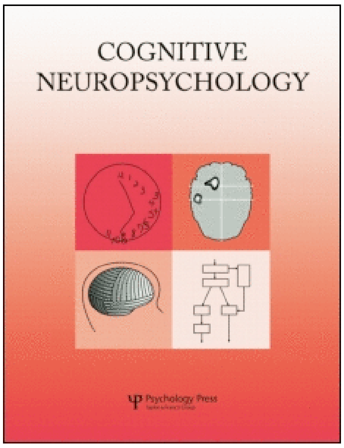

Cognitive Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713659042

The spatio-temporal dynamics of visual letter recognition

Daniel Fiset [a]; Caroline Blais [b]; Martin Arguin [b]; Karine Tadros [b]; Catherine Éthier-Majcher [b]; Daniel Bub [a]; Frédéric Gosselin [b]

[a] University of Victoria, Victoria, British Columbia, Canada [b] Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

First Published:February2009

To cite this Article Fiset, Daniel, Blais, Caroline, Arguin, Martin, Tadros, Karine, Éthier-Majcher, Catherine, Bub, Daniel and Gosselin, Frédéric(2009)'The spatio-temporal dynamics of visual letter recognition',Cognitive Neuropsychology,26:1,23 — 35

To link to this Article: DOI: 10.1080/02643290802421160

URL: http://dx.doi.org/10.1080/02643290802421160

COGNITIVE NEUROPSYCHOLOGY, 2009, 26 (1), 23–35

Psychology Press
Taylor & Francis Group

The spatio-temporal dynamics of visual letter recognition

Daniel Fiset
University of Victoria, Victoria, British Columbia, Canada

Caroline Blais, Martin Arguin, Karine Tadros, and Catherine Éthier-Majcher
Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

Daniel Bub
University of Victoria, Victoria, British Columbia, Canada

Frédéric Gosselin
Centre de Recherche en Neuropsychologie Expérimentale et Cognition, Département de Psychologie, Université de Montréal, Montréal, Québec, Canada

We applied the *Bubbles* technique to reveal directly the spatio-temporal features of uppercase Arial letter identification. We asked four normal readers to each identify 26,000 letters that were randomly sampled in space and time; afterwards, we performed multiple linear regressions on the participant's response accuracy and the space–time samples. We contend that each cluster of connected significant regression coefficients is a letter feature. To bridge the gap between the letter identification literature and this experiment, we also determined the relative importance of the features proposed in the letter identification literature. Results show clear modulations of the relative importance of the letter features of some letters across time, demonstrating that letter features are not always extracted simultaneously at constant speeds. Furthermore, of all the feature classes proposed in the literature, line terminations and horizontals appear to be the two most important for letter identification.

Keywords: Letter recognition; Letter features; Temporal processing; Classification images; Bubbles.

Most textbooks in cognitive psychology use alphabetic characters identification as a starting point in understanding the visual and cognitive mechanisms involved in visual object recognition (e.g., Lindsay & Norman, 1977; Matlin, 2005; Medin, Ross, & Markman, 2005; Neisser, 1967). As a microcosm, letters have many advantages over other more ecologically oriented categories

Correspondence should be addressed to Martin Arguin or Frédéric Gosselin, Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, H3C 3J7, Canada (E-mail: martin.arguin@umontreal.ca or frederic.gosselin@umontreal.ca).

We thank those who took part in this study. This research was supported by a grant from the Canadian Institute of Health Research (CIHR) to Martin Arguin, Frédéric Gosselin, and Daniel Bub; by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Martin Arguin and to Frédéric Gosselin; by a scholarship from the James S. McDonnell Foundation (Perceptual Expertise Network) and by a postdoctoral scholarship from the Fonds Québécois de Recherche en Nature et Technologies (FQRNT) to Daniel Fiset; and by a FQRNT graduate scholarship to Caroline Blais and to Karine Tadros.

http://www.psypress.com/cogneuropsychology
DOI:10.1080/02643290802421160

of visual objects (e.g., animals, houses, faces, etc.). They were designed as a limited set of objects composed of a limited assortment of traits, with the means to meet specific communication needs (i.e., reading and writing). These traits have relatively simple shapes and are typically displayed using two tones. For example, the uppercase letter "A" in the Arial font is composed of two slanted lines inclined at about 60 degrees that are joined at their superior extremity and that are connected by a horizontal line that touches both slants approximately in their centres. Of course, the trait description of the uppercase "A" may vary slightly across fonts but it remains sufficiently robust that the letter remains recognizable despite these small variations. Furthermore, letter identification as the starting point of word recognition and reading (Pelli, Farell, & Moore, 2003) has strong real-life relevance since a large proportion of our modern life activities involve this function. As the above illustration suggests, it is relatively easy to enumerate the letter traits necessary to describe each of the 26 uppercase letters of the alphabet in predefined fonts (e.g., Arial). Whether such traits are used by humans for letter identification, however, remains uncertain.

