Title
Feature contingencies when reading letter strings.

Authors
Coates, Daniel R
Bernard, Jean-Baptiste
Chung, Susana TL

Publication Date
2019-03-01

DOI
10.1016/j.visres.2019.01.005

Peer reviewed

Vision Research 156 (2019) 84–95

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Feature contingencies when reading letter strings

Daniel R. Coates[a,*], Jean-Baptiste Bernard[b], Susana T.L. Chung[b,c]

[a] College of Optometry, University of Houston, United States
[b] School of Optometry, University of California, Berkeley, United States
[c] Vision Science Graduate Group, University of California, Berkeley, United States

ARTICLE INFO

No. of reviewers - 2

Keywords:
Crowding
Letter recognition
Peripheral vision
Redundancy masking
Feature migration

ABSTRACT

Many models posit the use of distinctive spatial features to recognize letters of the alphabet, a fundamental component of reading. It has also been hypothesized that when letters are in close proximity, visual crowding may cause features to mislocalize between nearby letters, causing identification errors. Here, we took a data-driven approach to investigate these aspects of textual processing. Using data collected from subjects identifying each letter in thousands of lower-case letter trigrams presented in the peripheral visual field, we found characteristic error patterns in the results suggestive of the use of particular spatial features. Distinctive features were seldom entirely missed, and we found evidence for errors due to doubling, masking, and migration of features. Dependencies both amongst neighboring letters and in the responses revealed the contingent nature of processing letter strings, challenging the most basic models of reading that ignore either crowding or featural decomposition.

1. Introduction

Visually discriminating closely-spaced items is a challenging task that is accomplished seemingly effortlessly in everyday life. A paradigmatic example is reading, which involves the recognition of words comprising groups of letters that are often closely spaced together. A variety of theories exist to describe the mechanisms involved in processing complex visual information like text, generally proposing that information is processed independently and in parallel amongst constituent items such as the letters in a word. Here, we focus on characterizing the interactions and dependencies between adjacent letters and letter parts, believing these to be significant causes of recognition errors, which could in turn slow down reading.

Many theories of object recognition posit the use of spatial "features" for recognition. In the most basic formulation, features are simply spatial parts of compound objects like letters—for example the dot in a lower-case "i." These parts are distinctive, constituting the fundamental elements of object recognition processes. Items are recognized based on their constituent features, most likely in a parallel fashion for highly learned stimuli such as the arrays of letters that comprise words (Pelli, Burns, Farell, & Moore-Page, 2006).

Keen observations of the interaction effects between adjacent letters date back at least to Korte (1923). In this prescient treatise, Korte describes how letters and letter parts interact in a myriad of ways, such as migrations of features between letters, doubling of parts, and feature losses, as translated and summarized by Strasburger (2014). However, these observations were based on subjective verbal reports from subjects, lacking direct empirical support. These ideas are often invoked in a general sense by the contemporary literature on *crowding* (Pelli, Palomares, & Majaj, 2004; Levi, 2008), which is the modern term for interference effects from neighboring items, a phenomenon that worsens with increasing distance from the fovea (Bouma, 1970; for reviews, see Levi, 2008 or Pelli, 2008). Crowding is proposed to be a fundamental bottleneck limiting vision.

A stereotypical description of a crowded percept is one that has parts that are "jumbled" (Pelli et al., 2004; Levi, 2008). Details are present, but they cannot be assigned to the correct items, and have an indeterminate character. This description is in clear contrast to the deleterious effect of blur, which renders an object and its parts difficult to resolve. The proposed theoretical underpinning of the jumbled percept is a two-stage process involving first detection of features, and then their integration, which in crowded conditions may be compromised (Pelli et al., 2004). In the faulty over-integration hypothesis, identification errors are due to integration zones that are erroneously larger than a single item, causing features from adjacent items to become intermixed and leading to confusion about which features are in the target and which are in the flankers.

Despite the intuitive appeal and the widespread reference to the

* Corresponding author.
E-mail address: drcoates@central.uh.edu (D.R. Coates).

https://doi.org/10.1016/j.visres.2019.01.005
Received 4 July 2018; Received in revised form 5 January 2019; Accepted 9 January 2019
Available online 05 February 2019
0042-6989/

two-stage model, there is surprisingly scant experimental evidence supporting feature interactions, and some researchers have questioned the need for the feature abstraction. Since the early phenomenological descriptions of Korte (1923), only a few studies have sought to directly demonstrate feature migrations, or the spatial transfer of object parts to neighboring items. Wolford and colleagues performed the seminal experiments in a series of studies (Wolford, 1975; Wolford & Shum, 1980). One study used symbols containing features that comprised well-defined features of "tick" marks positioned on boxes in an array, and found positive evidence for feature perturbations, including an important directionality of errors *towards* the fovea.

However, these results have been called into question by Butler and colleagues (Butler & Morrison, 1984; Butler, Mewhort, & Browse, 1991), who ascribed the errors to migration of whole units, or guesses based on other members of a limited response set. Other authors have also noted that the results might be explained more parsimoniously by the migration of entire items (Hanus & Vul, 2013), based on a faulty location signal or "local sign," (Chung & Legge, 2009) or inadequately focused spatial attention (Strasburger, 2005). Resolving this debate is crucial for understanding precisely how object parts in arrays of items, such as letters in a word, are processed.

Analyzing the errors made while perceiving trials with multiple letters can provide the definitive evidence to answer this question. Letters constitute the ideal stimuli for this question, since they contain multiple parts, are highly learned, and are sufficiently heterogeneous to reveal complex processes involved in discrimination. Most crowding paradigms require that the subject report only a single item, such as the middle letter of a group, although there are several exceptions (Ağaoğlu & Chung, 2016; Harrison & Bex, 2015; Zhang, Zhang, Liu, & Yu, 2012), which have identified errors and mislocalizations at the level of entire characters (or single-feature items). To fully understand how features may interact, it is critical to understand the perception of each item in the presented array with a full-report response, such as in the present study. Otherwise, it is impossible to precisely determine the fate of parts in all positions.

Most recent studies of interactions between proximal items have typically been performed with elementary stimuli such as Gabors (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), T-like stimuli (Greenwood, Bex, & Dakin, 2009), oriented "clock-face" stimuli (Ester, Klee, & Awh, 2014; Ester, Zilber, & Serences, 2015), and rotated "C"s (Ağaoğlu & Chung, 2016; Harrison & Bex, 2015; Harrison & Bex, 2017). Importantly, in these studies stimuli varied along a single dimension, with one or two components that are present in every stimulus. In these cases, models based on "pooling" (Parkes et al., 2001), "averaging" (Greenwood et al., 2009), or the combination of proximal feature values in a population code (Harrison & Bex, 2015; van den Berg, Roerdink, & Cornelissen, 2010) have successfully captured the influence of neighboring items. Letter recognition, however, likely involves the detection and integration of features that are *binary* (elements may or may not be present in a stimulus) and that assume a variety of complex spatial relationships, making the relevance of pooling or averaging in letter recognition uncertain. It is difficult to apply the pooling and averaging theories to explain letter recognition errors. For instance, what would pooling or averaging predict as errors for "**b**" and "**j**," which each contain multiple non-overlapping features, such as an ascender and a round part, versus a descender and a dot?

One study (Põder & Wagemans, 2007) used stimuli comprising conjunctions of features on three independent feature dimensions (spatial frequency, orientation, and color), finding that flankers biased target responses, in a way consistent with feature and object mislocalizations and feature pooling. While this paradigm may more closely mimic letter recognition, the independence of letter features has been questioned (Townsend, Hu, & Evans, 1984), a topic returned to at length in the Discussion. Also, item composition based on three ever-present feature dimensions differs from the more variable nature of letter features. Furthermore, as stated previously, the single-item report paradigm precludes understanding of all aspects of multi-item recognition, such as whether features truly "migrate," rather than simply being doubled, and whether certain distinctive features are always noticed, even if mislocalized amongst flankers.

Determination of the distinctive spatial features of letters (even in an isolated context) also has a long history (for a review, see Grainger, Rey, & Dufau, 2008). The most well-known proposals for letter features are the simple line segment-based decomposition of capital letters, originating with Selfridge's Pandemonium model (Selfridge, 1958) and tested psychophysically by Gibson (Gibson, 1969); as well as the more complex analysis of lower-case letters by Bouma (1971). These two proposals, like most of the subsequent refinements, were based on patterns determined from the confusion matrices of letter identification experiments. Here, we continue this approach and extend the investigations of Bouma (1971) by showing how the lowercase letter features he proposed interact amongst displays comprising three adjacent letters (letter trigrams). Like Bouma, we use peripheral presentation to induce maximal errors, which also strengthens the effects of crowding between the letters, and provides insight about the factors limiting peripheral reading.

