Abstract: 102 words
Text: 3870 words
References: 39

**Features for uppercase and lowercase letter identification**

Daniel Fiset¹, Caroline Blais², Catherine Éthier-Majcher²,

Martin Arguin², Daniel Bub¹, & Frédéric Gosselin²

¹ Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, BC, V8W 3P5, CANADA

² Centre de Recherche en Neuropsychologie et Cognition, Département de Psychologie, Université de Montréal, CP 6128, succ. Centre-ville, Montréal H3C 3J7, CANADA

Address correspondence either to

Martin Arguin, Ph.D.
Département de psychologie
Université de Montréal
C.P. 6128, Succ. Centre-ville
Montréal, Qc
H3C 3J7
Fax: 514-343-5787
Phone: 514-343-2167
Email: martin.arguin@umontreal.ca

or to

Frédéric Gosselin, Ph.D.
Département de psychologie
Université de Montréal
C.P. 6128, Suc. Centre-ville
Montréal, Qc
H3C 3J7
Fax: 514-343-5787
Phone: 514-343-7550
Email: frederic.gosselin@umontreal.ca
ABSTRACT

The determination of the visual features mediating letter identification has a longstanding history in cognitive science. Many proposals have been made in this respect but none has yet been derived directly from empirical data. Here, we apply the Bubbles technique (Gosselin & Schyns, 2001) to reveal directly which areas of letters at five different spatial scales are efficient for the identification of lowercase and uppercase Arial letters. We provide the first empirical evidence that line terminations are the most important features for letter identification. We propose that these small features represented at several spatial scales help to discriminate among visually similar letters.
Considerable evidence indicates that words are recognized by letters rather than by the whole word shape (Legge, Mansfield & Chung, 2001; Paap, Newsome & Noel, 1984; Peroa & Rosa, 1995; but see Allen & Emerson, 1991; Hadley & Healy, 1991). Most convincingly, Pelli, Farell and Moore (2003) have shown that a word is unreadable unless its letters are separately identifiable. Using a noise masking procedure, they have demonstrated that the stimulus energy required for the recognition of a word increases linearly with the number of letters. Thus, whether a word can be identified is a straightforward function of the probability of identifying each of its individual letters. This is true even for the five most common three-letter words. Since these words have very familiar whole word shapes, the authors conclude that global shape exerts little or no impact on the probability of recognizing a word correctly. Given the above, it becomes obvious that a detailed knowledge of the mechanisms involved in visual letter identification is fundamental to our understanding of reading.

Echoing our introductory discussion about word recognition, a basic question about letter recognition is whether letters are perceived as global patterns or by features. A 45 year-old person, reading only one hour a day, will have identified more than 1,000,000,000 letters (Pelli, Burns, Farrell & Moore-Page, 2006). With this level of exposure, we would expect a high level of efficiency in letter recognition for literate human adults. To reach high levels of efficiency, perceptual strategy matters. In fact, using an ideal observer analysis, it is possible to demonstrate that the best solution – memory issues notwithstanding – to any object identification task is template matching rather than individual feature detection (Tjan, Braje, Legge & Kersten, 1995). Does the extensive practice of literate adults with letters lead them to use such an “optimal” strategy? Pelli et al. (2006) studied letter identification by measuring the contrast necessary for
identifying a letter embedded in visual noise. To isolate the visual constraints on letter identification, they compared human performance across different fonts (e.g. Helvetica, Sloan, Kunstler) and alphabets (e.g. Arabic, Chinese). They found that human efficiency was mainly explained by Attneave and Arnoult’s (1957) perimetric complexity (i.e. inside and outside perimeter squared divided by ‘ink’ area)—high complexity leading to low efficiency. This suggests that even though template matching may be the optimal solution in principle, normal human readers appear constrained to proceed by features, even for simple and highly trained stimuli such as letters. From this, the next question then becomes: What are the features upon which human letter recognition is based?