Seeking to identify a psychologically valid set of letter features (e.g., E. J. Gibson, 1969; Rumelhart & Siple, 1974), researchers in cognitive sciences have privileged the use of data from confusion matrices (Boles & Clifford, 1989; Bouma, 1971; Briggs & Hocevar, 1975; Gervais, Harvey, & Roberts, 1984; Geyer, 1977; Gilmore, Hersh, Caramazza, & Griffin, 1979; Loomis, 1982; Townsend, 1971; Van Der Heijden, Malhas, & Van Den Roovaart, 1984). A confusion matrix is constructed by measuring the human participant's ability to distinguish single letters in very demanding or special conditions so that errors frequently occur—typically as often as on 50% of trials. For example, some researchers examined the performance of children who had not yet integrated the exact visual form of letters (E. J. Gibson, Gibson, Pick, & Osser, 1962); others studied the performance of skilled readers when identifying letters presented for a brief duration (Townsend, 1971) or with extremely low contrast (Geyer, 1977). In these confusion matrices, errors in letter discrimination are thought to be helpful in defining the traits necessary for distinguishing letters from one another. Hence, it is commonly assumed that the frequent confusion between the uppercase "E" and "F" in these specific conditions validates the inferior horizontal line of the uppercase "E" as a diagnostic trait for the recognition of these letters. Even if this proposition makes sense, it does not tell us which part(s) of the bar help to discriminate between these two letters. For instance, it could be the intersection between the vertical and the horizontal bar, the termination of the horizontal bar, or the horizontal bar itself.

Much difficulty has been encountered when it comes to pinpointing the exact diagnostic areas for letter discrimination. We believe this originates from the vast gap between the letter confusion data that has been compiled and the letter features that have been proposed. In fact, we question whether letter-confusion matrices constitute the appropriate tool to provide a decisive set of data for determining the diagnostic features for letter identification. In particular, it is important to bear in mind that all the experimental manipulations required for the creation of letter-confusion matrices (low contrast or rapid presentation) exacerbate the relative importance of low spatial frequencies (e.g., Mazer, Vinje, McDermott, Schiller, & Gallant, 2002). Since this visual information is not optimal for human vision and leads to very high error rates, it may be inadequate for the discovery of the letter features underlying reading in daily life.

In this study, we used a classification image technique (e.g., Eckstein & Ahumada, 2002; Gosselin & Schyns, 2004) called *Bubbles* (Gosselin & Schyns, 2001) that uncovers more directly the letter components driving accurate recognition. The underlying logic of *Bubbles* is that if some piece of visual information is necessary to perform the task at hand, masking this information will impair performance, and revealing it will lead to a better performance. The plane of regression coefficients that is obtained through multiple linear regressions of performance as a function of the masks used to sample information on every trial, is called a *classification image*, and it reveals the effective information upon which the observed performances are based. We recently employed a version of *Bubbles* in order to reveal the potent visual features mediating uppercase and lowercase Arial letter identification across different spatial frequency bands (Fiset et al., 2008). The analyses conducted separately on each of the 26 uppercase and 26 lowercase letters confirmed that the spatial frequency information between 2 and 4 cycles per letter conveys the most potent visual information (Chung, Legge, & Tjan, 2002; Ginsburg, 1980; Legge, Pelli, Rubin, & Schleske, 1985; Majaj, Pelli, Kurshan, & Palomares, 2002; Parish & Sperling, 1991; Solomon & Pelli, 1994). To synthesize the large amount of data obtained from this experiment and to link it to the letter identification literature published during the years 1960–1980, we also determined the relative importance of the sets of features proposed in that literature as well as the relative importance of line terminations, a set of features that had not been considered in that literature. We found that terminations, relatively small features found at the extremities of lines, and horizontals, were the most effective in driving performance. To the best of our knowledge, this was the first empirical demonstration that line terminations are of the crucial importance for letter identification.

Here, we examine the space–time features for Arial uppercase letter identification by using a dynamic version of the *Bubbles* method (Blais, Fiset, Arguin, Jolicoeur, & Gosselin, 2004; Vinette, Gosselin, & Schyns, 2004). Extending the logic of the spatial *Bubbles* technique briefly discussed above to the time dimension amounts to saying that the probability of a correct answer should decrease if the information that is efficient for letter identification at a particular spatial location and moment is not revealed at that spatial location and moment and that it should increase if this information is revealed at that spatial location and moment. Therefore, in order to determine the efficient use of spatio-temporal information, we perform multiple linear regression between the participant's response accuracy and the space–time bubbles.

Method

Participants

A total of 4 students from the University of Montréal took part in this experiment. All had normal or corrected-to-normal visual acuity.

