# Quantifying crowded and uncrowded letter recognition

by

Daniel Robert Coates

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susana T. L. Chung, Chair
Professor Dennis M. Levi
Professor Thomas L. Griffiths

Spring 2015

**Quantifying crowded and uncrowded letter recognition**

## Abstract

Quantifying crowded and uncrowded letter recognition

by

Daniel Robert Coates

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Susana T. L. Chung, Chair

Despite decades of research, significant mysteries remain concerning two fundamental aspects of human visual processing: how we perceive moderately complex line shapes such as letters of the alphabet, and how peripheral vision differs from foveal vision. This dissertation comprises studies that address both these topics, using experiments with letter recognition and visual "crowding"—the deleterious influence of flanking objects on identification of a target.

Letters are well-studied psychophysical stimuli that strike a balance between degenerate patterns such as Gabor patches or gratings, with more uncontrolled stimuli such as natural images. Letters contain distinct spatial features that are more straightforward to quantify than the characteristics of natural scenes, yet have a sufficient richness of information to probe the general-purpose mechanisms of object recognition.

The last decade has seen a renewed interest in perception in the peripheral visual field, with a particular emphasis on the phenomenon of crowding. Many researchers acknowledge that this curious effect provides a unique window into the process of object recognition. Since there is such a marked difference between the periphery, where crowding is potent, and the fovea, where it is nearly absent, crowding sheds light on the dichotomy between visual processing in the high resolution fovea versus the information-limited periphery. Understanding crowding may illuminate general principles of biological visual processing.

# Contents

# List of Figures

# List of Tables

# Acknowledgments

First and foremost, I would like to thank my adviser Susana Chung. She supported me during my doctoral research, and served as an ideal role model for the design and execution of exceptional psychophysical experiments. Since learning psychophysics was my goal in moving from computational modeling, I was lucky to have such a competent mentor, and was doubly lucky to be her first doctoral student.

I was also fortunate to be part of a vibrant environment at Berkeley. One of the highlights of my academic career was our weekly crowding journal club, early in my studies, featuring Dennis Levi, Stan Klein, occasionally members of David Whitney's lab, and guests such as Christopher Tyler. I cannot imagine a more rewarding introduction to thoughtful, critical review of scientific research. Lastly, just as I was finishing, Gerald Westheimer became a remarkable personal inspiration.

I also want to thank the Vision Science graduate program, including the chair during the bulk of my tenure, Austin Roorda, and the fine administrators of Inez Bailey and Olga Lepilina, for their help, energy, and support. I am grateful for financial support from an institutional NEI training grant my first three years, and from an Ezell fellowship from the American Optometric Foundation. Our lab was fortunate to benefit from the help of optometry students like Jeremy Chin, who did phenomenal work on the project in Chapter 3.3, and many great undergraduate students.

The members of the lab: JB, Dion, and Girish, were wonderful friends and collaborators. The same goes for other members of the Berkeley vision science community, especially Will Tuten, Ram Sabesan, Paul Ivanov, and Alana Firl. Finally, I thank my family and my many friends scattered throughout the globe for their encouragement and support over these years.

# Chapter 1

# Introduction: Studying visual perception using letter recognition and crowding

## 1.1 Studying mid-level and low-level vision

It is hard to imagine studying anything but the visual system. Not only is it the most rich, vital sense for most people; investigating vision has provided a wealth of insight about the brain, connecting both psychology and neuroscience. My particular interest has tended towards low-level and mid-level vision. Specifically, I am drawn to those aspects that are the earliest, or closest to the "input" of the system. Ideally, at the earliest stages of the signal pathway the visual processing will be the most straightforward and comprehensible, less "corrupted" by higher cognitive processes. Coming from an engineering background, I yearn to make working models of early visual processes that can help explain the putative mechanisms in play, and predict how the brain is able to perform its impressive feats, one of which is interpreting the visual world.

### 1.1.1 Crowding

Crowding refers to the phenomenon whereby flanking objects impair identification of a target object. The description almost sounds banal, but this phenomenon has spawned a whole subfield of research. It is clear that objects in the periphery are harder to recognize than objects in our direct (foveal) gaze, but the reason why this is the case is not so clear. While trying to read words that are positioned on our peripheral retina, one can sense that something unusual is occurring. It is not just blurring from poor acuity that disturbs our percepts. Rather, there is some strange blending of the letters together that makes them resist individuation.

This exact effect has been my focus during doctoral research, with experiments mirroring the demonstration described above. We typically run psychophysics experiments where we ask subjects to identify peripheral crowded letters. With enough responses, we begin to see patterns in the types of letters that are confused and thus gain clues about what is limiting performance in crowded conditions. The cortex is known to have a very homogeneous architecture, and it is likely that crowding exposes generic aspects of information processing bottlenecks in the brain.

### 1.1.2 Letters

Letters are interesting psychophysical stimuli in their own right. In our opinion, they strike an ideal balance between natural scenes—difficult to quantify—and degenerate stimuli such as spots of light, gratings, or Gabor patterns. Experiments with the latter types of stimuli have been necessary to help our understanding of how simple patterns are *detected*, but are less useful in understanding *identification*—how complex shapes are discriminated from one other. Furthermore, letters have the advantage of being universal and well learned, and thus do not require extensive training.

On the other hand, it is not clear exactly how we discriminate letters from one another. Again, it sounds almost trivial to state this, exposing a nuance with the ubiquity of vision—it is hard to "step back" and realize the difficult problems our visual system solves each moment. A large part of my work has been to try to understand the components of visual recognition of single letters, across retinal locations. In this dissertation I will describe experiments used to study this topic, including a variety of statistical methods used for analyzing the results of letter recognition experiments.

### 1.1.3 Crowding with letters

My twin topics of crowding and letter recognition are synergistic. We want to understand basic facts about object recognition such as how object components ("features") are integrated together to form a coherent percept. Crowding, which is a breakdown in this process, gives clues. Letters are the ideal stimuli with which to probe crowding, since they also inform our understanding of featural processing.

I first discuss the analysis of letter identification experiments in Section 2.2. Incidentally, crowding errors will be analyzed, but crowding is not central to understanding the analysis methods described. Next, in Section 2.3, I discuss computational modeling of the letter recognition process. Here only single letters are considered. In Section 3, I present the results of several crowding experiments. The first experiment (Section 3.2) employs artificial "letter-like" stimuli, in order help advance our understanding of feature processing in simple shapes, both crowded and uncrowded. The second and third experiments (Section 3.3 and Section 3.4) discuss experiment with Tumbling-E stimuli, and manipulate aspects of the stimuli in various ways to test how the extent of the crowding interactions vary, which may reveal aspects of the neural mechanisms underlying crowding.

# Chapter 2

# Letter recognition

## 2.1 Approaches to the study of letter recognition

A primary goal of my work has been to study letter recognition, particularly aspects related to the legibility and visual similarity of letters. That is, how letters are identified utilizing only low-level, perceptual factors. It is useful to differentiate two aspects of the study of letter recognition that are often intertwined: analyzing letter identification confusions and modeling the process of letter recognition.

First, in Section 2.2 I will discuss how to analyze the *results* of letter recognition experiments. This can be done with the hope that studying letter identification may provide insight into general-purpose object recognition, which is still poorly understood. However, there may also be practical reasons: designing letter charts, making new fonts to mitigate discrimination difficulties (i.e., reading in patients with central vision loss, or developmental abnormalities), studying effects of font characteristics such as boldness, etc. On the other hand, there is also interest in understanding the *process* of letter recognition, which I discuss in Section 2.3. The purpose here is to try to better understand the mechanisms involved in object recognition by humans, at conceptual, algorithmic, and/or neural levels. The insights may be for purely scientific reasons, although improving computer vision algorithms is an incidental benefit that I will discuss.

Originally, the study of letter recognition was spearheaded by optometric researchers, who were prevalent in visual perception research. Others were psychologists interested in reading or development. More recently, theoretical researchers from psychology as well as applied computer vision researchers have tried to understand the mechanisms of letter perception. Clearly, however, there are many researchers and approaches that blur these distinctions. I situate our laboratory in this camp, trying to integrate theory and practical applications.

A tangential goal driving research from the computer vision field is how to build better computer vision algorithms using insights from the studies of biological recognition. Humans still outperform machines in many tasks—the success of the online image verification "CAPTCHAS" demonstrates this. Therefore, many computer scientists turn to inspiration from the human visual system for ideas on how to improve their algorithms. Some researchers come from a neuroscience or artificial intelligence background already (David Marr, Geoff Hinton, Fukushima), while others are seemingly driven purely by engineering concerns.

For both these aspects of letter recognition, my interest lies primarily in trying to understand low-level, purely perceptual factors, independent of other effects such as letter frequency, cognitive development, or word context. From a general standpoint, letters are useful stimuli because they may help illuminate the poorly understood task of object recognition. There is some element of memory involved, but since letters are so well learned (Pelli et al. [2006]), it can be assumed that the subject is primarily discriminating between familiar forms in a purely perceptual way. High level effects such as word context and letter superiority are minimized by using random strings, although subject biases may still be an important high-level cognitive factor in the recognition process. We account for this using the Luce biased choice model (Luce [1963]).

## 2.2 Letter recognition: analyzing the results

*Confusion matrices: everyone's got 'em, no one knows what to do with them.*

For quite some time, researchers have been interested in the types of errors made by subjects during letter identification tasks. In this chapter I present a historical overview of this endeavor, and an in-depth look at techniques that have been used to try to make sense of identification confusions.

I then document how I applied these techniques to letter recognition data we have collected in the lab. For our experiment, explained in more detail in Section 2.2.5, five subjects viewed letters which were presented at 10° eccentricity in the lower visual field, and embedded in a sufficient amount of noise to induce difficulty. There were two conditions, one in which letters were presented alone, and one in which letters were closely surrounded by two other letters—"crowded," which is known to impair performance. The basic idea is that the confusion matrix analysis techniques described can help illuminate the differences between the errors made in these two conditions.

### 2.2.1 Introduction

"Confusion matrices" are the preferred format to report results of identification experiments with stimuli having multiple alternatives. This method simply involves the presentation of responses in an NxN matrix (where there are N possible stimuli). The stimuli presented are denoted by rows, and observer responses are given in columns. Matrices can contain raw frequency counts, or can be normalized per stimulus presentation such that each row adds up to 1.0. The contents of each cell on the diagonal represents a correct answer, and off-diagonal entries correspond to errors.

An excellent review of classic confusion matrices has been written by Mueller and Weidemann ([2012]). Mueller has also built a website providing a wealth of data transcribed from historical papers (Mueller [2014]). I have downloaded and analyzed some of the classical confusion matrices he has made available. In the manuscript (Mueller and Weidemann [2012]), the authors highlight the difficulties that arise when trying to extract visual similarity from confusion matrices, including how the specific testing conditions influence results. Furthermore, it is shown that with carefully controlled experiments, three response factors can be revealed: the legibility of individual letters, the similarity between letters, and observer bias.

A primary goal of my research has been the analysis of these confusion matrices. More specifically, what reliable information can be gleaned from these complex matrices? When are whole matrices (or constituent cells) statistically distinguishable? What models can summarize (or usefully compress) the data contained in a large confusion matrix?

An important text for the analysis of confusion matrix for the social sciences is Wickens' book "Multiway Contingency Tables Analysis for the Social sciences" (Wickens [1989]). From a purely statistical perspective, the fundamental text is Bishop et al.'s "Discrete Multivariate Analysis" (Bishop, Fienberg, and Holland [1975]). While these textbooks offer good descriptions of the techniques for performing generic analyses on matrices (such as row/column association and independence, calculating measures of goodness of fit, etc.), their guidance is lacking for the types of operations we are interested in—extraction of "patterns" (i.e., groups of confusable letters, or summarizing relationships between the letters) from large (i.e., 26×26, 10×10) confusion matrices. Social scientists are typically more concerned with how continuous variables (factors) interact to influence a continuous outcome measure, as opposed to few variables with many categorical possibilities. If the number of dimensions is small (one or two) signal detection theory is applicable, but for three or more stimulus dimensions, the multidimensional extension of signal detection theory–Ashby and Townsend's general recognition theory (GRT) (Ashby and Townsend [1986])—is surely powerful, but appears impenetrable to non-experts.

Instead, the best guide in the landscape relevant to letter recognition that I have encountered is a 1992 paper by J. E. K. Smith entitled "Alternative biased choice models" (Smith [1992]). In this paper, Smith shows how, given several subject confusion matrices for an LCD letter identification task, exploratory analysis can advance from inspection of the raw confusion matrix (gross patterns of percent correct, etc.), to application of the Luce choice model to identify/remove bias, and even further to extensions relevant to a specific response hypothesis. It is refreshing that the author is honest about the exploratory nature of the analysis while proceeding systematically using principled methods.

### 2.2.2 Background

The earliest approach to studying letter recognition involved simply enumerating the errors that were made when viewing letters from a distance. Thus the errors manifested in this case are due to the reduced size of the characters. A pioneering study in this area was the work of Sanford ([1888]), who in 1888 tested identification of isolated letters in many different fonts and determined the "relative legibility" of letters by varying the viewing distance to find the farthest distance (i.e., smallest size) at which subjects could reliably identify each letter with a certain accuracy. He thus was able to characterize the relative difficulty of letters in the various fonts. He also listed the letters that were confused with each target letter. Since then, there have been many listings of letter confusion errors from a variety of perspectives, including other classic studies of font legibility (Tinker [1928]), to complex models of perceptual identification (many results summarized in Section 2.3, since Townsend ([1971a]). A voluminous reference on alphabetic confusion matrices is Mueller and Weidemann ([2012]), which lists many of the confusion matrices that have informed my dissertation. Furthermore, they have included relevant studies that measured letter similarity matrices using alternate methods, including subjective similarity ratings (Boles and Clifford [1989]; Keunnapas and Janson [1969]; Podgorny and Garner [1979]), response or saccade times (Jacobs, Nazir, and Heller [1989]; Podgorny and Garner [1979]) or a forced-choice task of discriminating letter pairs (Mueller and Weidemann [2012]).

Confusion matrix analysis reached a peak in the 1980s, with a variety of manuscripts publishing raw confusion matrices, and correlating empirical or model matrices under differing conditions (Townsend [1971a]; Townsend [1971b]; Gilmore [1985]; Watson and Fitzhugh [1989]; Coffin [1978]). More recently, and related to the current work, researchers have measured letter identification under the condition of "crowding," where adjacent letters worsen identification of the target. Several notable examples explicitly list crowded (Liu and Arditi [2001]; Bernard and Chung [2011]; Bedell et al. [2013]) or peripheral (Reich and Bedell [2000]) confusion matrices. Two further recent studies that published graphical crowded confusion matrices (two-dimensional matrix image plots without numbers) are Hanus and Vul ([2013]) and Freeman, Chakravarthi, and Pelli ([2012]).

Of the aforementioned confusion matrices, the last two (Hanus and Vul [2013]; Freeman, Chakravarthi, and Pelli [2012]) were used as components of recognition models, with upper case letters. Specifically, Hanus and Vul ([2013]) used the confusion matrix in a model for predicting crowded errors by randomly substituting letters using the measured confusion matrix. Freeman, Chakravarthi, and Pelli ([2012]), on the other hand, simply used the confusion matrix to identify the most similar and most dissimilar letter pairs for each target letter, and used these as input for a mixture model. Reich and Bedell ([2000]) counted and qualitatively identified significant confusion pairs of Sloan-like capital letters, while Liu and Arditi ([2001]) compared letter pairs between wide and narrow spacing conditions of upper case letters. Bedell et al. ([2013]) qualitatively noted the similarity and differences between confusion matrices under differing conditions of luminance and crowding with the 10 Sloan letters. Lastly, Bernard and Chung ([2011]) used the confusion probabilities of lower-case letters (which they called a "similarity score") and showed a positive correlation between flanker similarity (i.e., letter confusion probability) and an increase in error rate with crowded trigrams (three closely spaced letters).

Despite their utility, the validity of confusion matrix analysis has not been without controversy. Even as early as 1969, an article titled "Visual confusion matrices: fact or artifact?" (Fisher, Monty, and Glucksberg [1969]) challenged the utility of pursuing this path.

The non-agreement of confusion matrix correlations from different experiments presented difficulties for defenders of this approach (van der Heijden, Malhas, and Van Den Roovaart [1984]). A straightforward explanation, given by Gilmore ([1985]), is that the specific font is critical when comparing confusion matrices. Whether the effects of different stimulus degradations (short durations, low contrast, added noise, peripheral presentations, etc.) can be reliably differentiated has not been systematically explored. More recently, Denis Pelli has observed that using confusion matrices to attempt to identify features (an approach covered in the next chapter) has been a mostly fruitless venture (Pelli et al. [2006]). Nevertheless, we forge ahead!

### 2.2.3 Inducing errors

To yield usable confusion matrices, there must be sufficient identification errors to populate cells of the confusion matrix. This section summarizes some of the methods used to induce errors. One straightforward method is to reduce the size of the stimulus until some threshold level of identifiability is reached. This can be accomplished in one of two ways. One of the oldest methods is to change the distance between the observer and a fixed-size target, either physically (Sanford [1888])), or optically using lenses or mirrors. Alternatively, the stimulus itself may be manipulated, such as the shrinking letters on successively lower rows of a letter chart, or by scaling letters dynamically with a computer. With these size-based methods, the sampling resolution of the system is potentially being probed. This is almost certainly true for basic distinctions such as differentiating an “O” from a “C” at threshold. In this case, the subject is simply detecting a gap in a circle, so the threshold detectable gap will constitute a lower limit. With more complicated alphabetic distinctions, how exactly humans use the sampling-limited information is unclear.

Peripherally-viewed stimuli are subject to greater sampling limitations than foveally-viewed stimuli, yielding more errors. Are the letter errors induced in both conditions the same? Bouma studied identification of lowercase letters in exactly these two conditions: small foveal letters and larger letters at 7° to the left or right of fixation, and published the resultant confusion matrices (Bouma [1971]). While he did identify what appear to be distinct perceptual representations for the two conditions using a technique akin to multidimensional scaling (Section 2.2.4.8), he felt that the errors observed were “largely similar” and did not pursue the topic more, instead averaging the confusion matrices.

Direct evaluation of a possible difference between peripheral versus foveal letter recognition was performed by Higgins, Arditi, and Knoblauch ([1996]). These authors tested a hypothesis that mirror-image letters might be harder to discriminate in the periphery because of peripheral deficiencies in processing phase information (Bennett and Banks [1986]). Importantly, Higgins, Arditi, and Knoblauch ([1996]) scaled their stimuli large enough to account for sampling differences in the periphery versus the fovea (discussed further in Section 3.3.9.2). By comparing detection and identification, they found that differences in performance could be explained purely by sampling, refuting the phase hypothesis of foveal versus peripheral differences for letter recognition. Thus, peripheral stimuli scaled to the appropriate size exhibit the same errors as foveal stimuli. On a related note, Chung ([2010]) did find that *crowded* mirror-image letters evidenced a peripheral deficit versus non-mirror-image letters, even after size-scaling.

Shortening the duration is another means of inducing errors. Before computers were readily available, tachistoscopes were used to present letters briefly enough to achieve a certain level of errors. This was the method used by Townsend ([1971b]); Gilmore et al. ([1979]), while van der Heijden, Malhas, and Van Den Roovaart ([1984]) used an early PDP computer to achieve the same end. One might suspect that this approach to degrading the stimulus is related to lowering the contrast (or luminance) of each letter, another technique that has been used (Watson and Fitzhugh [1989]; Gilmore et al. [1979]; van der Heijden, Malhas, and Van Den Roovaart [1984]; van Nes and Jacobs [1981]). A related technique for degrading a stimulus is simply blurring the target, either optically or with a computer (Loomis [1990]).

A recent novel technique is that of "Bubbles" (Gosselin and Schyns [2001]). This technique limits information in the stimulus by opening Gaussian apertures of varying size and location at potentially different times. The proponents of this method applied it to alphabetic letters (Fiset et al. [2008]). Their goal was not to identifying letter confusions per se, but rather to see what spatial information led to subject responses, much like a classification image approach. When applied to upper and lower case letters, this method revealed that terminations are by far the most important cue used by observers in this task. However, the technique itself is not without controversy (Murray and Gold [2004]). It has been argued that the unique way in which selective information is revealed with this method may limit the generality of the results.

Spatial filtering is a further approach to increasing the difficulty of letter identification. This methods has been used extensively, for a variety of purposes, including characterization of the spatial channels involved in letter recognition (Chung, Legge, and Tjan [2002]). Confusion matrices resulting from spatial filtering have be studied preliminarily (Chung, Kumar, and Coates [2013]).

So far, all of the methods described involve limiting the information present in the stimulus. An alternative approach is to add noise on top of the stimulus. This method was first used with letters made of dots, with random dot arrays of varying density added (Uttal [1969]). More recent research has used added noise extensively, especially to characterize performance versus an "ideal observer" (Pelli, Farell, and Moore [2003]) or, using spatially filtered noise, to identify critical spatial frequency channels (Solomon and Pelli [1994]).

Lastly, and in some ways an extension of added noise, other stimuli present on a display contribute to task difficulty (Eriksen and Eriksen [1974]). Adding flanking letters sufficiently close causes crowding (Bouma [1970]). Thus, the studies listed earlier (Bernard and Chung [2011]; Liu and Arditi [2001]; Bedell et al. [2013]; Hanus and Vul [2013]; Freeman, Chakravarthi, and Pelli [2012]) belong in this category.

### 2.2.4 Confusion matrix analysis methods

Many in the vision community have been interested in comparing confusion matrices. The first real quantitative attempt at mathematically modeling data such as a 26x26 confusion matrix was by Townsend ([1971a]), whose manuscript continues to be an exemplar for the field, employing many of the methods I will discuss. Often, the goal will be to validate a model of recognition proposed by a researcher. Empirical and simulated matrices may simply be correlated, although it has been observed that the strength of the diagonal dominates the analysis, so it is important to also separately compute the correlation of off-diagonal entries (errors) (Gilmore [1985]; Watson and Fitzhugh [1989]). Sum of squared errors (Townsend [1971a]; Townsend [1971b]), and the percentage of variance explained (Geyer and Dewald [1973]) have all been used in this way.

There does not seem to be a rigorous metric for how "good" different values of correlation are. Instead, models can be chosen based on superior fitness. Likely the best method will be based on a Bayesian approach that incorporates the space of possible outcome matrices from a given model. Then the goodness-of-fit may yield an accurate estimate of how likely the model fits the data, versus possible alternatives. Nevertheless, a lingering difficulty with using a single measure of goodness-of-fit with a multivariate outcome such as a 26x26 matrix is that there is still an unavoidable trade-off between the contributions of small errors in many cells versus large errors in a few cells. The following is a list of possible measures that could be used to summarize a confusion matrix in one number, or several numbers. The last item has not been explored, as far as we know.

- Mean of (normalized) diagonal, corresponding to percent correct
- Count of small (zero or some small thresholded) cells
- Amount of symmetry across the diagonal (rows vs. columns) (Wickens [1989]): if completely symmetric, confusions between a given pair of letters is the same in both "directions." (i.e., the number of errors and false alarms are the same). Most empirical letter confusion matrices are not symmetric, having significant asymmetric contributions.
- Moments of the distribution of cell coefficients, on the diagonal, off-diagonal, or all entries, using either raw counts or normalized proportions. Is the distribution of confusion matrix elements peaky? Are there heavy tails? What distribution describes the coefficients best?

#### 2.2.4.1 Sum-squared error

The simplest metric for the agreement between two matrices is the sum of the squared errors between corresponding entries. Townsend ([1971a]) for example, presents several different proposals for modelling letter recognition, evaluated using the sum of squared deviations versus the empirical observations. Mathematically, the operation is performed on two N×N confusion matrices **P** and **Q** comprising elements of the form $p_{ij}$ and $q_{ij}$, respectively.

$$SSE = \sum (p_{ij} - q_{ij})^2 \tag{2.1}$$

#### 2.2.4.2 Correlation coefficient (Pearson's $r$)

A further possibility for the correspondence between two matrices is the correlation coefficient, also known as Pearson's $r$. As the following equation shows, this is the ratio between the covariance of corresponding entries and the product of the standard deviations of the matrices.

$$\rho_{P,Q} = \frac{cov(P,Q)}{\sigma_P \sigma_Q} \tag{2.2}$$

For two confusion matrices defined as above, this would be realized as the following, where $\bar{x}$ is the typical arithmetic mean.:

$$r = \frac{\sum (p_{ij} - \bar{p})(q_{ij} - \bar{q})}{\sqrt{\sum (p_{ij} - \bar{p})^2}\sqrt{\sum (q_{ij} - \bar{q})^2}} \tag{2.3}$$

#### 2.2.4.3 Pearson $X^2$ goodness of fit

A further possibility for comparing the fit of a model (with *expected* cell entries) to empirical data (with *observed* entried) is the Pearson X$^2$ statistic, equation 2.4, which is distributed as a $\chi^2$ distribution indexed by the degrees of freedom of the model.

$$X^2 = \sum \frac{(observed - expected)^2}{expected} \tag{2.4}$$

#### 2.2.4.4 Likelihood-ratio ($G^2$)

The likelihood ratio statistic (Wickens [1989]) is defined abstractly as Equation 2.5, and explicitly (for a confusion matrix P as defined above) in Equation 2.6, with the "hat" term $\hat{p}_{ij}$ representing the model predictions. The logarithm is the natural logarithm.

$$G^2 = 2\sum (observed) log(\frac{observed}{expected}) \tag{2.5}$$

$$G^2 = 2\sum p_{ij} log(\frac{p_{ij}}{\hat{p}_{ij}}) \tag{2.6}$$

#### 2.2.4.5 Matrix comparison with likelihoods

When comparing two different confusion matrices using a likelihood statistic, there is obviously an asymmetry if the numbers are directly substituted as either of the two variables using the previous two methods. Instead, a maximum likelihood estimate can be computed that represents something akin to an average of the two matrices (weighted by the row counts), since cells with smaller possible values (for example due to fewer presentations of a stimulus) should affect estimates less, due to uncertainty. This is a straightforward application of Bayes rule. The likelihood that the two matrices are the same is the sum of the $G^2$ statistics of the two matrices versus the maximum likelihood solution. A subtlety of using this statistic is how to deal with zero terms. The standard practice is to zero any individual terms for which the denominator is zero.