We performed a detailed analysis of the errors from five subjects reporting their full percepts of ten thousand peripheral trigrams, investigating at both a letter level and a feature level. To anticipate our results, we found strong evidence for feature-based interactions that defy simple letter-based models. The interactions revealed dependencies between items, such as a balanced loss and gain of spatial features, arguing against simple models based on independent recognition of each letter. Finally, we introduce several characteristic principles of letter feature interactions in peripheral letter trigrams: (1) features absent from a trigram are rarely "imagined", (2) presence of a feature category in a trigram is seldom entirely missed, (3) for adjacent letters with identical features, feature instances are often lost due to *redundancy masking* (Yildirim, Coates, & Sayim, 2019), and (4) both feature migrations and feature "doubling" may occur when features migrate between items.

Taken together, these results place strong constraints on the mechanisms involved in processing letter arrays, and help inform sophisticated models of both reading and crowding with realistic compound items like alphabetic letters.

2. Methods

2.1. Experimental details

Five subjects with normal or corrected-to-normal vision, aged 18–20, participated in this study. All subjects gave their oral and written consent before the commencement of data collection. This research was approved by the Institutional Review Board at the University of California, Berkeley, and was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Testing was performed binocularly in a dimly-lit room (standard office room light turned off).

Subjects viewed three-letter trigrams (random sequences of three letters) presented at 10° below a small fixation target in the lower visual field. We refer to these three items as the left, middle, and right letter (respectively), and refer to the left and right letters as the "outer letters."

A small square was presented as the fixation target at the center of the monitor and at the eye-level of each subject. An Eyelink II eye-tracker (SR Research Ltd.) was used to ensure subjects' fixation. Before each block of trials, the standard 9-point calibration routine was performed to calibrate subjects' eye positions, followed by a drift-correct. The experiment only proceeded after a successful calibration as defined by the Eyelink software.

On each trial, before a trigram stimulus was presented, subjects' eye positions while fixating at the fixation target were compared with calibrated values obtained during calibration. If the eye positions deviated from the fixation target by more than 1° in any direction, the trigram stimulus would not be presented and an audio tone alerted the subject to refixate properly. Trigrams, comprising lower-case letters rendered in Arial font, were presented for 50 ms or 200 ms, within separate blocks. Letter size had an x-height of 1.2°, with a center-to-center letter separation of 1.3° (despite the fact that Arial is a proportional-width font). This letter size was chosen such that subjects'overall performance for identifying letters was around 50%, which yielded a good number of error trials for analyses while keeping subjects sufficiently motivated for the task. The task of the subjects was to type the identities of the three letters (from left to right) using the keyboard.

Stimuli were generated using MATLAB (Mathworks, MA) and Psychtoolbox (Brainard, 1997; Pelli, 1997) on a MacBook Pro computer, and were displayed on a Sony GDM-F500R monitor at a resolution of 0.029 cm/pixel. Randomization was balanced such that each of the 26 letters of the alphabet appeared in the middle position exactly 200 times for each of the two durations, resulting in 10,400 trigrams over the course of the experiment for each subject. Less than 1% of trials were discarded due to eye movements or invalid (non-letter) responses.

2.1.1. Perimetric complexity for letters

To quantify the gain or loss of spatial parts of letters, we used a standard measure of the spatial content of each letter, the perimetric complexity (Attneave & Arnoult, 1956; Pelli et al., 2006), defined as the perimeter squared divided by the ink area. This measure has been found to be highly correlated with other measures of the spatial complexity of letters, such as the length of each letter skeleton (Bernard & Chung, 2011). Table 1 lists the perimetric complexity for each of the 26 letters.

2.1.2. Bouma (1971) letter features

To investigate the interactions of letter parts, we used features based on those defined by Bouma (1971). Bouma identified features based on letter confusions observed when subjects viewed isolated lower-case letters in the IBM "pingpong ball" Courier font. In his study, observers viewed letters in two conditions: eccentric (7°) and foveal (at a far distance), with little difference observed between errors in the two conditions. Table 2 shows the features present in each letter, for the five feature sets we used: descenders, ascenders, oblique, round, and arch. Note that a letter can belong to multiple sets (contain multiple features), such as "**b**," which has both an ascender and a round part. Also, not all letters are represented with this set of features, such as "**a**,**e**,**s**." Our goal was not to offer a comprehensive model of letter recognition in trigrams, but rather to show the dependencies between adjacent letters and features of letters. Furthermore, we are mostly interested in analyzing error patterns, rather than characterizing when trigrams are correctly identified.

2.2. Data analysis

As letter confusion data is heterogeneous and highly non-Gaussian, we relied heavily on Monte Carlo strategies to characterize statistical significance. Two simulations encoded constraints at different levels of analyses, in order to determine the magnitude of error patterns that could happen by chance. The simulations are described below, one generating simulated responses at the letter level (Section 2.2.1), and one generating random feature sets (Section 2.2.2).

2.2.1. Letter level independence model from confusion matrix simulations

The first simulation, at the letter level, was used to test hypotheses about independence between recognition of the three letters in the trigram (Section 3.1). To quantify dependencies between the letter positions, we tested the various analyses against simulated responses generated with the assumption of independence between the letter positions. That is, random letter responses were derived from the corresponding positional confusion matrices alone, independently of neighbors. This constitutes a baseline "null-model" with which to test hypotheses about whether recognition of a single letter is modulated by neighboring letters.

Specifically, we determined separate confusion matrices for each of the three letter positions (left, middle, right). Then, for each trial we generated a synthetic response using independent random draws from the corresponding distributions (including correct identification) for each of the three presented letters. Repeated simulations (representing one thousand complete experiments) were used in subsequent analyses throughout the paper.

2.2.2. Feature-level random letter sets to assess feature robustness

Section 3.2 describes different error patterns that suggest letter feature processing, such as the likelihood of an erroneous migration of a "descender feature" between a flanker and the middle letter. As the letter feature assignments are somewhat arbitrary, it is important to determine how likely the effects observed could be due to chance, rather than being characteristic interactions arising from the particular sets of letters sharing a feature. To determine chance levels, we generated random letter sets of the same cardinality as each feature set, and re-ran our analyses on the empirical data with one thousand random sets to assess the "significance" of each observed effect.

For example, for the descender set ("**g j p q y**"), one thousand random five letter sets were generated. The idea is that a completely random letter set (such as the meaningless set "**x q l s v**") should not exhibit any distinctive characteristics. To contrast with the descender set (which has a high incidence of flanker-middle mislocations), there should not be a particularly high incidence of migrations between those in the meaningless set (which presumably do not share a "feature"). Analysis using the empirical data and one thousand five letter sets provide a baseline for the likelihood of each statistical effect (here flanker-middle mislocation) by chance in arbitrary letter sets, and effects strongly outside those in the random range can be attributed to featural interactions. We simulated letter sets of size five for descenders and arch letters, size six for oblique letters, size seven for round letters, and size eight for ascender letters.

3. Results

Our results can be divided into two broad categories. First, we present analyses that were performed at the letter level, including characterization of letter transpositions, dependencies between letter positions, as well as the influence of letter complexity. After these, we show how decomposing letters into the features proposed by Bouma (1971) reveals evidence suggestive of featural effects that have previously been assumed, such as feature migrations between letters, and introduce several new principles governing feature interactions in crowding.