In the last three decades, insights about the features underlying letter recognition have come from experiments examining how recognition errors for each individual letter of the alphabet are distributed across the range of possible responses (i.e. the remainder of the alphabet; but see Petit & Grainger, 2002 for a different approach). Such confusability matrices however, can only be obtained under very unusual condition in order to elicit error rates that are sufficiently high (typically around 50%). For this purpose, some researchers have studied children who had not yet integrated the visual appearance of letters (Gibson et al., 1962). Others instead studied skilled readers but with target letters exposed very briefly (Townsend, 1971) or with extremely low contrast (Geyer, 1977). Sets of individual features have been proposed to predict these letter-confusability matrices (Briggs & Hocevar, 1975; Geyer & deWald, 1973; Gibson, 1969; Gibson et al., 1962; Laughery, 1971). In striking discrepancy with the findings of Pelli et al. (2006) reviewed above, template-matching models predict the outcome of letter confusability experiments better than feature-based models (Gervais, Harvey & Roberts, 1984; Holbrock,
An assumption that might resolve this lack of convergence is that the feature sets proposed by theories do not properly match those actually used by the human visual system. Indeed, such a mismatch would cause feature models to underperform in their prediction of human performance. This view highlights a fundamental weakness in the approach of using confusability matrices to indirectly reveal letter features; that is the vast gap between theories as to what constitutes a feature and the data used to test these theories. Here, we propose an alternative approach to determine more directly which components of letters drive their accurate recognition.

Recently, new tools referred to as classification image techniques have been developed to reveal aspects of a distal visual stimulus responsible for the measurable performance of observers in a specific task (e.g. Eckstein & Ahumada, 2002; Gosselin & Schyns, 2004). The underlying logic in all these methods is relatively simple. If specific visual information is necessary for the task at hand, depriving the observer of this information (using additive noise or a mask) will strongly impair his/her performance. In contrast, depriving the observer of non-diagnostic information will not substantially alter performance. For example, to correctly discriminate between an ‘F’ and an ‘E’, it seems obvious that the bottom part is necessary. Hiding this part with a mask should greatly impact the discrimination between the letters. Classification image techniques therefore allow a direct empirical examination of the diagnostic features used by human observers in letter identification.

Here, skilled adult readers attempted to identify letters randomly sampled at different spatial scales (see Fig. 1). Multiple linear regression was performed on the location of the samples and
accuracy scores to establish which regions of the stimulus mediate letter recognition, and at which spatial scales. This method, called Bubbles (Gosselin & Schyns, 2001; e.g. Adolphs et al, 2005; Nielson, Logothetis & Rainer, 2006; Smith, Gosselin & Schyns, 2006; Gibson et al., 2007), belongs to the general classification image approach. The results obtained using this direct and unbiased method were compared with the various features that have been proposed to explain letter recognition.

**General Methods**

*Participants.* Six graduate students from the Université de Montréal took part in this experiment. All had normal or corrected to normal acuity.

