fourth, we assumed that the ideal readers have perfect lexical knowledge. Fifth, we assumed that their low-level visual systems are imperfect and that they obey the same letter confusability matrix as the human visual system. Sixth and finally, we assumed that the attentional mechanisms of the ideal readers have limited processing capacity.

A number of studies suggest that human information processing capacity is limited. For example, the classic Miller (1956) study has shown that human short-term memory can only process seven chunks of information, plus or minus two. Studies of subitizing (Kaufman, Lord, Reese, & Volkmann, 1949) have shown that human subjects could rapidly and accurately report the numerosity of sets containing up to 3 or 4 elements (e.g., Dehaene & Cohen, 1994). Moreover, change blindness experiments have shown that subjects that were asked to notice changes in natural scenes do very poorly, even if the change involves very salient features of the scene (Levin & Simons, 1997; Simons & Rensink, 2005).

In the reading literature, the usual interpretation of the length effect observed with words containing more than 6 letters (Cohen et al., 2003; Fiset, Arguin & McCabe, 2006) is simply that long words overtax the limited processing capacities of the system. The results of the present study, however, imply that the maximum number of letters that can be processed simultaneously is inferior to five.

Given the limited capacity afforded to ideal readers, the optimal strategy—or rather the optimal family of strategies—is to scan the maximum number of letters within one clock cycle and to scan different groups of letters across clock cycles. One consequence of this—assuming that all the letters in a word have to be processed for identification—is that all words of less than m letters, m being the maximum number of letters that can be processed in a processing cycle, can be read at a glance. Conversely, all words of n letters, with $n > m$, have to be read in a number of processing cycles that increases linearly as a function of word length with a slope of n / m. Each member of the ideal reader family described in this article will differ from the others in terms of the number of letters it can process simultaneously. For each ideal reader, we will find the optimal attentional strategy, taking into consideration the lexical and visual constraints mentioned above.

The best attentional strategy S_c at clock cycle c is the attentional strategy K that minimizes the uncertainty regarding the target word:

$$S_c = K \Rightarrow \min\left(-\sum_i P_K\left(\text{word}_i \text{ on cycle}_{c+1}\right)\log_2\left(P_K\left(\text{word}_i \text{ on cycle}_{c+1}\right)\right)\right)$$

where P_K(word$_i$ on cycle $_{c+1}$) is the probability that each word in the lexicon is the target given a particular attentional strategy K on cycle c+1. The steps involved in finding the best attentional strategy S_c given a target word are summarized in the following pseudo-code:

```
Initialize the word probability matrix
Begin clock cycle loop
    [Generate a bubbles vector]
    Begin loop through all possible K's
        Temporary word probability matrix = word probability matrix
        Update the temporary word probability matrix according to K
        Compute a word probability vector using the temporary word probability matrix
        Compute uncertainty associated with K
        If uncertainty < than previous ones S_c = K
    End loop through all possible K's
    Update the word probability matrix according to S_c [and the bubbles vector]
If the maximum word probability > criteria, stop the clock cycle loop or else continue
```

Prior to the first processing cycle c, each letter (1/26) of each word has an equal probability of being the target. Every cell of the word probability matrix is therefore filled with 1/26. In other words, the ideal reader has not yet acquired any knowledge about the target word at this point in time. Its selection of the optimal attentional strategy will solely be based on its lexical and letter confusability knowledge.

When simulating a *Bubbles* experiment, a binary vector of length equal to that of the word being identified is randomly generated on each processing cycle. These bubbles cannot affect the choice of the optimal strategy because they do not last long enough (less than 50 ms, the minimum time required to plan and execute an attentional saccade) for the human reader to react to them. However this sampling does make a difference when the time comes to update the word matrix.

What warrants the optimality of the ideal reader is that it exhaustively searches the possible attentional strategies for the processing of cycle c+1. The number of candidate attentional strategies K to consider is the same at all processing cycles and is determined by $C_n^m = \frac{n!}{(n-m)!m!}$, where m is the number of letters that the model reader can process simultaneously and n is the number of letters in the word. For each of these candidate attentional strategies K, we calculate a vector containing the probabilities P_K(word$_i$ on cycle $_{c+1}$) that each word in the lexicon is the target on cycle c+1.

On cycle c+1, the probability of recognizing the letter or the combination of letters towards which attention is directed can be computed by using a letter confusability matrix derived through experimentation, in our laboratories, with human normal readers (see Table 1). A letter confusability matrix expresses the probability of identifying letter x given that letter y is presented, $P(x|y)$, with x and y spanning all letters of the alphabet.

These probabilities of identification/confusions can be obtained by putting high constraints on the visual system. The letter confusability matrix used herein was obtained by very briefly (33.3 ms) presenting letters that were embedded in a high level of noise (adjusted so as to maintain performance at 50% correct).

For example, if the target word was "javel" and if an ideal reader with an attentional capacity limited to one letter was to direct its attention toward the first letter of the target word, the probability of correctly identifying this letter as a 'j' would be 0.31. The probability of identifying it as an 'f' (as in the word "farce", for instance) would be 0.07 (i.e., the probability to confuse a 'j' for an 'f' is 0.07), and so on for all letters from the alphabet.

To pursue with this example, the next step in the search for the best attentional strategy S_c consists in updating a temporary word probability matrix: 0.31, 1/26, 1/26, 1/26, and 1/26 on the row of the target word "javel"; 0.07, 1/26, 1/26, 1/26, 1/26 on the row of the word "farce"; and so on for the rows of all the other words in the lexicon. This temporary word probability matrix is then used to construct a temporary word probability vector by multiplying all the elements in each row and by dividing the result by the sum of the 1,000 products (the total number of words in the model's lexicon).

The elements of this vector correspond to $P_K(\text{word}_i \text{ on cycle }_{c+1})$, or to the probability that each word in the lexicon is the target given that attention was directed to the first letter of the target word on cycle c+1. According to information theory, the uncertainty associated with this candidate strategy K is equal to $-\sum_i P_K(\text{word}_i \text{ on cycle}_{c+1})\log_2(P_K(\text{word}_i \text{ on cycle}_{c+1}))$. The zeros in the word probability vector were excluded from the computation.