Table 1
Perimetric complexity for Arial letters.

a	64.9	**b**	65.0	**c**	50.3	**d**	65.5	**e**	61.7	**f**	46.7	**g**	76.0	**h**	64.0	**i**	40.6
j	50.5	**k**	55.9	**l**	36.4	**m**	85.6	**n**	57.2	**o**	50.2	**p**	64.4	**q**	65.5	**r**	35.8
s	54.9	**t**	42.6	**u**	56.7	**v**	47.4	**w**	83.1	**x**	43.7	**y**	57.0	**z**	56.1		

Table 2
Feature decomposition, inspired by Bouma (1971). "**j**" = descender letters, "**l**" = ascender letters, "**x**" = oblique letters, "**o**" = round letters, "**n**" = arch letters.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
j							**g**			**j**						**p**	**q**								**y**	
l		**b**		**d**		**f**		**h**	**i**		**k**	**l**								**t**						
x											**k**											**v**	**w**	**x**	**y**	**z**
o		**b**	**c**	**d**			**g**								**o**	**p**	**q**									
n								**h**					**m**	**n**				**r**			**u**					

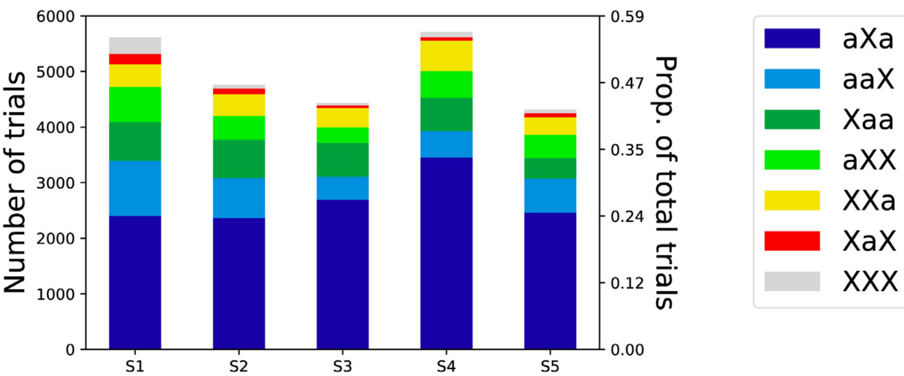

Fig. 1. Summary of errors for each subject. In the figure legend, an "a" indicates a correct answer in the corresponding position (order: left, middle, right), while an "X" indicates an error in that position. Incorrectly identifying the middle letter and correctly identifying the two outside letters was by far the most common error for all subjects, constituting nearly half of the error trials for each subject (or approximately a quarter of the total trials for each subject). Total number of trials are shown on the left axis, while the proportion of total trials are shown on the right axis. Each subject made at least one error on 42–55% of trials.

3.1. Letter level analysis

3.1.1. Gross error counts

Fig. 1 shows the counts of the different types of overall errors for each subject. Subjects made at least one error in 42–55% of trials. By far the largest type of error that occurred was the error in which the two outside letters were reported correctly, but the middle letter was misidentified. This is consistent with conventional observations of errors in crowded letter strings (Bouma, 1973) and accounted for nearly half of the total errors for each subject (or 22–33% of all trials for each subject). Trials in which only the left or right letter was misidentified occurred approximately equally (~6% of all trials). Two-letter errors with adjacent letters constituted approximately 3% of all trials, while two-letter errors with only the outside letters occurred less than 1% of the time. Finally, three-letter errors were rare (~1% of all trials).

Results from the two stimulus durations (50ms and 200ms) were not considerably different. The proportion of correct trials for the five subjects were 43%, 50%, 53%, 40%, 55%, respectively, for stimuli presented for 50 ms, and 48%, 58%, 61%, 50%, 60% for stimuli presented for 200 ms. Therefore, results for the two durations were combined for all analyses presented here.

3.1.2. Letter error confusion matrices

To implement the independence model, we computed individual confusion matrices for each subject at each letter position. The three positional confusion matrices, aggregated over all subjects, are shown in Fig. 2. As described earlier, rows from each confusion matrix defined the distributions that were used to generate synthetic responses for each letter in each position, including correct responses.

The diagonals, indicating correct identification of each letter, are not easy to distinguish in confusion matrices, so we have plotted the proportion correct for each letter in each position separately in Fig. 3, which also shows individual data. While there are some differences between the letters identified correctly for each subject, subjects were remarkably consistent in which letters were identified most easily ("**z**" and "**j**") and which letters were least correctly identified ("**t**" and "**i**" for example). There is some degree of asymmetry in identification rates ("**c**" has more errors in the left position, for example), which we discuss later.

3.1.3. Letter transpositions

The transposition of adjacent letters has been found to explain a significant portion of errors in crowding (typically 10-40% of errors), and are thought to be indicative of non-stimulus-related factors such as erroneous location coding (Chung & Legge, 2009). In the present experiment, however, transpositions of letters happened infrequently. We define letter transpositions simply as those trials in which responses for a pair of adjacent letters were exactly switched. For example, a trial with stimulus "**xdj**" and response "**xjd**" is classified as a transposition between the middle letter and the right letter. The middle letter was switched with one of the outer letters in only approximately 0.8% of all trials, with a roughly equivalent prevalence for switches to the left and right. Thus, approximately 1.6% of all errors were due to a letter mislocalization, much smaller than typically found.

If the source of these errors was simply an erroneous location signal, there should be no identifiable relationship between the transposed letters. That is, transpositions should involve random letter pairs, independent of the identity of the letters. We tested this in two ways. First, we analyzed the distributions of letters that were transposed, and second, we analyzed the transpositions simulated by the independence model. The distributions of transposed letters in the empirical data appeared highly non-uniform, suggesting a dependence on particular letters and letter pairs.

In the simulations, we found that transposition errors occurred much less frequently than in the empirical data, approximately one tenth of the occurrence (<0.1% of all simulated trials). Thus the empirical transpositions did not simply occur "by chance" from independent confusions in the two letters. Also, the simulated transpositions happened most frequently for the letters with the highest overall confusion rates, unlike in the empirical data. Taken together, these results show that transpositions were neither independent of the particular letters that were switched, nor did they simply result from the overall error susceptibility of the individual letters.

3.1.4. Error rate contingency on neighbor identification

Since the likelihood of transpositions was dependent on the identities of each element in an adjacent pair, this suggests that the overall accuracy rate for a letter might be correlated with the accuracy in identifying a neighboring letter. Indeed, we found that the probability of error occurrence at a given position was dependent on whether an error occurred in a neighbor.

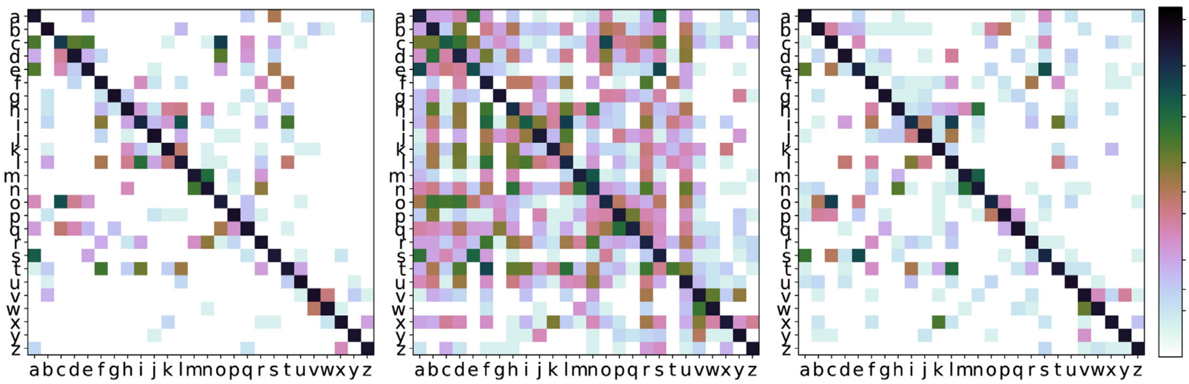

Fig. 2. Aggregate confusion matrices for left, middle, and right letters, respectively. Cells are colored by occurrence from light/pastel to dark (see the color bar on the far right for values of occurrence). Letters presented are shown in rows while responses are shown in columns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

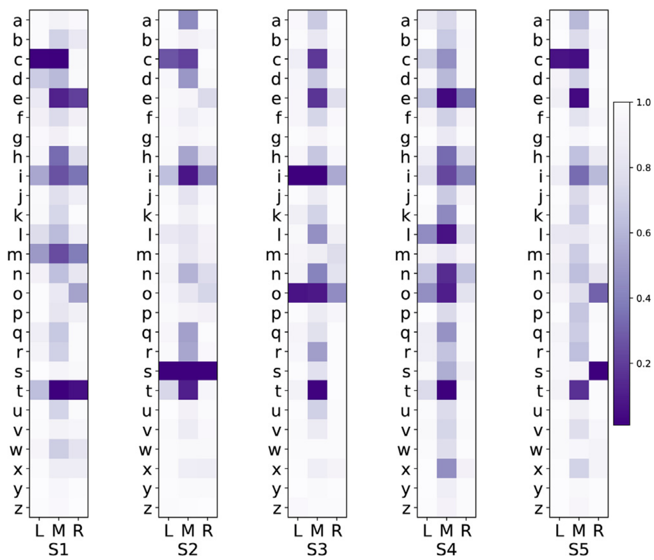

Fig. 3. Proportion correct for each subject identifying each of the 26 letters presented in each position. Proportion is colored as indicated by scale bar, with darker colors indicating worse performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3 shows the complete set of conditional probabilities for middle/outer pairs, for each subject individually as well as for the aggregated data. Each pair of columns shows the identification error rate of a letter position conditioned on whether the neighbor was erroneous, and when the neighbor was correctly identified. The main finding is that for all pairs of adjacent letters in every subject, the incidence of errors was significantly higher when a neighboring letter was incorrectly identified, with an elevation of 30% to 200%.