*Stimuli.* Stimuli were the 26 letters of the Roman alphabet displayed in lowercase (Arial 152 points) and uppercase (Arial 117 points), and printed in dark gray (luminance of 2.1 cd/m$^2$) over a light gray background (luminance of 57.3 cd/m$^2$). On average, both lowercase and uppercase letters horizontally subtended 1.35 degree of visual angle (64 pixels wide by 84 pixels high for uppercase, 64 pixels wide by 99 pixels high for lowercase). To reveal the visual features diagnostic for letter identification, we used Bubbles (Gosselin & Schyns, 2001). In a Bubbles experiment, stimuli information is randomly sampled, and multiple linear regression is performed on the samples’ locations and corresponding accuracy scores to reveal which parts of the stimuli, on the dimensions that were sampled, are correlated with performance. Here, we sampled letter stimuli in image space (x, y coordinates) and at varying spatial scales to uncover which letter parts are most correlated with letter identification.
The steps involved in the creation of an experimental stimulus were as follows: A letter stimulus is decomposed in five spatial frequency bands (128–64, 64–32, 32–16, 16–8, and 8–4 cycles/image or 32–16, 16–8, 8–4, 4–2, 2–1 cycles/letter; the remaining bandwidth served as constant background; see Fig. 1, first row), using the Laplacian pyramid (Burt & Adelson, 1983). The letter information at the five scales is then sampled using an opaque mask punctured by randomly located Gaussian holes (henceforth called “bubbles”) to avoid introducing spatial frequency artefacts. The size of the bubbles is adjusted according to frequency band to reveal 1.5 cycles of spatial information (Fig. 1, second row). Because the size of the bubbles increased as the spatial scale became coarser, the number of bubbles differed at each scale to keep the size of the sampled area constant across frequency bands. Finally, the five frequency bands of the letter stimulus partially revealed by the bubbles are fused to produce an experimental stimulus (Fig. 1, third row). For each letter stimulus, the total number of bubbles was adjusted on a trial-by-trial basis to maintain a correct identification rate of 52 percent (approximately halfway between chance level—3.85 percent—and perfect identification, which is optimal for the multiple linear regression).

Procedure. Each participant completed 100 blocks of 260 trials for each case, for a total of 52,000 trials per participant. Such a large number of trials is necessary to enable reliable
statistical inference about the correlations between letter parts and accuracy. The uppercase and lowercase letters were presented in separate blocks, and each participant completed all the blocks in a particular case before moving to the other case. Half the participants began with the lowercase letters.

Each trial began with a fixation cross displayed at the center of the screen for 494 ms. It was then immediately replaced by a “bubblized” letter (see the Stimulus in Fig 1 for an example), which remained on the screen for 200 ms. The participant had to identify the letter and then press the appropriate key (e.g. press the ‘a’ key when the letter is a or A) on a keyboard. No feedback was provided. The next trial started approximately 500 ms (i.e. a rough estimate of the time it took to compute the next stimulus) after the participant’s response.

The stimuli were displayed on a 21-inch monitor set with a refresh rate of 75 Hz and calibrated to allow a linear manipulation of luminance. The experiment was run on a PC-Pentium IV computer. The experimental program was written in Matlab, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The viewing distance was maintained constant at 107 cm using a chinrest.

**Results and Discussion**

On average, participants needed 54.2 bubbles to maintain their performance at 52% correct with lowercase letters and 30.9 bubbles with uppercase letters (see table 1 for the number of bubbles required for each letter on average). To pinpoint the features that different observers used to discriminate letters, we performed a least-square multiple linear regression on the bubble masks
and accuracy data. The plane of regression coefficients yielded by this operation is called a classification image. Here, a classification image is obtained by computing a Correct Plane by summing all the bubble masks (see Fig. 1, second row) that led to a correct answer, and an Incorrect Plane by summing all the bubble masks that led to an incorrect answer; and by subtracting the Incorrect Plane from the Correct Plane. One such classification image was computed per letter, per case and per frequency band. To estimate the mean and the standard deviation of the distribution of the null hypothesis (i.e., no correlation between accuracies and sampled stimulus information), we repeated this procedure on permuted accuracies. This mean and this standard deviation were used to calculate the Z scores of the classification images. To determine the letter information significantly correlated with accuracy, we applied the Pixel test to the Z-scored classification images ($p < .01$). The statistical threshold provided by this test corrects for multiple comparisons while taking the spatial correlation inherent to structured images into account (Chauvin et al., 2005). Fig. 2 shows this effective information, that is the statistically thresholded classification images, for all lowercase and uppercase letters, superimposed to the appropriate letter information to help with interpretation.