Materials and stimuli

Stimuli were displayed on a high-resolution Sony monitor at 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 at 57 cm by using a chinrest. Stimuli were uppercase letters printed in Arial font subtending on average 0.78 degrees of visual angle horizontally and 0.97 degrees of visual angle vertically. They appeared in dark grey (2.1 cd/m^2) over a light grey background (64.8 cd/m^2) and were sampled in space and time. More specifically, the "bubblized" movies consisted of a sequence of 12 successive frames, each presented on screen for a duration of 8.33 ms (for a total stimulus duration of 100 ms), displaying one letter of the alphabet sampled with Gaussian apertures (i.e., bubbles) randomly located in space–time (see Figure 1). Therefore, the spatial information (e.g., different groups of pixels in a letter) available to participants varied as a function of time within a trial, and the sequence of space–time bubbles also varied randomly across trials. Each bubble had a standard deviation of 0.1 degrees of visual angle (3 pixels) in the spatial domain and a standard deviation of 17.3 ms (2.08 frames) in the temporal domain. The temporal full width at half maximum of a bubble—40.7 ms—is less than the time required to plan and execute an attentional saccade, which ensures that participants are unable to shift their attention towards a particular bubble (i.e., estimates of the time needed to plan and execute an attentional saccade typically range between 50 and 85 ms; e.g., Wolfe, 1998; Wolfe, Alvarez, & Horowitz, 2000).

Procedure

Each participant performed 26,000 trials, each letter of the alphabet being presented an average of 1,000 times. The experiment was divided in

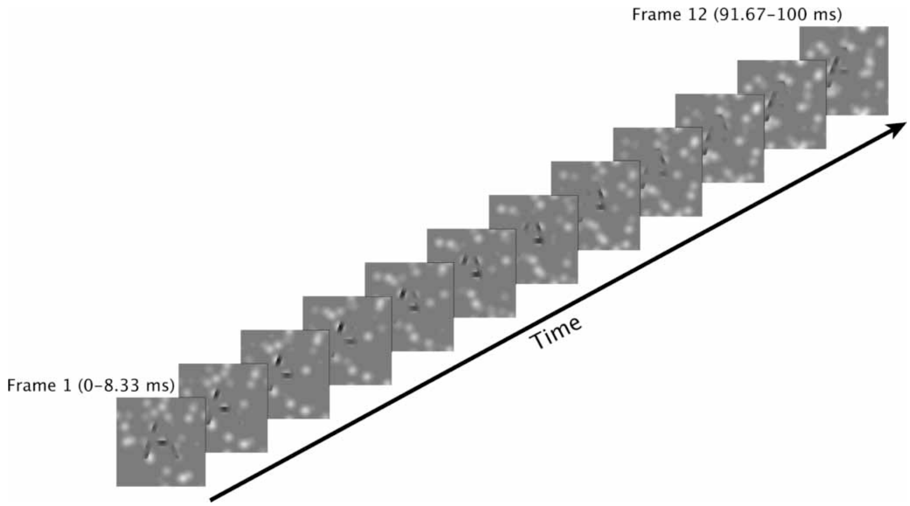

Figure 1. *Illustration of the spatio-temporal stimulus sampling. A total of 12 frames were presented successively at a rate of 120 Hz (i.e., 8.33 ms per image). In these frames, visual information was randomly sampled in space and time using Gaussian apertures (bubbles) with a spatial and a temporal extent. The standard deviation of the bubbles on the time dimension was chosen such that the duration of one bubble was shorter than the time required to plan and execute an attentional saccade.*

100 blocks of 260 trials each. The letter identification accuracy was maintained at 51% by adjusting the number of bubbles on a trial-by-trial basis using QUEST (Watson & Pelli, 1983). Since visual processing difficulty may vary across the alphabet, the number of space–time bubbles was adjusted independently for each letter. The initial number of bubbles for each letter was determined individually for each participant by administering two practice blocks, each composed of 50 trials, before the experimental blocks began.

On each trial, a homogenous grey screen was first displayed for 250 ms, accompanied by a 122-ms 1300-Hz pure tone to signal the beginning of the trial. The grey screen was immediately replaced by a bubblized letter movie that lasted 100 ms. This was immediately followed by a homogenous grey screen that remained visible until the participant responded. The task was to identify the target letter, and participants registered their responses by pressing the appropriate key on the keyboard. The next trial was triggered automatically after a 2-s intertrial interval. Participants received no feedback on their performance.

Results and discussion

The number of bubbles necessary to maintain performance at a 51% correct for each letter of the alphabet is reported in Table 1. The efficient use of the spatio-temporal information in the stimulus was determined by performing a multiple linear regression on the bubbles' volumes presented on every trial (explanatory variables) and the participants' response accuracy (predictor variable). That is, we constructed, for each participant, one regression coefficient volume (the two spatial dimensions and the temporal dimension) for each letter of the alphabet by subtracting, for a given letter, the sum of the bubbles' volumes that led to an incorrect response from the sum of the bubbles' volumes that led to a correct response. These volumes of regression coefficients are referred to as *classification movies*, which is a straightforward extension of classification images. The elements of these movies are referred to as *voxels* (by analogy to *pixels* in classification images).