#### 2.2.4.6 Confusion matrix symmetry

For many of the techniques about to be described, row/column *symmetry* is an important component. A matrix has this symmetry exactly if $p_{ij} = p_{ji}$ for all cells. In terms of psychophysical results, this means that for each letter pair (i,j), the proportion of “i” responses when “j” is presented is (approximately) the same as “j” responses to presentation of “i.” There is a formula for the amount of symmetry in a confusion matrix P, which is given in Equations 2.7–2.9. This equation splits the matrix into symmetric and asymmetric components, allowing the percent of asymmetry to be quantified.

$$M = \frac{P + P'}{2}, N = \frac{P - P'}{2}, \tag{2.7}$$

$$\text{Proportion symmetric} = \frac{\sum m_{ij}^2}{\sum p_{ij}^2}(i \neq j) \tag{2.8}$$

$$\text{Proportion asymmetric} = \frac{\sum n_{ij}^2}{\sum p_{ij}^2}(i \neq j) \tag{2.9}$$

Conversion to a symmetric matrix may be necessary, for example to apply multidimensional scaling or clustering. There are several numerical methods for symmetrizing a matrix. Simplest is taking the average of the two transposed cells. To convert a matrix P into a symmetric matrix Q, Equation 2.10 is an element-by-element expression for this average. The procedure can also be expressed concisely in matrix notation, as Equation 2.11.

$$q_{ij} = \frac{p_{ij} + p_{ji}}{2} \tag{2.10}$$

$$Q = \frac{P + P'}{2} \tag{2.11}$$

This formula was used by Reich and Bedell ([2000]), although the authors do discuss an alternative, given in Equation 2.12. The alternative method was introduced by Shepard

([1962]) in the context of multidimensional scaling (without the square root) and has been used in several relevant manuscripts (Gervais, Harvey, and Roberts [1984]). Interestingly, this expression for symmetrizing cells is identical to the "similarity" coefficients that result from a maximum-likelihood fit of the Luce choice model introduced in the next next section.

$$q_{ij} = \sqrt{\frac{p_{ij} \times p_{ji}}{p_{ii} \times p_{jj}}} \tag{2.12}$$

#### 2.2.4.7 Luce choice model (matrix *quasi-symmetry*)

The Luce choice model (Luce [1963]) is an attempt to isolate cognitive, higher-level decision factors (specifically, response bias) from purely low-level (i.e., perceptual similarity) aspects. In spirit it is both an analysis method and a model of human decision making (Luce [2005]). The method involves decomposing a matrix into two terms: a response bias vector, and a similarity matrix capturing the (symmetric) similarity between each pair of stimuli. The bias vector indicates an inherent preference towards giving a certain response, independent of the presented stimulus. The similarity matrix, on the other hand, is assumed to be entirely determined by the stimulus.

The Luce choice model for any confusion matrix can be computed in a maximum likelihood fashion with an exact solution using an algorithm described below. The vector of biases are called $\beta_j$s, and the remaining elements, which comprise the symmetric similarity matrix, are termed $\eta_{ij}$s. Note that from statistics this is known as a "quasi-symmetry model" (Bishop, Fienberg, and Holland [1975]) and thus has a solid statistical/mathematical underpinning, which has not always been appreciated.

Researchers have used the Luce choice model in several ways:

- Comparing the parameters ($\eta$s and $\beta$s), for (bias-free) parameter correlation (Gilmore [1985]) (versus comparing whole confusion matrices). We have utilized this method in Section 2.3.2.
- Using the decomposition as a pre-processing step, for further processing of the resultant symmetric similarity matrix, such as multidimensional scaling (Section 2.2.4.8) or additive clustering (Section 2.2.4.9).

In the past, researchers would use iterative nonlinear function fitting techniques (Gilmore [1985]; Townsend [1971a]; Townsend [1971b]) and arbitrary methods of estimating zero terms (Keren and Baggen [1981]), since it is crucial there are no zeros in the matrix. One way is to replace those numbers with an arbitrary constant such as, for example, 0.05, or to determine a value iteratively (Gilmore [1985]). This point was criticized by Smith ([1982]), commenting on Keren and Baggen ([1981]), as there is a single maximum likelihood solution determined by the model, a fact little understood previously. To determine the maximum likelihood estimate of the matrix involves transforming the original matrix into one satisfying the constraints of quasi-symmetry.

The fitting algorithm is derived from Iterative Proportional Fitting (IPF), and was described in Bishop, Fienberg, and Holland ([1975]), and was also didactically presented in Smith ([1982]). It is a convergent algorithm, rather than a search algorithm. Like IPF, a starting matrix is initialized with arbitrary values, often all ones. Then the first step of one iteration is to adjust the cells multiplicatively such that the row marginals are maintained. The next step is to adjust the cells such that the column marginals are maintained. Finally, the last step is to enforce that the underlying similarities are symmetric. These three steps are sufficient to enforce the model. The steps are repeated until there is no measurable change in the estimated values, and the resulting matrix represents the maximum likelihood fit of the model. The parameters can then be extracted using the following two equations, computed on the *expected* values $\hat{p}_{ij}$ (Smith [1982]).

$$\hat{\eta}_{ij} = \sqrt{\frac{\hat{p}_{ij} \times \hat{p}_{ji}}{\hat{p}_{ii} \times \hat{p}_{jj}}} \tag{2.13}$$

$$\hat{\beta}_j = \frac{1}{\sum_{k=1}^{N} \sqrt{\frac{\hat{p}_{jk} \times \hat{p}_{kk}}{\hat{p}_{kj} \times \hat{p}_{jj}}}} \tag{2.14}$$

#### 2.2.4.8 Multidimensional scaling

Multidimensional scaling (MDS) is another technique that constitutes both a method of analysis and a theory of human decision-making. Roger Shepard Shepard ([1958]) showed that analysis of the errors made during perception of Morse code symbols could yield useful insights about the possible “feature dimensions” that subjects use to perform discriminations in this modality. That is, mental representations of objects can be conceptualized as a vector of values that specify locations in some high dimensional space. When making a judgment about the identity of a stimulus at the threshold of performance, the values will be uncertain due to physical or neural noise, so the “nearest” stimulus is chosen (in cognitive space), determining which stimuli are more likely to be confused.

Since measures of distance in any space is symmetric, the contributing similarity (or dissimilarity) matrices must be symmetric. Tversky was one of the most vocal critics of spatial models (Tversky [1977]; Tversky and Gati [1982]), on the grounds that strong directed confusions (which he had collected many didactic examples of) were lost. To explicitly account for the observed asymmetries in empirical data, there have been several theoretical proposals, including those based on stimulus density in the cognitive space (Krumhansl [1978]). For the present analysis however, a principled approach is to leverage the Luce choice model as a “preprocessing” step, which is described in subsequent sections. The Luce choice procedure determines a bias vector as part of its computation, which accounts for the asymmetry. This combination is known as the “MDS-choice” model (Nosofsky [1985]), and has been utilized successfully in a number of situations.

To transform from a confusion matrix of cell counts or probabilities to a distance matrix requires an additional step. As originally proposed by Shepard (Shepard [1957]), the probability of confusions between stimuli is *exponentially* related to the psychological distance, as shown in Equation 2.15. It has been debated that the squared exponential distance could be more appropriate in some cases (Nosofsky [1985]), constituting a "Gaussian" form, but the exponential model is sufficient for our purposes.

$$\eta_{ij} = e^{-d_{ij}} \tag{2.15}$$

#### 2.2.4.9 Additive clustering

Traditional clustering is a common analytical technique, where each item is "labeled" in way that captures some proximity relationships in the data—items near each other should belong to the same cluster. Additive clustering is an extension of this technique. In additive clustering, an item can belong to multiple clusters, which gives it relevance to a feature-based decomposition (see next chapter). Specifically, in the additive clustering model, objects are identified by binary membership in a set of "classes." As opposed to the traditional model of clustering, which dictates that each object belongs to one class or cluster, in additive clustering objects can belong to multiple classes. Furthermore, there is no restriction to cluster membership, which distinguishes additive clustering from models such as hierarchical clustering (popularized in dendrogram trees). The formal mathematical definition is very simple. First, class membership is defined, for $m$ classes and $n$ stimuli, by an $n \times m$ binary matrix where $f_{ik} = 1$ if stimulus $i$ is in class $k$, and 0 otherwise. Furthermore, each cluster is assigned a "weight," which can be interpreted as importance or relevance to the subject. Therefore, the "similarity" of two stimuli $i$ and $j$ will be:

$$\hat{s}_{ij} = \sum_{k=1}^{m} w_k f_{ik} f_{jk} \tag{2.16}$$

In words, this equation states that the similarity is simply the sum of the weights of the classes in which both stimuli are members. For all other classes, one of the $f$ terms in Equation 2.16 will be zero. The mathematical simplicity of the model makes it very attractive, but a downside is that determining the values of the binary $f$ coefficients is a very difficult optimization problem. Elaborate methods have been proposed (Shepard and Arabie [1979]), but Lee recently introduced a simple and elegant method (Lee [2002]; Lee [2001]), inspired by genetic algorithms. The algorithm is stochastic, proceeding by "randomly" mixing together the binary matrices of optimal solutions, analogously to the process of biological reproduction whereby optimization is accomplished by evolutionary forces of natural selection. Similar algorithms have been successful in solving optimization problems which are difficult to solve with deterministic procedures (Mitchell [1998]).

| Subject | Condition | Stimulus contrast | Num. Trials | Prop. Correct |
|---|---|---|---|---|
| AXL | single | 0.28 | 9400 | 0.596 |
| KMA | single | 0.30 | 7000 | 0.562 |
| RXK | single | 0.35 | 6097 | 0.585 |
| RXL | single | 0.28 | 6000 | 0.638 |
| YTY | single | 0.2 | 5950 | 0.561 |
| AXL | crowded | 0.5 | 7000 | 0.528 |
| KMA | crowded | 0.5 | 7000 | 0.514 |
| RXK | crowded | 0.5 | 5200 | 0.375 |
| RXL | crowded | 0.475 | 6000 | 0.58 |
| YTY | crowded | 0.3 | 6120 | 0.361 |

Table 2.1: Subject stimulus contrast and performance.

#### 2.2.4.10 Summary of theoretical methods

The previous sections described a variety of techniques that have been used in the past to analyze letter identification results. This list is by no means exhaustive. For example, dendrograms resulting from hierarchical clustering have been used to visualize letter confusion results (Grainger, Rey, and Dufau [2008]). Methods known as "3-way" or "procrustean" have been used to capture individual differences in the context of MDS models (Cox and Cox [2000]).

The research area is not as vibrant as it once was, but there are still new methods being developed. For example, MJ Brusco recently published papers on how to extract subsets and concordant groups from confusion matrices (Brusco and Steinley [2006]; Brusco and Cradit [2005]). Machine learning, which is currently a highly active research area, will also likely contribute additional methods that could be applied to confusion matrices. Algorithms for additive clustering continue to be developed using more sophisticated techniques (Tenenbaum [1996]; Ruml [2002]), including Bayesian approaches (Navarro and Griffiths [2008]).

We turn now to an empirical application of these techniques. An experiment was run to acquire confusion matrices from several individuals under a pair of conditions. Ideally, for a given confusion matrix, these statistical procedures should be able to reveal patterns which are consistent across subjects, but different amongst the two conditions.

### 2.2.5 Experimental methods

As alluded to above, one goal in pursuing confusion matrix techniques is understanding the errors made when letters are subject to the phenomenon known as "crowding." Specifically, when letters are surrounded by other letters, particularly in the periphery, identification is hindered. We were interested in determining whether the particular patterns of errors in this condition could be differentiated from errors under other conditions, such as when letters

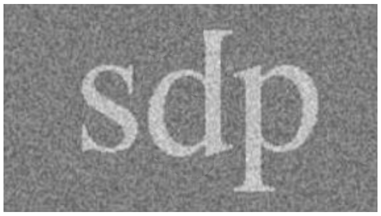


Figure 2.1: Example trigram stimulus in noise.

are presented alone in the periphery. The empirical study described next was performed to pursue this goal.

Five young subjects participated in the study. All had normal vision and used their standard optical correction. Subjects were seated in a dark room 50cm from the 15 in. CRT display, which used an electrical attenuator to provide finer contrast granularity (Pelli and Zhang [1991]), although this study did not demand accurate measurement of contrast threshold. Stimuli were rendered on a Macintosh computer using MATLAB and Psychtoolbox (Brainard [1997]). Each pixel subtended approximately 2.1 minutes of arc. The apparatus used in this study has been described previously in Chung, Levi, and Li ([2006]).

Stimuli were presented for 150 milliseconds at 10° eccentricity in the lower visual field. The stimuli comprised one or three lower case Times-Roman letters, depending on the condition. In the single condition, a single letter was presented at 10° below fixation, with a luminance low enough to induce sufficient errors. In the crowded condition, the target letter was flanked on the left and right side by other letters, different from the target. A higher stimulus intensity was used in the crowding condition to roughly equate the two conditions, since the flankers cause additional errors.

The background of a central 256×256 pixel image patch and stimulus letters were rendered using Gaussian noise, scaled to yield a specified average luminance level. Figure 2.1 is a depiction of a typical stimulus. The contrast of the letters varied per-subject to yield reasonably consistent levels of performance, as indicated by Table 2.1. The contrast of the background noise was fixed at 0.3 for all subjects and conditions. The letter size (in terms of the x-height, or the height of an "x" character) was fixed at 50 pixels, which corresponded to a retinal subtense of 1.7°. Letter spacing for the crowded condition was fixed at 1× the x-height of the letter.

### 2.2.6 Results

The results of each subject in the two conditions can be presented as a confusion matrix with ordinal cell counts as described above. This raw data is given in an Appendix on page 130, though normalized versions of the data will typically be used in order to compare across subjects. Figure 2.2 presents an example of a normalized confusion matrix for subject AXL—the raw cell counts are provided in Figure A.1. By convention, stimuli are denoted by the rows, from "a" (top) to "z" (bottom), and responses are given in the columns. The

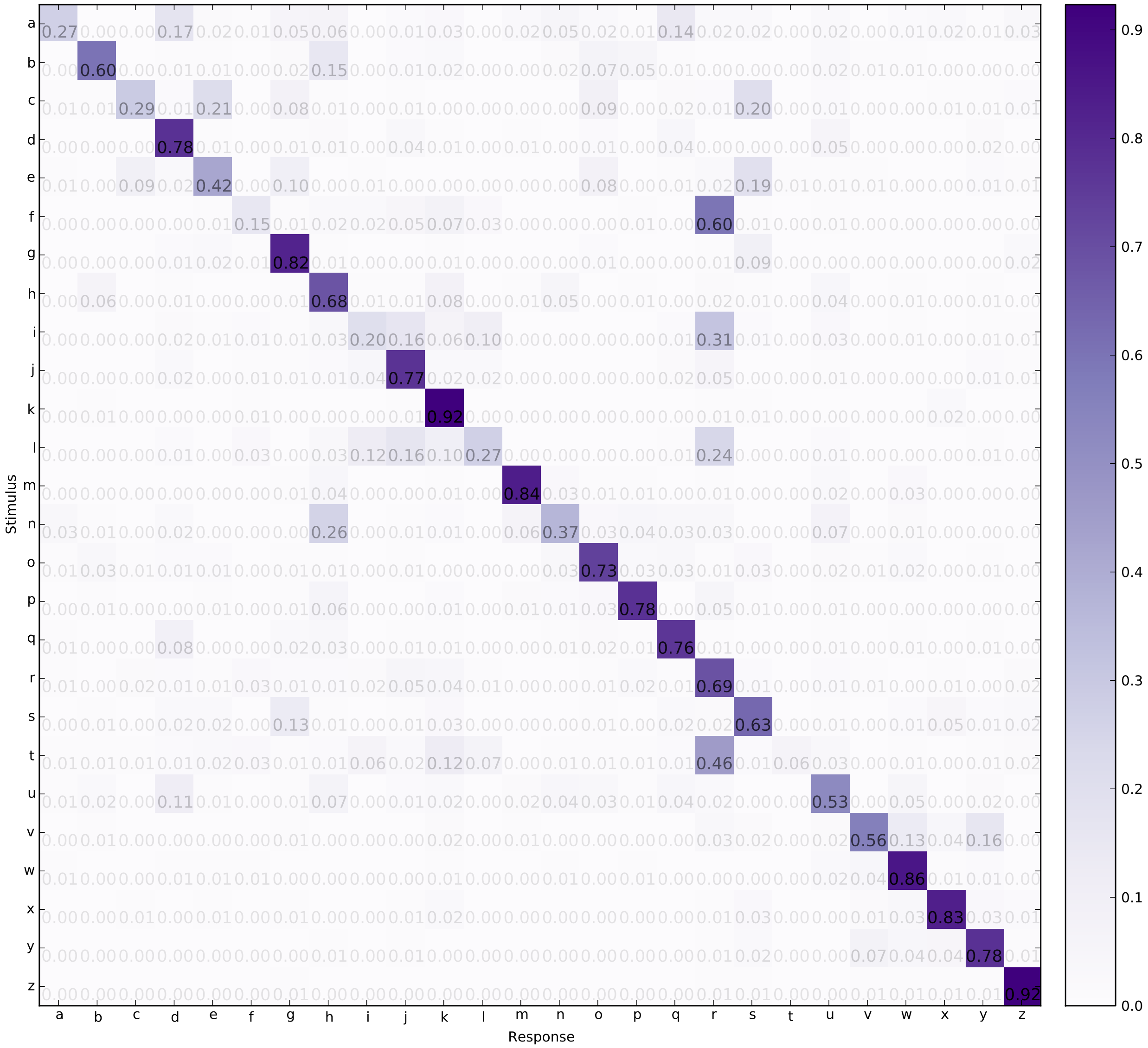
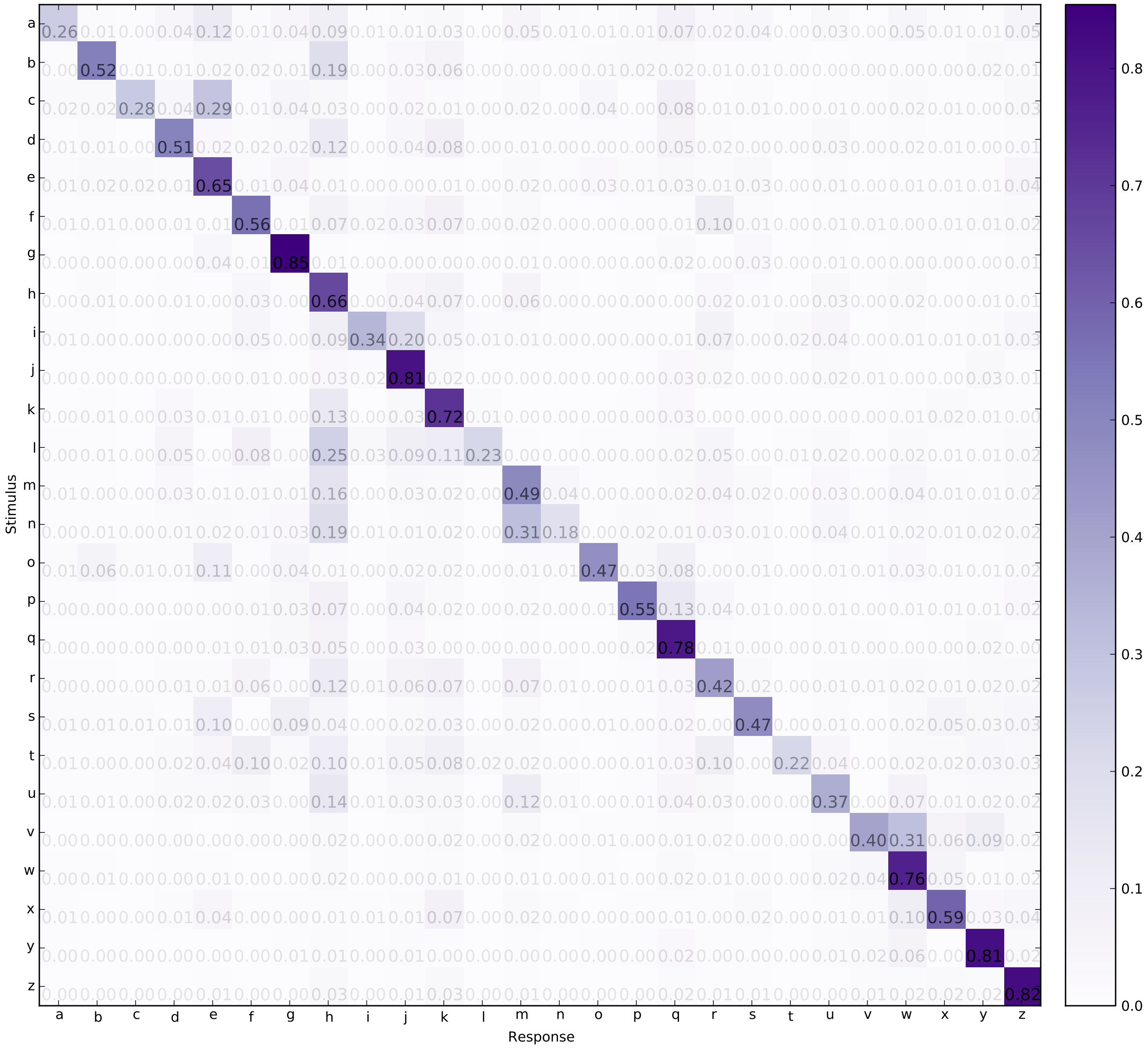


Figure 2.2: Normalized single-letter confusion matrix for subject AXL.



| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 0.26 | 0.01 | 0.00 | 0.04 | 0.12 | 0.01 | 0.04 | 0.09 | 0.01 | 0.01 | 0.03 | 0.00 | 0.05 | 0.01 | 0.01 | 0.01 | 0.07 | 0.02 | 0.04 | 0.00 | 0.03 | 0.00 | 0.05 | 0.01 | 0.01 | 0.05 |
| b | 0.00 | 0.52 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 | 0.19 | 0.00 | 0.03 | 0.06 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 |
| c | 0.02 | 0.02 | 0.28 | 0.04 | 0.29 | 0.01 | 0.04 | 0.03 | 0.00 | 0.02 | 0.01 | 0.00 | 0.02 | 0.00 | 0.04 | 0.00 | 0.08 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 0.00 | 0.03 |
| d | 0.01 | 0.01 | 0.00 | 0.51 | 0.02 | 0.02 | 0.02 | 0.12 | 0.00 | 0.04 | 0.08 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.05 | 0.02 | 0.00 | 0.00 | 0.03 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 |
| e | 0.01 | 0.02 | 0.02 | 0.01 | 0.65 | 0.01 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.03 | 0.01 | 0.03 | 0.01 | 0.03 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.04 |
| f | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.56 | 0.01 | 0.07 | 0.02 | 0.03 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.10 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.02 |
| g | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.01 | 0.85 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.03 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| h | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.03 | 0.00 | 0.66 | 0.00 | 0.04 | 0.07 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.03 | 0.00 | 0.02 | 0.00 | 0.01 | 0.01 |
| i | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.09 | 0.34 | 0.20 | 0.05 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.07 | 0.00 | 0.02 | 0.04 | 0.00 | 0.01 | 0.01 | 0.01 | 0.03 |
| j | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.03 | 0.02 | 0.81 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.02 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.03 | 0.01 |
| k | 0.00 | 0.01 | 0.00 | 0.03 | 0.01 | 0.01 | 0.00 | 0.13 | 0.00 | 0.03 | 0.72 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.01 | 0.00 |
| l | 0.00 | 0.01 | 0.00 | 0.05 | 0.00 | 0.08 | 0.00 | 0.25 | 0.03 | 0.09 | 0.11 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.05 | 0.00 | 0.01 | 0.02 | 0.00 | 0.02 | 0.01 | 0.01 | 0.02 |
| m | 0.01 | 0.00 | 0.00 | 0.03 | 0.01 | 0.01 | 0.01 | 0.16 | 0.00 | 0.03 | 0.02 | 0.00 | 0.49 | 0.04 | 0.00 | 0.00 | 0.02 | 0.04 | 0.02 | 0.00 | 0.03 | 0.00 | 0.04 | 0.01 | 0.01 | 0.02 |
| n | 0.00 | 0.01 | 0.00 | 0.01 | 0.02 | 0.01 | 0.03 | 0.19 | 0.01 | 0.01 | 0.02 | 0.00 | 0.31 | 0.18 | 0.00 | 0.02 | 0.01 | 0.03 | 0.01 | 0.00 | 0.04 | 0.01 | 0.02 | 0.01 | 0.02 | 0.02 |
| o | 0.01 | 0.06 | 0.01 | 0.01 | 0.11 | 0.00 | 0.04 | 0.01 | 0.00 | 0.02 | 0.02 | 0.00 | 0.01 | 0.01 | 0.47 | 0.03 | 0.08 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.03 | 0.01 | 0.01 | 0.02 |
| p | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.04 | 0.02 | 0.00 | 0.02 | 0.00 | 0.01 | 0.55 | 0.13 | 0.04 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 |
| q | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 | 0.05 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.78 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |
| r | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.06 | 0.00 | 0.12 | 0.01 | 0.06 | 0.07 | 0.00 | 0.07 | 0.01 | 0.00 | 0.01 | 0.03 | 0.42 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 | 0.02 | 0.02 |
| s | 0.01 | 0.01 | 0.01 | 0.01 | 0.10 | 0.00 | 0.09 | 0.04 | 0.00 | 0.02 | 0.03 | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.47 | 0.00 | 0.01 | 0.00 | 0.02 | 0.05 | 0.03 | 0.03 |
| t | 0.01 | 0.00 | 0.00 | 0.02 | 0.04 | 0.10 | 0.02 | 0.10 | 0.01 | 0.05 | 0.08 | 0.02 | 0.02 | 0.00 | 0.00 | 0.01 | 0.03 | 0.10 | 0.00 | 0.22 | 0.04 | 0.00 | 0.02 | 0.02 | 0.03 | 0.03 |
| u | 0.01 | 0.01 | 0.00 | 0.02 | 0.02 | 0.03 | 0.00 | 0.14 | 0.01 | 0.03 | 0.03 | 0.00 | 0.12 | 0.01 | 0.00 | 0.01 | 0.04 | 0.03 | 0.00 | 0.00 | 0.37 | 0.00 | 0.07 | 0.01 | 0.02 | 0.02 |
| v | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.40 | 0.31 | 0.06 | 0.09 | 0.02 |
| w | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.02 | 0.04 | 0.76 | 0.05 | 0.01 | 0.02 |
| x | 0.01 | 0.00 | 0.00 | 0.01 | 0.04 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.01 | 0.01 | 0.10 | 0.59 | 0.03 | 0.04 |
| y | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.06 | 0.00 | 0.81 | 0.02 |
| z | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.01 | 0.03 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.02 | 0.82 |

Figure 2.3: Normalized crowded confusion matrix for subject AXL.