This uncertainty equation outputs the number of *bits* contained in a sequence of probabilities associated with events. It reaches a maximum when all events are equiprobable (equal to 9.9658 bits for a word probability vector of 1,000 elements such as the list of five-letter words used in the *Bubbles* experiment), and it approaches zero when one event has a probability approaching 1.

The candidate attentional strategy K associated with the least uncertainty is the best attentional strategy S_c at processing cycle c and it is perforce the attentional strategy chosen by the ideal reader. Note that on the rare occasions that this computation leads to ties, one of the winning strategies is selected at random.

In all the processing cycles that follow the first one, the same computations are performed, and for each cycle all the knowledge acquired on the preceding cycles is taken into account. This is achieved by updating the word probability matrix according to either S_c or, when simulating a *Bubbles* experiment, according to the intersection between S_c and the revealed letter positions.

Going back to the ongoing example, suppose that the best attentional strategy on the first processing cycle was to allocate attention to letter position 1. Remember that the word probability matrix after the first processing cycle was 0.31, 1/26, 1/26, 1/26, and 1/26 on the row of the word "javel"; 0.07, 1/26, 1/26, 1/26, and 1/26 on the row of the word "farce"; and so on for the rows of all the other words in the lexicon. If, for example, the candidate attentional strategy loop on the second cycle processes the first letter position again, the temporary word probability matrix would look like: $.31^2$, 1/26, 1/26, 1/26, and 1/26 for the word "javel"; $.07^2$, 1/26, 1/26, 1/26, and 1/26 for the word "farce"; and so on for all other words in the lexicon.

This is slightly counterintuitive because paying attention twice to the same letter position should increase certainty and yet $.31^2$ is *smaller* than .31. We have to bear in mind that what matters for the computation of uncertainty is the relative difference between the various probabilities in the word probability matrix. In this case, there is a decrease in uncertainty from the first to the second cycle because $.31^2/.07^2 = 19.6$ is greater than $.31/.07 = 4.4$. This last observation also stands in cases where the word probability is smaller after attending to one letter position than before attending to it. In fact, letter positions that were never attended to are somewhat arbitrarily given a weight set to 1/26. However, since all the words in the word probability matrix have the same weight on the letter positions that have not been attended to, this weight can be factored out. In other words, what matters is what varies between words.

When one of the words reaches a probability of being the target that is greater than an arbitrary criterion, the run is terminated. To obtain results comparable to the thresholded classification movie presented in Figure 2, we used a stopping criterion of .51, namely the mean accuracy of human participants. Finally, we averaged the S_c's associated with each word in the five-letter-word lexicon considered as the target ten times (i.e., 1,000 words * 10 repetitions = 10,000 simulations).

Figure 3 shows the space-time deployment of attention for the combination of free parameters that lead to the greatest r^2 between the predictions of the models and human data. The human data focuses on the interaction of time and the left-right plane, which was projected on the floor of Figure 2. A first point to note in the results of the computational analysis is the high diagnosticity of letter positions 1, 3, and 4, from most to least diagnostic, during the initial processing cycles. An examination of the results obtained in the *Bubbles* experiment reveals that these three letters are also the most useful for the tested human readers, albeit not in the same order. Also, the results of all the simulations show two bursts of activity over time, just as those of humans (see pinkish time histograms next to the prediction planes); this is particularly clear for ideal reader number 4.

Goodness of Fit Between Reading Models and Human Data

We will now quantitatively compare the results of the *Bubbles* experiment reported above with all the models concerned with the deployment of attention during reading, including the four ideal readers introduced in the present article. In order to conduct this comparison, we have made a number of assumptions and simplifications that we will specify carefully so as to avoid any misinterpretation. Since none of the models make any commitments about the use of letter ascenders, bodies, or descenders, we collapsed the up-down dimension in the thresholded classification volumes. We also assumed that the whole letter was the spatial atom and that a slice of 8.33 ms was the temporal atom. Therefore, we compared the theoretical proposals with the averaged human data within these 120 space-time atoms (5 letter positions x 24 frames = 120 space-time atoms).

The various reading models were best-fitted to the human data as follows: the fully parallel model had two free parameters (i.e., a beginning and an end); the left-to-right model had six free parameters (i.e., a beginning for the leftmost letter as well as for the four next letter positions and an end for the rightmost letter); the outside-in model had four free parameters (i.e., a beginning for outermost letter positions as well as for the next two groups of letter positions and an end for the innermost letter); and each one of the four ideal readers—i.e., those without any built in constraint as to letter extraction order and with a processing capacity of one, two, three, and four letters in parallel, respectively—had two free parameters (i.e., a beginning and an end).

Figure 3 shows the space-time deployment of attention for the combination of free parameters that led to the greatest r^2 between the predictions of each model and human data. To summarize the predictions along the time and along the left-right dimensions, we flanked the prediction planes with pinkish histograms.

Insert Figure 3 about here

Of all the model readers proposed in the literature, the left-to-right reader is the one that explains the most data variance ($r^2 = .2239$). A closer look at the results of this model, however, reveals that this relatively good fit is essentially due to the use of information at letter position 4 on frames 6 to 11. This model does not capture the global aspect of the behavioural data, namely the use of letters 3 and 4 before letter 1, as well as the very late use of letter 2.

Ideal reader 1, which is only capable of processing a single letter at once, explains more data variance ($r^2 = .2578$) than any other model discussed in this article. The difference between the r^2 of ideal reader 1 and the left-to-right reader might appear small but a likelihood ratio, corrected for the number of free parameters,

$$\frac{\left(1-r^2_{\text{left-to-right}}\right)}{\left(1-r^2_{\text{ideal reader 1}}\right)} e^{\left(k_{\text{left-to-right}}\frac{n}{\left(n-k_{\text{left-to-right}}-1\right)}-k_{\text{ideal reader 1}}\frac{n}{\left(n-k_{\text{ideal reader 1}}-1\right)}\right)},$$

where k_x is the number of free parameters in model x and n is the number of data points (n = 24 frames * 5 letters = 120), indicates that the human data is 1,096.8 times more likely to have occurred if ideal reader 1 were true than if the left-to-right reader were true (Glover & Dixon, 2004).