If all $64 \times 64 \times 12$ voxels were of equal importance for successful letter identification,

Table 1. *Average number of bubbles required to maintain performance at 51% correct at the end of the experiment*

	Number of bubbles
A	50.0
B	84.6
C	78.2
D	58.2
E	68.2
F	82.1
G	58.8
H	38.2
I	132.0
J	94.7
K	38.9
L	92.4
M	47.8
N	43.7
O	108.2
P	88.7
Q	91.4
R	72.9
S	59.1
T	47.9
U	62.2
V	59.4
W	39.1
X	40.6
Y	46.4
Z	45.4

they would have uniform regression values. 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 spatial central horizontal strip in the classification movies ($40 \times 64 \times 12$ voxels) where the letters were located. The strips above and below were used to estimate the mean and the standard deviation of the null distribution and to transform the classification movies into Z-scores. A total of 26,000 trials might seem like a lot but in this case it is not enough to obtain one classification movie per letter for each participant. We thus summed the classification movies transformed into Z-scores across the 4 participants and, to transform the resulting group classification movies into Z-scores, divided them by $\sqrt{4}$. Finally, to determine the letter space–time information significantly correlated with accuracy, we conducted a one-tailed Pixel test (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005) on the group classification movies transformed into Z-scores ($S_r = 30{,}720$ voxels; full-width half maximum $= 2.66$; i.e., the geometric mean of the spatial and the time full-width half maximums; $Z_{crit} = 4.46$; $p < .001$). The statistical threshold provided by this test corrects for multiple comparisons while taking the spatial and temporal correlation inherent to our technique into account.

Movies directly representing the space–time use of letter information are available on http://www.mapageweb.umontreal.ca/gosselif/dynamic_letters. Figure 2 depicts the same results in two dimensions while losing as little information as possible. In a statistically thresholded classification movie, some significant voxels are connected together, and some are not. For example, on the bottom termination of the letter "I" (see Figure 2), it is likely that more than one voxel will be useful, and that most of the useful voxels on that letter feature will be connected in space or in time. In contrast, the voxels located on the top termination of the letter "I" may not be connected with those of the bottom termination since both groups are far away from each other. We contend that each cluster of connected significant voxels is a letter feature. The temporal dimension of classification movies also informs us about the order in which these letter features are acquired. Therefore, we divided the significant voxels into space–time clusters of connected significant voxels. More precisely, we searched for so-called "26-connected" voxels (i.e., adjacent either in one of the six cardinal directions or in one of the 20 oblique directions). However, a cluster of significant voxels could contain as little as one voxel. We then collapsed the time dimension of the classification movies to represent the three dimensions of our results on a two-dimensional figure; each cluster was reduced to its spatial silhouette. We depicted the different cluster silhouettes in different colours to facilitate interpretation. The times of onset and offset of every cluster is indicated in white next to the cluster.

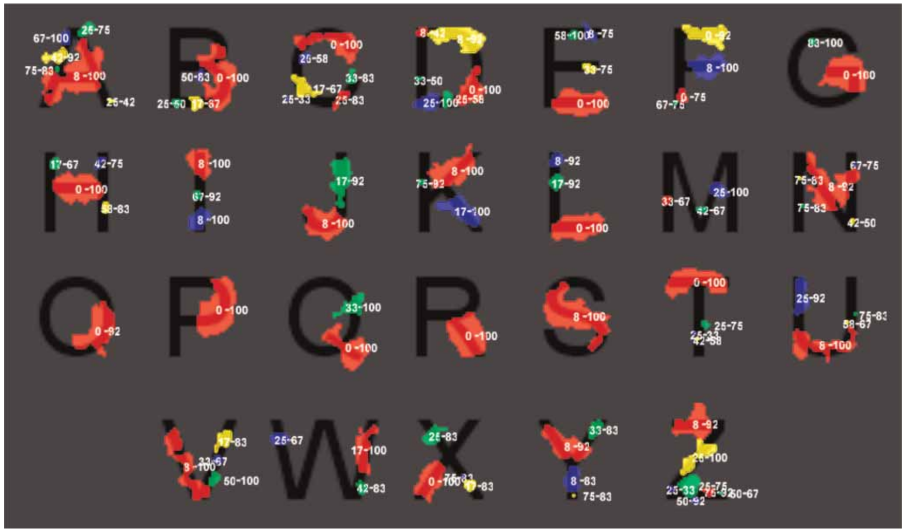

Figure 2. *Colours show space–time clusters (collapsed on the time dimension) significantly correlated with correct letter identification ($p \leq .001$) superimposed upon the appropriate letter. Four colours were used to help cluster segregation. The numbers in white near each cluster indicate the beginning and end of this cluster relative to stimulus onset.*

As mentioned in the Introduction, the literature on letter identification has already proposed various sets of letter features that are assumed to underlie identification. The issue of the relative impact of these different features on letter identification has generated significant interest. Because the *Bubbles* method is pixel or voxel based, it does not require an a priori definition of what the features for letter identification are. As we have described above, the classification movies as well as Figure 2 reveal, by themselves, the shape and the position of the efficient features—clusters of significant voxels. However, it is possible to decompose the classification movies into any set of features so as to assess the degree to which it accounts for recognition performance (e.g., B. Gibson, Lazareva, Gosselin, Schyns, & Wasserman, 2007). To bridge the gap between the letter identification literature and the experiment reported in this