A first glance at the leftmost columns of Fig. 2a and 2b may give the impression that letters are almost completely revealed. However, this is far from being the case. Only 32% and 24% of uppercase and of lowercase letter ink area is depicted, respectively. The impression of
completeness is due to the fact that what is revealed are the most informative regions of letters. We believe that this is a rather impressive demonstration that the experiment has indeed succeeded in revealing the letter regions that effectively drive recognition performance (see also Fig. 3).

A comparison of the number of pixels that are significantly useful for letter identification at each frequency band reveals a clear advantage for the information between 2 and 4 cycles per letter (see Fig. 4). This analysis was performed by calculating the proportion of significant pixels that fell on each frequency band. Note that only the pixels falling directly on letter “ink” were included in this analysis and all those that follow.

The classification images allowed us to re-evaluate the various proposals regarding potent letter features using our classification images. The letters were decomposed 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. We also included
terminations, a feature that had not been previously considered. We created, for each letter of the alphabet and each case, 213 masks containing the following 10 features: vertical, horizontal, slant tilted left or right, curves opened up, down, left or right, terminations, and intersections. For example, the uppercase letter ‘A’ was decomposed in eight masks that is, one slant tilted left, one slant tilted right, one horizontal, two terminations, and three intersections. 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. For example, in the uppercase letter ‘A’, the two slants and the horizontal did not contain the pixels of the three intersections and the two terminations. These proposed features make no explicit commitment to spatial frequencies. To compare them directly with the classification images, we thus disposed of the spatial frequency dimension by collapsing it prior to smoothing, thereby retaining only the $x, y$ dimensions. We thresholded these bidimensional classification images to keep only the portions of the stimulus corresponding to the highest five percent of the regression coefficients. We then calculated the proportion of the total number of these pixels that fell on each feature and divided this proportion by the total number of pixels in that feature, thus normalizing for feature size (not adjusting for feature size only amplifies the effects reported below). We conducted this analysis for each letter separately and then combined the results by summing the proportions calculated for each feature class across the 26 letters of a case, and subsequently dividing these summed proportions by the number of occurrences of the feature class within the alphabet. In order to reveal the relative importance of the features for the correct identification of letters, the grand sum across feature classes was normalized to 1. The relative importance of each feature class thus obtained is summarized in
Figure 5. Terminations are, by far, the most important features for both letter cases, with scores of .30 and .35 for upper- and lowercase letters, respectively. In fact, terminations are respectively 1.5 and 1.8 times more important than horizontals, which constitute the next most important feature class for both upper- and lowercase letters.

To compare the features used by human observers for letter identification to those that would be obtained assuming an optimal use of diagnostic information for letter identity, we built an ideal observer model, which uses all the visual information available for uppercase and lowercase letter identification. The ideal observer was submitted to the same experiment as the human participants and performed the same number of trials per letter as human participants. The number of bubbles for each letter was set to the average of the number of bubbles used by the human participants for each letter independently. An adjustable quantity of white Gaussian noise was added to the letters prior to sampling them with Gaussian apertures to equate human and model performance (i.e. 52 % accuracy for each letter). On each trial, the model determined the Pearson correlation between the sparse input (i.e. the noisy letter revealed by bubbles) and each of the 26 possible original letters of a case revealed by the same bubble mask. The categorization response was the letter with the highest correlation with the stimulus.

The results of the feature analysis of the ideal observer are presented in Figure 5 along with those of the human feature analysis. Interestingly, the profile of usefulness of the different feature
classes is flatter for the ideal observer than for humans. The difference between the first and second most important features for the ideal observer is of 2% for both lowercase and uppercase letters whereas it is of 16% and of 10% for lowercase and uppercase letters, respectively. Most importantly, the usefulness of the terminations is much lower for the ideal observer (ranked 5th and 6th out of 10 for lowercase and uppercase letters, respectively) than for human observers. This indicates that the importance of the terminations for human participants is largely attributable to constraints imposed by properties of the human visual system rather than by constraints that are exclusively determined by the stimulus set.