In sum, the best model considered in this article is ideal reader 1, followed by ideal reader 2 (the data is more than 14.8087 times as likely to have occurred if ideal reader 1 were true than if ideal reader 2 were true). The other models follow thusly: ideal reader 4, ideal reader 3, the left-to-right reader (the data is 9.5932 times as likely to have occurred if ideal reader 3 were true than if the left-to-right model were true), the outside-in reader, and finally the fully parallel reader, which is trailing far behind.

Ideal reader 1 offers a simple and straightforward explanation as to why letters would be extracted in a specific order. Specifically, it suggests that skilled readers take advantage of the statistical properties of words (i.e., some letter positions are more diagnostic than others), and therefore allocate more attention to the letter positions that increase the probability of recognizing the target word. Surprisingly, the completely sequential version of the model explains more data variance than the partially parallel version. This supports a letter-by-letter process rather than a fully or even a partially parallel strategy of visual information extraction.

General Discussion

First, we used The *Bubbles* technique to examine the extraction of visual information over time in a five-letter word identification task. In a nutshell, it was found that the use of letter positions 3 and 4 begins around 42 ms and ends around 133 ms. The early space-time use of letter position 1 starts at approximately 42 ms and finishes around 75 ms. Letter position 1 is used effectively a second time, together with letter position 2, between 142 ms and 175 ms.

Second, we examined the optimal strategy of letter information extraction for systems with limited processing capacity and perfect lexical knowledge. A family of four ideal readers, respectively capable of processing one, two, three or four letters simultaneously, were run on the five-letter word lexicon used in the *Bubbles* experiment. These novel ideal readers provide a much better fit to the *Bubbles* results than the three main models proposed in the literature: the fully parallel, the left-to-right and the outside-in models. Indeed, the worst of the ideal readers (ideal reader 3) still provides a better fit than the best of the models from the literature (the left-to-right reading model). Ideal reader 1, with a processing capacity of one letter, is the one that best fits the human data. These results suggest that human readers use a nearly optimal strategy of letter extraction, and process letters one at a time.

The behavior of ideal reader 1 is nonetheless different from human behavior and it accounts for only about 25% of the variance. In particular, the unfolding over time of the human data is not exactly what is predicted by the ideal reader 1. It goes without saying that, reading performance can be influenced by numerous factors not taken into account by the models examined herein: lexical frequency, imageability, hemispheric asymmetry, visual acuity at different eccentricities, etc. These are also likely to influence how participants extract information through time. For example, it is possible that the left hemisphere bias for verbal stimuli leads to an easier visual processing of the letters that fall within the right visual hemifield or that the letters that are further from the fixation point are slightly harder to recognize because of decreased visual acuity. Of course the order in which letter information is extracted would take these various factors into account. The aim of our computational analysis was not to exactly reproduce human performance, but rather to examine and better understand the nature of the task that had faced our human participants.

It should be added that the models have nothing to say about the observed variations in the use of information along the up-down dimension. We have already pointed out that the upper portion of the letters appears to be much more important than the lower parts (Figure 2). This is consistent with the findings of Huey (1908), which showed that reading is slow and effortful when the top part of words is removed whereas the effect is minor when the bottom part is removed. Future models will have to address this bias. Likewise, the models do not predict the within-letter left-right modulations captured in the classification volume of Figure 2. For example, the left half of letter position 2 and the right half of letter position 3 seem to be especially useful for word recognition.

The word-length effect revisited

The present findings directly contradict the quasi-universal notion that normal word recognition rests upon parallel letter processing, an idea essentially based on the fact that the number of letters in a word hardly has any effect on the time required to read it (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Henderson, 1982; Weekes, 1997). Apart from constituting a null effect, this observation must be considered inconclusive because particular forms of serial letter processing can accommodate the absence of a length effect.

For instance, one—unlikely—possibility is that serial processing would be exhaustive (i.e., the complete set of letters in the word is examined) but that its rate increases proportionately with word length. A more likely possibility is that serial letter processing is self-terminating—i.e., it terminates when sufficient evidence has been gathered to reliably permit word identification. This type of processing would be a viable and economically sensible strategy, especially considering that the total number of words of a particular length actually constitutes a small subset of the possible letter combinations of that length. The serial letter processing strategy demonstrated by the present experiment is entirely compatible with this hypothesis since only subsets of the letters comprised in the words appear to have been processed to a significant degree. Within this context, if it is further supposed that the number of letters that actually need to be processed for word recognition remains constant across word lengths, then the lack of a word length effect can be reconciled with serial letter processing.

To assess this hypothesis in a relatively direct manner, we submitted ideal reader 1 to all the French words without diacritic marks that contain from 4 to 7 letters. This analysis revealed that the average number of processing cycles to reach the .99-accurate stopping criteria remained approximately constant across word lengths (i.e., 4.66, 4.71, 4.95, and 4.91 cycles for words containing 4, 5, 6, and 7 letters, respectively). Most interestingly, the letter positions that are the most informative during the first cycles (i.e., letters 1, 3, and 4—see Table 2) remain the same for all word lengths. This latter observation is congruent with studies addressing the effect of word length on optimal viewing position. Indeed, it has been shown that the optimal viewing position, which is very near the center of the stimulus for five-letter words, shifts towards the left—i.e., closer to where letters 1, 3, and 4 are located—for longer words (Brysbaert, Vitu & Schroyens, 1996). Given the above, we argue that this leftward gaze shift may be explained by the need to keep the most informative letters within the high-acuity portion of the retina.