Figure 3. *Templates of letter features used for the a priori feature analysis. One template was created for each feature (i.e., intersections, horizontals, verticals, slants tilted right, slants tilted left, curves opened at the top, curves opened at the bottom, curves opened on the left, curves opened on the right, and terminations) present in each letter of the alphabet. The pixels comprised in each template are depicted in red.*

article, we determined the relative importance of the features proposed in the letter identification literature. More specifically, we conducted a priori feature analyses for all the letters of the alphabet grouped together (similarly to Fiset et al., 2008) as well as for each letter separately. We first created 111 templates by decomposing each letter into the full complement of local features that have been proposed in the literature except for global features such as symmetry, cyclic change, and parallelism, which were not considered—vertical, horizontal, slant tilted left or right, curves opened up, down, left or right, and intersections (Figure 3). We also included terminations, a feature that has only been considered by Fiset et al. (2008). The terminations and intersections were defined as letter ink within a radius of 13 pixels of the centre of the feature, as identified by the authors. To make sure that the masks for these features were independent from those of other features, we subtracted the area corresponding to the terminations and intersections from the other feature masks. We then calculated, for each letter and frame, the proportion of significant voxels falling on each feature. Only the voxels falling directly on letter print were included in the analysis. The vectors of 10 proportions were each normalized to 1 in order to reveal the relative importance of all features for each letter. The results of this analysis for each individual letter are presented on Figure 4. Error bars were computed via bootstrap: 1,000 group classification movies were computed by summing four classification movies made of pixels sampled randomly, with replacement, from the four classification movies of the participants. In fact, the video clips available on http://mapageweb.umontreal.ca/gosselif/dynamic_letters depict the sum of these bootstrap classification movies. Bright red means that the pixels are present on 100% of classification movies, black means that they are present on 0.1%, and gray letter and white background means that they are present in none of the classification movies. These video clips thus indicate the between-subject variability. To compute the error bars in Figure 4, the group classification movies obtained via bootstrap were analysed in exactly the same way as the empirical group classification movie: They were smoothed, were transformed into Z-scores, were submitted to Pixel tests, and underwent feature analyses. The error bars correspond to 1.96 times the standard deviation observed in the simulated feature analyses—95% confidence intervals. Most of the between-subject variability occurs at the beginning and at the end of stimulus duration, which could be due to intersubject variability in stimuli onset and offset estimations.

Different spatio-temporal patterns may be observed. For some letters, one feature remains useful from the beginning to the end of the stimulus presentation. This is the case, for example, with the terminations in letter “I” and with the horizontal in letter “G”. A second spatio-temporal pattern that may be observed in the results is the simultaneous presence of two or more letter features. For example, in letter “G”, the relative usefulness of the terminations, the horizontal bar, and the intersection is approximately constant across time. Finally, for other letters, one feature appears early in the classification image, then disappears, and, sometimes, another feature appears. For example, in letter “U”, the curve opened at the top are essentially the only useful features from 17 to 42 ms; in letter “W”, terminations are the most useful feature from 25 to 42 ms after stimulus onset, and then slants tilted right become the most useful feature from 50 to 100 ms after stimulus onset.

To compare our results with those of Fiset et al. (2008), we combined the results across the 26 letters of the alphabet and across time, for each feature class, and divided that grand total by the number of letters containing a given feature. This resulted in a vector of 10 numbers that was normalized to 1 in order to reveal the relative importance of all features across all the letters of the alphabet (see Figure 5; error bars in Figure 5 were computed via bootstrap like the error bars in Figure 4). Terminations and horizontals are the most important features for uppercase Arial letter identification.