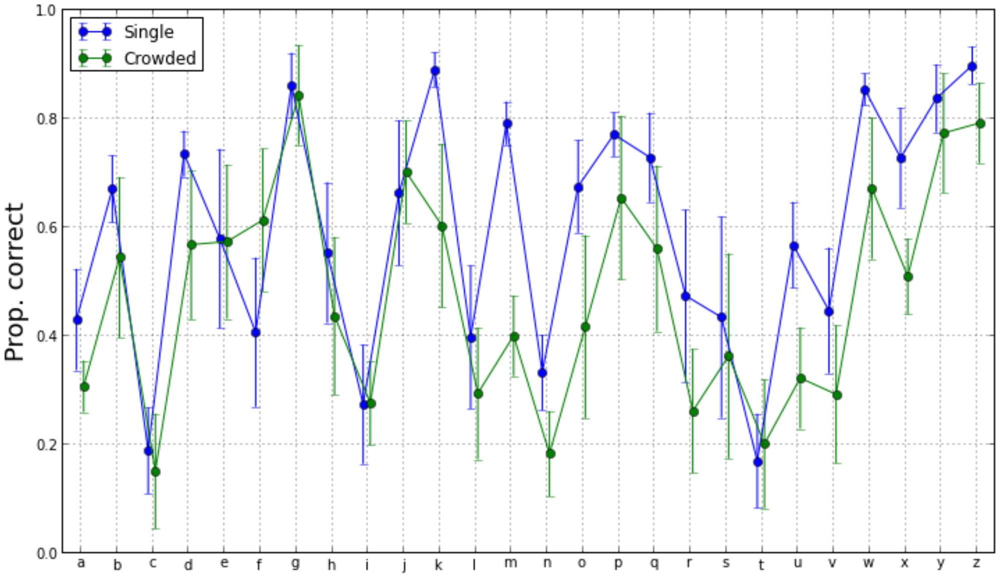


Figure 2.4: Proportion correct for each letter in the two conditions, averaged over the five subjects. Error bars are standard deviation across subjects.

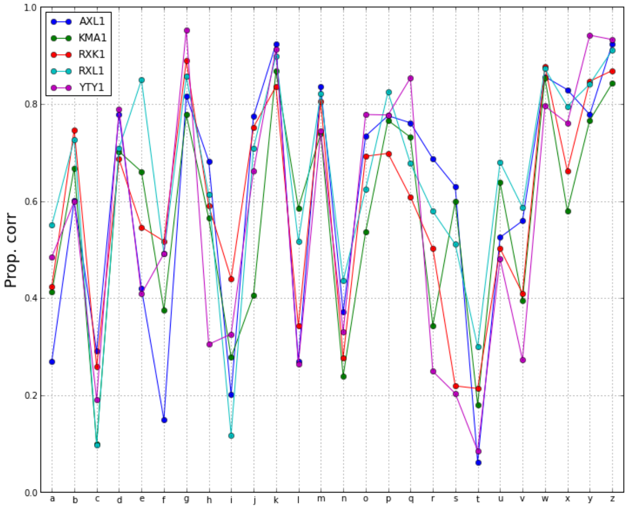


Figure 2.5: Proportion correct for each letter in the single condition, for each of the 5 subjects.

diagonal entries indicate correct results, while the off-diagonal entries indicate errors. For comparison, the crowded confusion matrix for this same subject is provided in Figure 2.3.

The proportion correct (values along the diagonal) for each of the 26 letters is given in Figure 2.4 in each of the two conditions. Note that trends are relatively consistent across the 5 subjects, as shown in Figure 2.5 and Figure 2.6. The utility of Figure 2.4 is that it nicely highlights the relations between proportion correct in the single-letter and crowded

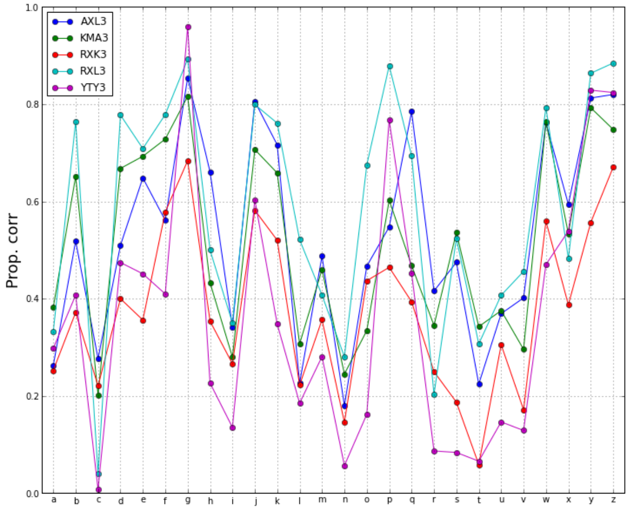


Figure 2.6: Proportion correct for each letter in the crowded condition, for each of the 5 subjects.

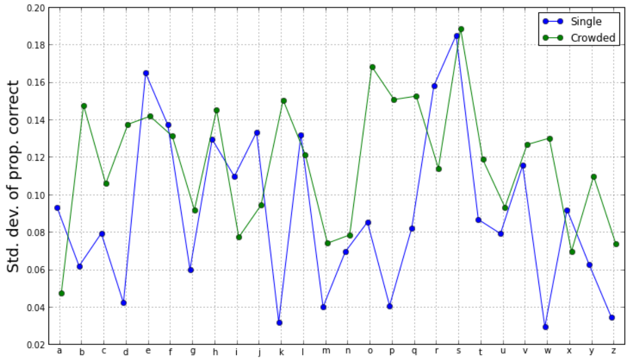


Figure 2.7: Standard deviation of proportion correct across each subject for the two conditions.

condition. First, the variability across subjects is typically higher in the crowded condition, indicated by the greater extent of the green error bars (higher standard deviation across subjects). This fact is shown further in Figure 2.7, which plots the standard deviation of the proportion correct (across subjects). Several letters are in good agreement across subjects in the single condition, specifically d, g, k, m, p, w, and z. Those are also the letters that subjects typically performed the best on, meaning that the subjects agreed with one another when the letters were easily identifiable (except for m). For other letters, there were more individual differences concerning proportion correct.

Another way to format these data is to show the mean and standard deviation on a single plot, as in Figure 2.8. Each letter and condition is indicated with a colored character, with a gray line connecting the two so the relationship between the single and crowded conditions

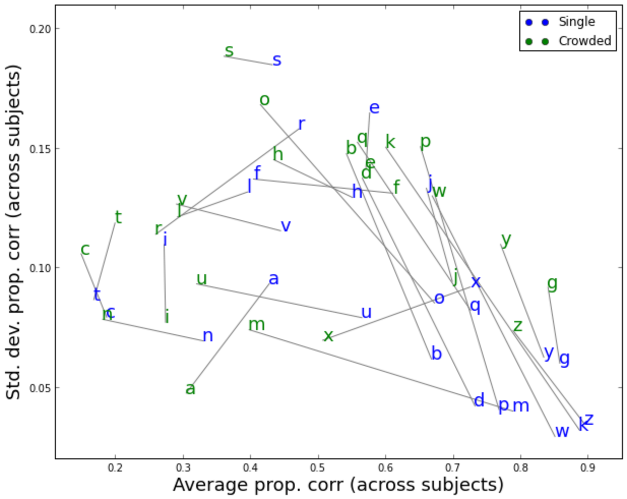


Figure 2.8: Standard deviation versus average of correct letter identification.

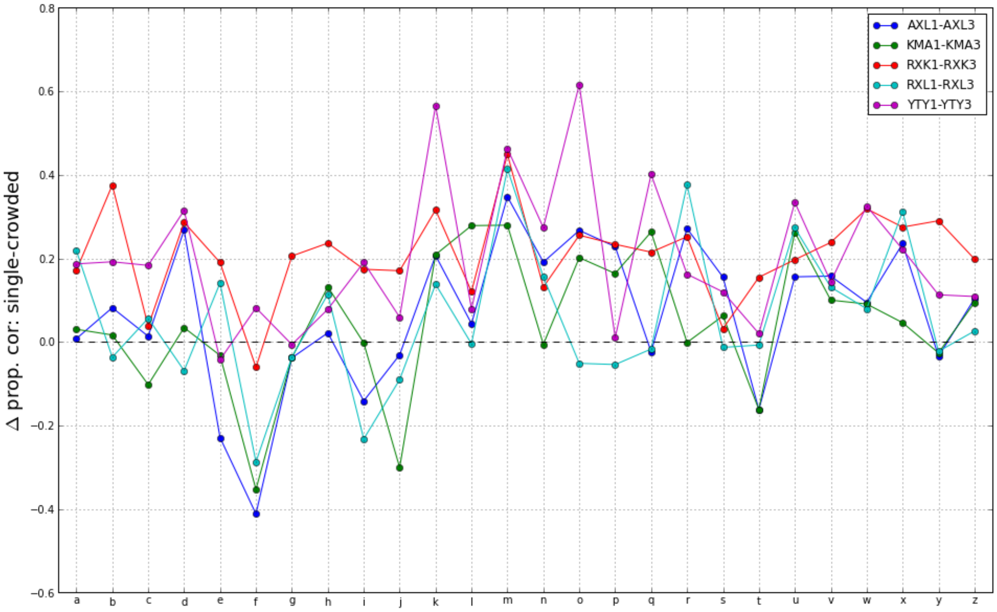


Figure 2.9: Difference between the single and crowded proportion corrects for each letter and each subject.

is clear. The general trend is that letters became slightly harder to identify (shift left), with a concomitant large increase in the variability between subjects (shift up). The letters with lower overall proportion correct (both single and crowded) differed the most. Several letters were not significantly different between the two conditions: c, t, i, s, and e. Some letters became much worse when crowded (u, and m). Lastly, for the two letters t and f, performance improved (relative to the average performance) when crowded.

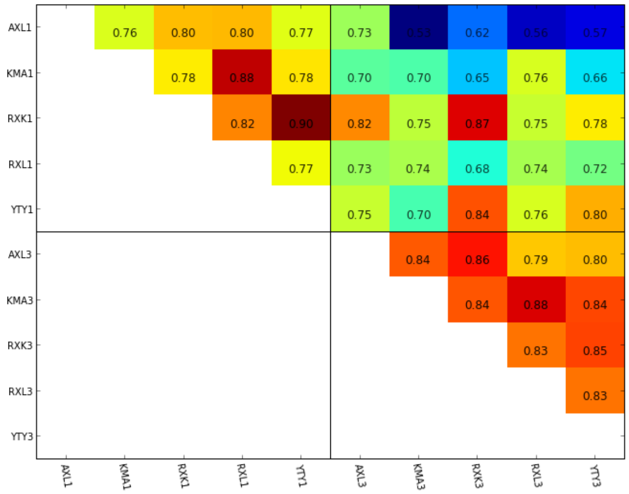


Figure 2.10: Correlation coefficients between the diagonals of each of the 10 (5 subjects, single and crowded) normalized confusion matrices. The center dividing lines demarcate the sets of single (left and upper five, ending in “1”) and crowded (right and lower five, ending in “3”) results.

Figure 2.8 is revealing, but are the general trends consistent within subjects? Figure 2.9 plots the per-letter difference between the proportion corrects in the two conditions for each subject. Transforming to Z-scores before subtraction did not yield qualitatively different results. Figure 2.9 confirms the unique status of the letter f. Four of the subjects performed better on this letter (relative to the mean performance) in the crowded condition. Generally, subjects identified single letters better, most consistently with a, h, k, l, m, n, o, r, s, u, v, w, x, and z. For the remaining letters there were significant individual differences. The challenge now is thus, is there a way to capture the patterns in these individual differences? Can insight be gained on the specific type of degradation that crowding incurs? Clearly there is a difference between the errors in the two conditions tested.

To compare the subjects to one another as well as between the two conditions, we computed the correlation coefficient of the diagonal entries (the proportion correct of each letter). This is presented in Figure 2.10 in the form of an upper-triangular matrix. Generally, the correlation is higher between the subjects on the same condition, versus comparing across conditions, although the difference is not large. What if the whole matrix is compared? Figure 2.11 presents this analysis. The correlations are quite large (0.91-0.97 for the single condition, and 0.85-0.95 for the crowded condition), and the difference between the two conditions is not well differentiated (0.77-0.94). This result is simply an artifact of the structure of the confusion matrix. Specifically, the diagonal entries dominate the computation in a way that is consistent for all confusion matrices. It is more informative to analyze the error terms separately. Figure 2.12 plots the correlation of all terms except the diagonal for each of the 10 confusion matrices (5 subjects × 2 conditions). The correlations are much lower, and seem to exhibit fewer systematic patterns. For example, YTY3 agrees nearly as well

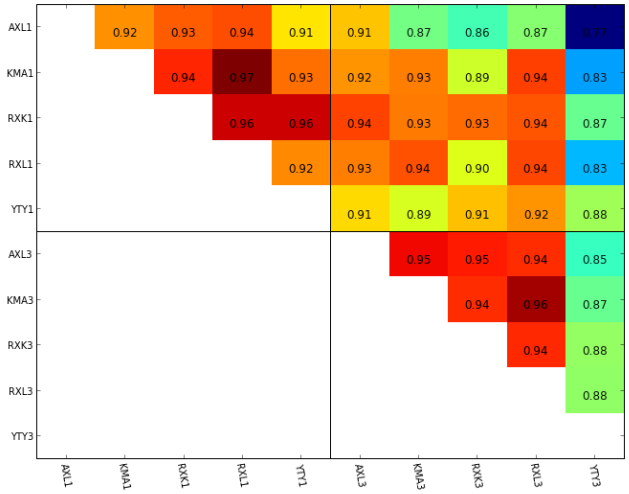


Figure 2.11: Correlation coefficients between each of the 10 (5 subjects, single and crowded) normalized confusion matrices. The center dividing lines demarcate the sets of single and crowded results.

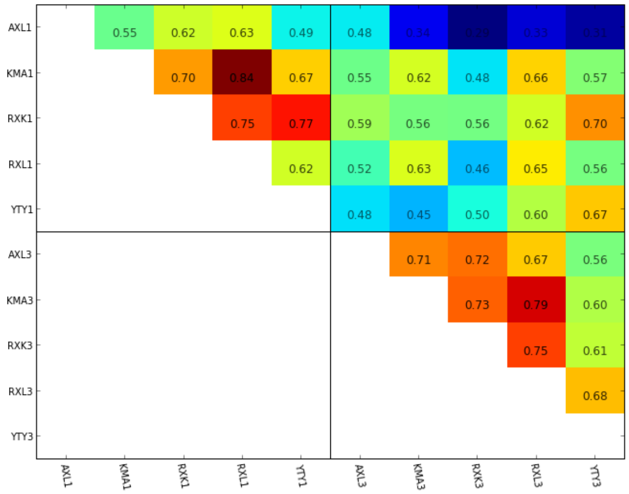


Figure 2.12: Correlation coefficients between the off-diagonal entries (errors) of each of the 10 (5 subjects, single and crowded) normalized confusion matrices. The center dividing lines demarcate the sets of single and crowded results.

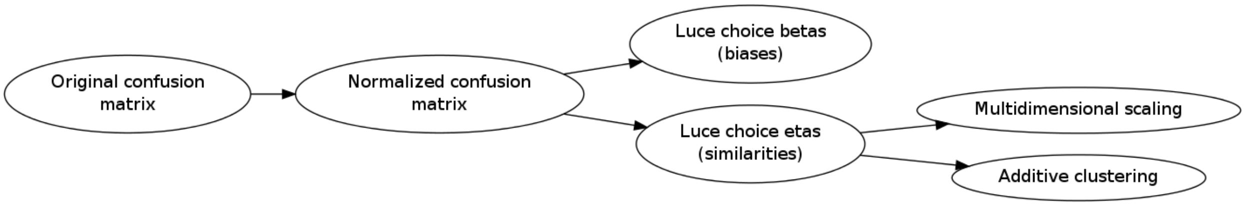

Original confusion matrix
Normalized confusion matrix
Luce choice betas (biases)
Luce choice etas (similarities)
Multidimensional scaling
Additive clustering

Figure 2.13: Processing pipeline of analyses performed.

with the uncrowded errors of the other subjects as with the crowded errors, though many of the other subjects show an identifiable difference between the crowded and uncrowded conditions.

#### 2.2.6.1 Luce choice model

To isolate possible per-subject response biases in the data, as well as to transform the asymmetric confusion matrix into a symmetric "similarity" matrix, the Luce choice model was employed. The confusion matrix for each subject in the two conditions was transformed using the Luce choice model, using the methods described in Section 2.2.4.7. Furthermore, the parameters estimated from the model were analyzed as described earlier. The letter similarities were analyzed with multidimensional scaling (MDS-choice model), and additive clustering. Figure 2.13 depicts a cartoon of the processing pipeline that will be described henceforth.

The outcome of each transformation, the vector of biases $\boldsymbol{\beta}$ and triangular matrix of similarities $\boldsymbol{\eta}$ is given in Section A.2. The model fits the data well, which is not surprising given that the model has so many parameters. Is the model truly capturing "biases" and "similarities" in a meaningful way? Ideally, the biases for a given subject will be similar across the two conditions. This would hold if the vector indeed represents purely cognitive biases that are unrelated to the visual input. The similarities, on the other hand, should reflect perceptual appearance variations between the two conditions. Furthermore, when comparing *across* subjects, the difference based on condition (crowded versus uncrowded) would ideally have commonalities.

To explore these ideas, the Luce choice $\beta$ parameters were plotted in two different ways. Figure 2.14 shows the $\beta$ parameters for the 5 subjects on a separate plot for each condition. There does appear to be some consistency amongst subjects for a given condition, but it is far from perfect. Certain trends are clear. While all subjects have a bias for "k" in the single condition, this pattern is reduced for the crowded condition, while other letters see systematic increased magnitude when crowded, such as "j" and, to a lesser extent, "u" and "w." Figure 2.15 illustrates the *within subject* agreement across the two conditions by plotting the data in a different way. The $\beta$s are in good correspondence between conditions for subjects AXL, RXK and (to a lesser extent) YTY.

The correlation plot of Figure 2.16 enables direct examination of these relationships. First, the observations in the previous paragraph hold. The correlation between the betas

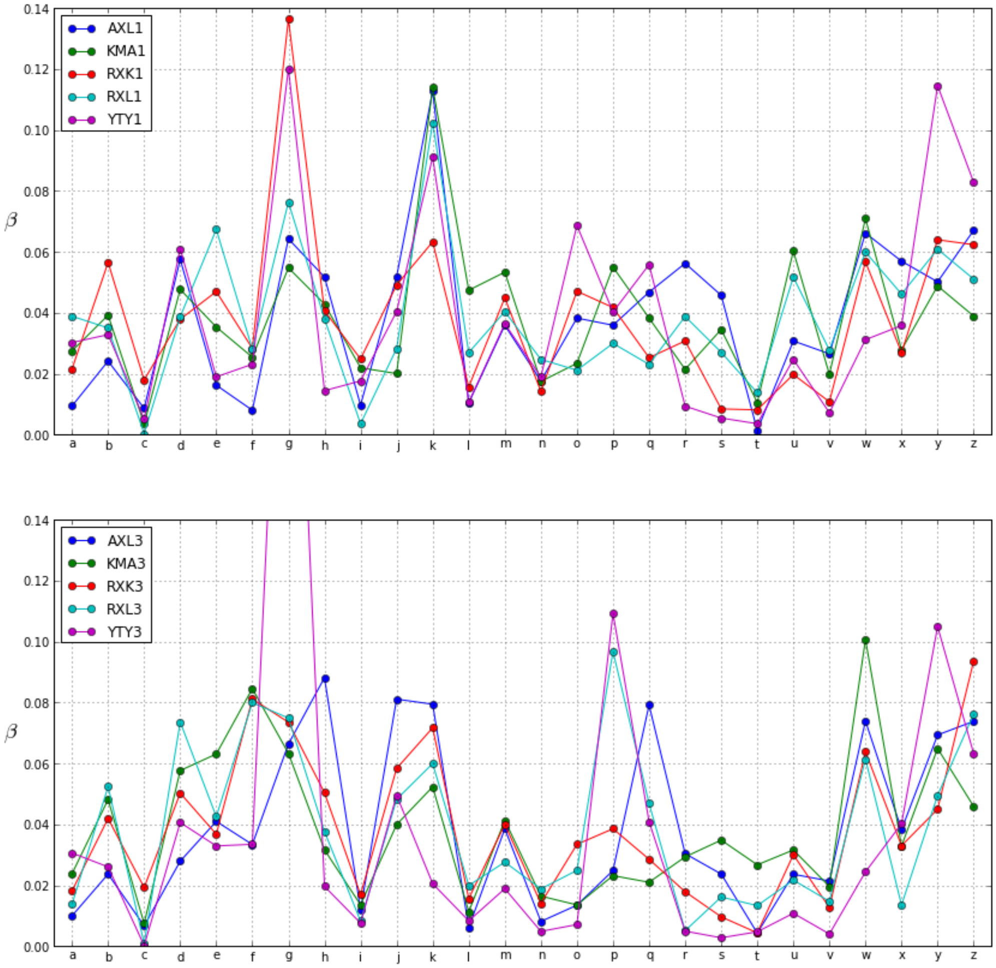


Figure 2.14: Luce choice $\beta$ parameters for each subject, separated by condition. Upper plot shows the single condition, lower plot shows the crowded condition. The value for "g" (exceeding limits of lower plot) for subject/condition YTY3 is 0.29

in the crowded and uncrowded conditions are the greatest for subjects AXL, RXK, and YTY, at 0.74, 0.71, and 0.71, respectively. The other two subjects only achieved correlations of 0.46 and 0.45. Interestingly, these other two subjects (KMA and RXL) correlated well in the uncrowded condition (0.8), but not in the crowded condition (0.63). Generally the across-subject, across-condition correlations (upper right of Figure 2.16) were weaker (<0.65), except for the curiously strong RXK1 versus YTY3 correlation of 0.83.