Incidentally, it should be noted that the hypothesis of serial self-terminating letter processing proposed here might to a large degree explain the occurrence of a substantial length effect in pseudoword reading (Weekes, 1997). Indeed, the number of legal pseudowords of a given length is necessarily greater than the total number of words of that length. Because of this, one should expect the need for the examination of a greater number of letters in pseudowords than in words. Furthermore, it should also be expected that this increase will be magnified with letter strings of increased length.

References

Anderson, J. R. (1991). The adaptative nature of human categorization. *Psychological Review, 98*, 409-429.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier. J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*, 235-260.

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

Brysbaert, M., Vitu, F., & Schroyens, W. (1996). The right visual field advantage and the optimal viewing position effect: On the relation between foveal and parafoveal word recognition. *Neuropsychology, 10*, 385-395.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*, 659-667

Cohen, L., Martinaud, O., Lemer, C., Lehéricy, S., Samson, Y, Obadia, M., et al. (2003). Visual word recognition in the left and right hemispheres: Anatomical and functional correlates of peripheral alexias. *Cerebral Cortex, 13*, 1313-1333.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). A computational dual route model of reading aloud. *Psychological Review, 100*, 589-606.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception & Performance, 20*, 1197-1211.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90*, 551-566.

Dehaene, S. & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 958-975.

Deubel, H. & Schneider, W. X. (2003). Delayed saccades, but not delayed manual aiming movements, require visual attention shifts. *Oculomotor and Vestibular Systems: Their Function and Disorders, 1004*, 289-296.

Eckstein, M. P., & Ahumada, A. J. (Ed.). (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, *2*, i-i.

Fiset, D., Arguin, M., Bub, D., Humphreys, G.W., Riddoch, J. (2005). How to Make the Word-Length Effect Disappear in Letter-by-Letter Dyslexia. Implications for an Account of the Disorder. *Psychological Science, 16*, 535-555.

Fiset, D., Arguin, M., & McCabe, E. (2006). The breakdown of parallel letter processing in letter-by-letter dyslexia. *Cognitive Neuropsychology, 23*, 240-260.

Fiset, D., Blais, C., Ethier-Majcher, Arguin, M., Bub, D., & Gosselin, F. (submitted). Features for lowercase and uppercase letter identification.

Forster, K.I., & Chambers, S.M. (1973). Lexical access and naming time. *J ournal of Verbal Learning and Verbal Behaviour*, *12*, 627-635.

Frederiksen, J.R., & Kroll, J.F. (1976). Spellling and sound: Approaches to the internal lexicon. *J ournal of Experimenta l Psychology: Human Perception and Performance*, *2*, 361-379.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological Review, 96*, 267-314

Geisler, W. & Murray, R. (2003). Practice doesn't make perfect. *Nature, 423*, 696-697.

Glover, S.R., & Dixon, P.(2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11*, 791-806.

Godijn, R. & Theeuwes, J. (2003). Parallel allocation of attention prior the execution of saccades sequences. *Journal of Experimental Psychology: Human Perception & Performance, 29*, 882-896.

Gold, J. M. & Shubel, E. (2006). The spatiotemporal properties of visual completion measured by response classification. *Journal of Vision*, *6*, 356-365.

Gosselin, F., & Schyns, P. G. (2001a). Bubbles: A technique to reveal the use of information in recognition. *Vision Research, 41*, 2261-2271.

Gosselin, F. & Schyns, P. G. (2001b). Why do we SLIP to the basic-level? Computational constraints and their implementation. *Psychological Review, 108*, 735-758.

Gosselin, F., & Schyns, P. G. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Science, 6,* 70-77.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of memory representations. *Psychological Science*, 14, 505-509.

Gosselin, F., & Schyns, P. G. (2004a). No troubles with Bubbles: A reply to Murray and Gold. *Vision Research, 44*, 471-477.

Gosselin, F., & Schyns, P. G. (Ed.). (2004b). A picture is worth thousands of trials: Rendering the use of visual information from spiking neurons to recognition. *Cognitive Science, 28*, 141-146.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518-565.

Harm, M. W. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia : insights from connectionist models. *Psychological Review, 106*, 491-528.

Henderson, L. (1982). *Orthography and word recognition in reading*. London: Academic Press.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison- Wesley.

Huey, E. B. (1908/1968). The psychology and pedagogy of reading. Cambridge, MA: MIT Press.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. L. A. Baddeley (Ed.), *Attention and Performance IX*. Hillsdale, NJ: Erlbaum.

Jordan, T. R., Patching, G. R., & Milner, A. D. (2000). Lateralized word recognition: Assessing the role of hemispheric specialization, modes of lexical access, and perceptual asymmetry. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 1192-1208.

Jordan, T. R., Patching, G. R., & Thomas, S.M. (2003).Assessing the role of hemispheric specialisation, serial-position processing, and retinal eccentricity in lateralised word recognition. *Cognitive Neuropsychology, 20*, 49-71.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.

Kwantes, P. J., & Mewhort, D. J. K. (1999). Evidence for sequential processing in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 25*, 376-381.

Lamberts, K. (2005). Interpretation of orthographic uniqueness point effects in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 14-19.

Legge G.E., Hooven, T.A., Klitz, T.S., Mansfield, J.S., & Tjan, B.S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research, 42*, 2219-2234.

Legge G.E., Klitz, T.S., & Tjan, B.S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review, 104*, 524-553.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*, 501-506.

Mamassian, P., Landy, M. S. and Maloney, L. T. (2002). Bayesian Modelling of Visual Perception. In R. Rao, B. Olshausen and M. Lewicki (Eds.) *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13-36). Cambridge, MA: MIT Press.

McCabe, E., Blais, C. & Gosselin, F. (2005). Effective categorization of objects, scenes, and faces through time. In C. Lefebvre et H. Cohen (Eds.) *Handbook of Categorization in Cognitive Science* (pp. 767-791). Amsterdam: Elsevier.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*, 293-299.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Miller, G. A. (1956). The magical number seven, plus or minus Two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97

Murray, R. F., & Gold, J. M. (2004) Troubles with bubbles. *Vision Research, 44,* 461-470.

Neri, P., & Heeger, D. (2002) Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience, 5,* 812-816.