As expected, in simulations based on the independence model, these contingencies were completely absent, with performance at a given position independent of performance at other positions. Simulated error proportions were centered around the mean of the empirical neighbor-error and neighbor-correct values with little variation. A demonstrative example is shown graphically in Fig. 4. The two empirical data points (shown by the green and red lines, indicating conditioning on neighbor-correct versus neighbor-error, respectively) are well outside the range of values predicted by the simulations, shown as the colored histograms. Thus, the differences between neighbor-correct and neighbor-error shown in Table 3 are highly significant based on these Monte Carlo confidence intervals.

3.1.5. Conservation of complexity

As the previous analysis showed that errors in neighboring positions were correlated, we next investigated whether balanced gain and loss of spatial parts of each letter might underlie the concomitant errors. While a more detailed analysis based on feature decomposition will come later, we first examined an overall measure of the spatial content of each letter based on the perimetric complexity of each stimulus and response.

We define the change in perimetric complexity (Δ complexity) as the complexity of the response minus the complexity of the stimulus. Therefore, a positive value means that complexity is gained in the response, while a negative value means complexity is lost. Fig. 5 plots the summed complexity change for the outer letters versus the complexity

Table 3
Errors contingent on performance in neighboring positions for each of the subjects. Each column indicates conditional probability of error occurrence, as specified in column label. For example, the first column represents the probability of a left letter error when there is a middle letter error, while the second column shows the probability of a left letter error when the middle letter is identified correctly. See also Fig. 4 for an illustration of this example demonstrating robustness based on simulations.

	$\cancel{L} \mid \cancel{M}$	$\cancel{L} \mid M$	$\cancel{R} \mid \cancel{M}$	$\cancel{R} \mid M$	$\cancel{M} \mid \cancel{L}$	$\cancel{M} \mid L$	$\cancel{M} \mid \cancel{R}$	$\cancel{M} \mid R$	$\cancel{M} \mid \cancel{R,L}$	$\cancel{M} \mid R,L$
S1	19%	13%	25%	18%	44%	35%	44%	34%	62%	35%
S2	14%	11%	15%	11%	37%	31%	37%	30%	40%	31%
S3	12%	9%	9%	7%	37%	32%	40%	32%	43%	32%
S4	14%	11%	13%	9%	50%	44%	52%	43%	62%	44%
S5	12%	6%	15%	10%	46%	31%	41%	31%	45%	32%
Aggregate	14%	10%	15%	11%	43%	34%	43%	34%	54%	35%

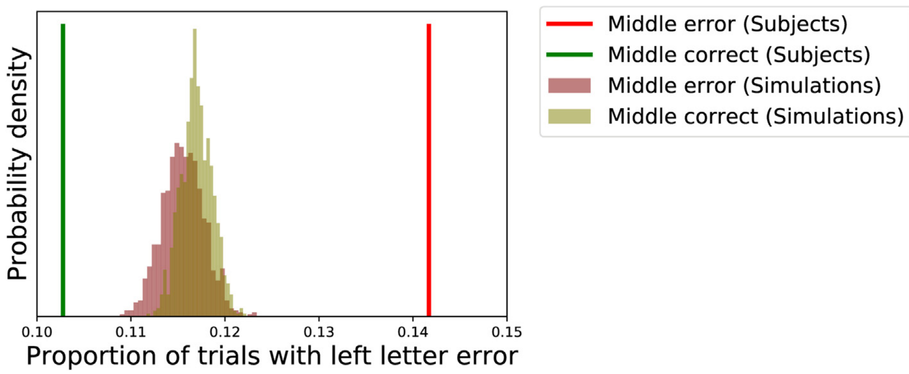

Fig. 4. Illustration of Monte Carlo approach to statistical robustness determination of error contingencies (Section 3.1.4). The conditional likelihood of a subject making a left letter error is shown by the two vertical lines, conditioned on whether the middle letter was correctly identified (green vertical line) or erroneous (red vertical line). A left letter error was about 40% more likely when a middle letter error occurred, and simulations from the independence model (colored histograms) show that this difference was extremely unlikely to happen by chance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

change for the middle letter, for trials with an error in the middle letter and at least one of the outer letters.

There was a negative correlation ($r = -0.28$) between the complexity changes in the middle and outer positions, indicating that when the middle letter lost (gained) features, the outer letters gained (lost) features. To determine the robustness of this statistic, we also computed the correlation for simulated trials with the three independent confusion matrices for each subject. In simulations from the independence model, the absolute value of the correlation never exceeded 0.05, demonstrating that the observed correlation we found was extremely unlikely to occur by chance.

As shown by the marginal histograms in Fig. 5, the changes in complexity in both positions were roughly centered around zero. However, the mean change of the middle letter was slightly positive (0.13), indicating a gain in complexity, while the mean change of the outer letters was negative (-0.28), a small loss.

3.1.6. *Error dependence on letter confusability*

It is already well known that crowding is stronger when flanking items are similar to a central target, a result found along basic stimulus dimensions like color and shape (Kooi, Toet, Tripathy, & Levi, 1994) as well as the confusability of letters (Bernard & Chung, 2011; Freeman, Chakravarthi, & Pelli, 2011). To expand these findings, we also tested the influence of similarity, but with the added ability to examine errors in the outer positions in addition to the middle letter. Like Bernard and Chung (2011), we assumed that the confusability of a pair of letters (defined empirically by their confusion prevalence) is indicative of their visual similarity, and thus used confusion proportions directly as the measure of visual similarity.

Our analysis determined whether the number of errors in each adjacent letter pair depended on the confusability of the two letters. The incidence of letter confusions at the middle position was used to define the confusability of each pair of letters. We performed this analysis separately for the left and right pairs as follows. First, we binned trials by the number of letters correctly reported: 0, 1, or 2. Then, for each of

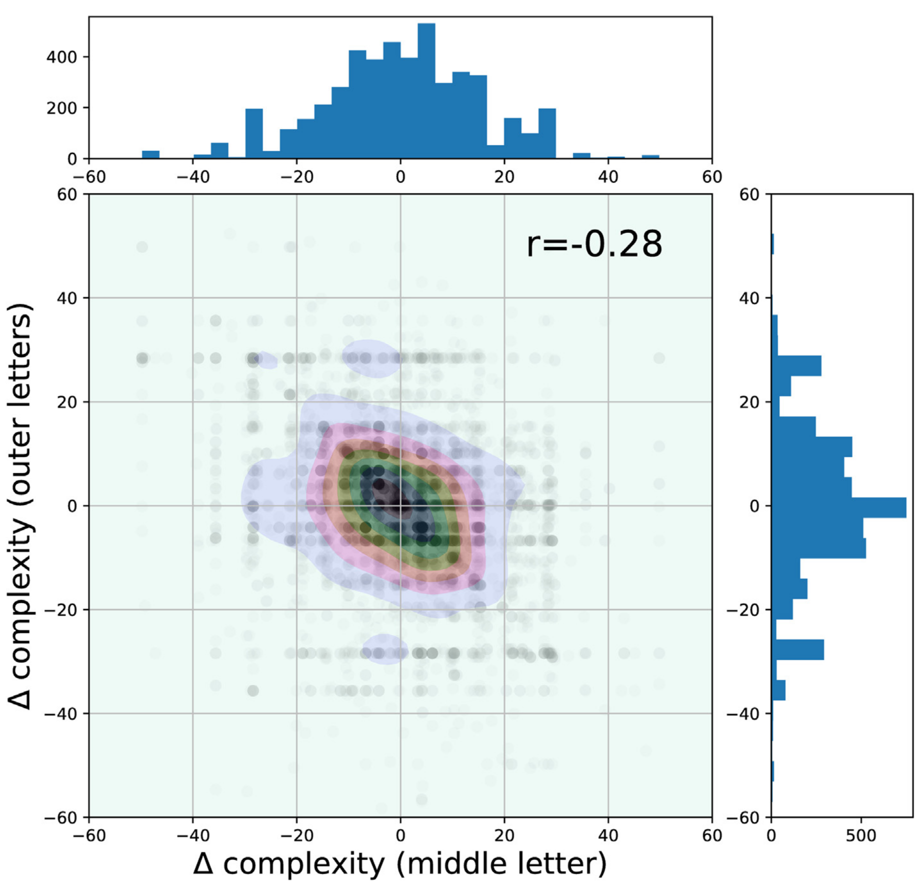

Fig. 5. For trials with an error in the middle position and at least one outer letter, the summed change in complexity of the outer letters is plotted against the change in complexity of the middle letter. A positive number indicates that complexity is gained in the response. Each point represents a single trial. The correlation coefficient, given in the upper right corner, indicates a negative correlation—complexity gain/loss is balanced between letter positions.