#### 2.2.6.2 Classical multidimensional scaling

As stated previously, multidimensional scaling is a procedure to reduce the dimensionality of a matrix of stimulus similarities to some smaller number of dimensions that captures "distances" between stimuli. We performed classical multidimensional scaling to the simi-

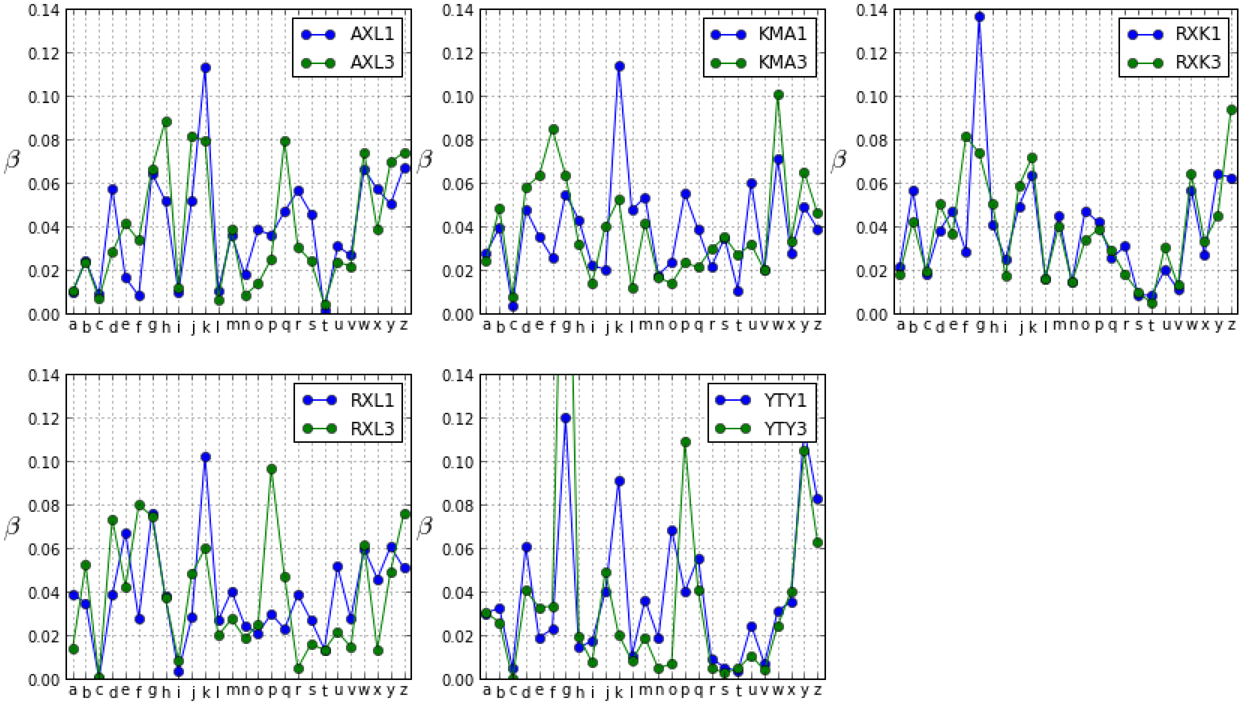


Figure 2.15: Same data as Figure 2.14. Luce choice $\beta$ parameters, arranged by subject, to show difference between conditions.

larity matrices resulting from the Luce choice decomposition of Section 2.2.6.1. With the classical method of MDS, determining the relative contribution of each of the dimensions of the decomposition is straightforward from examination of the eigenvalues of the similarity matrix. A plot of the sorted eigenvalues is known as a "stress plot," and can be used to determine a "sweet spot" for the appropriate number of dimensions, minimizing variance while not over-fitting. Figure 2.17 presents the stress plot for all 10 similarity matrices. While methods vary for interpreting these plots, intuitively 5-10 dimensions would be optimal for these data—where the curves have a "knee" such that there is diminishing returns for adding additional dimensions.

Nevertheless, since our goal is illustrative, we will present the letter proximities using only the first three dimensions. Two or three dimensions are the only possibilities for reasonable graphical display. Figure 2.18 gives the three dimensional decomposition for all 10 matrices. Note that the horizontal axis is the first (or strongest) component of the decomposition. As such, for the uncrowded solution, the first dimension generally captures the letters that are closest to a single ascender–i.e., tall and straight letters like "l,i,t,f." The second, next-significant dimension, tangent to the plane of the page, typically discriminates the small round letters like "e,c,o,s." Finally, the third dimension does vary across subject. For subject RXL, this dimension subsumes both the late-in-the-alphabetic letters "w,x,y,z" while also concerning ascender letters like "h" and "b." For all four other subjects, this dimension

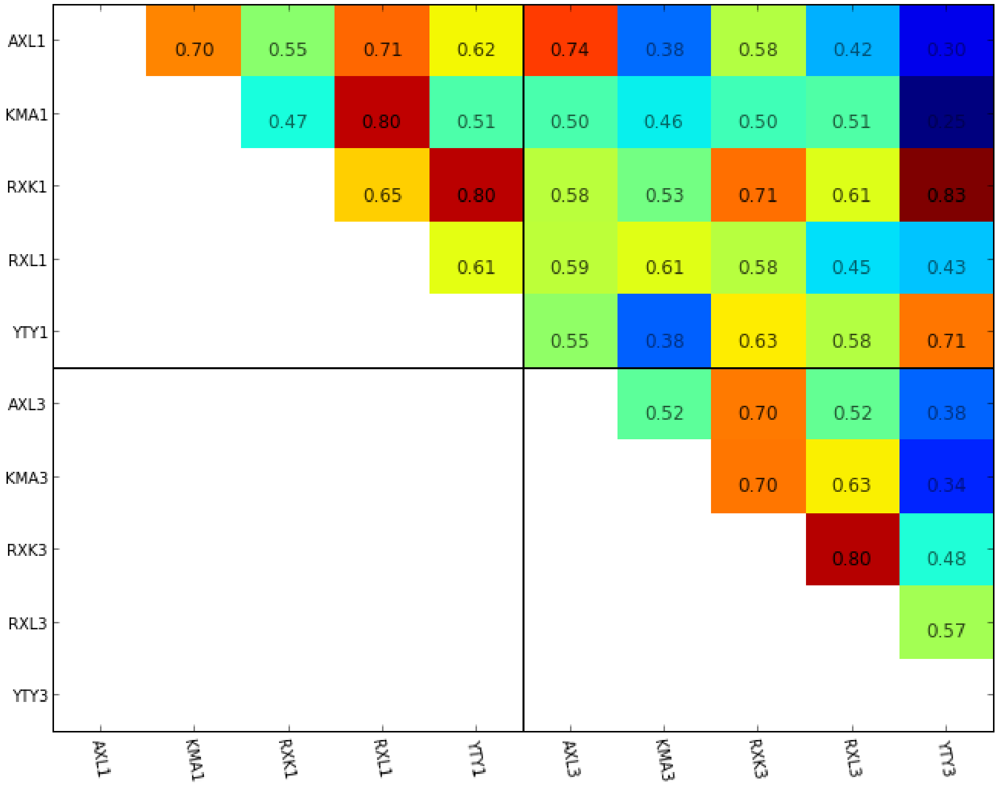


Figure 2.16: Correlation of beta parameters of Luce decomposition for all 10 confusion matrices.

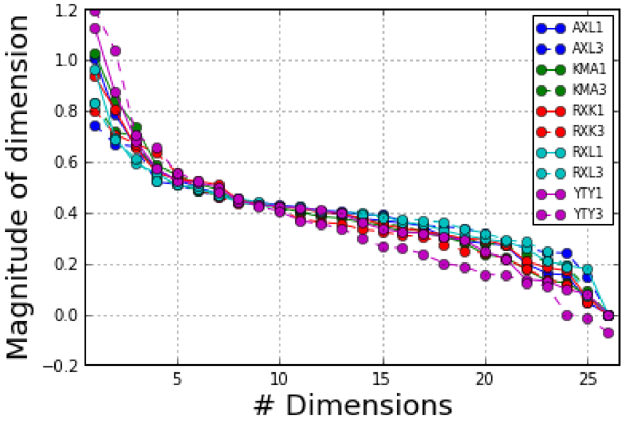


Figure 2.17: "Stress plot" showing influence of each dimension in classical MDS of Luce similarities, for all 10 confusion matrices. Solid lines show single-letter condition, dashed lines show crowded condition for each subject.

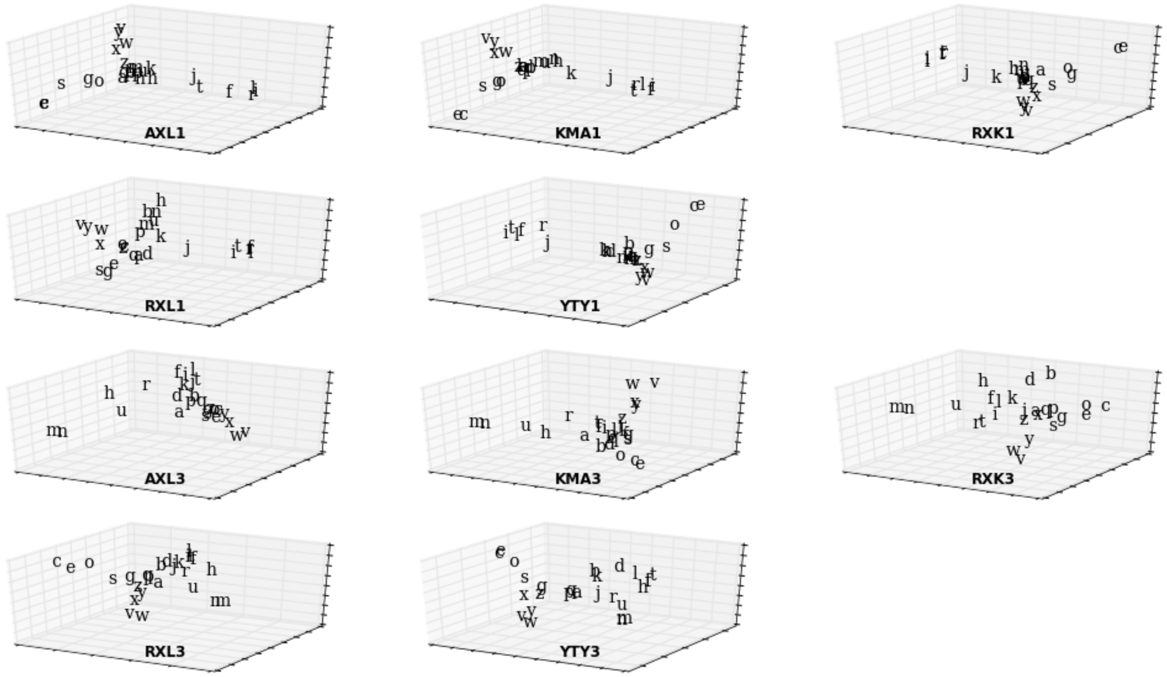


Figure 2.18: Three-dimensional classical MDS of letter similarities for each subject and condition. Ordered by significance are the X (horizontal), Y (tangent to page), then Z (vertical) axis. Top two rows indicate single letter condition, while bottom two rows indicate crowded condition.

almost solely discriminates the later letters from the rest of the laters. A large clump of letters in the "middle" of the plot indicates the need for additional dimensions in order to separate these in space.

The results for the crowded condition (lower two rows of Figure 2.18) are quite different. For this condition, "m" and "n" almost always appear together in an extreme location of the perceptual space, and are noticeably more proximal to "u" and "h" (letters with an arch) in the crowded condition. There is generally more spread in the spatial locations, although the small round letters have a similar proximity and extremal location in both conditions.

#### 2.2.6.3 Additive clustering

As a further attempt to discriminate the pattern of errors in the two conditions (crowded and uncrowded), we applied the additive clustering method described in Section 2.2.4.9 to the ten confusion matrices. Since it is unclear how many clusters are optimal, we first determined the goodness-of-fit (in terms of variance accounted for), of optimal solutions for different numbers of clusters. Figure 2.19 plots the variance accounted for as a function of number of clusters, averaged across subjects, for both the single and crowded conditions. As always, additional clusters will account for more variance, with the risk of overfitting.

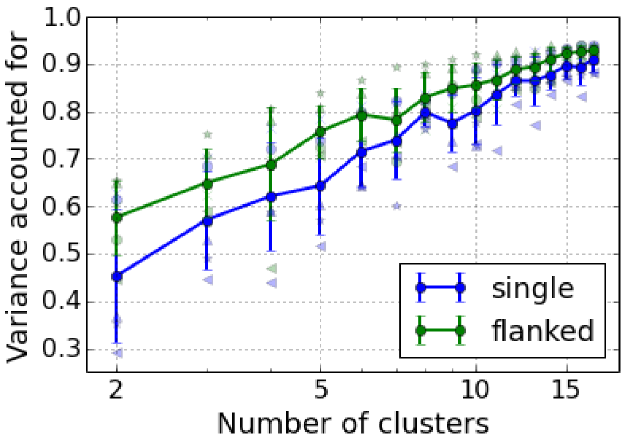


Figure 2.19: Variance accounted for as a function of number of clusters used for additive clustering.

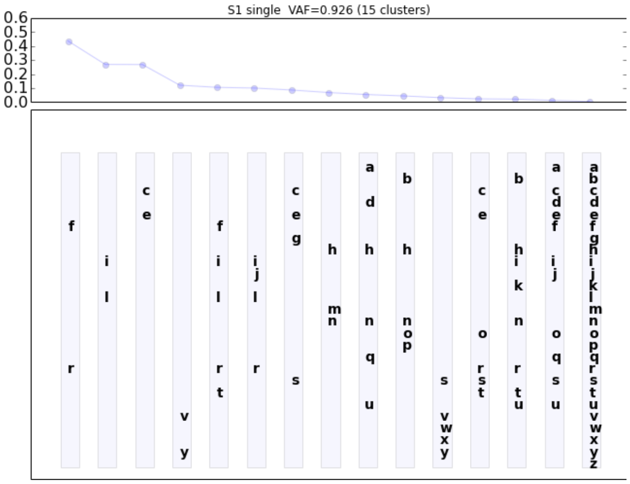


Figure 2.20: Additive clustering solution for single letter. The top curve shows the calculated weights on each cluster. Cluster membership is indicated by letter occurrence in each of the vertical bars, which show the 15 clusters. Clusters are sorted from left to right based on weights.

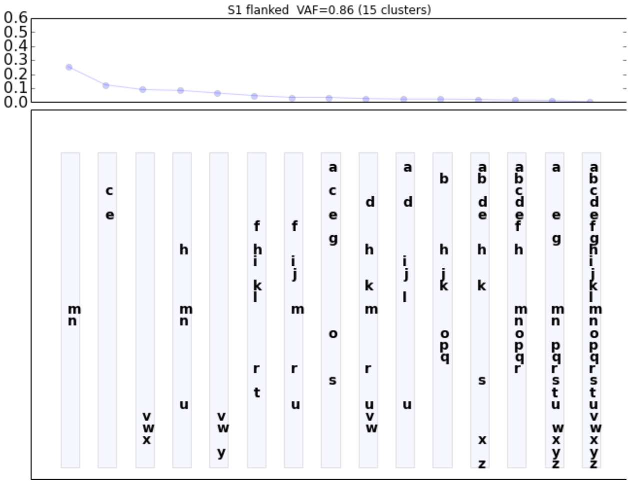


Figure 2.21: Additive clustering solution for crowded letter. See Figure 2.20 for explanation.

Figure 2.20 presents the additive clustering decomposition of one subject in the single condition, with 15 clusters (one of which is the "global" cluster comprising all letters.) Each cluster is indicated by a vertical bar, which are sorted from left to right by decreasing magnitude (or weight). The magnitude is indicated by the line top plot on the top graph. Figure 2.21, on the other hand, shows the decomposition of the crowded confusion matrix for the same subject. There are obvious differences between the two decompositions, which nicely echo the qualitative observations arising from our exploration of the multidimensional scaling method in Section 2.2.6.2. Specifically, the most dominant clusters in the single condition comprise straight-ish letters like "f,r,i,l," while the dominant cluster when crowded (by far) is "m,n." The cluster containing "c" and "e" is common to both conditions, and relatively strong. The group containing letters with arches ("m,n,h,u") has a greater magnitude in the crowded condition, and "u" does not appear when letters appear alone.

To make these observations concrete, Figure 2.22 plots the top twelve clusters that appear in the uncrowded or crowded analysis for subjects. To determine which clusters to include, the total magnitudes of each cluster in the 10 results were summed and thresholded. The ordering of the clusters vertically is arbitrary, an attempt to group them by common members. The observations of the previous paragraph from the single subject are validated in the group data. Clusters containing "c" and "e" are strong in both conditions, but are more pronounced in the single-letter condition. The cluster comprising "m" and "n," on the other hand, are strong primarily in the crowded condition for all subjects. The reason that few other clusters appear is due to individual differences. Clusters that appear for the different subject in the crowded condition are significantly variable.

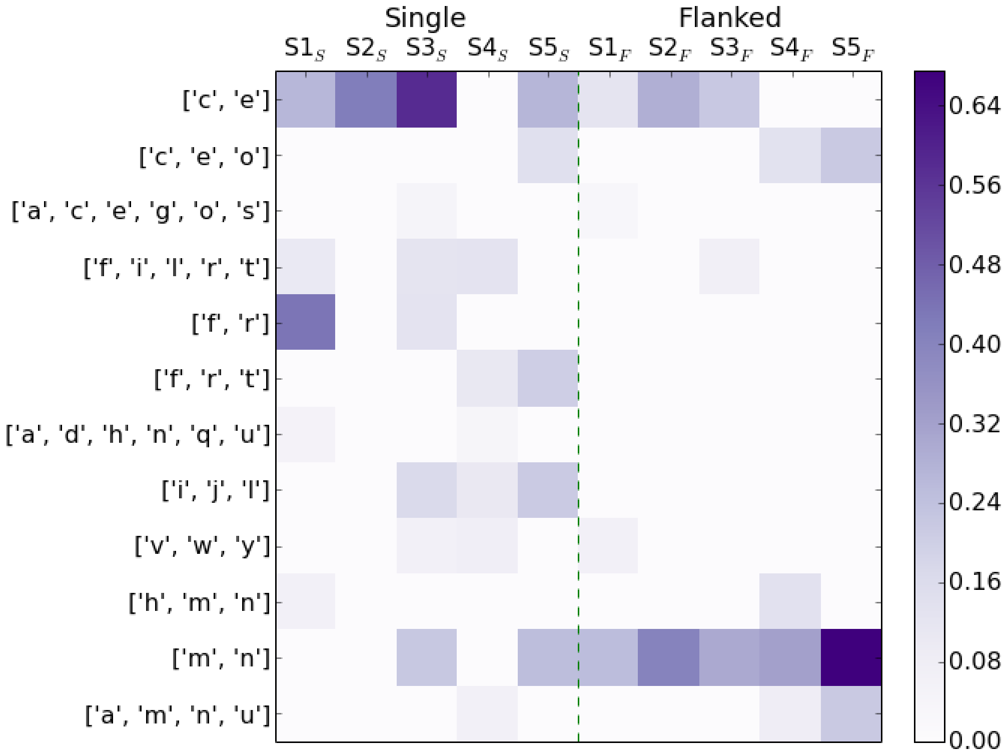


Figure 2.22: Clusters common to both single and crowded condition. Color indicates the magnitude of that cluster for a given subject and condition.

### 2.2.7 Summary

This section summarized some of the key techniques that have been used to analyse error patterns in letter confusion matrices.

The associated experiment and analysis characterized letter confusion error patterns in two conditions: peripheral (10° in the lower visual field) single letters in significant Gaussian noise, and peripheral crowded letters (10° in the lower visual field), with horizontally-oriented (i.e., tangential) flanking letters in modest noise. We looked for differences in: confusion matrices, Luce choice model coefficients, low-dimensional representations of Luce similarity coefficients, and additive clustering analyses of Luce similarity coefficients. The pipeline of analyses is shown in schematic form in Figure 2.13.

There were clear qualitative differences between the two conditions. The crowded confusion of “m” and “n” demonstrates a well-known characteristic of crowding: in a crowded stimulus comprising multiple objects (letters), it is difficult to attribute object parts to the correct object (letter). The fact that these letters have very different overall shape envelopes (Bouma [1971]) means that there were few instances of this confusion cluster in isolated peripheral vision (Figure 2.22). However, when crowded, these errors become commonplace, since the errors induced by crowding expose a different mechanism. Loss of high-frequency detail occurs in crowding as well as uncrowded vision, as evidenced by the "c" and "e" confusion that occurs in both the crowded and uncrowded condition.

## 2.3 Letter recognition: modeling the process

### 2.3.1 Introduction and background

Modeling how letter recognition is impaired in the crowded periphery is a central interest of our lab. Age-related macular degeneration (AMD) is a primary concern, and patients with AMD report difficulty reading as a fundamental complaint. Since these patients rely on their periphery for all visual functions, understanding peripheral impairments could help inform assistive technologies, such as selective enhancement of contours in images for more efficient processing (Kwon et al. [2012]), or font design for resisting the deleterious effects of crowding (Bernard, Aguilar, and Castet [2014]).

A precursor to understanding faulty letter recognition in crowding is an understanding of "normal" letter recognition. Presumably, this would be associated with optimal conditions of foveal viewing, where acuity is best and crowding is not a limiting factor. Unfortunately, there has been no single satisfactory model that explains the letter recognition process. In this chapter I review the menagerie of proposals, focusing on those that I believe to be the most promising.

#### 2.3.1.1 Template matching

A "null model" for letter recognition is the use of static letter image templates that are directly compared to the input image, pixel-by-pixel. In fact, this is how "ideal observer" models of letter recognition function (Parish and Sperling [1991]). The simplest form of this idea is easy to criticize, since the procedure is too sensitive to image manipulations that humans ignore—our vision is remarkably invariant to many types of stimulus manipulations and distortions, like rotation, warping, and more (Neisser [1967]). Furthermore, image templates are able to utilize information from all relevant pixels in a way very different from human recognition. Humans are poor at absolute spatial localization. We excel instead on relative computations, signaled by *differences* between nearby units. The point has been made by Westheimer, and constitutes the basis for our ability to perform hyperacuity tasks (Westheimer [1979b]). On the other hand, there are extensions of template matching that extend its applicability and plausibility, of course. Deformable templates represent the most straightforward extension (Jain, Zhong, and Lakshmanan [1996]; Amit, Grenander, and Piccioni [1991]). In this extension, smooth transformations are allowed on the templates when matching an image, allowing a greater range of stimulus invariances.

Even in its basic form and with its known limitations, template matching has strong proponents (Watson and Ahumada [2012]; Watson and Fitzhugh [1989]; Gervais, Harvey, and Roberts [1984]; Loomis [1990]). The debate of template models versus "feature" models (described later) is concisely summarized in a recent manuscript by Watson and Ahumada ([2012]). These authors remind the reader that the specific putative features necessary for a feature-based model remain elusive, and hence the statistical success of template matching schemes (such as theirs) precludes rejection as a viable model of the letter identification process.

#### 2.3.1.2 Fourier decomposition

Spatial frequency-based Fourier decomposition of letters has been proposed. The most extreme form of a Fourier-based letter recognition approach proposes that the visual system computes a two-dimensional Fourier transform of the input image, and compares the resultant power spectrum to the spectra of stored letter templates. This operation would be one way to achieve positional invariance, and follows from the theoretical ideas put forth by DeValois and DeValois ([1988]). Providing evidence against a pure Fourier approach was Coffin ([1978]), who found that the similarities between frequency spectra of letter templates did not correlate with human errors. Gervais, Harvey, and Roberts ([1984]), however, later showed evidence for the Fourier approach and pointed out several possible oversights of Coffin. Gervais and collaborators believed that Coffin's used of raw correlation values was inappropriate, preferring a Euclidean distance metric instead. Furthermore, Coffin compared results across different fonts and letter sizes, which would affect the results considerably.

The truth of visual processing is likely closer to a compromise between frequency-based approaches and the localized edge filters of Hubel and Weisel. One solution is the usage of Gabor filters, which sample information limited in both space and frequency. These filters have been successful in modeling the responses of cortical simple cells (Marčelja [1980]; Daugman [1985]), and yield efficient codes for images (Daugman [1985]; Daugman [1988]; Olshausen and Field [1996]). Olshausen and Field ([1996]) showed that the filters could emerge naturally from the process of learning a sparse code for natural scenes. Most neurally inspired models of the visual system, such as Neocognitron (Fukushima [1980]) and HMAX (Riesenhuber and Poggio [1999]) include Gabor-like detectors as a first stage of processing.

**2.3.1.2.1 Spatial frequency channels** A related concept is that of analysis based on spatial frequency channels. A groundbreaking result in the psychophysical study of letters was the determination that observers seem to use specific spatial frequency bands for recognition (Parish and Sperling [1991]; Solomon and Pelli [1994]; Chung, Legge, and Tjan [2002]; Majaj et al. [2002]; Gold, Bennett, and Sekuler [1999]). Evidence for this comes from either spatial frequency filtering of the letter stimuli (Parish and Sperling [1991]; Chung, Legge, and Tjan [2002]; Gold, Bennett, and Sekuler [1999]), or noise masking experiments (Solomon and Pelli [1994]; Majaj et al. [2002]) using filtered noise added to the letters. The threshold elevation revealed by the sundry experiments all show that the filters used to identify letters have a band-pass shape, centered somewhere between 1-3 cycles per letter. Importantly, when the same procedure was applied to a "white-noise ideal classifier" by Solomon and Pelli ([1994]), a low-pass shape emerged. In the context of letter recognition, the ideal classifier is defined as a pixel-based template matcher (Parish and Sperling [1991]).

#### 2.3.1.3 Feature-based models

It is natural to hypothesize that letters are made up of simple "features" that are independently detected early in the visual system, then combined (pooled) in some fashion. The most notable theoretical proposal is the early "Pandemonium" model of Selfridge and Neisser (Selfridge [1958]; Selfridge and Neisser [1960]). Their model comprised a stage with "demons" that looked for simple features like edges of a certain orientation. After this stage, a different set of decision demons combined together the responses of the earlier stage to determine which letter was most likely. Their motivation in developing this model was to overcome the problems with the simplest forms of template matching. They recognized that using a hierarchical system such as theirs would enhance recognition by allowing: (1) more invariance in the relative locations of the individual parts, (2) resistance to global deformations, and (3) the recognition of a greater variety of letter shapes (such as optotypes in different fonts).

Eleanor Gibson added careful empirical support (Gibson et al. [1962]), originating from her study of development. She showed how children generalize letter shapes in a way that illuminates the specific components used for recognition. Based on these results, she proposed a set of likely features used to recognize capital letters, which she enumerated in her classic book on perceptual learning (Gibson et al. [1962]). Since then, several researchers have presented abstract feature lists like hers, usually validated using some form of confusion matrix comparison (Geyer and Dewald [1973]; Briggs and Hocevar [1975]; Keren and Baggen [1981]).