Nobre, A. C., Allison, T. & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*, 260-263.

Pashler, H. (1998). *The psychology of attention.* Cambridge: MIT press.

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

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

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423*, 752–756.

Posner, M. I. (1980). Orienting of attention. *Quaterly Journal of Experimental Psychology, 32*, 3-25.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618-660.

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

Rayner, K., & Johnson, R.L. (2005). Letter-by-Letter Acquired Dyslexia Is Due to the Serial Encoding of Letters, *Psychological Science, 16*, 530-534.

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage, 20*, 1609-1624.

Sereno, S. C., & Rayner, K. (2000). The when and where of reading in the brain. *Brain and Cognition, 42*, 78-81.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye-movements and event-related potentials. *NeuroReport, 9*, 2195-2200.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present and future. *Trends in Cognitive Sciences, 9*, 16-20.

VanRullen, R. & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*, 207-213.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes! *Cognitive Science, 28*, 289-301.

Watson, A. B. (1998). Multi-category classification: template models and classification images. *Investigative Opthalmology and Visual Science, 39*, S912.

Watson, A. B. & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113-20.

Watson, A. B., & Rosenholtz, R. (1997). A Rorschach test for visual classification strategies. *Investigative Opthalmology and Visual Science, 38*, S1.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology, 50A*, 439-456.

Whitney, C. (2001) How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8*, 221-243.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*, 33-39.

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature, 406*, 691.

Author Note

Caroline Blais, Martin Arguin, Pierre Jolicoeur, and Frédéric Gosselin are at the Département de psychologie of the Université de Montréal. Daniel Fiset is at the Department of Psychology at the University of Victoria.

This work was supported by a grant from the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT) to Martin Arguin, Frédéric Gosselin, and Pierre Jolicoeur; by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Caroline Blais; and by a post-doctoral scholarship from the Fonds Québécois de Recherche en Nature et Technologies (FQRNT) to Daniel Fiset. We thank Benoit A. Bacon for proofreading a draft of this manuscript.

Correspondance concerning this article should be addressed either to Frédéric Gosselin or to Martin Arguin, Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, Canada, H3C 3J7. Electronic mail may be sent to: frederic.gosselin@umontreal.ca or to martin.arguin@umontreal.ca.

Table 1. Confusability matrix (part 1). Lines represent the stimuli presented and columns represent the probability of response for each letter of the alphabet.

	a	b	c	d	e	f	g	h	I	j	k	l	m
a	0.33	0.02	0.025	0.0275	0.085	0.0025	0.08	0.035	0.015	0.005	0	0	0.0175
b	0.005	0.5475	0.03	0.0225	0.015	0	0.015	0.11	0	0	0.0075	0.0075	0.005
c	0.0075	0.02	0.555	0.0275	0.04	0.01	0.075	0.0075	0.005	0.0025	0	0	0.01
d	0.025	0.01	0.085	0.5725	0.0225	0	0.0425	0.015	0.005	0.015	0.0025	0	0.0075
e	0.05	0.0325	0.1325	0.0125	0.345	0	0.0875	0.01	0.005	0.0075	0	0	0.01
f	0.015	0.0025	0.01	0.015	0.0075	0.31	0.0175	0.0075	0.1075	0.05	0.025	0.0925	0.005
g	0.0075	0.0125	0.025	0.0025	0.01	0.0025	0.6375	0.0075	0.0025	0.0325	0.0025	0.0025	0
h	0.0025	0.0575	0.0125	0.005	0.005	0.01	0.01	0.5875	0.0025	0.005	0.0525	0.0025	0.0125
I	0.01	0.0075	0.0225	0.0075	0.0025	0.08	0.02	0.045	0.1525	0.0825	0.0375	0.1075	0.015
j	0.0125	0.0025	0.01	0.0175	0.005	0.0725	0.05	0.0275	0.11	0.3125	0.0075	0.1225	0.005
k	0.0075	0.0075	0.0075	0.0025	0.005	0.0075	0.01	0.0275	0.0025	0	0.7625	0.015	0
l	0.0175	0.0125	0.01	0.0175	0	0.075	0.02	0.035	0.1275	0.105	0.03	0.15	0.01
m	0.0075	0	0.0075	0	0.0025	0.005	0.0075	0.0125	0.005	0	0.0075	0.005	0.82
n	0.0275	0.0025	0.0175	0.01	0.015	0.005	0.005	0.12	0.005	0.005	0.0075	0	0.0375
o	0.03	0.0375	0.1325	0.0525	0.0325	0	0.075	0.02	0.0025	0.0125	0.005	0	0.0075
p	0.0075	0.0225	0.0125	0.01	0.0225	0.01	0.0325	0.0125	0.005	0	0.005	0	0
q	0.015	0.005	0.06	0.035	0.015	0.0025	0.1175	0.02	0.0025	0.005	0.0025	0.0025	0.0075
r	0.005	0.0025	0.0575	0.0075	0.005	0.1225	0.0175	0.015	0.0575	0.0275	0.0325	0.0375	0.005
s	0.0325	0.02	0.0475	0.01	0.0575	0.0025	0.0825	0.0075	0.0125	0.0125	0.015	0.0025	0.0075
t	0.005	0.0075	0.0275	0.01	0.01	0.145	0.02	0.035	0.12	0.0325	0.0325	0.0975	0.0075
u	0.0225	0.02	0.0175	0.0175	0.01	0.0025	0.045	0.0175	0.0025	0.0125	0.0125	0.0125	0.02
v	0	0.005	0.005	0.01	0	0.0075	0.01	0.0025	0.005	0.005	0.0075	0.0025	0.02
w	0.005	0.0025	0	0	0.0075	0	0.0025	0.005	0	0	0.005	0	0.005
x	0.01	0	0.0125	0.0025	0	0.01	0.005	0.005	0.015	0.01	0.075	0.0125	0.0075
y	0.0025	0	0	0.0025	0.0025	0.015	0.0025	0	0	0.0125	0.005	0.0025	0.0025
z	0.0025	0.0025	0.0025	0	0.0125	0.015	0.035	0.0125	0.005	0.0125	0.0125	0.015	0.005