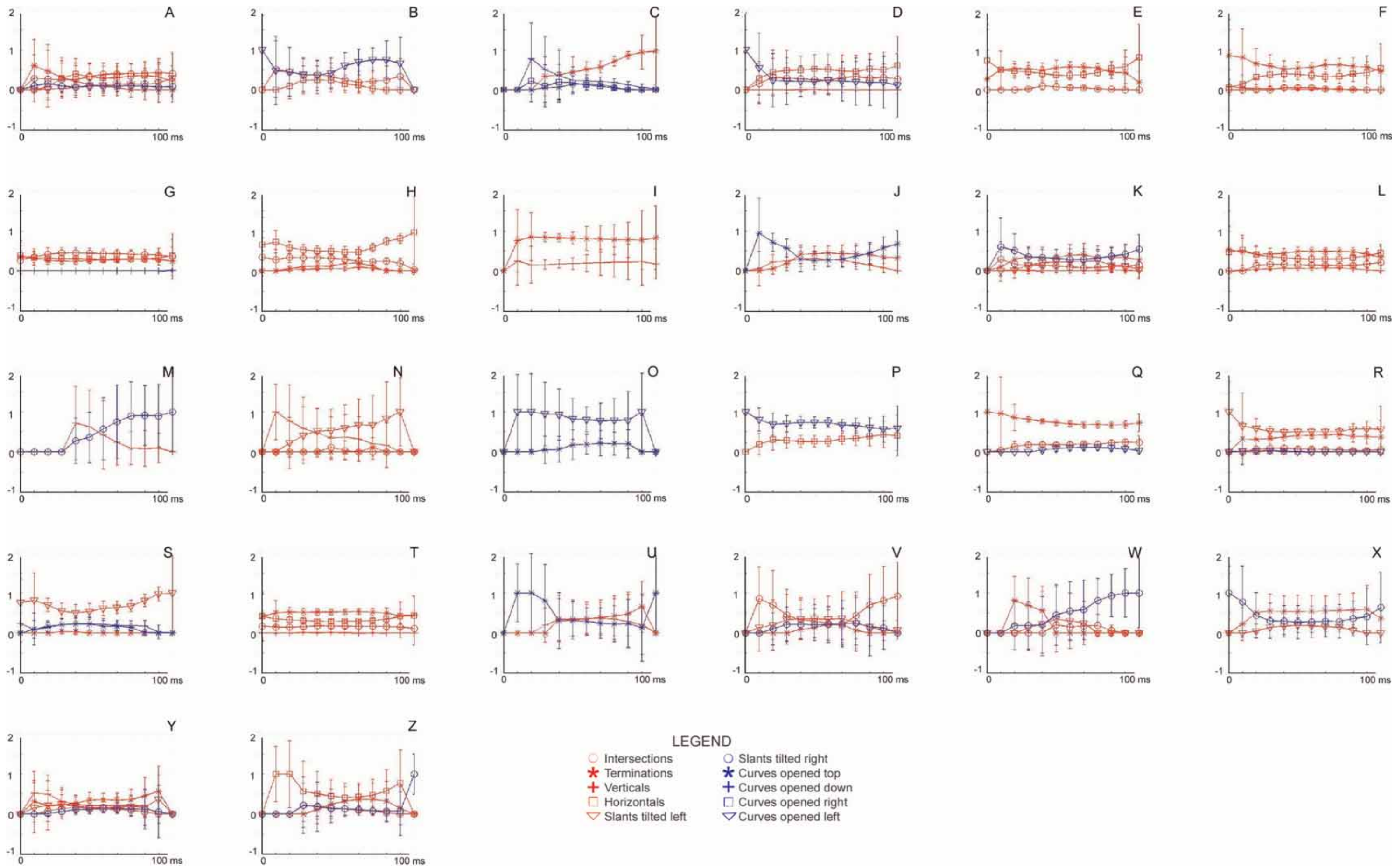

Figure 4. *Results of the a priori feature analysis performed on each letter and each frame. Each graph shows the relative importance of the features comprised in each letter of the alphabet. Note that if no significant pixel fell on one of the feature comprised in a letter (e.g., no significant pixel fell on the vertical bar in letter "B"), there is no curve corresponding to this feature in the graph. Error bars indicate 95% confidence intervals.*

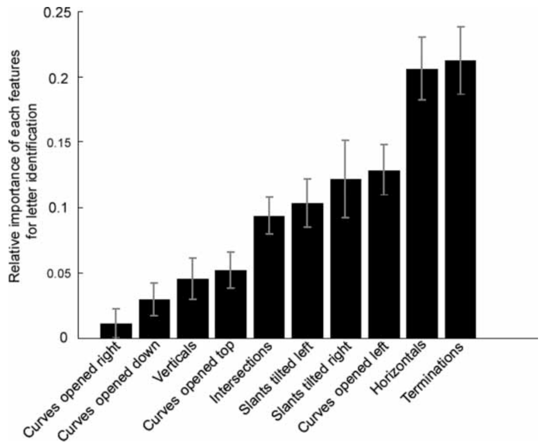

Figure 5. *Results of the overall a priori feature analysis, all letters and all frames confounded. Error bars indicate 95% confidence intervals. To view a colour version of this figure, please see the online issue of the Journal.*

GENERAL DISCUSSION

We used *Bubbles*, a classification image technique, to reveal the letter areas responsible for the accurate identification of uppercase Arial letters in space–time. The space–time clusters that are significantly correlated with letter identification are shown for every letter in Figure 2 (movies are available from http://mapageweb.umontreal.ca/gosselif/dynamic_letters). These, we claimed, are the space–time features for letter identification. Nonetheless, to create a link with the literature, we examined the relative importance of 10 feature classes that have been proposed to be important for the letter identification (i.e., intersections, horizontals, verticals, slants tilted right, slants tilted left, curves opened at the top, curves opened at the bottom, curves opened on the left, curves opened on the right, and terminations). In the "General discussion", we focus on these a priori feature analyses because they suffice for the arguments put forth and because they should make the arguments more concise.