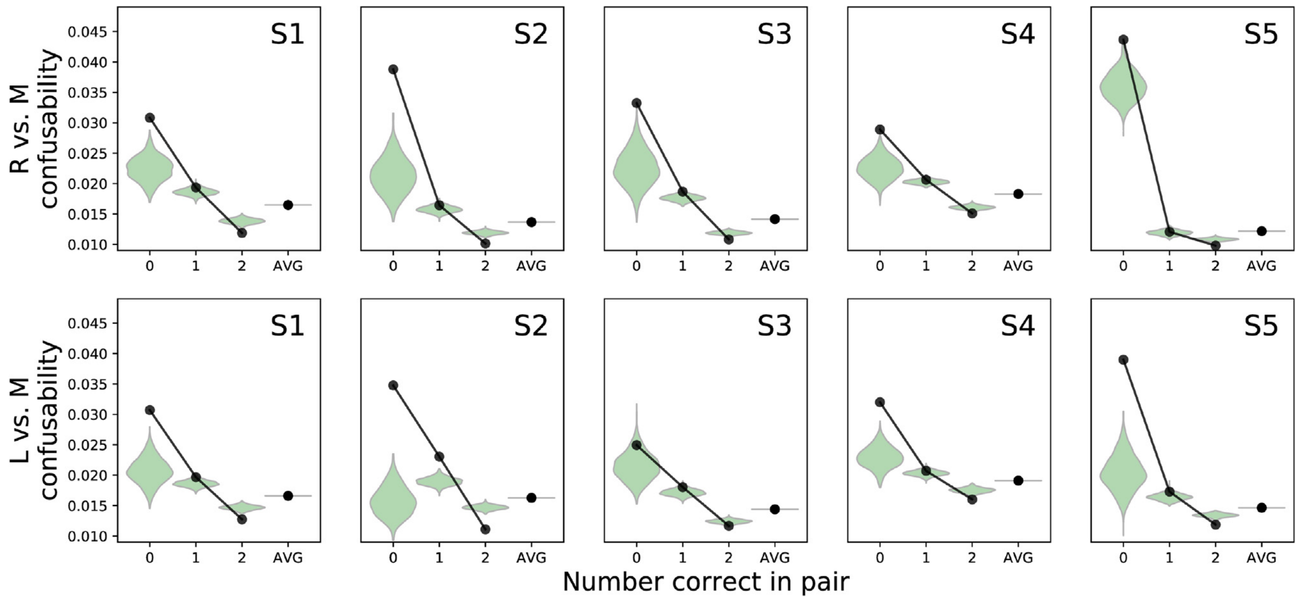

Fig. 6. The average confusability of letters for each pair of middle and outer letters, plotted against the number of correct letters reported in the pair. Points and lines show the empirical data, while colored "violin plots" show the distributions resulting from 1000 simulations. Correct reports are negatively correlated with overall confusability, in both the data and the simulations.

these subsets, we computed the average confusability of the letters in each pair across trials. As a baseline we also performed this operation on simulations from the independence model.

The results are plotted in Fig. 6. Empirical results are shown with the points and lines. Clearly, the number of letters correctly reported is inversely correlated with the confusability of the two letters comprising the pair. However, simulations showed that this trend also is present in the independence model, shown as the distributions on each plot. On the other hand, complete errors in which both the outer and the middle letter are incorrectly identified are well outside the 95% confidence interval predicted by the model for eight out of ten of the (subject, side) groups, all except (S5, right/middle) and (S3, left/middle). Thus, errors in the response are correlated, in a stimulus-dependent way that cannot be predicted by the individual confusion matrices.

3.2. Principles of letter feature interactions

The previous analyses showed that identification of a particular letter in a group of closely-spaced letters is strongly dependent on properties of the adjacent letters: both stimulus properties as well as correlated response patterns. One hypothesis for what might cause this interference is the existence of interactions between particular spatial parts ("features") of adjacent letters. To test this hypothesis, we used the feature decomposition outlined by Bouma (1971), and performed a detailed micro-analysis of feature presence and absence in the confusion results from empirical data and simulations.

Several characteristic patterns suggestive of letter feature sensitivity emerged from the empirical results: (1) features were seldom "hallucinated" if they were completely absent from a trigram; (2) if present anywhere in a trigram, a feature was rarely missed; (3) with multiple instances of a feature within a trigram, the quantity of the feature in the response was often underestimated; and (4) features could either "migrate" or be doubled in adjacent locations. We describe each of these principles in more detail using the descender set as an illustrative example.

The first two error patterns are observed globally, at the level of feature classes in the entire trigram, which we report in the next two sections. First, we describe the rarity of global false alarms (few "ghost features"), and then we describe the paucity of global misses. Following these, we turn to the final two error patterns, which are revealed by feature false alarms and feature misses in the middle letter. Specifically, we show how the latter two types of item errors can often be explained by the presence of a feature in a flanking letter, which interferes with recognition of the middle letter. For the final two principles we focus on the middle letter errors due to their prevalence, as outer letter errors were much rarer. For all subsequent analyses, only trials with at least one error were included.

3.2.1. Lack of ghost features

When no instances of a feature occurred in a trigram, it was quite rare that this feature was reported anywhere in the response. Hallucination of such a "ghost" feature would constitute a global false alarm, an error that was quite rare in these data. For error trials without any descenders, the probability of reporting a descender at any position was 2.7%, much less than the nearly 44.3% probability of reporting a descender across all trials. These statistics are shown as the first numeric column of the first two rows of Table 4. We quantified the robustness of this result using both Monte Carlo approaches. First, we determined the percentage of hallucinations in simulations from the independent confusion matrices model, which resulted in estimates between 4.1% and 4.8% (95% confidence interval). Then, we computed the probability of hallucinations for randomly-generated 5-letter sets, which resulted in a 95% confidence interval of 9.8%–41.6%, making the observed false alarm percentage well outside the bounds expected by chance. Hallucinations of oblique features, shown in the third numeric column, were also strikingly small (1.5%). Round features were hallucinated more often (9.6%), but still outside the 95% confidence intervals predicted by both types of simulations. Thus the absence of ghost features was outside of the 95% confidence intervals for all feature types except ascenders and arch letters. We consider the arch letters further in Section 3.2.4. Ascenders included the most letters, and false alarms had a variety of causes, including confusions between descenders and ascenders (there was a descender in stimuli resulting in ghost ascender trials 64% of the time.)

In summary, the feature letter sets we evaluated had unique characteristics. It was rare for a subject to respond with a letter containing a feature if there were no instances of the feature in the trigram. Neither the confusion matrix simulations nor the random feature letters could reproduce this behavior. We propose that the distinctiveness of these letter features made their complete absence salient and unmistakable.

3.2.2. Few global misses

The previous section showed that features were rarely invented if entirely absent from a trigram, but what about when features were present—how often were they missed? If at least one instance of a feature occurred in a trial, there was a very low probability that the subject would report zero instances and entirely "miss" the feature. Descenders were missed in only about 2.2% of the error trials, much less than the 55.6% probability of not reporting a descender across all trials (Table 4, first numeric column of the third and fourth numeric rows.)

The 95% confidence region from random letter sets was 9.5–10.6%, and from the independent confusion matrix simulations was 11.6%–40.4%, as shown in Table 4. All five of the features displayed a significantly small number of misses compared to the simulations. Again the descender and oblique (3.1%) had the strongest effect. Existence of a feature in the trigram nearly always resulted in a report of the feature in at least one of the three positions. Ascenders and round letters were also rarely missed, though their percentages were higher (5.2% and 6.0%, respectively), likely due to the larger set-size of each feature set. Arch letters also achieved significantly fewer global misses (8.3%) than either simulation predicted. As before, we propose that the salience of these features caused their presence to be unambiguous, resulting in a response (either correct or incorrect) with at least one feature instance.