Table 1. Confusability matrix (part 2).

	n	o	p	q	r	s	t	u	v	w	x	y	z
a	0.1025	0.0625	0.025	0.0125	0.0125	0.055	0.0075	0.03	0.0025	0.0125	0.015	0.0025	0.0175
b	0.0575	0.025	0.065	0.005	0.0025	0.015	0.0075	0.0275	0.015	0.0025	0.005	0.0075	0
c	0.035	0.0775	0.0225	0.0125	0.035	0.0075	0.0075	0.015	0.015	0.005	0.0025	0.0025	0.0025
d	0.03	0.0525	0.0075	0.02	0.005	0.005	0.005	0.0525	0.01	0.005	0	0.0025	0.0025
e	0.0375	0.0875	0.02	0.0175	0.045	0.015	0.0025	0.04	0.0125	0.02	0.0025	0.0075	0
f	0.0125	0.0075	0.02	0	0.14	0.005	0.0575	0.0075	0.0125	0.0025	0.0075	0.055	0.0075
g	0.015	0.0325	0.0375	0.05	0.0125	0.04	0	0.05	0.005	0.005	0	0.005	0.0025
h	0.165	0.0075	0.01	0	0.02	0.0025	0.0075	0.0025	0	0.0075	0.0025	0.01	0
I	0.04	0.01	0.0175	0.0125	0.095	0.0125	0.0475	0.0225	0.0325	0.02	0.0125	0.08	0.0075
j	0.015	0.01	0.025	0.005	0.035	0.01	0.0425	0.0125	0.02	0.005	0.01	0.0525	0.0025
k	0.0075	0.0025	0.0175	0	0.0275	0.005	0.0075	0.0075	0.0125	0.0075	0.04	0.01	0
l	0.0425	0.005	0.0225	0.0075	0.14	0.0025	0.05	0.01	0.04	0.0125	0.0125	0.035	0.01
m	0.065	0.0075	0.005	0	0.025	0	0.0025	0.005	0.005	0.005	0	0	0
n	0.6	0.0325	0.0275	0.005	0.04	0.0025	0.0025	0.0175	0.0025	0.01	0	0	0.0025
o	0.05	0.3225	0.0675	0.055	0.0075	0.0125	0.0075	0.045	0.0125	0.005	0	0.005	0.0025
p	0.0525	0.025	0.7125	0.0075	0.0275	0.0025	0.0125	0.005	0.0075	0.005	0	0.0025	0
q	0.0575	0.03	0.0125	0.525	0.015	0.015	0.0025	0.035	0.0075	0.0025	0.0025	0.0025	0.0025
r	0.04	0.005	0.015	0.0025	0.3925	0.0025	0.045	0.0025	0.0125	0.005	0.01	0.065	0.01
s	0.035	0.0225	0.035	0.0025	0.0175	0.5125	0.0075	0.0275	0.01	0.005	0.0075	0.005	0.0025
t	0.0175	0.005	0.0175	0.005	0.14	0.015	0.1325	0.0025	0.02	0.0075	0.0175	0.0575	0.0125
u	0.04	0.0275	0.02	0.0225	0.01	0.0125	0.005	0.565	0.0425	0.035	0	0.0075	0
v	0.0075	0.005	0.0075	0	0.01	0.0025	0	0.01	0.485	0.03	0.0225	0.34	0
w	0.0075	0.0025	0.005	0	0	0	0.0025	0	0.0525	0.875	0.0025	0.02	0
x	0.01	0.0025	0.005	0.0025	0.0375	0.0075	0.005	0.0025	0.03	0.0125	0.5325	0.1725	0.015
y	0.005	0	0	0	0.005	0	0	0	0.06	0.005	0.0075	0.8675	0.0025
z	0.0025	0.01	0.0075	0	0.0375	0.005	0.005	0.005	0.0025	0.0025	0.04	0.025	0.725

Table 2. Median activity obtained with ideal reader 1 for each letter position in words containing 4, 5, 6 or 7 letters. For five-letter words, list (a) contained all five-letter words without diacritics whereas list (b) contained the 1000 words used in the bubbles experiment reported in the present article.

	Letter position						
Word length	**1**	**2**	**3**	**4**	**5**	**6**	**7**
4 letters	1.4479	2.6082	2.7020	2.4689			
5 letters (a)	1.5107	3.0702	2.3223	2.4952	3.0415		
5 letters (b)	1.4103	3.0697	2.1373	2.3352	2.9281		
6 letters	1.5393	3.2689	2.2985	2.4078	2.6729	3.0000	
7 letters	1.4584	3.1411	2.2476	2.4237	2.8387	2.6546	3.0241

Figure Captions

Figure 1. The French word "bulle" sampled using 332 bubbles, which was the average number of bubbles used by the participants in the first half of the experimental sessions. Only the horizontal strip of the stimulus that contains letter signal is displayed. Each one of the 24 stimulus frames has a duration of 8.33 ms, for a total stimulus duration of 200 ms. The magnified portion of the stimulus shows a complete bubble cycle.

Figure 2. A thresholded classification movie. The space-time voxels reaching statistical significance are depicted in bright red in the center of the figure and are overlaid on the word “javel”. The numbers within or nearest to each of the seven blobs indicate the size of these blobs in voxels. The voxels were projected onto three bidimensional planes: the back wall—to isolate the spatial left-right and up-down dimensions; the floor—to isolate the time and the left-right dimensions; and the right wall—to isolate the time and the up-down dimensions. The number of significant voxels projected onto a single pixel on the planes is represented by red saturation (see legend). The dim grey lines delimit the 24 frames on the time dimension and the three different regions of the five letters on the space dimensions (i.e., body, ascenders, descenders).