In the first feature analysis, we computed the importance of each feature class for every letter and frame (see Figure 4). If human observers processed the features of letters simultaneously at different but constant speeds—henceforth we speak about *simple parallel observers*—the relative importance of the features would be invariant across frames in their classification movies (McCabe, Blais, & Gosselin, 2005). To illustrate, consider the following toy problem: A simple parallel observer is exposed for the duration of two frames (f_1 and f_1) to a pseudoletter composed of two parts (p_1 and p_2) each sufficient to identify the pseudoletter. Bubble masks can be represented as 2×2 matrices:

$$\begin{pmatrix} p_1f_2 & p_2f_2 \\ p_1f_1 & p_2f_1 \end{pmatrix},$$

where a cell is equal to 1 when there is a bubble and to 0 otherwise. Suppose that the observer requires only one frame to process p_1 and two frames to process p_2. A total of 13 bubble masks (out of a possible 15) would lead to correct responses:

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \text{and} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix};$$

and the remaining two would not:

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{and} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The classification movie computations described in the "Results and discussion" section—in the ongoing example, it rather is a classification image—would consist in summing up all bubble masks weighted by plus or minus 1/(number of bubbles), respectively, if the mask led to a correct or an incorrect letter identification:

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$
$$+ \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$
$$+ \frac{1}{3}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$
$$+ \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 3.75 & 1.75 \\ 3.75 & 1.75 \end{pmatrix}.$$

The relative importance of the two parts is the same across frames, and this is always true of simple parallel observers.

Our results do not fully support the hypothesis that humans are simple parallel observers. Indeed, there are modulations of the relative importance of the feature classes across time in some letters (see Figure 4). This is particularly clear for letters “C”, “F”, “M”, “U”, “W”, and “Z”. But the hypothesis does fit with the results obtained for some letters. In letter “G”, for example, the relative usefulness of the terminations, the horizontal bar, and the intersection is approximately constant across time. Interestingly, these three features are next to each other spatially in letter “G” (see also Figure 2). This may result from the fact that these features fall within the “spotlight” of attention, which permits their simultaneous processing. Note, however, that our analysis does not allow us to infer whether two regions (or more) that are simultaneously above statistical threshold in our classification movies are actually processed simultaneously. Thus, it could be that on some trials, participants used one region, and, on other trials, they used the other region.

In the second a priori feature analysis, we computed the importance of each feature class with all letters and frames confounded (see Figure 5). The most important outcome of this analysis is the confirmation of the prime importance of terminations and horizontals. The case of terminations is especially surprising because no research team has ever suggested that terminations could be key features for letter identification except Fiset et al. (2008). Indeed, Fiset et al. conducted a *Bubbles* experiment to uncover the spatial features for uppercase and lowercase Arial letter identification at different spatial scales—they did not sample time. The results of the a priori feature analyses in the two studies are strikingly similar ($r = .96$). Apart from intersections, which came in fourth position in Fiset et al. and are in sixth position here, the order of importance of the nine remaining features is exactly the same. This really is striking given that the stimulation parameters used in Fiset et al. differed greatly from those used in the present experiment: Fiset et al. used spatial bubbles with a standard deviation of 0.72, 0.36, 0.18, 0.09, and 0.045 letter width, respectively, from the lowest to the highest spatial frequency band along which stimuli were filtered (1–2, 2–4, 4–8, 8–16, and 16–32 cycles per letter), instead of randomly located space–time bubbles with a standard deviation of 0.13 letter width across space and with a standard deviation of 40.8 ms across time; Fiset et al. displayed the letters for 200 ms instead of 100 ms; and Fiset et al. used letters with an average width of 1.35 degrees of visual angle instead of 0.78 degrees of visual angle (reducing letter size is known to induce a shift in the use of spatial information toward lower spatial frequencies; e.g., Majaj et al., 2002). This suggests that the results obtained in the experiment reported in this article are robust to parameter changes and that they generalize to different experimental conditions. However, it could also be that the results are an artefact of the a priori feature analysis and letter statistics. Fiset et al. also report an a priori feature analysis applied on an *ideal observer* classification images. An ideal observer optimally uses all the information available to perform the task at hand (e.g., Solomon & Pelli, 1994). The purpose of such a model is not so much to fit human data but to understand how the human data diverge from an optimal implementation that uses all the available information, without constraint. For the ideal observer, the terminations ranked 5th and 6th out of the 10 feature classes, and the horizontals ranked 7th and 4th for lowercase and uppercase letters, respectively. The correlation between the relative importance of the features for human observers and ideal observer is quite low ($r = .16$). On this basis, we are confident that the prime importance of the terminations and horizontals, in particular, and the relative importance of the other features for human participants is due to constraints imposed by the human visual system rather than by constraints imposed by the stimuli or analyses.