3.2.3. Feature false alarms and inward migrations

While false alarms were rare for features that did not occur in a stimulus at any position, feature false alarms in the middle letter did occur frequently. It can be inferred from the lack of global hallucinations that the false alarms for the middle item were generally due to occurrence of the feature in a flanking letter, a fact we now demonstrate. Statistics are given as rows of Table 5, referenced by a row number. Columns represent the different feature types evaluated. In this section the first column, corresponding to the descender feature, is fully described. Each cell contains both the conditional statistic determined from the empirical results (in bold), as well as 95% intervals resulting from either the independent confusion matrix simulations (the parenthesized range immediately below the value) or the random letter feature set simulation (the subsequent parenthesized range).

For descenders, half of the errors with a descender response in the middle position contained a descender false alarm (row 1: 49%). This result in itself is not notable, since this percentage is within the range of both the independent confusion matrices and the random letter set simulations. However, the presence of a feature in a flanking outer position strongly coincided with middle feature false alarms. Row 2 shows that 61% of all middle letter descender false alarms contained a descender in at least one flanker, much more than predicted by either of the simulations (32–39% or 30–45%). Furthermore, when an outer flanking letter contained a descender, the likelihood of a middle letter descender false alarm almost quadrupled (row 5: 3.6×), increasing from 11% (row 4, no flanking feature) to 41% (row 3, flanking feature), again far outside the range from either of the simulations, suggesting a role for flanking features in feature false alarms.

To determine the fate of the features from the outer flankers, we analyzed whether outer feature losses varied depending on middle letter feature false alarms. Specifically, a flanker feature loss concomitant with a middle letter feature gain can be taken as evidence of an inward feature mislocation. For descenders, the percentage of trials with conditional flanker feature loss was 36% in the presence of a middle descender false alarm (row 6), significantly higher than predictions from either simulation. Without the middle descender false alarm conditional, the percentage of error trials with a loss of a descender feature was much smaller (row 7: 6%). Thus, the feature false alarm increased the likelihood of an outer loss by six times (row 8), strongly suggesting the presence of inward feature mislocations.

Moving to the other feature columns, this conditional ratio was even higher for the oblique feature (12.9×), mainly due to an even smaller unconditional inward mislocation rate (2%). The remaining three features (arch, ascender, and round) had smaller ratios, but in excess of the 95% confidence intervals from the simulations. These features also had weaker ratios of feature false alarms contingent on flanker features, compared to descender and oblique, yet were significantly associated with a feature in the flanker, except in the case of the arch letters. The arch letters were especially prone to false alarms, as seen previously with ghost feature errors. Many types of stimuli appeared to contribute to these confusions, especially round letters and ascenders in both the middle letter and the outer letters.

3.2.4. Feature loss: repeated features "merge" due to redundancy masking

Lastly, we examined trials in which a feature in the middle position was lost, which often resulted from multiple instances of the feature occurring in a trigram. To isolate these cases, we analyzed trials in which the middle letter contained a feature and the response omitted the feature (a middle position feature "miss"). For the descender, feature loss comprised 61% of all the error trials involving a middle descender letter. To determine the cause of these losses, we analyzed feature contingencies within all three trigram letters in the stimulus and response. When a middle descender loss occurred, there was a descender in at least one of the outer positions of the presented trigram a remarkable 77% of the time, far outside the confidence intervals from both the independent confusion matrices and random letters: 27–32% and 26–37%, respectively, as shown in Table 6.

Analyzed in a different way, the probability of losing a descender in the middle letter was five times as great when an outer letter contained a descender than when it did not (59% vs. 11%, respectively). Interestingly, however, the probability of losing a descender in the middle letter was only slightly higher when both outer letters contained a descender (66%: not shown); one flanking feature appeared to be enough to induce a feature loss. In summary, a flanking descender significantly increased the likelihood of losing a middle descender. Additionally, in almost all cases (98–99% for all feature types), the feature was retained in the outer letter. Therefore, the flanking feature survived crowding, while the middle feature was lost. The two instances of the feature effectively merged together, which we ascribe to redundancy masking.

The other feature error of interest was feature migration toward the flanking letters, which would be evidenced by a feature gain in the outer letters concomitant with a loss in the middle position. While these errors were more rare than those due to interference from existing flanking features, these feature migrations could be clearly identified. First, whereas descender gains in flanking letters happened on approximately 14% of the trials when there was a middle feature loss, these descender gains happened on only 0.6% of error trials without a middle error loss, making an outward mislocation 24 times more likely in the presence of the middle feature loss.

Most of the other feature types followed a similar pattern to the descenders, although only the oblique errors were as dramatic, having a similar percentage of errors corresponding to feature losses (57%), and a similarly large increase in outward mislocations due to middle feature loss errors (15.7×). With oblique flankers, the increase in middle feature losses conditioned on the presence of outward features tripled (3.2×). Ascenders had a smaller percentage of errors that were due to losses (28%), meaning that a majority of errors with ascenders retained the ascender feature, unlike the previously described features. However, like the other error types, a flanking ascender was likely to be present when a middle ascender was lost (61%). The round and arch features had weaker effects, but only the ratio of outward mislocations with arches did not exceed 95% confidence intervals. Interestingly, in 80% of the trials in which an arch letter in the middle was lost, an ascender letter was reported in the middle, regardless of the flankers. We hypothesize that only one of the vertical components was correctly identified, causing errors much like the redundancy masking reported earlier, but at the sub-feature level.

4. Discussion

To explain how groups of letters are processed, theories often make assumptions about underlying mechanisms, such as the parallel and independent nature of letter processing, the use of spatial features for recognition, and interactions (such as migrations) of features between letters. In this study we tested these assumptions by analyzing an extensive dataset of behavioral responses from subjects identifying all three letters in thousands of trigrams presented in their lower visual field. We performed analyses at both the letter and feature level, finding evidence for both inter-letter dependencies and feature-level interactions.

Note that our goal was not to provide a complete model of letter processing. All 26 letters were not represented in the feature sets, for example. Rather, our intent was to reveal broad characteristics of crowded letter confusions. We demonstrated several forms of inter-letter dependence, including the effects of confusability on adjacent letter errors, conservation of complexity, and a response contingency in errors amongst neighbors. These findings challenge models that process each letter individually in parallel, and suggest that at least a portion of letter recognition processing in multi-letter arrays requires dependent units of more than a single letter. General processing deficiencies such as inattentiveness on certain trials cannot explain our results, due to the strong stimulus dependencies observed.

Nor was our goal to construct a full model of feature-based crowded item processing. As stated in the Introduction, several previous studies have used arrays of homogeneous items on continuous feature dimensions to quantitatively characterize how multiple features in the crowding zone may interact. Due to the heterogeneity of letter features (including the featural differences we report in Tables 4–6), it appears challenging to construct generic rules for predicting the behavior of quantities of non-specific "letter features," although some authors have made precise proposals (Pelli et al., 2006). Our analyses were not designed to test hypotheses about specific numbers of features (such as the number of oblique or cardinal strokes). Instead, we simply evaluated the detection of "obliqueness" amongst the three letters in each trigram. Decomposing letters into specific quantities of features would be a straightforward extension of our analysis and several historical models have proposed exactly these types of featural decompositions for letters (Geyer & DeWald, 1973). However, such an approach should incorporate the findings of Townsend and colleagues (Townsend et al., 1984), who found that detection dependencies can exist even amongst the strokes within single characters.

We observed letter mislocations much less frequently than typically reported in the literature, which is typically greater than 15% of errors (Huckauf & Heller, 2002; Strasburger, 2005; Freeman et al., 2011; Zhang et al., 2012; Strasburger & Malania, 2013). The cause for this difference is unclear, although our use of a full-report paradigm is a likely contributor. In a previous experiment reporting all three letters of a trigram presented in the lower visual field, we observed < 10% of errors to be letter mislocations (Chung, Li, & Levi, 2012). Furthermore, in all of the previous studies showing a large proportion of letter transpositions, letter strings were presented in the left or right visual field, in a radial arrangement with respect to fixation (Huckauf & Heller, 2002; Strasburger, 2005; Strasburger & Malania, 2013; Zhang et al., 2012), whereas ours were presented in the lower visual field, with the letters tangentially arranged with respect to fixation. Despite their small number, the mislocations observed did not reflect the occasional transpositions of random, unrelated letters pairs, as might be expected from a faulty position signal (Chung & Legge, 2009). Such an effect should be stimulus-independent, whereas we found that the mislocation rate was correlated with the confusability of adjacent letters.