**General Discussion**

We used *Bubbles*, a classification image technique, to reveal the letter areas responsible for accurate letter identification. Although this method has been efficiently applied in the face recognition field, it had never been applied to letter identification. Figure 4 clearly shows that the 2-4 cycles per letter frequency bandwidth conveys the most potent visual information. This is congruent with results obtained by others (Ginsburg, 1980; Legge et al, 1985; Parish & Sperling, 1991; Solomon & Pelli, 1994; Chung et al., 2002; Majaj et al, 2002) who found that letters subtending 1.35 degrees of visual angle, as in the present study, are optimally masked by noise containing spatial frequencies around 3.1 cycles per letter. However, the 260 classification images of Fig. 2 tell a more subtle and complete story. For example, the uppercase letters ‘C’ and ‘G’ share three curves that have convexities directed upwards, to the left and downwards. The classification image shows that one of these features (the one opened right) is efficiently processed in low-spatial frequency bandwidths. But this is not the end of the story. The identification of these letters also requires small features represented throughout several spatial-
frequency bandwidths that is, intersection and horizontal bar for ‘G’, and lower part termination for ‘C’. All spatial frequencies, including higher spatial frequencies, are necessary to resolve the edges of these small features. In fact, in our determination of the relative importance of letter features, we found that relatively small terminations were the most distinctive (Fig. 5). For example, the inferior termination of the uppercase letter ‘C’ clearly allows the discrimination of this letter from the uppercase letters ‘G’, ‘Q’, and ‘O’, and is in fact sufficient for its correct identification.

A similar feature analysis was performed on the classification images of the ideal observer (see Figure 5). Terminations ranked 5th and 6th out of the 10 potential features considered, for lowercase and uppercase letters, respectively, which is markedly different from human observers. The stimulus areas effective for the ideal observer are those that are most diagnostic for the unique identification of the target letter—i.e. for its discrimination against the remainder of the alphabet—without any other constraint applying. The divergence in the effective features for the ideal observer and humans is thus (at least partially; see below for another potential determinant) attributable to the fact that, on top of the need to use features that can indeed discriminate among letters, feature use by humans is also determined by the constitution and organization of their visual system. This suggests that, for instance, the great importance that terminations hold for human participants rests on an interaction of the relative diagnosticity of this feature and a disproportionately strong disposition of the human visual system to encode it (relative to other features). Some neurons in the primary visual cortex of monkeys respond to terminations that is, when a properly oriented line-end is centered in the receptive field but not when a line of edge extends across it. These are a subset of the cells with strongly end-stopped receptive fields.
(hypercomplex cells; Hubel and Wiesel, 1968). Moreover, it is believed that a number of these special cells converge to single neurons in V2 (von der Heydt & Peterhans, 1989), providing a possible early mechanism for letter identification. Just as significantly, the poor use by humans of features such as verticals and curves opened up, which are highly effective for the ideal observer (see Fig. 5), suggest that the human visual system may be poorly equipped to process such features.

The analysis by feature conducted in this paper is the first, to the best of our knowledge, to demonstrate the crucial importance of terminations for letter identification in humans. Historically, Gibson (1969) had suggested that discontinuities were important for letter recognition, and terminations can be construed as discontinuities, but her proposal remained rather vague and not substantiated empirically.