The basic idea of a hierarchical model has been bolstered by findings from the mammalian visual system such as Hubel and Wiesel's groundbreaking work in cat and monkey visual cortex (Hubel and Wiesel [1962]; Hubel and Wiesel [1974]). They found that in the primary visual cortex (V1 in primate), neurons that retinotopically tile the visual field seem most sensitive to basic visual forms like edges or bars. A further class of neurons, complex cells, introduce invariance by pooling spatially proximal simple cells with compatible properties. This basic idea of sampling and pooling seems to be repeated further along in the visual system, with the "preferred" features for neurons becoming progressively more complicated.

However, questions have become more subtle as researchers have empirically tested ideas. Firstly, can a single set of features be conclusively identified? Fiset et al. ([2008]) sought to identify features of uppercase and lowercase letters using the "Bubbles" method (Gosselin and Schyns [2001]). With this method, stimulus details are limited by selectively applying Gaussian apertures to the image. By looking at which apertures lead to which responses, the information used by observers is revealed, much like classification images (Eckstein and Ahumada [2002]). Fiset et al. ([2008]) found that terminations were by far the most diagnostic feature that subjects used to identify letters. Similar results were obtained by  Lanthier et al. ([2009]) using a different technique. They manually deleted pixels from capital letter images and found that vertices were the most important cue, as proposed for general object recognition by Biederman ([1987]). In the context of letters, these vertices are typically located at the terminations, validating the findings of Fiset et al. ([2008]).

**2.3.1.3.1 Featural independence** A detail concerning the sufficiency of feature models is whether constituent features are processed independently and uniformly. This is an important issue, since it is a requirement for almost all feature-based models, starting with Pandemonium (Selfridge [1958]), and continuing to (for example) recent models of feature combination in crowding (Greenwood, Bex, and Dakin [2009]). This hypothesis has been tested directly for simple forms made up of line segment features. Featural independence has been shown by some researchers (Wandmacher [1976]; Townsend, Hu, and Ashby [1980]; Townsend, Hu, and Ashby [1981]), but has been challenged by more recent experiments from Townsend and colleagues (Townsend, Hu, and Evans [1984]), who showed interdependencies between feature detection in even simple figures. They also showed that features were not detected uniformly, and in a later study (Townsend, Hu, and Kadlec [1988]) determined that a “payoff” (reward) on individual features biased their detection. Townsend has been by far the most vigorous researcher in studying models of feature-based letter processing.

**2.3.1.3.2 Time course dynamics** There have been proposals about the temporal dynamics of perceptual processing. The prominent theme is that processing starts from the coarse aspects of visual input, moving to progressively finer details, a theory reviewed by Hegdé ([2008]). A model for the dynamics in terms of letter-like shapes was proposed and tested by Lupker ([1979]). A more quantitative validation was developed by Townsend, Hu, and Kadlec ([1988]) with a feature-based alphabet comprising figures made from straight and diagonal line segments.

**2.3.1.3.3 Global/Gestalt features** An enhancement to feature-based models is the use of global or holistic, rather than parts-based attributes. The Gestalt school of thought proposed that parts-based theories cannot explain all visual phenomena—some phenomena, such as phi motion, demonstrate that perception of “wholes” can have primacy over the constituent parts (Wagemans et al. [2012]). A modern proponent of this idea is James Pomerantz, who has shown that *emergent* features of objects, such as closure or other figural relations, pop-out in ways that challenge feature-based models (Pomerantz, Sager, and Stoever [1977]; Pomerantz and Pristach [1989]). Less radically, some researchers simply add global features to the lists of features used by abstract models. Many of the proposed alphabetic feature lists, including those of Gibson et al. ([1962]); Geyer and Dewald ([1973]); Keren and Baggen ([1981]) include Gestalt properties such as symmetry, closure, or parallelism in their list of distinctive features used in letter recognition. However, it is unclear how these emergent properties could arise from a functional model operating on the pixels of an image, so concrete implementations leveraging Gestalt features are lacking.

**2.3.1.3.4 Interactive Activation** The interactive activation model of McClelland and Rumelhart ([1981]) was one of the first operative models of recognition using letters built up from well-defined constituent features. Their features were line segments making up LCD display-like artificial letters from Rumelhart and Siple ([1974]). Unique to their model is the attempt to account for context effects such as the word superiority effect (Reicher [1969]), where knowledge of possible words are known to bias detection of constituent letters. To enable this effect, top-down connections between words units (endowed with a dictionary) modulated the feature detectors units. A limitation of this system is that its alphabet and well-defined feature set have not been extended to lower-case letters.

### 2.3.2 Convolutional neural network model of letter recognition

Feature-based models are the most interesting class of models to us, since they: (1) have some physiological support, (2) can be extended naturally to model (for example) crowding, (3) have intuitive appeal from theory and developmental studies, and (4) can be extended to general-purpose object recognition. Issue (2) arises since many models of crowding assume that features are independently detected, then wrongly attributed between target and flankers (Levi [2008]; Nandy and Tjan [2007]; Pelli, Palomares, and Majaj [2004]).

There is a class of networks known as "convolutional neural networks" that have become very popular in the computer vision community (LeCun et al. [1998]). They descend from the biologically-inspired Neocognitron (Fukushima [1980]), satisfy the criteria from the previous paragraph, and provide several further advantages. First, these networks have already seen successful use in letter and number recognition, so their practicality has led to a number of efficient implementations. Second, they typically comprise multiple layers of processing whose preferred "features" can either be hand-tuned, or can emerge through optimal learning. This conveniently addresses the debate about arbitrary determination of features.

Although these networks are inspired by the mammalian visual system, comparison of their results versus human performance has been (so far) absent. Most researchers are solely interested in achieving maximum performance on a benchmark object recognition dataset. Our goal, on the other hand, was to take an implementation of a convolutional neural network and subject it to a series of "virtual psychophysics" experiments. We were interested in whether the errors the network makes resemble the errors made by humans. To quantify the correspondence, we used the statistical tools of confusion matrix comparison developed in Section 2.2.

Most generally, the name "convolutional neural networks" refers to a class of neural networks that are a subset of traditional neural networks in several key ways that are inspired by biological plausibility. First, the simple cell/complex cell dichotomy observed by Hubel and Wiesel ([1962]) is formalized in discrete layers of the network, and connections are restricted to be spatially proximal, rather than the all-to-all connectivity of traditional neural networks. Typically, multiple layers of simple and complex cell arrays are stacked, a caricature of the observed hierarchical structures of visual cortex (Felleman and Van Essen [1991]).

Figure 2.23 is a cartoon depiction of a convolution network. The pixels of a target image are used directly as input. The first layer contains nodes akin to simple cells, typically exhibiting edge-like preferences. An array of these nodes tiles the space (matching retinotopic organization) in grid-like fashion, and convolve the input image at each location with their small receptive fields. The output of the array of cells with a given feature is thus a 2D

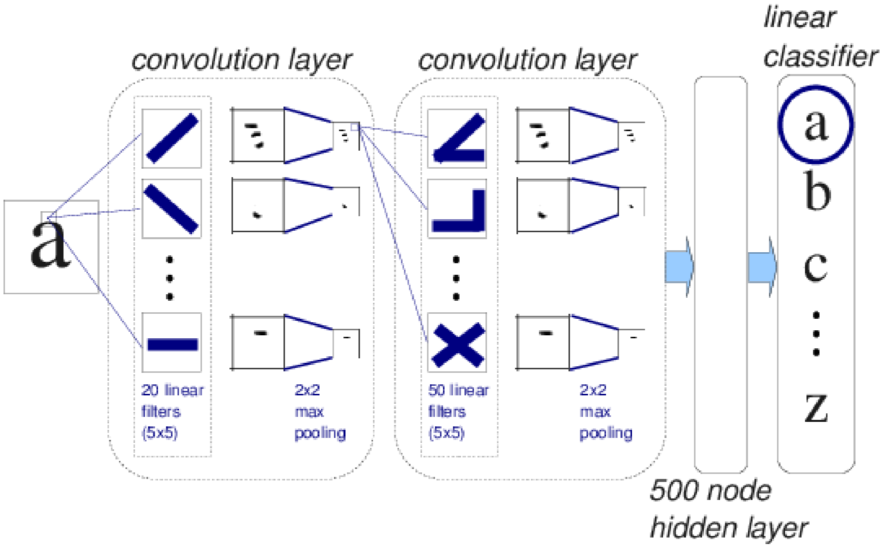


Figure 2.23: Schematic diagram depicting convolutional network architecture. See text for details.

representation of how well the image matches its preference at each location. For example, one set of detectors would represent the amount of "horizontal" energy across the image.

Groups of simple cell outputs that are spatially proximal (and potentially proximal in feature space, i.e., similar orientations) are then pooled together by the next layer, representing the complex cell operation. A single output is determined for the aggregate using an arbitrary operation such as a maximum or average. This step both introduces spatial invariance and incorporates the non-linear operation necessary for interesting behavior.

The number of cells and spatial density in each layer is arbitrary. It could be hand-tuned to the particular input, for example, or matched to physiology. An arbitrary number of these convolution+pooling layers are stacked, with the number of cells in each area decreasing. A side-effect of this structure is that the spatial extent of neurons in each successive layer will increase in terms of the "retinal" input space. This matches observations from neurophysiology (Freeman and Simoncelli [2011]) of the change in receptive field sizes from V1, V2, to V4.

Lastly, a decision stage must determine the output of the whole network for a given pattern. For this network an additional layer of "hidden nodes" connects the output of the last convolution later to each possible output, 26 letters in our case. The network is trained using substantial labeled training data, with connection weights in the entire network modified to optimize performance in the task of letter recognition. The hidden nodes are typically not justified or analyzed–rather they represent an opaque mechanism whereby the convolution layers outputs are optimally utilized, diverging somewhat from biology.

#### 2.3.2.1 Biological IMplausibilities

Before proceeding, I believe it is important to acknowledge the ways in which these models differ from known details of biological visual processing, and will briefly list a few of these.

**2.3.2.1.1 Feedforward architecture.** It is well known that a huge number of connections in the brain are feedback or lateral connections. For example, retinal input constitutes only 10% of the connections to the LGN (lateral geniculate nucleus), which is typically described as simply a "buffer" before visual cortex. The remaining 90% are back-projections from the cortex and other areas, with unknown function (Guillery and Sherman [2002]). Adding feedback or lateral connections greatly complicates the operation of networks, due to the introduction of temporal dynamics. The strict hierarchical nature of cortex is also becoming increasingly challenged as a sufficient description of cortical architecture (Hegde and Felleman [2007]).

**2.3.2.1.2 Temporal dynamics.** Temporal dynamics is ignored in typical convolutional neural networks. That is, all computations are performed at once, in a deterministic way. Many realistic models of neural dynamics involve different time courses of inhibition and excitation, which could contribute to center/surround interactions, and/or coarse-to-fine processing. The original Neocognitron (Fukushima [1980]) was a "relaxation" network, involving complex dynamical interactions that eventually settled into a stable state. The most famous relaxation network is the Hopfield Network (Hopfield [1982]). The Interactive Activation model of McClelland and Rumelhart ([1981]) is a particularly relevant example. Finally, the family of "ART" models developed by Grossberg (Grossberg [1987]) represent a unique class of models that balance top-down and bottom-up influences. The success of application-specific neural networks from the computer vision community is partly due to the greater controllability of static networks.

**2.3.2.1.3 Neural details.** Although the stereotype simple/complex cell dichotomy is maintained, there is greater heterogeneity of cells in actual brains, including the following details that are typically overlooked in neural networks, including convolutional ones. It is unclear if any of these effects will turn out to be crucial for a full understanding of neural processing.

- Inhibitory versus excitatory signals
- Laminar architecture
- Spike-timing dependent plasticity
- Information in spikes versus simple rate coding
- Receptive field sizes grow with eccentricity. Heterogeneity would complicate the model, and hinder parallel-processing implementations.

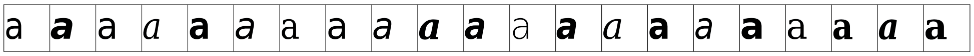


Figure 2.24: Lower-case letter "a" in the 21 Deja family fonts used.

- Significant redundancy/overlap in cortical columns. There have been some interesting recent proposals that introduce overlap into convolutional neural networks (Ngiam et al. [2010]; Hyvärinen and Hoyer [2001]; Kavukcuoglu et al. [2009]). Intriguingly, results have shown the development of topographic organization resembling the pinwheels of V1. Further, as overlap is increased, models begin to resemble population code methods (Averbeck, Latham, and Pouget [2006]).

## 2.3.3 Methods

Despite a few divergences from known neuroscience, the convolutional neural network still has the advantages that it matches the gross architecture and function of the visual system reasonably well, and has been very successful in identifying letters and numbers, even handwritten ones. I will now describe our experiments comparing the neural network against human results. This involved testing the network with a variety of stimuli, and evaluating its "behavior." That is, we analyzed the results of its operation, especially identification errors. There were two types of analyses performed. We evaluated the critical frequency band used for identification by adding filtered noise to the network's test input. We also compared the confusion matrices of the convolutional neural network with several human confusion matrices.

### 2.3.3.1 Synthetic stimuli

To verify that the network recognized single letters in a translation-invariant way, we introduced stimulus variations that would confound typical schemes such as pixel templates. Stimulus variations during training also make the network more robust to letters it hasn't seen. Our generated stimuli had multiple degrees of variations. First, the font was randomly chosen from members of the set of "DejaVu" fonts, high-quality free fonts available on Linux systems. Members of this set include monospace, serif, sans-serif, and bold, italics, and bold-italics versions of each character, shown in Figure 2.24. For these experiments, we restricted font choice to either serif or `monospace` versions of the letter in training and testing, allowing **bold**, *italics*, and ***bold-italics*** variations of each. The size of each letter was random, between 17 and 26 pixels. Each letter was randomly located in the image array (with care taken to avoid cutting off any parts of a letter). Each letter was rotated a random angle between -25 and +25 degrees. Lastly, random noise of restricted spatial frequency content was added to the image. Pseudo-code of the stimulus generation process is provided below. Example of stereotypical letter images are presented in Figure 2.25.

For training, a large set of letter images (72,000 stimuli) was generated and presented to the network. The network connections were then "frozen" and testing commenced, using multiple small test sets (2,600 stimuli each), generated with per-condition parameters. Most training parameters were the same for all experiments, with the variations described above. For the critical band experiments, two sets of additional experimental parameters were manipulated for the testing phase, the noise characteristics and target versus background pixel intensities: see Section 2.3.4 for details.

Given: Number of stimuli, set of fonts, letter size range, noise standard deviation, noise spatial filter type and frequency, letter contrast (versus noise).

Generate ideal filter in the Fourier domain. A 28×28 image is created, with zeros at all radii for frequencies that are stopped, and ones for all frequencies that are passed. Bandpass filters were one octave wide.

For each stimulus:

1. Render random character ("a"-"z") in random font at random (x,y) location in 28×28-pixel square bitmap, ensuring that no pixels of the letter are cut off.
2. Rotate letter image to a random angle within range specified.
3. Generate a noise image mask as specified by experiment, filtered in frequency domain. Create an array of 28×28 normally-distributed numbers. Multiple the FFT of the noise by the noise filter created previously, then perform inverse FFT to yield spatial noise image.
4. Scale letter image to desired numerical intensity (using contrast amount specified for that experiment) and add noise image.

#### 2.3.3.2 Network implementation

To implement the convolutional neural network, I used Python, with Numpy and Scipy mathematical extensions (Oliphant [2007]), PIL (Python Imaging Library), and the Theano library (Bastien et al. [2012]; Bergstra et al. [2010]). Theano provides a pedagogical convolutional network example known as "LeNet," which I used as a starting point. Theano, developed at the University of Toronto, offers several very attractive advantages over competing frameworks. First, the core system has been built with high performance in mind. Network models run automatically on a GPU without any modification. Since operations on an array of nodes (such as the neurons in a layer) are homogeneous, this particular application is ideal for GPU-based execution. Implementation of neural networks using Theano merely involves specifying the network architecture, connection patterns, and learning rules. Thus, execution of the network model is handled internally by a highly optimized mathematical

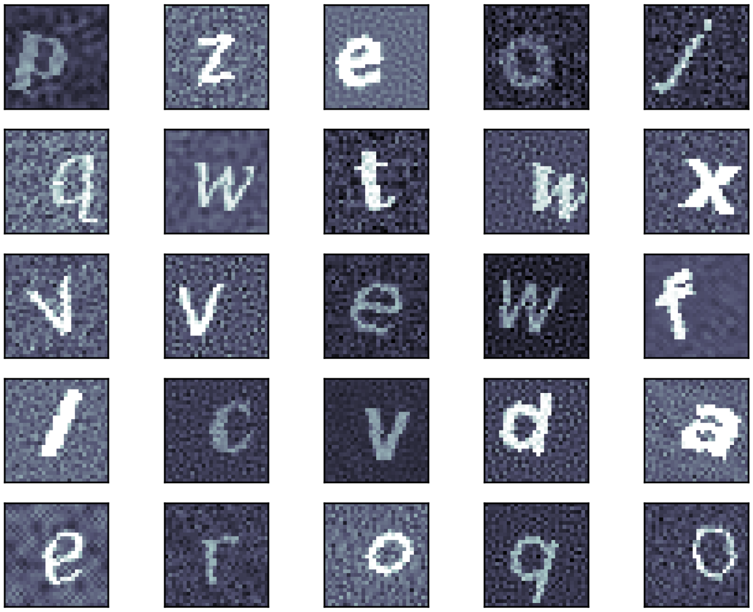


Figure 2.25: Example stimuli generated for training and testing convolutional network

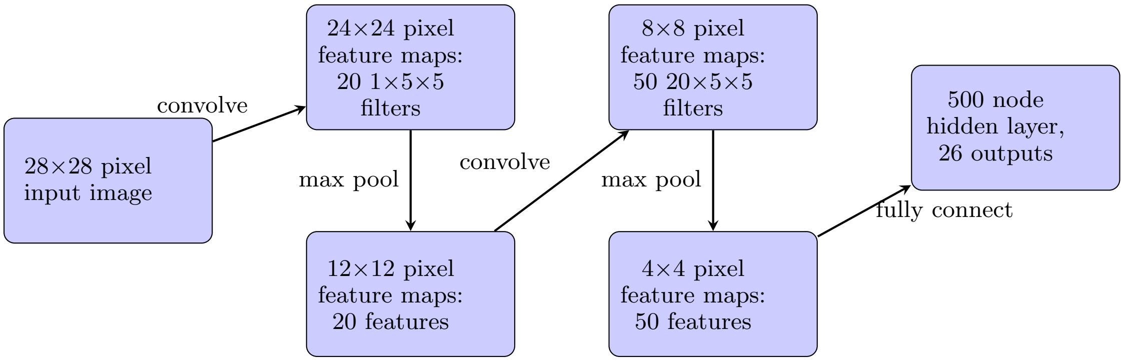

28×28 pixel input image
convolve
24×24 pixel feature maps: 20 1×5×5 filters
max pool
12×12 pixel feature maps: 20 features
convolve
8×8 pixel feature maps: 50 20×5×5 filters
max pool
4×4 pixel feature maps: 50 features
fully connect
500 node hidden layer, 26 outputs

Figure 2.26: Schematic showing implementations details of convolutional neural network.

kernel.

Specifics of the system architecture details are given in Figure 2.26. The network consists of two stacked "simple+complex" layers. Each of these first contains 5×5 filters that convolve over spatial locations on their input. There are 20 filter types in the first layer and 50 filter types in the second layer. The output of each of these convolution layers are then downsampled, with a "max" taken on each 2×2 cell neighborhood. The resultant feature map images are thus quartered. The output of the second convolution layer is then sent to

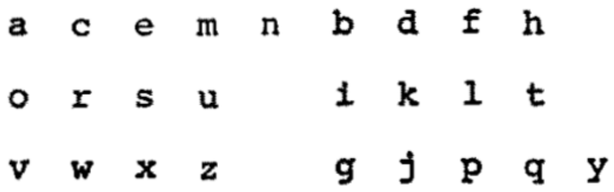


Figure 2.27: Lower-case characters in Courier font used by Bouma ([1971]).

a fully-connected 500 node hidden layer, a traditional Perceptron-like neural network layer with a **tanh** nonlinearity. Finally, these 500 nodes are connected to 26 letter outputs, in effect performing logistic regression to yield a letter identity.

During the training procedure, labeled letters images are shown to the network. Parameters are updated using stochastic gradient descent with a batch size of 500. That is, 500 letter images (and their identities) are given to the network at once, and all parameters are updated simultaneously. This includes filter weights for the convolutional layers, feature biases for the pooling layers, input weights for the hidden layer, and weights from the hidden layer to the 26 outputs.

#### 2.3.3.3 Validation with empirical confusion matrices

To test whether errors made by the convolutional neural network were similar to the errors made by human observers, we compared model confusion matrices against human data, using the methods of Section 2.2. We used classic confusion matrices from the literature and experiments from our laboratory. Specifically, we used the two confusion matrices published in Bouma ([1971]), obtained by averaging the confusion matrices of ten subjects in two different conditions, involving lower-case letters in a Courier font, shown in Figure 2.27. One condition involved foveal viewing of small letters. The height of a lower-case "x" was 2 minutes of arc, and the stimulus presentation duration was three seconds. In the other condition letters appearing to either the left or right of fixation randomly, with an x-height of 13.5 minutes and a duration of 200 msec.

The human empirical data we utilized came from the experiment already described in Section 2.2.5. Briefly, five subjects viewed either single or crowded letters in the Times font located at ten degrees in the lower visual field. When crowded, flankers appeared to the left and right of the target at 1.2× the x-height of the letter. The stimulus duration was 150 ms, and errors were induced by adding Gaussian pixel noise to the letters. Different amounts of noise were added to each subject to keep percent correct relatively consistent between 50% and 60%, as shown in Table 2.1. To anticipate the Results section, analysis was performed both *within* and *across* experimental conditions. Since five observers participated in our experiments, their individual data can be compared/contrasted to each other within a given experiment. The two confusion matrices from Bouma ([1971]) are evaluated in the same way.

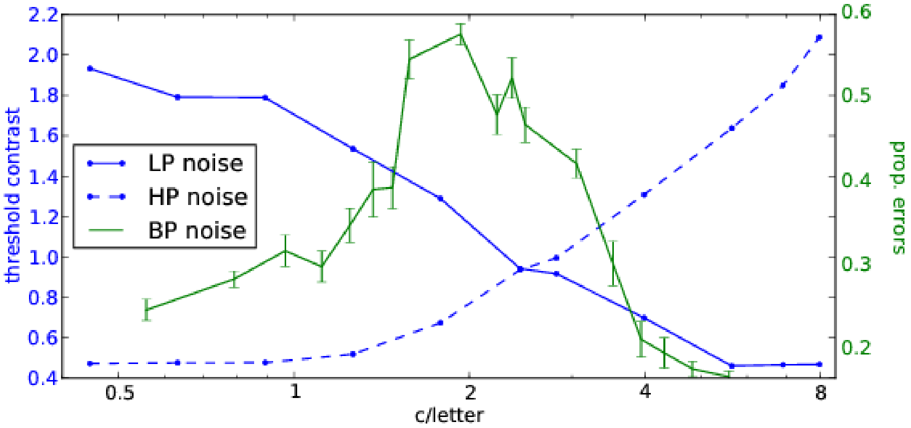


Figure 2.28: Bandpass "critical band" revealed using noise masking. Each curve indicates a different type of noise masking, low-pass (LP), high-pass (HP), or 1-octave band-pass (BP), as indicated by figure legend. For the low-pass and high-pass noise, threshold target contrast for identification is plotted (see text for details). For the band-pass noise (green curve), average percent correct was evaluated in ten separate runs—error bars indicate standard deviation across runs.

### 2.3.4 Results: critical spatial frequency channel

As discussed in Section 2.3.1.2.1, the use of tuned spatial channels for letter identification by human observers has broad empirical and theoretical support. The shape of the spatial tuning filter for the convolutional neural network was thus of prime interest to us. For the network to exhibit human-like performance, it should be subject to similar limitations as human observers, and one of these limitations is the band-pass spatial filter that has been identified.

To determine the critical band of the network, we used spatially filtered noise in two ways. First, following Solomon and Pelli ([1994]), we added either low-pass or high-pass filtered noise as a mask. White noise (unique to each trial) was filtered in the frequency domain using an ideal filter centered at one of 11 logarithmically spaced cutoff frequencies, using the methods described in Section 2.3.3.1. Each simulation run tested a given pixel intensity level (versus the background noise). There were 9 levels logarithmically spaced in arbitrary pixel intensity units. A cumulative Gaussian was fit to the average percent correct results, and a threshold pixel intensity (or threshold "contrast") was identified for each spatial frequency band. The results are shown as the blue curves in Figure 2.28.