The analysis of letter errors in terms of a feature decomposition was conceived as a method to investigate the validity of the hypothesis that features may float freely between letters. This has remained a topic of some debate (Strasburger & Malania, 2013), especially since transpositions of entire letters are often observed, including in the early formulation of Wolford (1975). With accuracy at a reasonable level (50–60%), letter mislocalization errors were rare, yet confusions did not appear haphazard. Instead, specific error patterns revealed the influence of neighboring letters on identification.

The characteristics observed reflect the nonlinear and heterogeneous aspects of letter processing. The principles of distinctiveness (few global feature hallucinations and misses) illustrate the nonlinear nature most easily. One explanation of letter recognition represents letters as numerical values along several separable feature axes, determined for example using multidimensional scaling (Kuennapas & Janson, 1969). However, with such a formulation, it is difficult to explain how the existence of a feature in the stimulus can so strongly influence the outcome in a dichotomous fashion. Rather, our empirical results suggest that feature presence or absence is binary rather than graded. On the other hand, the exact location of these features can be indeterminate, as proposed by crowding theories, a result demonstrated

Table 4
Feature detection at the global (trigram) level. For each feature(in columns), each cell indicates the percentage of error trials falling into the categories described in Section 3.2.1 and Section 3.2.2. The two ranges below each statistic indicate the 95% confidence intervals from simulations: the first row from the independent confusion matrices model, and the second row from random n-character letter sets. Global feature errors (false alarms and misses of feature categories in the entire trigram) are distinctive, and can not be explained by chance.

	Descender ($95\%_{indep}$) ($95\%_{letters5}$)	Ascender ($95\%_{indep}$) ($95\%_{letters8}$)	Oblique ($95\%_{indep}$) ($95\%_{letters6}$)	Round ($95\%_{indep}$) ($95\%_{letters7}$)	Arch ($95\%_{indep}$) ($95\%_{letters5}$)
Global feature hallucinations	**2.7%** (4.1%–4.8%) (9.8%–42%)	**19%** (22%–24%) (17%–57%)	**1.5%** (2.4%–2.9%) (12%–46%)	**9.6%** (11%–12%) (15%–51%)	**14%** (12%–13%) (9.8%–42%)
Unconditional feature response	**44%** (41%–42%) (41%–57%)	**73%** (72%–73%) (61%–75%)	**47%** (47%–47%) (48%–64%)	**63%** (62%–63%) (55%–70%)	**50%** (50%–50%) (41%–57%)
Global feature misses	**2.2%** (9.5%–11%) (12%–40%)	**5.2%** (7.8%–8.6%) (8.8%–27%)	**3.1%** (5.8%–6.6%) (11%–36%)	**6%** (8.1%–9%) (10%–31%)	**8.3%** (9.6%–11%) (12%–40%)
Unconditional feature omittance	**56%** (58%–59%) (43%–59%)	**27%** (27%–28%) (25%–39%)	**53%** (53%–53%) (36%–52%)	**37%** (37%–38%) (31%–45%)	**50%** (50%–50%) (43%–59%)

Table 5
Feature contingencies for middle letter feature false alarms (FAs). For each feature column, each row shows the percentage of error trials satisfying the given condition described in Section 3.2.3. The two ranges below each statistic indicate the 95% confidence intervals from simulations: the first row from the independent confusion matrices model, and the second row from random n-character letter sets.

	Descender ($95\%_{indep}$) ($95\%_{letters5}$)	Ascender ($95\%_{indep}$) ($95\%_{letters8}$)	Oblique ($95\%_{indep}$) ($95\%_{letters6}$)	Round ($95\%_{indep}$) ($95\%_{letters7}$)	Arch ($95\%_{indep}$) ($95\%_{letters5}$)
(1) Percentage of feature errors that are feature: FAs `p(FA|feature_error)`	**49%** (46–52%) (59–97%)	**34%** (33–35%) (48–88%)	**33%** (30–35%) (53–95%)	**35%** (34–37%) (53–92%)	**51%** (50–52%) (59–97%)
(2) Pct. of feature FAs having feature in flanker: `p(flanker_feature|FA)`	**61%** (33–40%) (31–46%)	**60%** (53–56%) (49–61%)	**65%** (38–47%) (37–51%)	**56%** (48–52%) (43–56%)	**27%** (35–39%) (31–46%)
(3) Pct. of feature FAs, given feature in flanker: `p(FA|flanker_feature)`	**41%** (19–22%) (38–75%)	**33%** (27–28%) (35–65%)	**24%** (9.7–12%) (38–71%)	**31%** (24–26%) (37–67%)	**40%** (37–40%) (38–75%)
(4) Pct. of feature FAs, given no feature in flanker: `p(FA|¬flanker_feature)`	**11%** (26–32%) (43–78%)	**22%** (28–30%) (37–68%)	**6.3%** (13–18%) (42–76%)	**21%** (26–28%) (39–71%)	**41%** (41–45%) (43–78%)
(5) Increased likelihood of feature FA given flanker feature: (`Row 3/Row 4`)	**3.6×** (1.3–1.5) (0.9–1.3)	**1.5×** (1.0–1.1) (1.0–1.2)	**3.9×** (1.3–1.6) (0.9–1.2)	**1.5×** (1.1–1.1) (0.9–1.2)	**1.0×** (1.1–1.2) (0.9–1.3)
(6) Pct. of flanker feature losses, given FA: `p(flanker_feat_loss|FA)`	**36%** (1.8–2.1%) (2.1–16%)	**14%** (3.6–4.1%) (3.9–18%)	**28%** (0.7–0.9%) (2.8–17%)	**13%** (2.6–3%) (3.5–18%)	**12%** (1.1–1.4%) (2.1–16%)
(7) Pct. of flanker feature losses: `p(flanker_feat_loss)`	**5.8%** (0.3–1.8%) (3.7–14%)	**7.5%** (1.5–2.6%) (5–14%)	**2.2%** (0–1.3%) (4.3–14%)	**5.7%** (0.9–2.1%) (4.7–14%)	**3.7%** (0.3–1%) (3.7–14%)
(8) Increased likelihood of flanker feature loss given FA: (`Row 6/Row 7`)	**6.1×** (0.2–0.9) (0.6–2.7)	**1.9×** (0.4–0.7) (0.6–1.7)	**12.9×** (0.0–1.5) (0.6–2.0)	**2.2×** (0.3–0.7) (0.6–1.8)	**3.2×** (0.3–0.8) (0.6–2.7)

by the feature gain and loss analyses. In fact, the ability for features to mislocalize confirms the early observations of Korte (1923) concerning free-floating features (Strasburger, 2014), which has also been invoked by Pelli et al. (2004) to characterize crowding.

The heterogeneous nature of letter processing was also revealed with this analysis. While confusion matrices capture the overall confusability of pairs of letters, we have shown that there are additional distinctive patterns at the level of particular sets of letters, in this case those sets of letters defined by "features." Here, the features had a clear interpretation in terms of distinctive spatial parts, although this was not a necessity for our analysis. Several of the feature principles we observed have been previously studied, but not in the context of crowding, nor with lowercase letters. Townsend and colleagues studied the detection of discrete line segments in square uppercase letters and an artificial alphabet in a series of studies (Townsend, Hu, & Ashby, 1981; Townsend et al., 1984). Using conditional probability, they found evidence for dependence between features of compound items, rather than independent, uniform detection of features. One differing result is their finding of what they termed "ghost features," which correspond to our hallucinations. They did find evidence for ghost features (Townsend et al., 1984), which they used to reject high threshold feature sampling assumptions. Since they presented items foveally, very short presentation times were needed to induce errors (< 10 ms), which could explain the difference observed. Imagination may become more engaged with weak stimulus information.

We did not consider how adjacent spatial features in certain dispositions could conjoin, leading to errors resulting from direct spatial combination of constituent features, such as an "**l**" combining with an "**o**" to its right to form a "**b**." This type of error was suggested for groups of tiny, closely spaced capital letters in the fovea (Liu & Arditi, 2000). The asymmetries seen in the individual letter accuracy plots (such as a "**c**" to the left being less accurate) could also suggest this type of interaction. Nevertheless, while this type of error may have been the source of some portion of feature errors, we were able to capture many errors by simply assuming that features are combined in pairs or in the entire trigram without regard for particular spatial arrangements (i.e., a "bag of features" approach).

We found that a major factor driving the disappearance of a feature in the middle position was the occurrence of the feature in a neighboring letter. This is reminiscent of the phenomenon of *redundancy masking*, whereby identical items in close proximity cannot be properly individuated: the repeated element may be identified correctly, but the exact number is uncertain (Sayim & Taylor, in revision; Yildirim, Coates, & Sayim, 2017; Yildirim et al., submitted). Here, we found that a pair of adjacent features may effectively "merge" together, creating a percept containing only one instance of a feature, which was typically reported in the flanker rather than the more heavily-crowded middle letter. With feature false alarms in the middle position, a similar pattern emerged. The most typical report was doubling of a feature that already existed in a flanker, although migration from the outward letter into the center letter was also observed.