Why are small features such as terminations so important for letter identification in humans? Terminations are clear discontinuities in bars and curves and thus provide reliable information about the absence of intersections – also called coterminations in the object recognition literature – in the target letter. An informal examination of other fonts suggests that the presence and the relative location of terminations and intersections might be ‘font invariant’ properties of letters. In fact, it is relatively obvious that we can create an ‘A’ without any slant. However, the two central intersections and the two terminations are critical properties of the letter ‘A’. This analysis suggests that a novel font would remain identifiable as long as a subset of these small features is available for visual extraction. It also points to another possible determinant of the divergence in feature use by humans and the ideal observer discussed above; namely the invariant diagnosticity
of terminations across fonts, to which our ideal observer was obviously insensitive. Previous attempts to improve reading speed in individuals with low-vision by bandpassing word images in the mid to high spatial frequency range led to equivocal results (e.g. Fine & Peli, 1995). These failures may be attributable to the fact that, beyond sheer energy enhancement, the diagnosticity (with respect to letter identity) of the visual features of letters in the critical mid to high spatial frequency range was not improved by this manipulation. The results presented in this article should allow the creation of a font where the diagnosticity of the most effective features for the identification of letters is enhanced. It remains to be seen whether such a font would lead to faster letter recognition and in turn to faster word recognition in normal readers and individuals with letter-by-letter dyslexia.
ACKNOWLEDGEMENTS

We thank the subjects 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 post-doctoral scholarship from the Fonds Québécois de Recherche en Nature et Technologies (FQRNT) to Daniel Fiset; and by a FQRNT graduate scholarship to Caroline Blais.
REFERENCES


Table 1. Average number of bubbles required to maintain performance at 52% correct at the end of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>Uppercase</th>
<th>Lowercase</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16</td>
<td>45</td>
</tr>
<tr>
<td>B</td>
<td>38</td>
<td>44</td>
</tr>
<tr>
<td>C</td>
<td>42</td>
<td>84</td>
</tr>
<tr>
<td>D</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>E</td>
<td>37</td>
<td>54</td>
</tr>
<tr>
<td>F</td>
<td>41</td>
<td>55</td>
</tr>
<tr>
<td>G</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>H</td>
<td>33</td>
<td>55</td>
</tr>
<tr>
<td>I</td>
<td>21</td>
<td>133</td>
</tr>
<tr>
<td>J</td>
<td>32</td>
<td>45</td>
</tr>
<tr>
<td>K</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>L</td>
<td>33</td>
<td>128</td>
</tr>
<tr>
<td>M</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td>O</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>P</td>
<td>45</td>
<td>44</td>
</tr>
<tr>
<td>Q</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>R</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td>S</td>
<td>32</td>
<td>57</td>
</tr>
<tr>
<td>T</td>
<td>32</td>
<td>61</td>
</tr>
<tr>
<td>U</td>
<td>34</td>
<td>64</td>
</tr>
<tr>
<td>V</td>
<td>23</td>
<td>45</td>
</tr>
<tr>
<td>W</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>X</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Y</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>Z</td>
<td>23</td>
<td>37</td>
</tr>
</tbody>
</table>
Figure legends

**Fig. 1.** Illustration of the stimulus-generation process: Each original letter was first decomposed into five spatial frequency bandwidths of one octave each (top row). Each bandwidth was then independently sampled with randomly positioned Gaussian windows (i.e. bubbles), thus revealing sparse information at each spatial frequency bandwidth (middle row). The sum of information samples across scales produced a “bubblized” letter (bottom row).

**Fig. 2.** The first column of the left and right halves, respectively, show all Arial letters in lowercase and in uppercase revealed by their effective filters. The next five columns display in red the significant pixels for each spatial frequency bandwidth from fine to coarse.

**Fig. 3.** Oscar Wilde’s famous quote “The truth is rarely pure and never simple.” revealed by effective (p<.05; see Fig. 2) and by ineffective filters. Both filter types show the same ink area but at the two tails of the distribution.

**Fig. 4.** Relative use of the five sampled spatial-frequency bandwidths for Arial letter recognition in uppercase and in lowercase.

**Fig. 5.** Relative use of ten letter-feature classes for Arial letter recognition in uppercase and lowercase in humans (gray) and in ideal observer (black) using image location x scale samples. The letter-feature classes are ordered from least (left) to most (right) useful to humans.
Features for letter identification
The truth is rarely pure and never simple.
Features for letter identification