Alternatively, in separate simulations we added bandpass filtered noise (1 octave wide) at one of 18 different spatial frequencies, and recorded the percent correct outcome across the test set. This result is shown in Figure 2.28 as the green curve. Results from both techniques revealed a similar critical filter, approximately 1-2 octaves wide, with center frequency between 1.5 and 3 cycles per letter, nearly identical to the previous results from

| Identifier | Font | Individual | Notes |
|---|---|---|---|
| UCB1 | Times | 5 subjects | single letter in noise, 10° lower visual field |
| UCB3 | Times | 5 subjects | crowded letters in noise, 10° lower visual field |
| `Bouma` | Courier | 2 matrices | foveal, and 7° left or right VF, 10 subjects averaged |
| `ModelC` | Courier | n/a | n/a |
| ModelT | Times | n/a | n/a |

Table 2.2: Summary of five result-sets used for confusion matrix comparison

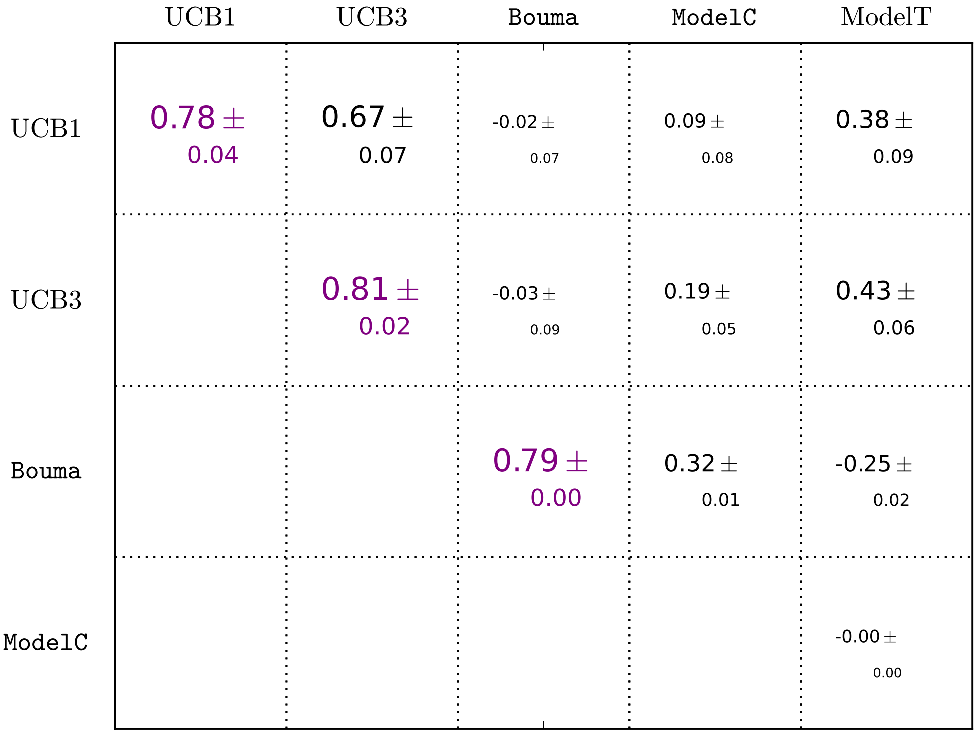
human psychophysics.

### 2.3.5 Results: confusion matrix comparison

The results described in Section 2.3.4 showed that the network broadly uses the same information in the stimulus as human observers, and has similar limitations in terms of spatial frequency. However, can a more detailed comparison be made? The strongest comparison is to look at the specific errors made by the network, and compare to the errors made by human observers. This section presents a comparison of the confusion matrices of the convolutional network to several results from human experiments. The fonts used in each of the experiments is of particular interest. Specifically, we compared five sets of confusion matrices for lowercase letters. They are listed in Table 2.2.

The first two confusion matrix sets were from new experiments from our lab described in Section 2.3.3.3. The next set of matrices came from Bouma ([1971]). Since these three result-sets included multiple individual instances (our five different subjects for each of the two conditions) or separate conditions (Bouma's two conditions), these matrices were also compared within-experiment. We evaluated two variants of the convolutional network model. One version of the network, denoted "ModelT", was trained and tested with a Times serif font, and the other one, "`ModelC`" was trained and tested with a monospace Courier-like font. The convolutional neural network confusion matrices were not compared within-experiment.

Figure 2.29 shows how the diagonals of the five confusion matrices (correct responses) correlate both *within* and *between* the five experiments. These values are the standard correlation coefficients of the 26-entry vector indicating proportion correct for each letter. The purple entries on this figure show the within-experiment correlation. The remaining entries denote correlations of average results across experiments. The size of each number is based on the magnitude of the value, for visualization. First, note that the correlations of the subjects within experiments was quite high: approximately 0.8. Between experiments, only the two Berkeley experiments had a large (0.67) correlation. Fittingly, the two Berkeley experiments (using Times font) correlated better with the Times model (0.38, 0.43) than the Courier model (0.09, 0.19), while the Bouma experiment (Courier font) correlated better with the Courier model (0.32) than the Times model (-0.25).



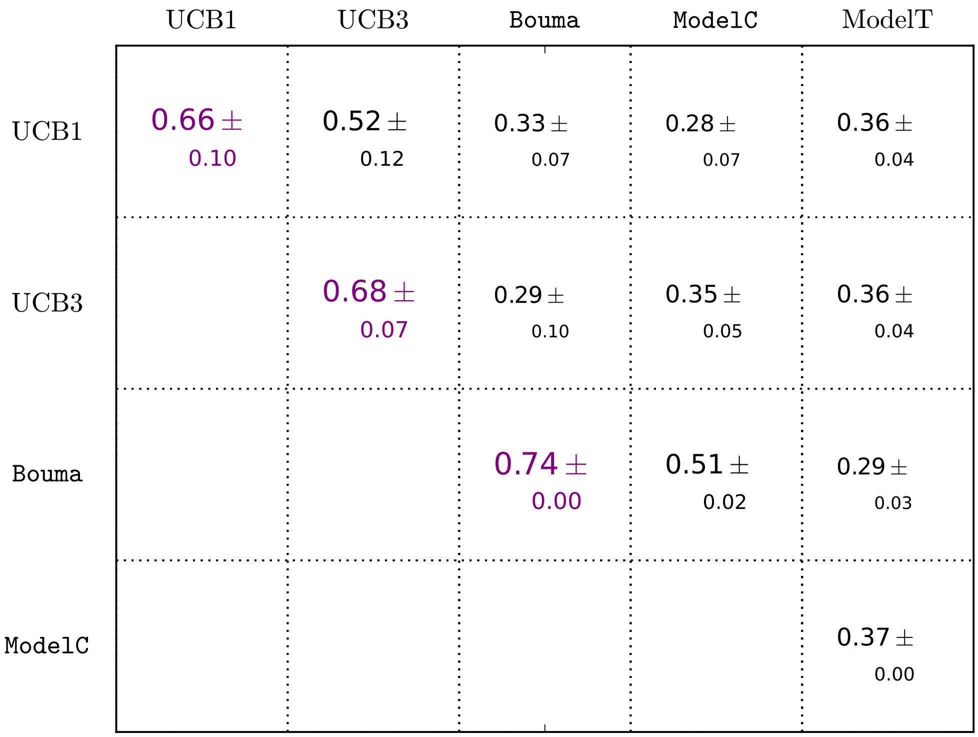
| | UCB1 | UCB3 | Bouma | ModelC | ModelT |
|---|---|---|---|---|---|
| UCB1 | 0.78 ± 0.04 | 0.67 ± 0.07 | -0.02 ± 0.07 | 0.09 ± 0.08 | 0.38 ± 0.09 |
| UCB3 | | 0.81 ± 0.02 | -0.03 ± 0.09 | 0.19 ± 0.05 | 0.43 ± 0.06 |
| Bouma | | | 0.79 ± 0.00 | 0.32 ± 0.01 | -0.25 ± 0.02 |
| ModelC | | | | | -0.00 ± 0.00 |

Figure 2.29: Correlation of diagonal entries (correct responses) of confusion matrices. Entries in purple indicate *within- experiment* correlations, while black entries shows average correlation *between* experiments. The size of each number is scaled according to its magnitude.

Figure 2.30 repeats this procedure for the off-diagonal entries of the five confusion matrices, which comprise the error cells. Again the within-experiment correlations were large (0.66-0.74), but the between-experiment correlations did not segregate by font quite as well as in Figure 2.29. Errors from the two Berkeley experiments correlated only slightly better with the Times model (0.36, 0.36) than the Courier model (0.28,0.35). The Bouma errors correlated much better with the Courier model (0.51) than the Times model (0.29).

Next we applied the Luce biased choice model (Section 2.2.4.7) to the five matrices, and examined the correlation of the fit parameters. First, the correlation of the bias parameters is shown in Figure 2.31. Interestingly, the within-experiments correlations are still relatively large (0.32, 0.57, 0.64), but smaller than the within-experiment correlations for the raw matrix entries. The only other large correlation of bias parameters was between the two Berkeley experiments. The rest of the correlations were small (<0.16). The strong anti-correlation between the Bouma results and the Times model (-0.46) is curious. The non-correlation is satisfying, if the parameters truly reflect subject response biases rather than perceptual attributes.

Finally, we correlated the similarity parameters, presented in Figure 2.32, which ideally



| | UCB1 | UCB3 | Bouma | ModelC | ModelT |
|---|---|---|---|---|---|
| UCB1 | 0.66 ± 0.10 | 0.52 ± 0.12 | 0.33 ± 0.07 | 0.28 ± 0.07 | 0.36 ± 0.04 |
| UCB3 | | 0.68 ± 0.07 | 0.29 ± 0.10 | 0.35 ± 0.05 | 0.36 ± 0.04 |
| Bouma | | | 0.74 ± 0.00 | 0.51 ± 0.02 | 0.29 ± 0.03 |
| ModelC | | | | | 0.37 ± 0.00 |

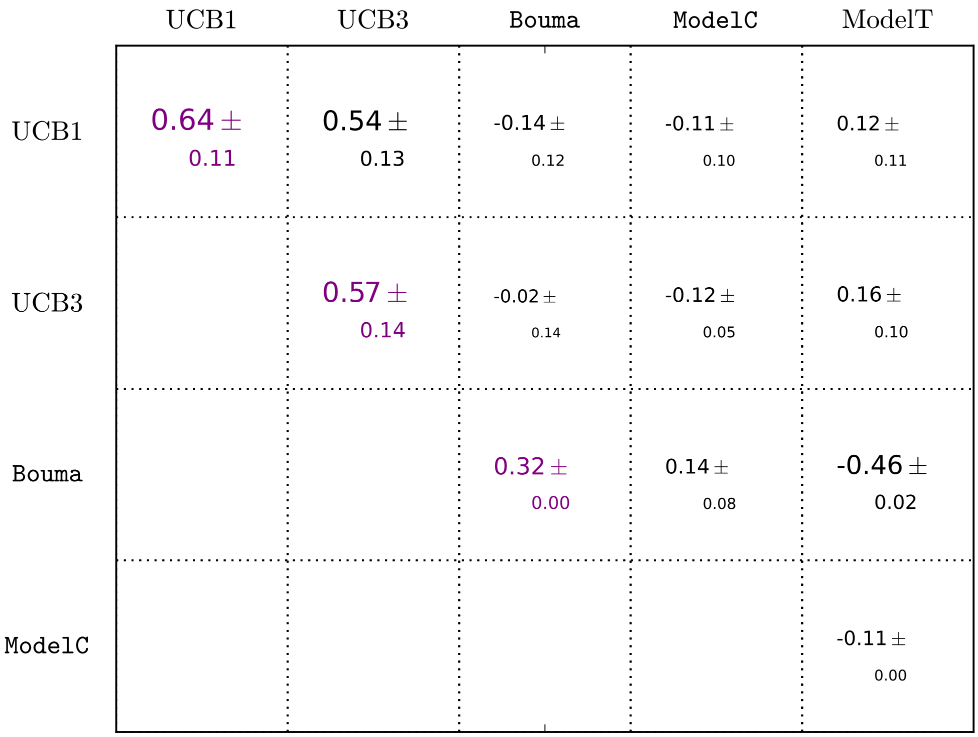
Figure 2.30: Like Figure 2.29, but for off-diagonal entries of confusion matrices only.

will be purely based on the stimuli, without bias. The highest within-experiment correlations so far are observed: 0.79-0.85. The Berkeley experiment with single letters correlates better with the Times model (0.61) than the Courier model (0.4). The crowded Berkeley experiment, however, correlates approximately equivalently with both models (0.62, 0.58). The Bouma results correlate much better with the Courier model (0.61) than the Times model (0.38). These results will be discussed further in the next section.

### 2.3.6 Discussion

The good match of the spatial frequency channels to human data was a success. Like Solomon and Pelli ([1994]), we created a visualization of noise band and contrast, similar to the demonstrations that show contrast sensitivity using contrast-modulated grating of different spatial frequencies. Figure 2.33 plots example letters in low-pass noise stimuli, with spatial frequency represented by the columns. The target letter contrast decreases from bottom to top. The overlaid color contour lines show model performance (proportion correct), smoothed to fit the two-dimensional profile. The message of the demo is intuitive—stimuli which we find harder to discriminate challenge the network as well.

An important outstanding issue is the source of the band-pass effect in the convolutional network. Further experiments are needed to identify if the critical band reflects information

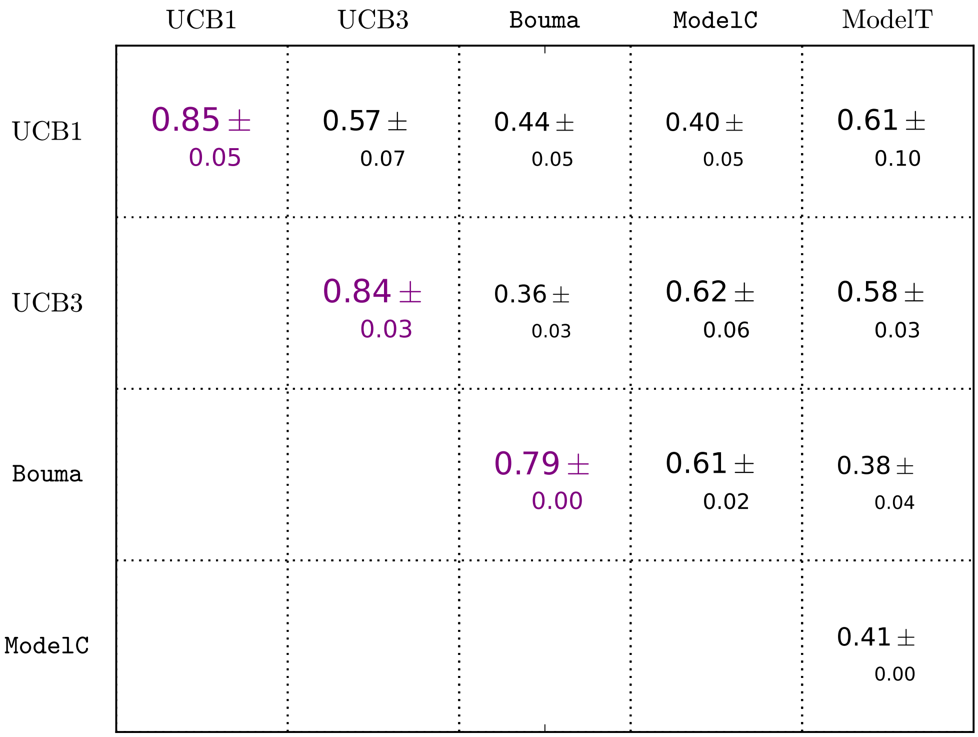


| | UCB1 | UCB3 | Bouma | ModelC | ModelT |
|---|---|---|---|---|---|
| UCB1 | 0.64 ± 0.11 | 0.54 ± 0.13 | -0.14 ± 0.12 | -0.11 ± 0.10 | 0.12 ± 0.11 |
| UCB3 | | 0.57 ± 0.14 | -0.02 ± 0.14 | -0.12 ± 0.05 | 0.16 ± 0.10 |
| Bouma | | | 0.32 ± 0.00 | 0.14 ± 0.08 | -0.46 ± 0.02 |
| ModelC | | | | | -0.11 ± 0.00 |

Figure 2.31: Correlation between and across experiments of Luce choice fit bias ($\beta$) parameters.

in the stimulus, the V1-like receptive fields of the network, or if it has some other origin. To address the first possibility, Solomon and Pelli ([1994]) stated that the low-pass nature of the ideal classifier proved that the critical band reflected limitations inherent in the visual system. Later however, Chung, Legge, and Tjan ([2002]) showed that the ideal-observer for letter recognition actually has a band-pass shape, called the *letter sensitivity function*, or ideal LSF. The shape of the LSF determines which frequencies are most informative for making discriminations concerning letter identity. The origin of its shape may be related to the frequency content in the letter stimuli themselves. To test this, Põder ([2003]) analyzed the spatial frequency content of letter images, and indeed found band-pass energy, instead of broad-band or 1/f characteristics. Fuller understanding of the spatial frequency channels of the convolutional neural network could be obtained by testing a greater range of letter sizes, with respect to the overall image size and first layer filter sizes.

Although the convolutional neural network we used was a biologically plausible model of visual cortex, no effort was made to model other aspects of the visual pathway, such as the contrast threshold to identify gratings, quantified as the human contrast sensitivity function (CSF) (Campbell and Robson [1968]). A full model of the letter identification process should incorporate additional stages of processing (Beckmann and Legge [2002]), and it has been proposed that the CSF (plus the LSF described in the previous paragraph) could best account



| | UCB1 | UCB3 | Bouma | ModelC | ModelT |
|---|---|---|---|---|---|
| UCB1 | 0.85 ± 0.05 | 0.57 ± 0.07 | 0.44 ± 0.05 | 0.40 ± 0.05 | 0.61 ± 0.10 |
| UCB3 | | 0.84 ± 0.03 | 0.36 ± 0.03 | 0.62 ± 0.06 | 0.58 ± 0.03 |
| Bouma | | | 0.79 ± 0.00 | 0.61 ± 0.02 | 0.38 ± 0.04 |
| ModelC | | | | | 0.41 ± 0.00 |

Figure 2.32: Correlation between and across experiments of Luce choice fit similarity ($\eta$) parameters.

for the exact band-pass shape of the critical band (Chung, Legge, and Tjan [2002]). It is unclear what effect incorporating the CSF would have on a convolutional network.

Several observations can be made about the confusion matrix comparison results. First, the correlations of the parameters within an experiment (between subjects or conditions) were smaller for the bias parameters than both of the raw matrix correlations. This, along with the small between-experiment bias parameters, suggest that the biases could be less dependent on the stimuli, and indeed reflect internal response biases, a stated goal of the Luce choice model. Interestingly, the convolutional neural network exhibited "bias." Besides studying the confusion matrices, an examination of the connections, especially in the final decision layer, could reveal that certain letters have stronger activations when the stimuli are highly ambiguous. Presumably more complicated letter shapes require more hidden units than simple letters, which may over-represent responses for certain letters.

The similarity parameters showed a good correspondence between experimental font (from the human experiments) with the appropriate computational model. Correlation of the confusion matrix diagonal entries was reasonably effective in discriminating based on font, but this method ignores the error entries in the matrices. Correlation of the off-diagonal entries, on the other hand, failed to discriminate the Times results. Thus, by explicitly modeling the bias using the Luce choice model (which resulted in little correlation between

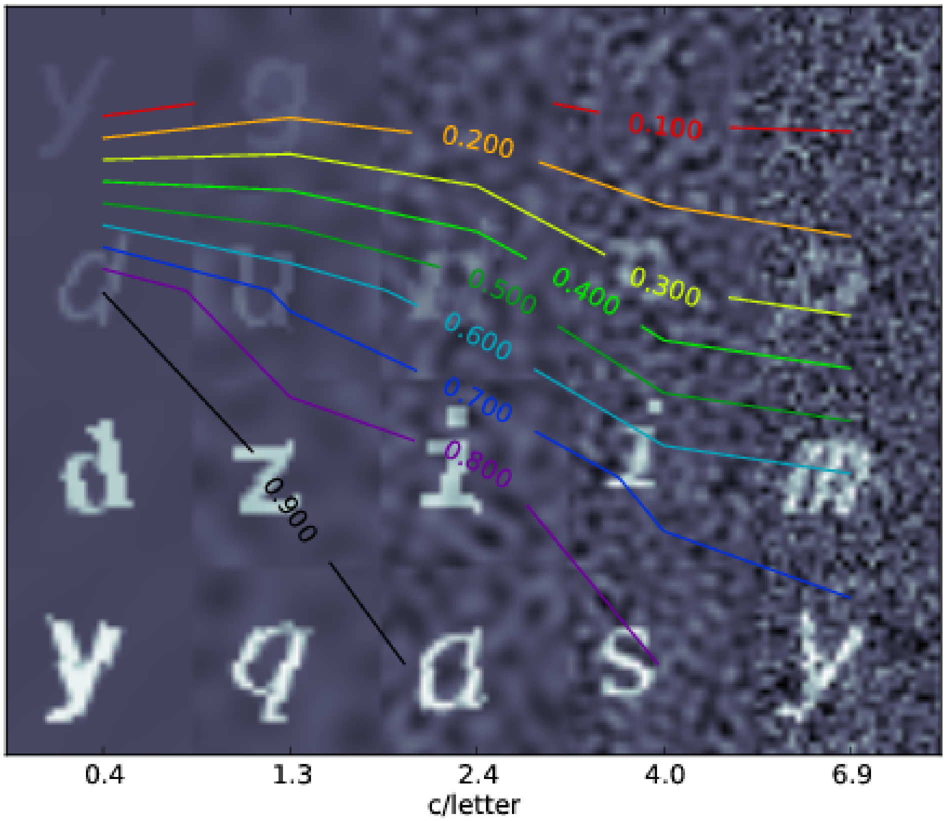


Figure 2.33: Demonstration showing stimuli with model results overlaid as colored lines. Rows represent different contrast for the letter target, and columns indicate a different background noise. Colored contour lines indicate model proportion correct.

experiments), the remaining similarity parameters captured the resulting font differences. The only exception was the crowded Times experiment, which correlated just as well with the Courier model. However, we already know from Section 2.2 that crowded confusions differ from non-crowded confusions, so it is not surprising that this result indicated a pattern beyond mere font dependence.

### 2.3.7 Summary

This section discussed a variety of models that have been proposed for letter recognition, and presented a functional convolutional neural network implementation we tested extensively with a variety of fonts and letter distortions. As expected from the literature, the network's performance was impressive, especially given the stimulus degradations. The identification of a critical frequency band similar to that observed in humans, as well as the good correspondence of the Luce choice model similarity parameters with corresponding psychophysical experiments, suggest that this model could be a useful tool for understanding human letter identification.

# Chapter 3

# Crowding

## 3.1 Introduction

*Crowding is an enigma wrapped in a paradox and shrouded in a conundrum.* (Levi [2008])

The first report of the effect known as "crowding" was almost 100 years ago. In 1923 Korte ([1923]) described the difficulties identifying letters and short German words in the periphery, noting several curious phenomena concerning what his subjects reported. He categorized several stereotypical effects, such as the following: "*. . . a feature of a letter or a whole letter is added to another letter, or a detail becomes so dominant that it absorbs everything else.*" (Korte [1923]), as translated by Strasburger, Rentschler, and Jüttner ([2011]). Although his description was highly qualitative, attempts to isolate the specific characteristics of the deleterious effects of flankers on identification of a target letter remain a vital concern of researchers in this area.

There are practical interests in crowding, such as helping patients with macular degeneration. These patients rely solely on their periphery for visual function, and thus may theoretically be limited by crowding, although recent work has called into question whether crowding is the primary limiting factor (Chung [2014]). Deficits in amblyopia have also been likened to those in the crowded periphery (Levi, Song, and Pelli [2007]). On the other hand, crowding may also have deep theoretical implications. It potentially reflects aspects of cortical organization and hierarchical information flow (Chaney, Fischer, and Whitney [2014]), perhaps providing insights about how our visual system makes sense of the "blooming, buzzing confusion" that our eyes are continually bombarded with.

Since our interest is especially crowding with letters, in the first section of this chapter I present experiments designed to illuminate the nature of feature interactions in crowding. The difficulties in isolating the features of letters was discussed previously in Section 2.3.1.3, so we designed custom stimuli in an attempt to characterize how well-defined features might be improperly combined in crowded characters. To fully understand crowding, it is necessary to explore the effect of various stimulus conditions, such as the contrast of the target and flanker. In the second section, I present an experiment characterizing acuity of crowded Tumbling-E optotypes as a function of eccentricity, contrast, and flanker spacing. Finally, I present a related experiment in the third section. Instead of simply manipulating the contrast of a crowded stimulus, we generated the conditions for isolation of a specific visual channel, the "konio" pathway which originates from S-cone signals in the retina. Due to the segregation of this pathway's signals in the brain, this was an attempt to advance our understanding of the possible physiological locus of crowding, which remains elusive.

### 3.1.1 Crowding proposals

The literature on crowding is vast and diverse, and there are now a plethora of good overviews and reviews (Levi [2008]; Pelli and Tillman [2008]; Whitney and Levi [2011]), so I will restrict this short introduction to the areas that I find most relevant. My primary interest is how, specifically, crowding affects visual perception, and the neural mechanisms involved. Modeling interactions and making predictions are the most convincing proof that the behavior is sufficiently understood. I share the concern of Tyler and Likova ([2007]) that too often models of crowding are so vague that they cannot make predictions or be linked to findings from neuroscience. However, my opinion is that the key difficulty is that visual recognition *itself* is poorly understood (see Section 2.3), thus complicating our understanding of the breakdown in recognition that is caused by crowding.