We believe these results shed light on fundamental aspects of the process of recognizing letters that are in close proximity, which is often the case in text. We find evidence for the hypothesis that letters are composed of features, and that these features have certain identifiable signatures that statistically emerge from the behavioral responses. Features can mix amongst neighboring items, likely leading to the identification errors that are observed with crowded text. These feature errors happen even in the absence of letter transpositions, bolstering the free-floating feature hypothesis, though with additional important dependencies such as adjacent letter confusability. All in all, these results present a clear challenge to the simplest models of textual processing that posit independent detection and integration at a letter level. Instead, more sophisticated models of reading should include a visual

Table 6
Feature contingencies for middle letter feature losses. For each feature type (in columns), each row shows the percentage of error trials satisfying the given condition described in Section 3.2.4. The two ranges below each statistic indicate the 95% confidence intervals from simulations: the first row from the independent confusion matrices model, and the second row from random n-character letter sets.

	Descender (95%$_{indep}$) (95%$_{letters5}$)	Ascender (95%$_{indep}$) (95%$_{letters8}$)	Oblique (95%$_{indep}$) (95%$_{letters6}$)	Round (95%$_{indep}$) (95%$_{letters7}$)	Arch (95%$_{indep}$) (95%$_{letters5}$)
(1) Pct. of feature errors that are feature losses: `p(loss|feature_error)`	**61%** (59–64%) (63–98%)	**28%** (27–29%) (48–89%)	**57%** (55–59%) (54–95%)	**36%** (35–37%) (53–93%)	**42%** (41–44%) (63–98%)
(2) Pct. of feature losses having feature in flanker: `p(flanker_feature|loss)`	**77%** (27–33%) (26–37%)	**61%** (47–52%) (45–55%)	**67%** (34–40%) (33–44%)	**60%** (41–45%) (39–49%)	**46%** (29–33%) (26–37%)
(3) Pct. of feature losses, given feature in flanker: `p(loss|flanker_feature)`	**59%** (32–39%) (43–76%)	**27%** (21–23%) (34–68%)	**48%** (28–33%) (40–74%)	**33%** (23–26%) (38–71%)	**44%** (29–33%) (43–76%)
(4) Pct. of feature losses, given no feature in flanker: `p(loss| ¬flanker_feature)`	**11%** (26–29%) (38–81%)	**17%** (21–24%) (32–72%)	**15%** (21–24%) (37–77%)	**20%** (24–27%) (34–74%)	**26%** (29–32%) (38–81%)
(5) Increased likelihood of feature loss given flanker feature: (`Row 3/Row 4`)	**5.2×** (1.1–1.4) (0.8–1.4)	**1.6×** (0.9–1.0) (0.8–1.3)	**3.2×** (1.2–1.5) (0.8–1.4)	**1.7×** (0.9–1.0) (0.8–1.4)	**1.7×** (0.9–1.1) (0.8–1.4)
(6) Pct. of flanker feature gain, given loss: `p(flank_feat_loss|loss)`	**14%** (0.3–1.3%) (4.4–13%)	**12%** (1.5–2.8%) (6.1–15%)	**7.4%** (0.1–0.9%) (4.9–14%)	**5.4%** (0.4–1.3%) (5.7–14%)	**3.4%** (0.3–1.2%) (4.4–13%)
(7) Pct. of flanker feature gain, given no loss: `p(flank_feat_loss|¬loss)`	**0.6%** (0.6–0.9%) (1.6–8%)	**2.1%** (1.8–2.3%) (2.7–10%)	**0.5%** (0.4–0.6%) (2–8.3%)	**0.9%** (0.6–0.9%) (2.5–9.1%)	**0.9%** (0.6–0.8%) (1.6–8%)
(8) Increased likelihood of flanker feature gain given loss: (`Row 6/Row 7`)	**24.0×** (0.4–1.8) (1.2–4.9)	**5.7×** (0.7–1.4) (1.2–3.2)	**15.7×** (0.2–2.0) (1.2–3.6)	**5.6×** (0.5–1.7) (1.2–3.5)	**3.6×** (0.5–1.8) (1.2–4.9)

word processing module that reflects the optimal use of distinctive letter features gated by the bottleneck of crowding.

Acknowledgments

Supported by NIH/NEI research grant R01-EY012810 (STLC) and the University of California, Berkeley, Undergraduate Research Apprentice Program. The authors acknowledge the assistance of Girish Kumar for writing the code to run the experiments.

References

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*, 452.
Ağaoğlu, M. N., & Chung, S. T. L. (2016). Can (should) theories of crowding be unified? *Journal of Vision, 16* 10–10.
Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision, 11*, 1–16.
Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.
Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research, 11*, 459–474.
Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research, 13*, 767–782.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
Butler, B. E., Mewhort, D. J. K., & Browse, R. A. (1991). When do letter features migrate? A boundary condition for feature-integration theory. *Perception & Psychophysics, 49*, 91–99.
Butler, B. E., & Morrison, I. R. (1984). Do letter features migrate? A note of caution. *Psychological Research, 46*, 223–236.
Chung, S. T. L., & Legge, G. E. (2009). Precision of position signals for letters. *Vision Research, 49*, 1948–1960.
Chung, S. T. L., Li, R. W., & Levi, D. (2012). A fuller report on mislocation errors in visual crowding. *Journal of Vision, 12* 332–332.
Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1022–1033.
Ester, E. F., Zilber, E., & Serences, J. T. (2015). Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *Journal of Vision, 15* 4–4.
Freeman, J., Chakravarthi, R., & Pelli, D. G. (2011). Substitution and pooling in crowding. *Attention, Perception, & Psychophysics, 74*, 379–396.
Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception & Psychophysics, 14*, 471–482.
Gibson, E. J. (1969). *Principles of perceptual learning and development.* East Norwalk, CT, US: Appleton-Century-Crofts.
Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12*, 381–387.
Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences, 106*, 13130–13135.
Hanus, D., & Vul, E. (2013). Quantifying error distributions in crowding. *Journal of Vision, 13* 17–17.
Harrison, W., & Bex, P. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology, 25*, 3213–3219.
Harrison, W. J., & Bex, P. J. (2017). Visual crowding is a combination of an increase of positional uncertainty, source confusion, and featural averaging. *Scientific Reports, 7*, 45551.
Huckauf, A., & Heller, D. (2002). What various kinds of errors tell us about lateral masking effects. *Visual Cognition, 9*, 889–910.
Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8*, 255–279.
Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen [On the apprehension of gestalt in indirect vision]. *Zeitschrift für Psychologie, 93*, 17–82.
Kuennapas, T., & Janson, A.-J. (1969). Multidimensional similarity of letters. *Perceptual and Motor Skills, 28*, 3–12.
Levi, D. M. (2008). Crowding – An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*, 635–654.
Liu, L., & Arditi, A. (2000). Apparent string shortening concomitant with letter crowding. *Vision Research, 40*, 1059–1067.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739–744.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology, 18*, 445–451.

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., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12), 1136–1169.
Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision, 7*(2), 1–12.
Sayim, B., Taylor, H. (in revision). Letters lost: Capturing appearance in crowded peripheral vision reveals a new kind of masking.
Selfridge, O. (1958). Pandemonium: A paradigm for learning. *HMSO,* 513–526.
Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision, 5*(11), 1024–1037.
Strasburger, H. (2014). Dancing letters and ticks that buzz around aimlessly: On the origin of crowding. *Perception, 43*, 963–976.
Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *Journal of Vision, 13* 24–24.
Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research, 43*, 259–275.
Townsend, J. T., Hu, G. G., & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception & Psychophysics, 36*, 35–49.
van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLOS Computational Biology, 6*, e1000646 .
Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review, 82*, 184–199.
Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception & Psychophysics, 27*, 409–420.
Yildirim, F. Z., Coates, D. R., Sayim, B. Crowding impairs subitizing. Poster session presented at the European Conference on Visual Perception 2017, Berlin, Germany.
Yildirim, F. Z., Coates, D. R., Sayim, B. (submitted). Redundancy masking: The loss of repeated items in peripheral vision.
Zhang, J.-Y., Zhang, G.-L., Liu, L., & Yu, C. (2012). Whole report uncovers correctly identified but incorrectly placed target information under visual crowding. *Journal of Vision, 12* 5–5.