But why exactly are terminations and horizontals so important for human letter identification? Regarding terminations, one possible hypothesis is that, because they are located on the extremities of letters, they are less likely to suffer from visual crowding than other features. This hypothesis is supported by the results of Fiset et al. (2008) with lowercase letters. Indeed, although terminations were the most important feature for both letter cases, they were more important relative to the other features for lowercase letters than to those for uppercase letters. Since the distance between the terminations and the rest of the letter is, on average, larger in lower- than in uppercase letters (i.e., because of their extensions), their sparing from crowding should be more important. This hypothesis also predicts that terminations should become even more important in word recognition, where crowding is further increased by adjacent letters. Interestingly, Chung, Tjan, and Lin (2008) showed that the extremities of lowercase letters are very useful for the correct identification of the middle letter in random triplets of letters. Regarding horizontals, it is possible that their importance also comes from a reduction of crowding when they are part of a letter. In fact, horizontals create some space either around the letter (e.g., in letter T) or within the letter (e.g., letter H), therefore reducing the crowding between letters in a letter string, or between features in an isolated letter. Other explanations are possible, however, and additional research will be necessary to understand why terminations and horizontals are important for letter identification.

One other important question that remains is: Why are the letter features extracted in the particular temporal order found here (see Figure 4)? We have examined different hypotheses but none has yet proven effective in correctly accounting for the findings reported above. For instance, we have examined whether the order of feature extraction follows a systematic spatial pattern. That is, were the different spatial locations (e.g., upper left quadrant, upper middle quadrant, upper right quadrant, etc.) processed in a systematic order? On each frame, we found a similar number of significant pixels across the stimulus areas, which led to a rejection of this hypothesis. We have also implemented an optimal sequential model—somewhat similar to “Mr. Chips” (Legge, Klitz, & Tjan, 1997)—to reveal the “optimal” order of feature extraction. In this model, on each time frame, a Gaussian window (we tried window sizes of standard deviations ranging from 2 to 9 pixels but this made little difference) was moved across the spatial extent of the letter to find which group of pixels would maximally decrease the uncertainty (i.e., minimize the entropy) about the target identity given the information that had already been accumulated before. A feature analysis across all letters revealed that this optimal sequential model did not primarily use terminations. Moreover, when similar features were used by the ideal and the human observers for a given letter, the order in which they were used usually differed. For letter “E”, for example, our optimal model first used the middle horizontal bar and then moved to the lower one, whereas for the human observers, this order was reversed. Overall, then, a straightforward account of the order in which the letter features become useful for normal human readers has yet to be uncovered. Future studies will be needed to further our understanding of the dynamics of letter identification.

REFERENCES

Blais, C., Fiset, D., Arguin, M., Jolicoeur, P., & Gosselin, F. (2004). Space-time spread of attention during a lexical decision task [Abstract]. *Journal of Vision*, *4*(8):525, 525a. Retrieved from http://journalofvision.org/4/8/525/, doi:10.1167/4.8.525.

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*, 579–586.

Bouma, H. (1971). Visual recognition of isolated lowercase letters. *Vision Research*, *11*, 459–474.

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

Briggs, R., & Hocevar, D. J. (1975). A new distinctive theory for upper case letters. *The Journal of General Psychology*, *93*, 87–93.

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.

Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, *42*, 2137–2152.

Chung, S., Tjan, B., & Lin, Y. (2008). Feature maps for letters [Abstract]. *Journal of Vision*, *8*(6):970, 970a. Retrieved from http://journalofvision.org/8/6/970/, doi:10.1167/8.6.970.

Eckstein, M. P., & Ahumada, A. J., Jr. (Ed.). (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, *2(1)*, i-i. Retrieved from http://journalofvision.org/2/1/i/, doi:10.1167/2.1.i.

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D. N., & Gosselin, F. (2008). *Features for uppercase and lowercase letter identification*. Manuscript submitted for publication.

Gervais, M. J., Harvey, L. O., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 655–666.

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*, 487–490.

Gibson, B., Lazareva, O. F., Gosselin, F., Schyns, P. G., & Wasserman, E. A. (2007). Non-accidental properties underlie shape recognition in mammalian and non-mammalian vision. *Current Biology*, *17*, 336–340.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Meredith.

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. A. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative & Physiological Psychology*, *55*, 897–908.

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception and Psychophysics*, *25*, 425–431.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. *Proceedings of the SID*, *21*, 219–227.

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

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

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

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research*, *25*, 239–252.

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing: An introduction to psychology*. New York: Academic Press.

Loomis, J. M. (1982). Analysis of tactile and visual confusion matrices. *Perception and Psychophysics*, *31*, 41–52.

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*, 1165–1184.

Matlin, M. W. (2005). *Cognition* (6th ed.). Hoboken, NJ: Wiley.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences, USA*, *99*, 1645–1650.

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

Medin, D. L., Ross, B. H., & Markman, A. B. (2005). *Cognitive psychology* (4th ed.). New York: John Wiley & Sons.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal. *Vision Research*, *31*, 1399–1416.

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

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

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review*, *81*, 99–118.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*, 395–397.

Townsend, J. T. (1971). Alphabetic confusion: A test of models for individuals. *Perception and Psychophysics*, *9*, 449–454.

Van Der Heijden, A. H. C., Malhas, M. S. M., & Van Den Roovaart, B. P. (1984). An empirical interletter confusion matrix for continuous-line capitals. *Perception and Psychophysics*, *35*, 85–88.

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

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.