#### 3.1.1.1 Spatial localization and mislocalization of objects and features

It is not disputed that the spatial localization of objects is worse in the periphery than in the fovea (Westheimer [1982]; Klein and Levi [1987]; Hess and Field [1993]). These peripheral deficiencies have been measured for both absolute location judgments and fine relative position estimation. How to make the leap from these findings to interactions *between* multi-feature objects in multi-element displays is unclear, since a model of detection, integration, and decision mechanisms becomes necessary. The simplest view is that flanking objects are confused for target objects, which is called the "substitution" model of crowding (Chung and Legge [2009]; Strasburger and Malania [2013]; Ester, Klee, and Awh [2014]; Zhang et al. [2012]). Probably everyone would agree that substitution is a necessary component of crowding, but it is currently an area of fierce debate whether substitution is a *sufficient* characterization of crowding (Ester, Zilber, and Serences [2015]).

It seems probable that errors other than substitution errors arise from crowding. Unfortunately, the particular nature of the experimental stimuli used complicates the issue. Different types of stimuli are likely identified using disparate mechanisms, and how these mechanisms interact when under the influence of crowding likely varies widely. This is discussed further in Section 3.2.1. Suffice it to say, there does seem to be interactions between *parts* of objects, since conjunctions between letters are observed—see especially Section 3.1.1.3 below on Feature Integration Theory. How the object *parts* are mislocalized, "pooled", etc., depends on the stimuli. A common assumption for parts-based interactions in crowding is that features are *detected* in a first stage, then combined together in a second stage to support recognition (Pelli et al. [2006]). A logical conclusion is that it is the *second* stage that is susceptible to crowding, but not the first stage (Chung, Levi, and Legge [2001]; Neri and Levi [2006]). Therefore, *detection* is unaffected by crowding (Pelli, Palomares, and Majaj [2004]).

Other descriptions of crowding, such as an overly coarse spatial resolution of attentional mechanisms (Intriligator and Cavanagh [2001]) are hard to discriminate from the proposals described above. It is certainly conceivable that attentional factors cause the observed substitutions, feature mislocalization, or pooling. On this we will stay agnostic, and instead focus on the particulars of the interactions.

#### 3.1.1.2 Textures versus objects

There are recent proposals that describe peripheral vision in terms of summary statistics and texture processing (Balas, Nakano, and Rosenholtz [2009]; Freeman and Simoncelli [2011]), with an accompanying computational model originating from computer vision (Portilla and Simoncelli [2000]). This viewpoint is gaining popularity, but we believe it is less applicable to our interest in letter processing. While the overall texture of lines of printed text are well captured by the model, important structural details of individual letters are lost. The initial developers of the texture model (Portilla and Simoncelli [2000]) acknowledged that it could not capture all images equally well. Faces or other large single objects are unrealistically distorted, defying summarization in terms of texture patterns. We believe that a parts-based approach, involving relations between well-defined visual features, to be a more relevant model.

#### 3.1.1.3 Feature Integration Theory

Anne Treisman's "Feature integration theory" (FIT), is one of the foremost attempts to formalize a functional model of the object recognition process in terms of features and objects (Treisman and Gelade [1980]). She and her collaborators proposed that sets of "features" are detected early in the visual system pre-attentively in parallel, then integrated together to form distinct visual percepts. Breakdowns in the process of binding features to their constituent objects, such as the incorrect migration of features to adjacent objects when under challenging conditions such as peripheral viewing or brief displays, are known as "illusory conjunctions." These phenomena have been heavily studied by Treisman and colleagues (Treisman and Schmidt [1982]; Treisman and Paterson [1984]; Prinzmetal [1981]). Crowding and illusory conjunctions have sometimes been compared and contrasted. Pelli, Palomares, and Majaj ([2004]) proposed that some of the effects reported in the illusory conjunction literature could be due to attentional load or memory, as opposed to crowding, which is typically defined to be primarily perceptual.

Some of the experiments of FIT are quite relevant, particularly those that involve purely spatial features. Others, of less relevance to us, studied false conjunction of different feature dimensions such as color and letter identity. One of the earliest papers by Treisman on

FIT (Treisman and Gelade [1980]) had an experiment where the set of letters could allow a spatial conjunction ("P" and "Q", which could form an "R" if the diagonal of the capital "Q" abandoned its host), and found an effect in search asymmetry. The results in Treisman and Paterson ([1984]) offer a more direct test, with displays comprising "S" characters and vertical lines. In some cases, dollar signs ("$") appeared, providing evidence for the migration of spatial features between objects. William Prinzmetal performed similarly relevant experiments, such as one in which circles with either a horizontal or vertical line could combine to form a circle containing a plus sign (Prinzmetal [1981]). While some of these studies were mentioned in Pelli, Palomares, and Majaj ([2004]), a comprehensive reappraisal of this line of research, placed in the context of crowding, could be a valuable addition to the modern canon of crowding research.

## 3.2 Experiments with crowded letter-like line segment symbols

### 3.2.1 Introduction

It is well known that crowding has a deleterious influence on identification of targets (such as letters) when the target is flanked by other symbols (Levi [2008]; Pelli and Tillman [2008]), but the exact nature of the degradation is still unknown. Crowded stimuli have been colorfully described as "a jumble," "inchoate smudge," "feature perturbations," "illusory conjuction," "invalid feature combination," etc. No doubt aspects of all of these descriptions are accurate, and surely a loss of location and/or identity information is involved, but how to quantify these effects?

Most would agree that object recognition consists of detection of low-level features, with some process to bind these together to form distinct shapes. Attempts to study feature interaction in crowding have typically been limited to artificial symbols whose identity has few possibilities, such as Gabors with several possible orientations (Petrov and Popple [2007]), Landolt Cs or Tumbling Ts (Hariharan, Levi, and Klein [2005]; Levi, Hariharan, and Klein [2002]; Levi, Klein, and Hariharan [2002]; Danilova and Bondarko [2007]), or symbols that vary along a continuous axis (Greenwood, Bex, and Dakin [2009]; Dakin et al. [2010]; Ester, Klee, and Awh [2014]). Fewer studies have looked at crowding with *conjunctions* of two feature dimensions. Those that have (Neri and Levi [2006]; Põder and Wagemans [2007]) used orientation as one dimension and color (or contrast) as the other. This type of conjunction may differ from ones based purely on multiple spatial features (Duncan and Humphreys [1989]). For features that vary on a continual axis (such as the orientation of Gabor), valuable characteristics such as feature averaging have been identified (Greenwood, Bex, and Dakin [2009]; Dakin et al. [2010]), and challenged—(Ester, Klee, and Awh [2014]; Ester, Zilber, and Serences [2015])), but how well do these principles generalize to letters? In contrast to the previously described stimuli, the most likely representation of alphabetic characters is a binary code where the presence or absence of features is what differentiates letters from one another (Gibson [1986]; Keren and Baggen [1981]), rather than one-dimensional values on a continuum.

Studies with more complex symbols, such as English or Chinese characters, have characterized interactions with more general descriptions of each character, such as "complexity" as defined by ink area or skeleton length (Bernard and Chung [2011]; Zhang et al. [2009]; Pelli et al. [2006]; Wang, He, and Legge [2014]), or using letter similarity (discriminability) as derived from a confusion matrix (Bernard and Chung [2011]; Freeman, Chakravarthi, and Pelli [2012]; Hanus and Vul [2013]). To identify properties of feature interactions in crowding that may be relevant to a task such as letter identification, we created artificial symbols where the "alphabet" consisted of symbols that were the combinations of one or two line segments, similar to stimuli used previously in an unflanked context (Townsend, Hu, and Evans [1984]; Townsend, Hu, and Ashby [1981]; Wandmacher [1977]).

Using such stimuli, Townsend, Hu, and Evans ([1984]) clarified several important principles of feature perception with unflanked symbols. For example, they found that features in their characters are not detected independently of one another. That is, some features could be detected more (or less) reliably in the presence of another feature. Townsend, et al. also found that different features have different detection probabilities. Testing specific properties such as these in a crowded context could help add quantitative constraints to models that posit feature detection and integration. Furthermore, there are interactions specific to multi-symbol stimuli (such as the closely spaced letter trigrams used for crowding) which have been proposed, such as effects leading to "feature migration", "mislocalization", and "illusory conjunctions." By analyzing the error distributions of the psychophysical reports on crowded percepts, primarily using the tools of conditional probability, we set out to address several questions concerning the interaction of target and flanker features.

## 3.2.2 Experimental Methods

In order to bypass questions about what, exactly, the critical features for letter recognition are, we designed a custom alphabet using stimuli made from distinct line segments features, shown in Figure 3.1. An important aspect of our stimuli is that they have differing levels of legibility, both crowded and uncrowded, much like letters of the alphabet. Our "alphabet" had distinct clusters of confusable symbols, much like the groups of confusable letters in latin alphabets, such as the lower case vowels (Bouma [1971]). In order to more closely mimic letter recognition, in the experiments described below any of the 10 symbols could appear in any of the letter positions, without constraints. This allowed us to identify effects across stimulus classes and types that might have remained unexposed if the trials had limited sets in given blocks.

### 3.2.2.1 Stimuli

**3.2.2.1.1 Features.** The four features used to construct each character of the alphabet included two cardinally-oriented line segments: horizontal ( – ) and vertical ( ı ), as well

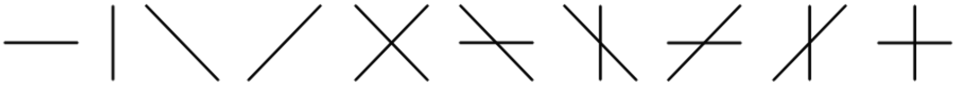


Figure 3.1: Ten symbol "alphabet" used in these experiments.

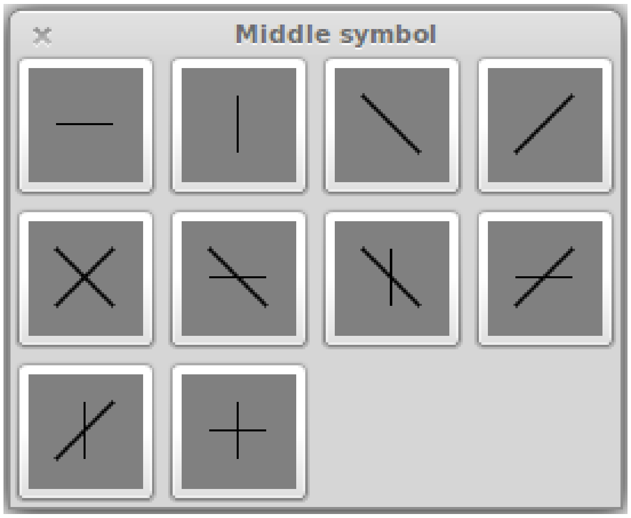


Figure 3.2: Response screen. Subjects indicated response by clicking on symbol with mouse.

as two oblique line segments: top titled left ( ⟍ ), and top titled right ( ⟋ ). The endpoints of each of the line segments were positioned on a square bounding box around the center of each symbol. The oblique segments connected the far corners of the box and were thus longer (by $\sqrt{2}$ times) than the cardinal segments.

**3.2.2.1.2 Symbols.** Each character of the alphabet was composed of one or two features. There were four *singleton* symbols (one for each feature) made up of a single feature. There were six *compound* symbols: × , + , and four symbols we denote "scissors" ( ⍀ , ⟍ , ⌿ , ł ). The scissors symbols were composed of one cardinal segment and one oblique segment.

### 3.2.2.2 Experimental Setup

There were four experiments. For all experiments, stimuli were presented at 10° in the lower visual field. In both Experiments 1 and 2, the symbol bounding box was 0.6°×0.6°. Experiments 3 and 4 used larger symbols, 1.0°×1.0°. In Experiments 1 and 3, the stimuli were single symbols, while in Experiments 2 and 4 they were crowded—surrounded closely on the left and right by a flanking symbol. The center-to-center spacing of the symbols was 1.2 times the height (and width) of each symbol.

For all experiments, stimuli were presented at 10° in the lower visual field. In Experiment 1, the stimuli were single symbols, while in Experiment 2 they were crowded—surrounded closely on the left and right side by a flanking symbol. The center-to-center spacing of the symbols in Experiment 2 was 1.2 times the height (and width) of each symbol. For both Experiment 1 and 2, the symbol bounding box was 0.6°×0.6°. Experiment 3 and Experiment 4 used larger symbols, 1.0°×1.0°. Experiment 3 tested unflanked symbols while Experiment 4 tested flanked symbols.

A trial proceeded as follows. A fixation circle was presented in the center of the screen. As soon as the subjects hit a key, a stimulus was presented at 10° below fixation, then erased after a duration of 150 ms. A response screen then appeared, looking exactly like the screenshot in Figure 3.2. The subject used a mouse to select their response. For Experiment 1 and Experiment 3, subjects reported the target letter, and for Experiment 2 and Experiment 4 they were instructed to report only the center letter of the trigram.

#### 3.2.2.3 Subjects

Subjects were seated in darkened room, and viewed the stimulus using their normal correction with their right eye occluded with an eye patch. All had 20/20 or better vision. Five young adults (ages 19–21, all female) participated in Experiment 1 and Experiment 2. Each subject completed 1000 trials for Experiment 1 and 3000 trials for Experiment 2. Blocks were intermixed between experiments. Four adults (ages 18–28, one male), different from the subjects in Experiment 1 and Experiment 2, participated in Experiment 3 and Experiment 4. Each subject completed at least 2000 trials for Experiment 3 and 3000 trials for Experiment 4. Blocks were intermixed between experiments.

### 3.2.3 Statistical Methods

Most of the methods used in this chapter rely on comparing confusion matrices, typically the 10×10 confusion matrices for the whole "alphabet." Often the whole matrix will be "conditioned" on a value, such as the occurrence of a certain flanker. To generate an appropriate confusion matrix simply involves extracting those cases that satisfy the condition. This may reduce the number of cases, so the number of trials for proportions in each cell is used in the p-value estimation techniques described below.

#### 3.2.3.1 Comparing confusion matrices

We utilize several methods to compare confusion matrices, typically reporting the results of the disparate methods. Most simply, given two confusion matrices, a count can be made of which cells have statistically different values. To do this, a maximum likelihood (ML) estimate is made based on the number of trials and proportions from each confusion matrix. Then, to determine a p-value, 1000 Monte Carlo simulations generate pairs of synthetic confusion matrices based on the estimate. From the simulations, a distribution of differing cell counts is formed. Finally, the p-value is used to determine where in this distribution the observed results fall.

A similar procedure is used for an alternative means to quantify confusion matrix differences, the maximum Z-score of the difference between corresponding cells. For each cell, a Z-score can be estimated as described below in Section 3.2.3.2. The maximum Z-score is a simple measure of how different pairs of cells are between the two matrices. The Monte Carlo procedure described above can again be used to generate p-values. Lastly, measures of matrix agreement such as $X^2$ or the likelihood ratio $G^2$ (both described in Chapter 2.2) can be used to compare whole matrices.

#### 3.2.3.2 Comparing proportions

To test the null hypothesis that two proportions are the same, a standard equation which yields a Z-score is used:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \tag{3.1}$$

The variables $\hat{p}_1$ and $\hat{p}_2$ are the estimates of the two proportions (from the empirical measurements), with $n_1$ and $n_2$ denoting the number of relevant trials for each proportion. $\hat{p}$ represents the estimate given the null hypothesis that the two proportions are the same:

$$\hat{p} = \frac{h_1 + h_2}{n_1 + n_2} \tag{3.2}$$

The number of "hits" are $h_1$ and $h_2$, out of $n_1$ and $n_2$ trials, respectively. Intuitively, given the null hypothesis that $\hat{p}$ is the actual proportion being measured by both conditions, Equation 3.1 describes the difference between the two normal variables, which is itself a normal distribution centered at the difference between the two proportions, with a standard deviation given by the sum of the standard deviations of the two distributions. The variances of the two proportions are $\frac{\hat{p}(1-\hat{p})}{n_1}$ and $\frac{\hat{p}(1-\hat{p})}{n_2}$, respectively, assuming the null hypothesis.

#### 3.2.3.3 Confidence intervals for proportions

To get the 95% confidence intervals shown in the plots, a standard estimate for proportions is used. Specifically, Formula 3.3 presents the estimate for an empirical proportion of $\hat{p}$ for n trials, with significance testing at a level of $z$ standard percentile units. For example, for 95% confidence intervals (two-tailed), $z$=1.96. If not stated otherwise, all figures in this subsection with error bars can be assumed to be calculated using this method.

$$\hat{p} \pm z\sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} \tag{3.3}$$

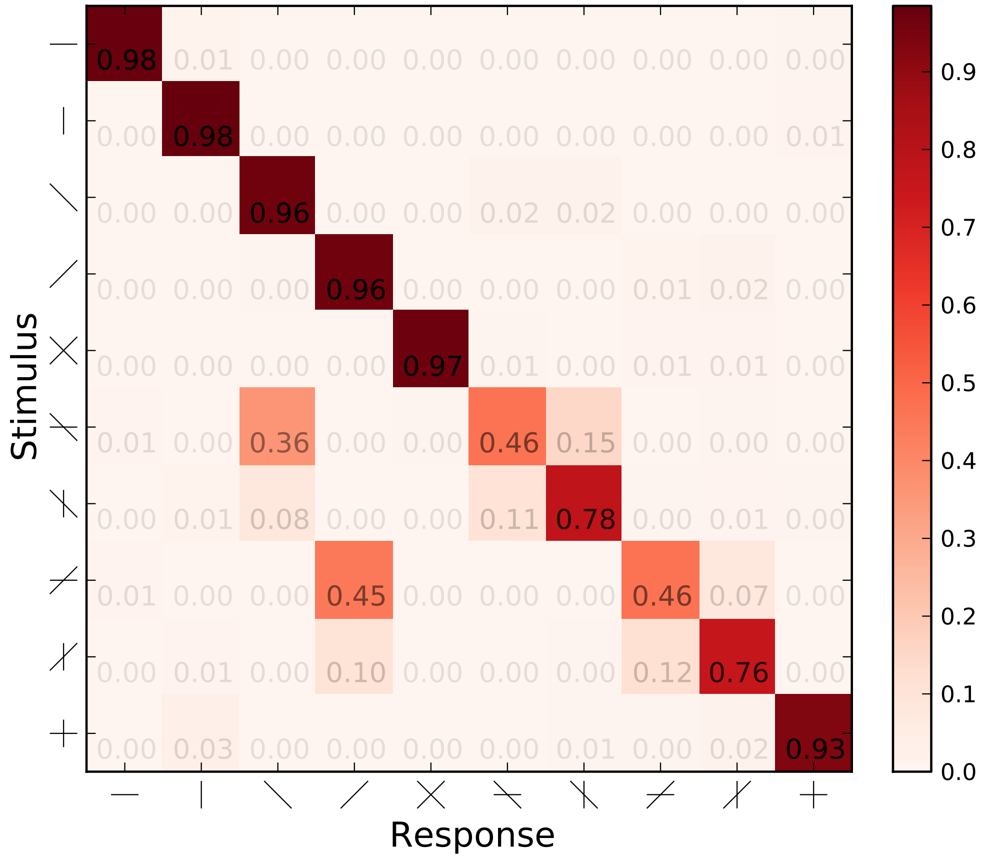


Figure 3.3: Single symbol confusion matrix, averaged across observers

### 3.2.4 Results: Experiment 1 and Experiment 2

#### 3.2.4.1 Experiment 1

Experiment 1 measured identification for symbols presented in isolation, thus providing a baseline for within-symbol feature interactions. The results, presented in a 10×10 confusion matrix, are shown in Figure 3.3, averaged across observers.

**3.2.4.1.1 Uncrowded feature dependencies** We first evaluated feature dependencies for the uncrowded stimuli, following the analysis of Townsend, Hu, and Evans ([1984]). First, we determined hit and false alarm rates, shown in Table 3.1. Notably, the hit rate for the horizontal feature is poor, while the hit rate for the vertical feature is slightly greater, followed by high hit rates for the two oblique features. The cardinal directions had 5% false alarm (FA) rates, as opposed to miniscule FA rates for the oblique features. Feature false alarms are defined as the report of a singleton or compound symbol containing a feature that is not present in the target symbol. It is clear from the confusion matrix shown in Figure 3.3 that

| | − | \| | ⟍ | ⟋ |
|---|---|---|---|---|
| Hit rate | 0.721 | 0.889 | 0.979 | 0.986 |
| FA rate | 0.0505 | 0.0513 | 0.00692 | 0.00846 |
| d’ | 2.23 | 2.85 | 4.5 | 4.59 |

Table 3.1: Uncrowded feature hit and FA rates, based on all trials

| | $P_H$( − \|...) | $P_H$( \| \|...) | $P_H$( ⟍ \|...) | $P_H$( ⟋ \|...) |
|---|---|---|---|---|
| $P_H$(...\| − ) | 0.721 | 0.986 | 0.978 | 0.986 |
| $P_H$(...\| \| ) | 0.945 | 0.889 | 0.968 | 0.978 |
| $P_H$(...\| ⟍ ) | 0.48 | 0.803 | 0.979 | 0.986 |
| $P_H$(...\| ⟋ ) | 0.475 | 0.772 | 0.977 | 0.986 |

Table 3.2: Uncrowded conditional feature hit rates—probability of a hit (column feature), conditioned on presence in stimulus of (row feature)

these false alarms did not happen except in the context of a scissors symbol. For example, false alarms for the horizontal feature came primarily from just two cells: a response of ⍀ for a ⫮ stimulus, and a response of ⌿ for a ∤ stimulus.

A feature hit or false alarm that is dependent on other features can be formalized using conditional probability. The conditional feature hit rates are shown in Table 3.2. There are significant differences between the conditional and unconditional proportions. For example, the hit rate for the horizontal feature, when conditioned on the left oblique present—$P_H$( − | ⟍ )=0.48, is significantly lower than the hit rate for the feature alone $P_H$( − )=0.721. The existence of obliques modulates the hit rate of the cardinals. Even further, this level of conditional alone is not sufficient to describe the results, since this rule fails to capture the fact that the horizontal miss proportion is smaller when the oblique was retained. That is, to adequately describe this error requires an expressions such as P( ⍀ | ⫮ ), which is essentially a term at the symbol level, or a cells from the 10×10 confusion matrix of Figure 3.3.

Similar conclusions follow from Table 3.3, the conditional analysis for false alarms. An interesting aspect of this table is the higher rate of conditional false alarms involving the two cardinal symbols (0.841 for $P_F$( | | − ) and 0.846 for $P_F$( − | | )). This is essentially a feature-level confusion between these two symbols. However, importantly this feature confusion is dependent on the presence of an oblique in the target, again clear from Figure 3.3. This observation strengthens the point that a model at the *symbol* level is necessary to capture these results, especially for scissors symbols.

To study the types of errors identified in the previous paragraphs, the errors resulting from scissors stimuli will now be analyzed in isolation. This analysis focuses on the four relevant rows of the confusion matrix shown in Figure 3.3. To see the pattern of errors in this confusion matrix subset more clearly, the columns for each row can be reordered to reveal

| | $\mathrm{P}_F(\,-\,\vert\ldots)$ | $\mathrm{P}_F(\,\shortmid\,\vert\ldots)$ | $\mathrm{P}_F(\,\diagdown\,\vert\ldots)$ | $\mathrm{P}_F(\,\diagup\,\vert\ldots)$ |
|---|---|---|---|---|
| $\mathrm{P}_F(\ldots\vert\,-\,)$ | 0 | 0.0841 | 0.00667 | 0.0123 |
| $\mathrm{P}_F(\ldots\vert\,\shortmid\,)$ | 0.0846 | 0 | 0.00923 | 0.0123 |
| $\mathrm{P}_F(\ldots\vert\,\diagdown\,)$ | 0.0508 | 0.0651 | 0 | 0.00872 |
| $\mathrm{P}_F(\ldots\vert\,\diagup\,)$ | 0.0518 | 0.0364 | 0.00769 | 0 |

Table 3.3: Uncrowded conditional feature false-alarm rates—probability of a false alarm (column feature), conditioned on presence in stimulus of (row feature)

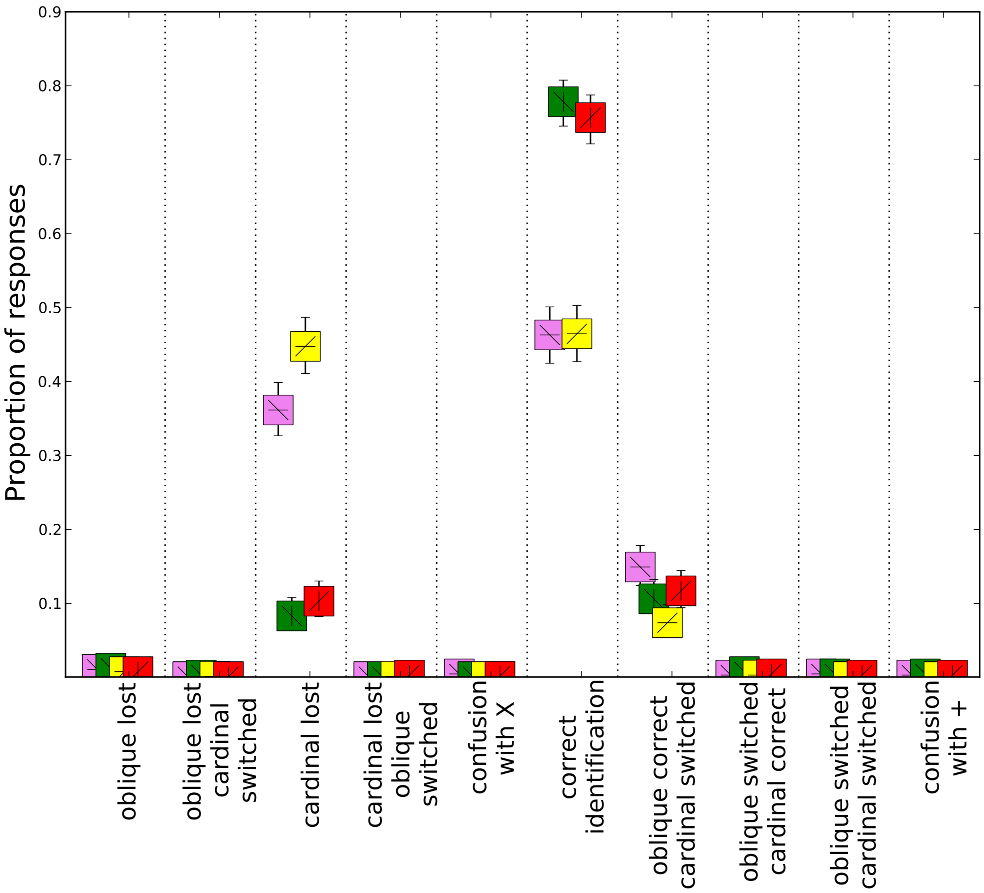


Figure 3.4: Pattern of responses for unflanked scissors stimuli, grouped by the type of error. Icons indicate target stimulus presented.

consistent effects based on the target stimulus. For example, the switching of the oblique feature is one stereotypical error that could occur. Visualizing this error based on Figure 3.3 is difficult, since it involves four seemingly unrelated cells. Figure 3.4 thus presents the four rows of the confusion matrix, reordered based on the class of error, and presented graphically instead of in tabular form. The proportion of occurrences is plotted on the ordinate, for an error type indicated by the abscissa, as shown on the tick labels. Most error types involve a transformation of features of the target, such as "cardinal lost," except for particular errors such as "confusion with × ."