Figure 3. Space-time deployment of attention for the combination of free parameters that lead to the greatest r^2 between human behavioural data and the predictions of the seven models considered in the article. The number of free parameters and r^2 are specified for each model reader. To summarize the predictions along the time and the left-right dimensions, histograms flank the prediction planes. The models are

ordered from best (top left) to worst (bottom right) according to their likelihood ratios corrected for the number of free parameters.

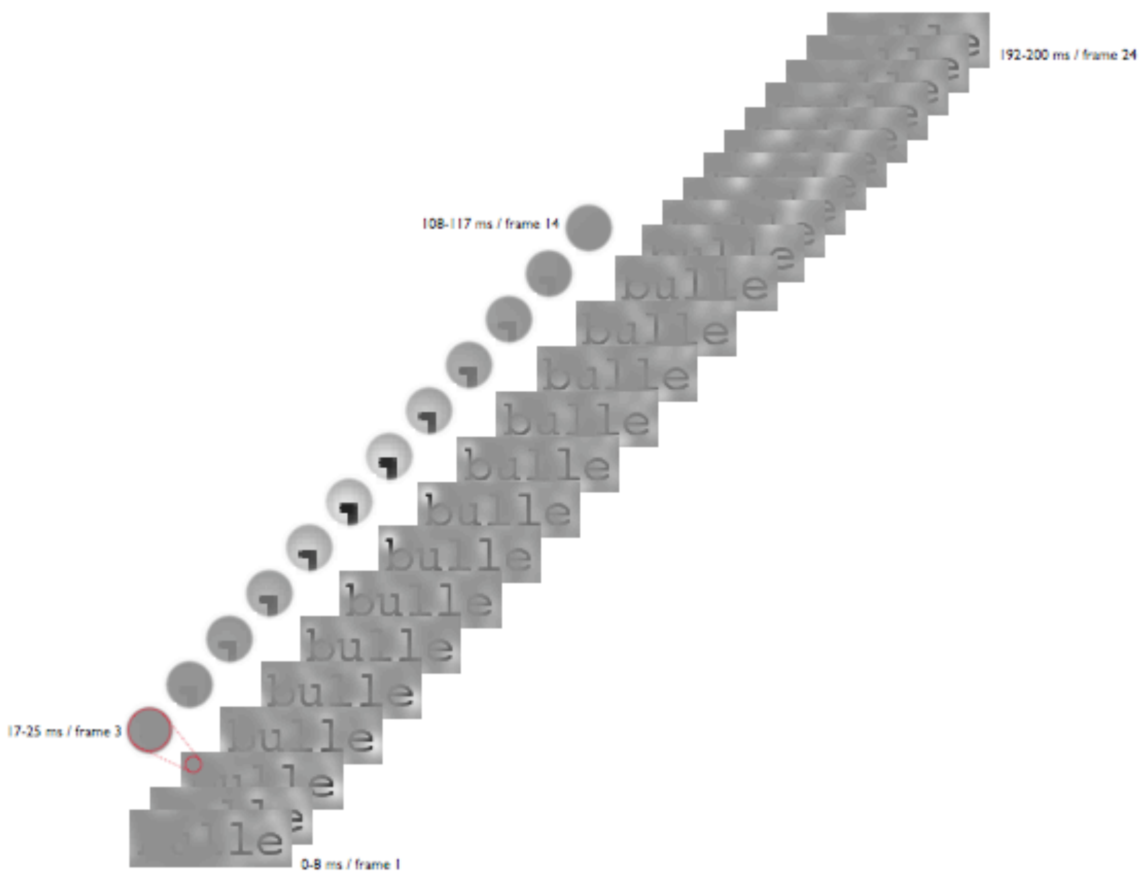
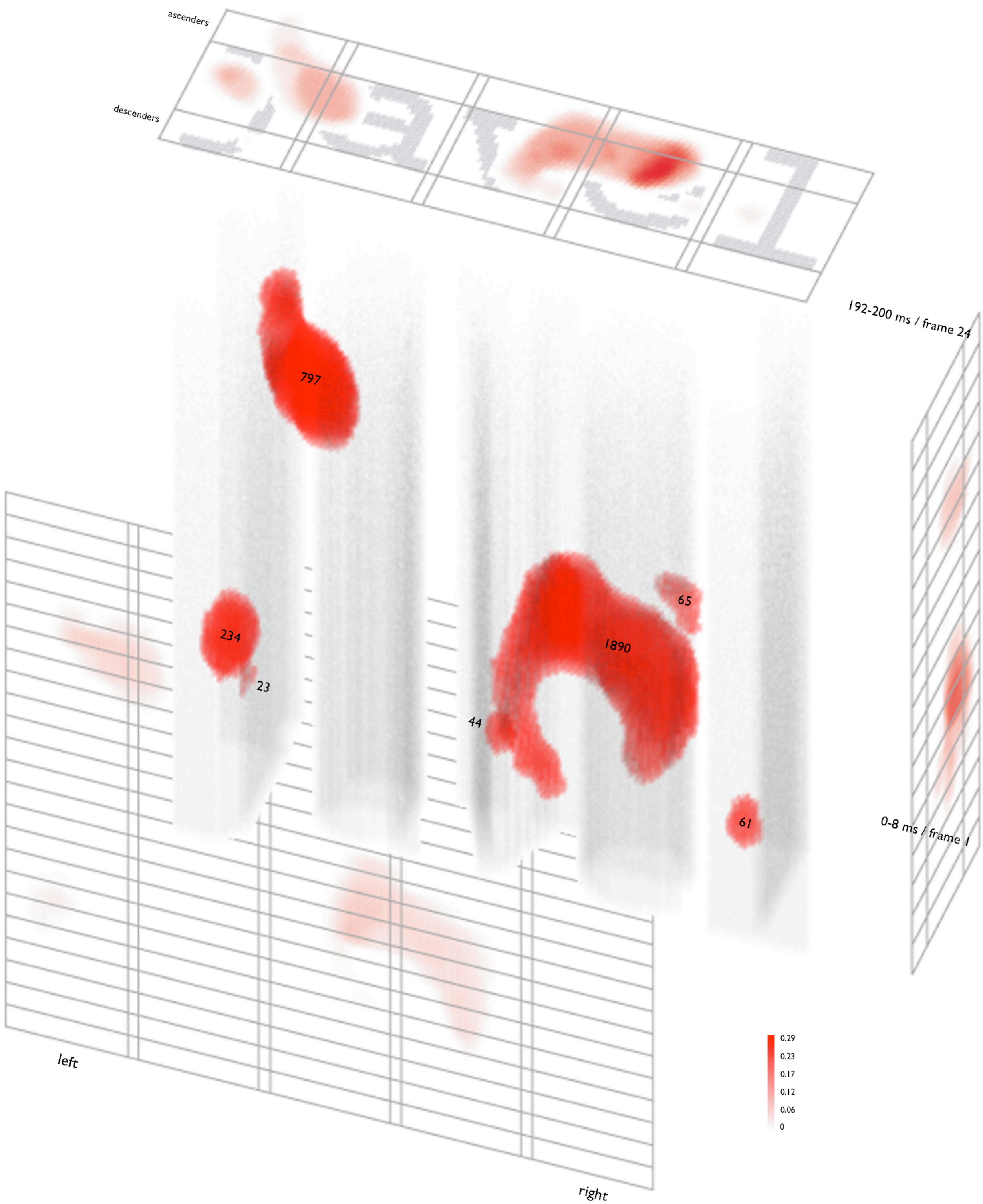
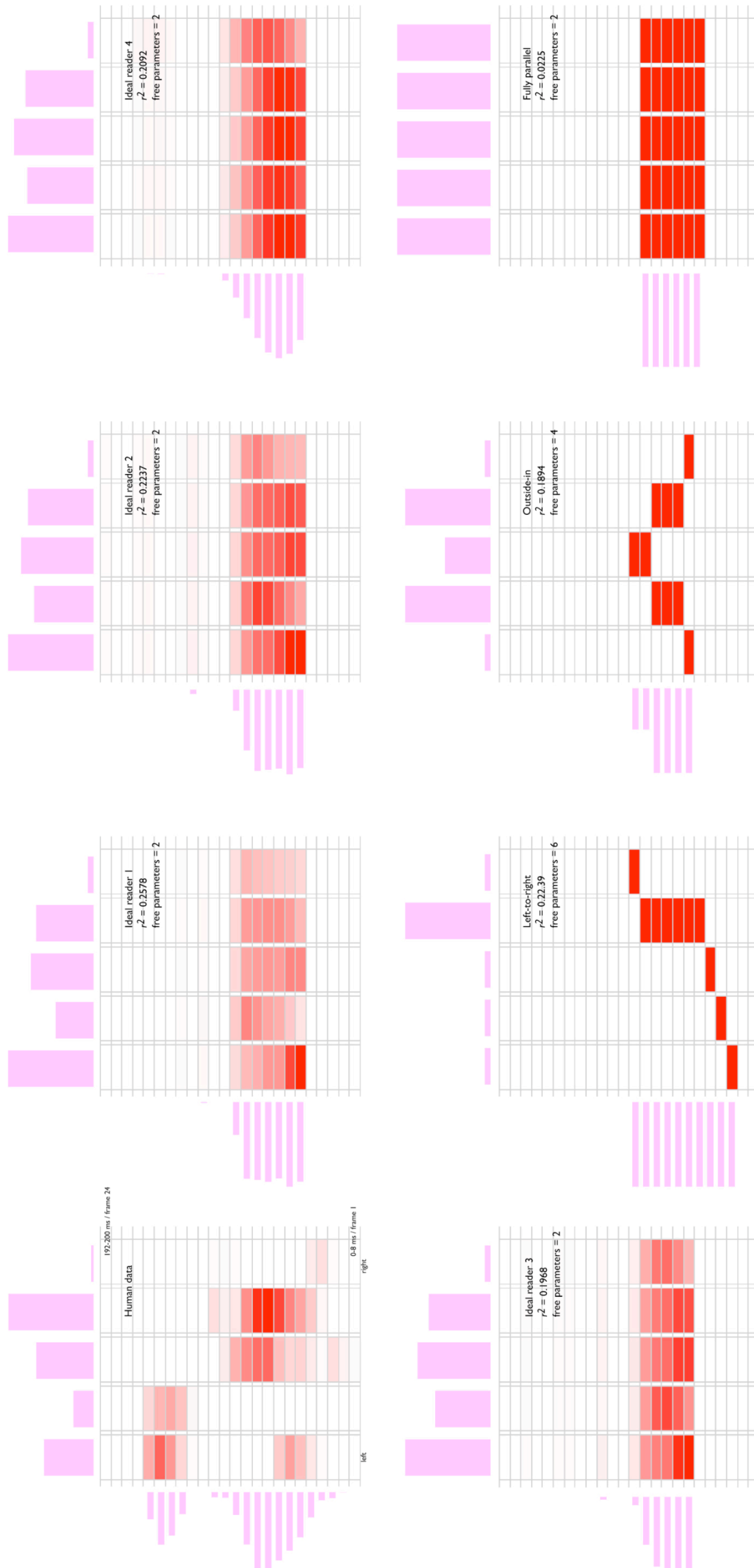


Figure 1. Blais, Fiset, Arguin, Jolicoeur & Gosselin


ascenders
descenders
192-200 ms / frame 24
797
234
23
1890
65
44
61
0-8 ms / frame 1
left
right
0.29
0.23
0.17
0.12
0.06
0



Human data
192-200 ms / frame 24
0-8 ms / frame 1
left
right
Ideal reader 1
$r^2 = 0.2578$
free parameters = 2
Ideal reader 2
$r^2 = 0.2237$
free parameters = 2
Ideal reader 4
$r^2 = 0.2092$
free parameters = 2
Ideal reader 3
$r^2 = 0.1968$
free parameters = 2
Left-to-right
$r^2 = 0.2239$
free parameters = 6
Outside-in
$r^2 = 0.1894$
free parameters = 4
Fully parallel
$r^2 = 0.0225$
free parameters = 2