There are two distinct groups of responses: those involving vertical scissors ( ⫮ , ∤ ), and those involving horizontal ones ( ⍀ , ⌿ ). From Figure 3.3, it is already clear that vertical scissors are more likely to be correctly identified than horizontal ones—Figure 3.4 breaks down the specific cases. The proportion correct groups based on the cardinal feature (vertical scissors are correct more often), and the cardinal lost is more prevalent for horizontal scissors. The cardinal features are switched (but the obliques are retained) similarly for the four symbols, approximately 10% of the time for each. There is virtually no difference between errors involving the left and right obliques. These similarities and differences will be utilized for validation in subsequent analyses. In summary, we find, in agreement with Townsend, Hu, and Evans ([1984]), that:

- The four features have different detection probabilities and interaction effects. A notable exception is the two oblique symbols, which are nearly indistinguishable from each other statistically.

- The probability of detecting a feature is not independent of the other features present in the stimulus, nor are the individual response probabilities independent of what other features are reported.

- The probability of a feature "lost" was 17% (with 98% of those coming from scissors stimuli), while the probability of a feature gained (called by Townsend a "ghost feature") was just 2%. This disproportion between features gained and features lost agrees well with the literature.

#### 3.2.4.2 Experiment 2

Experiment 2 measured identification for the center symbols of a closely spaced (1.2× symbol size center-to-center spacing) trigram of symbols. Figure 3.5 shows the 10×10 confusion matrix for the central letter, combined over all flankers and subjects. The errors from Experiment 1 (primarily the errors with scissors stimuli) shown in Figure 3.4 were maintained, as well as many new types of errors added. How to characterize the complex pattern of errors induced by crowding? We first proceed by micro-analyses like those performed for Experiment 1.

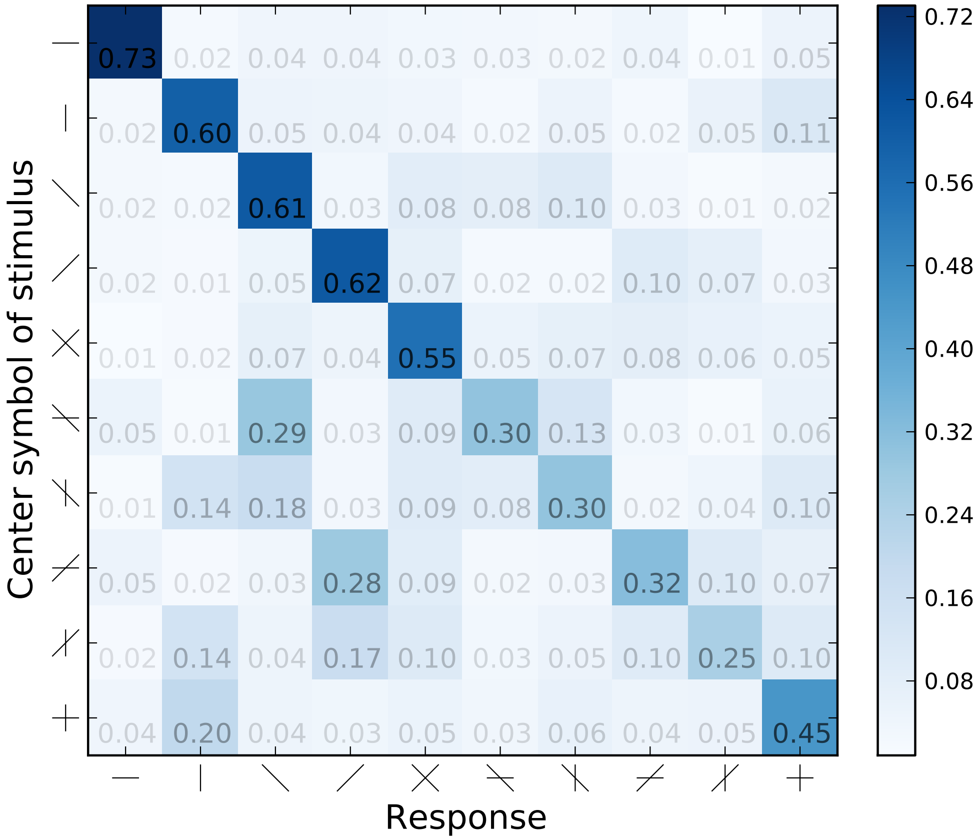


Figure 3.5: Confusion matrix for crowded symbols.

| | P$_H$( – ∣...) | P$_H$( ǀ ∣...) | P$_H$( ⟍ ∣...) | P$_H$( ⟋ ∣...) |
|---|---|---|---|---|
| P$_H$(...∣ – ) | 0.577 | 0.754 | 0.809 | 0.789 |
| P$_H$(...∣ ǀ ) | 0.56 | 0.673 | 0.651 | 0.621 |
| P$_H$(...∣ ⟍ ) | 0.439 | 0.585 | 0.769 | 0.731 |
| P$_H$(...∣ ⟋ ) | 0.46 | 0.542 | 0.743 | 0.75 |

Table 3.4: Crowded conditional feature hit rates

**3.2.4.2.1 Crowded feature hit and false alarm rates** The conditional feature and false alarm rate are given in Table 3.4 and Table 3.5. There is less heterogeneity in these tables than in the unflanked tables, showing more consistency between rows.

**3.2.4.2.2 Scissors error grouping** To more clearly see the errors made with scissors stimuli, we grouped the relevant errors like we did for the uncrowded confusions. The results are shown in Figure 3.6. There were more incidences of different types of errors overall than the unflanked results from Figure 3.4. Differences between vertical and horizontal scissors

| | $\mathrm{P}_F$( − \|...) | $\mathrm{P}_F$( \| \|...) | $\mathrm{P}_F$( ⟍ \|...) | $\mathrm{P}_F$( ⟋ \|...) |
|---|---|---|---|---|
| $\mathrm{P}_F$(...\| − ) | 0 | 0.171 | 0.158 | 0.152 |
| $\mathrm{P}_F$(...\| \| ) | 0.21 | 0 | 0.189 | 0.172 |
| $\mathrm{P}_F$(...\| ⟍ ) | 0.185 | 0.186 | 0 | 0.165 |
| $\mathrm{P}_F$(...\| ⟋ ) | 0.198 | 0.179 | 0.18 | 0 |

Table 3.5: Crowded conditional feature false-alarm rates

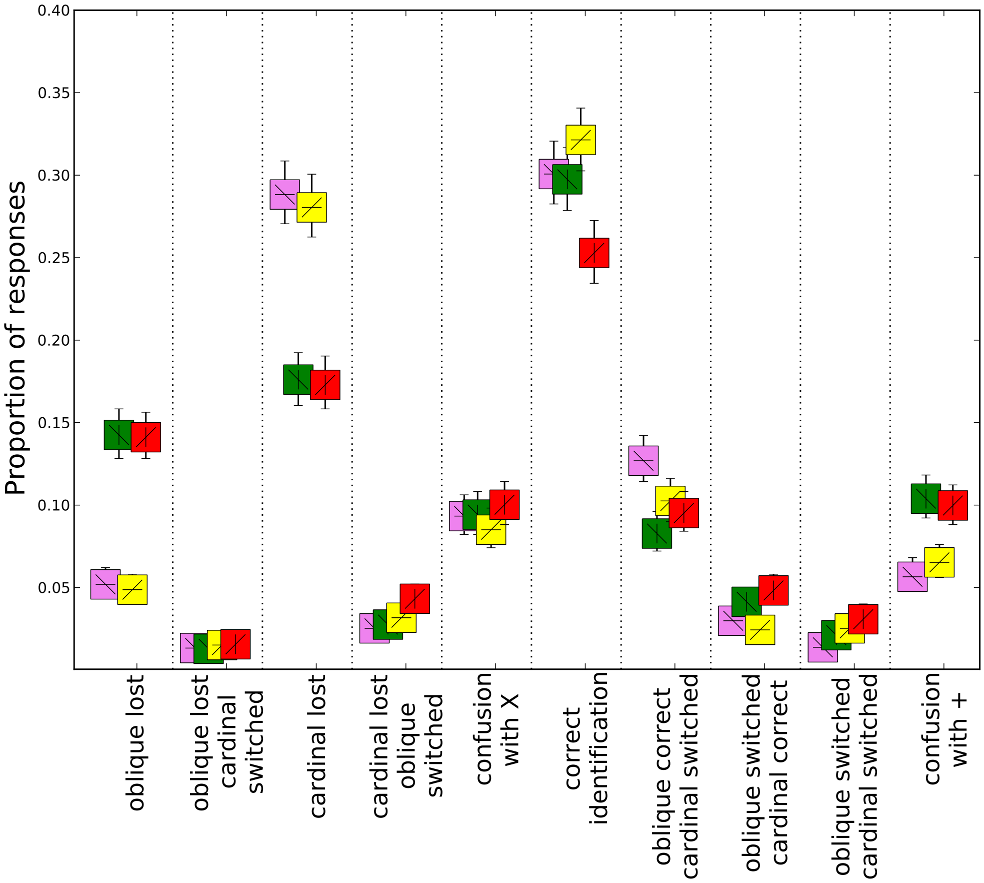


Figure 3.6: Pattern of responses for flanked scissors stimuli, grouped over error type.

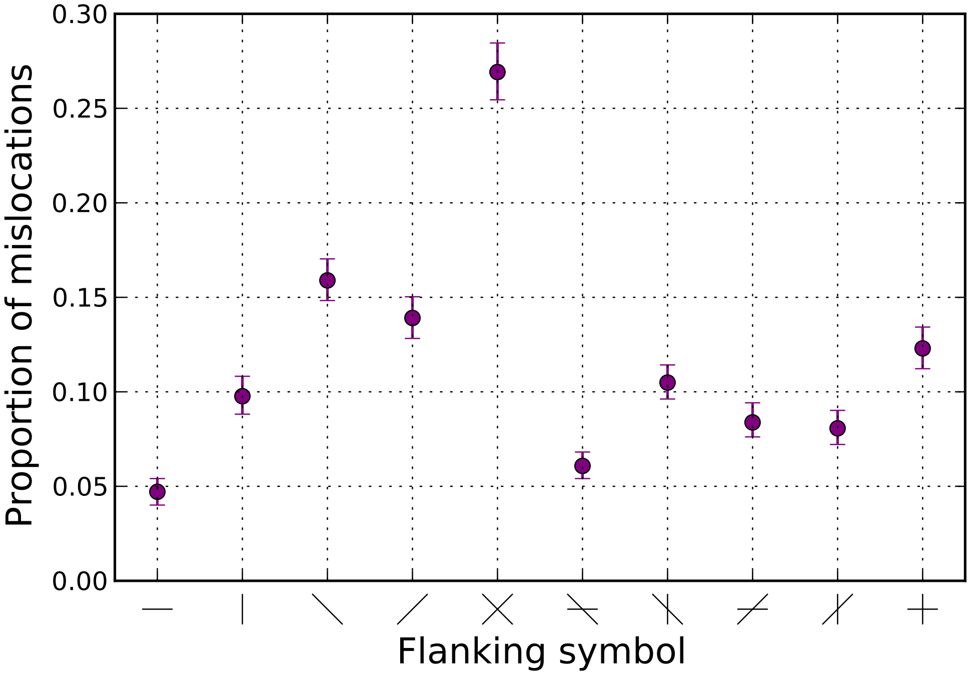


Figure 3.7: Overall rate of mislocation errors for each of the symbols in Experiment 2. Symbol on the abscissa was either of the flankers (but not the target), and was erroneously reported as the target.

were still present. Like for unflanked symbols, the horizontal was "lost" much more often than the vertical (0.30 versus 0.17, respectively). The oblique was lost more often for the crowded vertical scissors (0.15) than for horizontal scissors (0.05). Interestingly, correct identification was more similar between horizontal and vertical scissors (proportions of 0.25–0.30) when the symbols were crowded, unlike the large difference between horizontal and vertical scissors when the symbols were presented in isolation. Confusions with × and switching of the cardinals were errors unique to crowding that had approximately the same occurrence proportion (∼0.10 for both types of errors, regardless of original target type). Confusion with + , on the other hand, happened slightly more often with vertical scissors (0.09) than with horizontal (0.06). The remaining types of errors were small but nonzero, and not significantly different between vertical and horizontal scissors.

**3.2.4.2.3 Proportion of mislocation errors** Erroneously reporting a flanker symbol in place of the target is known as a "mislocation error." The proportion of mislocation errors are plotted in Figure 3.7 as a function of flanker symbol. That is, for each type of symbol, we extracted those trials where one, or both, of the flankers, (but not the target) was the corresponding symbol. The mislocation error rate was then computed as the proportion of these trials in which the symbol was erroneously reported as the target. A more detailed

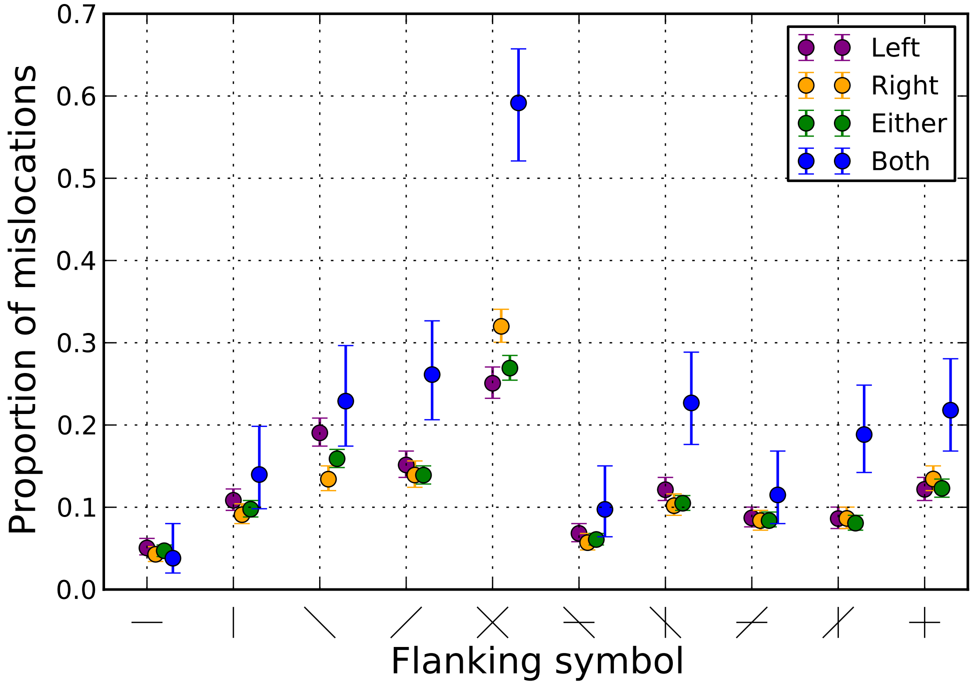


Figure 3.8: Mislocation error rates for each of the symbols in Experiment 2, for flanker occurrence as specified in the legend.

view is given in Figure 3.8, which separates the cases by left and right flanker, as well as the case where one symbol is repeated in both flankers. For the compound symbols, mislocation rates were approximately double when a symbol appeared in both flankers.

**3.2.4.2.4 Symmetry of L and R flankers** Does a flanker feature or flanker symbol located to the left of the target letter have the same effect on target confusions as that feature or symbol on the right? Most would assume so, and almost all models take it as a given. To directly test this assumption, we compared conditional confusion matrices based on the two cases. Figure 3.9 presents an example of two conditional confusion matrices. In this example, these matrices are conditioned on a horizontal feature present in either the left flanker (left panel) or right flanker (right panel). For each confusion matrix, 8700 trials were included, distributed approximately equally across the ten stimulus rows.

Results of comparing left and right matrices are shown in Table 3.6 and Table 3.7, for features and whole symbols, respectively. P-values less than 0.05 indicate that the hypothesis that the confusion matrices are the same must be rejected. Based on a featural analysis using the number of statistically significant cells, for all features except / , the number of cells that differed between the matrices are well within the 95% confidence intervals expected by chance. If quantifying based on individual Z-scores, there are cells in the matrix with a larger difference for both oblique features.

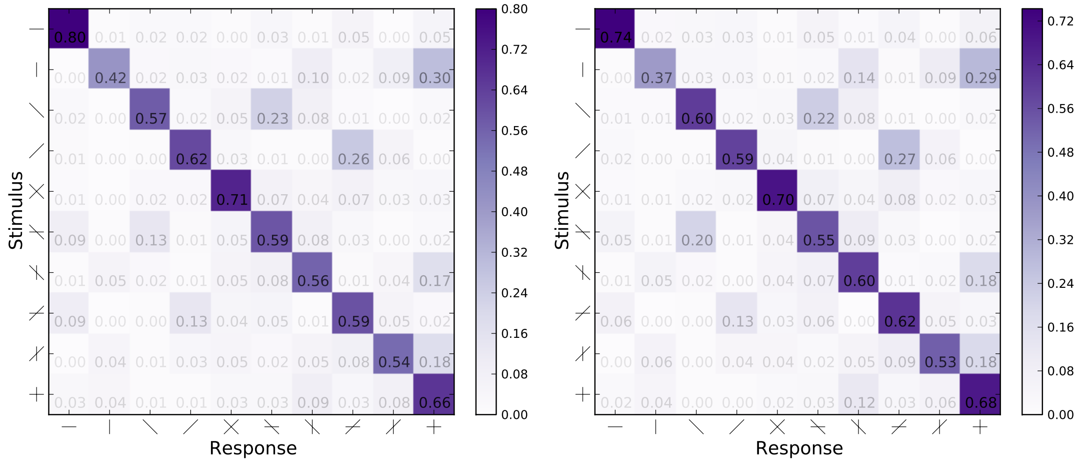


Figure 3.9: Confusion matrices for Experiment 2, conditioned on left flanker containing a horizontal feature (left panel), or right flanker containing a horizontal feature (right panel).

| Feature | # diff | p-val | max Z | p-val |
|---|---|---|---|---|
| – | 1 | 0.998 | 2.027 | 0.991 |
| \| | 4 | 0.711 | 2.764 | 0.43 |
| ∖ | 7 | 0.219 | 3.767 | **0.017** |
| / | 10 | **0.031** | 3.843 | **0.013** |

Table 3.6: Comparison of confusion matrices conditioned on a flanker *feature* located in the left versus right flanker. Second column indicates number of cells (out of 100) that differ significantly between the two conditions. P-value of this cell count is determined using 1000 Monte Carlo simulations. Fourth column is maximum Z-score of difference between corresponding cells in the two matrices. P-value is determined using 1000 Monte Carlo simulations. Bold values indicate statistical significance of p-values.

Based on a symbol-level analysis, six out of the ten symbols have statistically significant number of different cells (up to 13 different cells). However, the relatively small maximum Z-scores shown in the right columns indicated that the individual cell values are not hugely different from what could be expected by chance. Thus, there is an effect, but it is not particularly strong in any one cell.

**3.2.4.2.5 Features versus symbols (as flankers)** To examine whether features of flankers act independently on the targets, we compared the conditional confusion matrices for: a full symbol on either side of a target, versus the features for that symbol split amongst the left and right flankers. For example, consider the test for the ⍀ symbol. First, a confusion matrix was computed for those trials where either the left or right flanker was ⍀. Then, a separate confusion matrix was computed that included trials with either a – feature

| Symbol | # diff | p-value | max Z | p-value |
|---|---|---|---|---|
| − | 8 | 0.075 | 2.688 | 0.367 |
| \| | 10 | **0.025** | 3.008 | 0.159 |
| ╲ | 12 | **0.001** | 3.023 | 0.144 |
| ╱ | 10 | **0.021** | 3.029 | 0.135 |
| × | 12 | **0.003** | 2.997 | 0.151 |
| ⍀ | 12 | **0.003** | 3.385 | **0.038** |
| ⑊ | 7 | 0.198 | 3.217 | 0.08 |
| ⌿ | 13 | **0.002** | 4.119 | **0.001** |
| ⫽ | 6 | 0.31 | 2.65 | 0.471 |
| + | 3 | 0.857 | 2.576 | 0.53 |

Table 3.7: Comparison of confusion matrices conditioned on a flanker *symbol* on the left, versus flanker symbol on the right

| Symbol | # diff | p-value | max Z | p-value | $G^2$ | p-value |
|---|---|---|---|---|---|---|
| × | 10 | **0.01** | 4.23 | **0.01** | 121 | <**0.001** |
| ⍀ | 1 | 0.94 | 2.35 | 0.64 | 55.9 | 0.06 |
| ⑊ | 1 | 0.96 | 2 | 0.94 | 44.8 | 0.32 |
| ⌿ | 3 | 0.53 | 2.67 | 0.32 | 48.8 | 0.19 |
| ⫽ | 1 | 0.95 | 2.12 | 0.83 | 42.5 | 0.41 |
| + | 9 | **0.02** | 3.38 | **0.02** | 98.2 | <**0.001** |

Table 3.8: Conditional confusion matrices for features vs. symbols decomposition. See text for details.

on the left and a ╲ on the right; or, a ╲ on the left and a − on the right. The resultant two confusion matrices were then compared using the methods described. To ensure equal stimulus complexity, only stimuli comprising three compound symbols were used for this analysis (4706 trials). The results are shown in Table 3.8. For the non-scissors compound symbols ( × and + ), the confusions induced by the constituent features of each symbol differ significantly from the results of the whole symbol. For the four scissors symbols, on the other hand, the results are nearly indistinguishable between a feature and symbol level.

## 3.2.5 Discussion of Experiments 1 and 2

### 3.2.5.1 Asymmetries

We found significant symbol inhomogeneities in the data. Importantly, the × and + symbols were the most likely to be identified correctly, and were most conducive to mislocations when acting as flankers. The scissors stimuli, on the other hand, tended to be either confused with each other, or confused with the × and + . Within the scissors, there was a difference between those featuring a horizontal feature and those featuring a vertical feature. The horizontal was more likely to be "lost" than the vertical.

#### 3.2.5.2 Symbols versus features

Could we decisively show that processing of our stimuli involved whole symbols, versus features? Our results were equivocal. For the scissors symbols, the hypothesis of independent feature processing could not be rejected (Table 3.8). However, for the other two compound symbols ( × , + ) there was a significant difference from independent feature processing.

### 3.2.6 Experiment 3 and Experiment 4

We wondered if one reason for the inhomogeneities observed in Experiment 1 and Experiment 2 was the difficulty of the task. From Figure 3.3, the four scissors were identified correctly only 61.5% of the time, even when unflanked. The remaining symbols were all identified correctly greater than 93% correct. When flanked, scissors were identified correctly only 25-30% of the time (Figure 3.5). Is it possible that because the scissors were so hard to identify, subjects would avoid these symbols, leading to the results we saw for the × and + symbols? To explore this possibility, we repeated the experiments, this time using a larger target size: $1^\circ \times 1^\circ$. Four subjects participated in this experiment. Except for the size of the symbols, the experimental setup was the same as Experiment 1 and Experiment 2. Experiment 3 tested unflanked symbols (like Experiment 1), and Experiment 4 tested flanked symbols (like Experiment 2).

### 3.2.7 Results

#### 3.2.7.1 Overall confusions

Figure 3.10 shows the unflanked confusion matrix for the larger letters, combined across subjects. As desired, subjects had little difficulty identifying the isolated symbols. Proportion correct was 0.96 overall, including 0.89-0.97 for the scissors. One of the only errors was loss of the horizontal cardinal, which had a proportion of 0.06. Figure 3.11 shows the flanked confusion matrix for the larger letters, marginalized over all flankers and combined across subjects. First note that the proportion correct for the scissors is much more reasonable for the larger letters: 0.51-0.59. Identification for the singletons was 0.47-0.73 and for × and + was 0.72 and 0.63, respectively. There was less difference in identifiability between the scissors and non-scissors.

#### 3.2.7.2 Scissors errors

The types of errors made for scissors targets for Experiment 4 are shown in Figure 3.12. Note that for this plot a log ordinate is used to more clearly see the breakdown of errors,

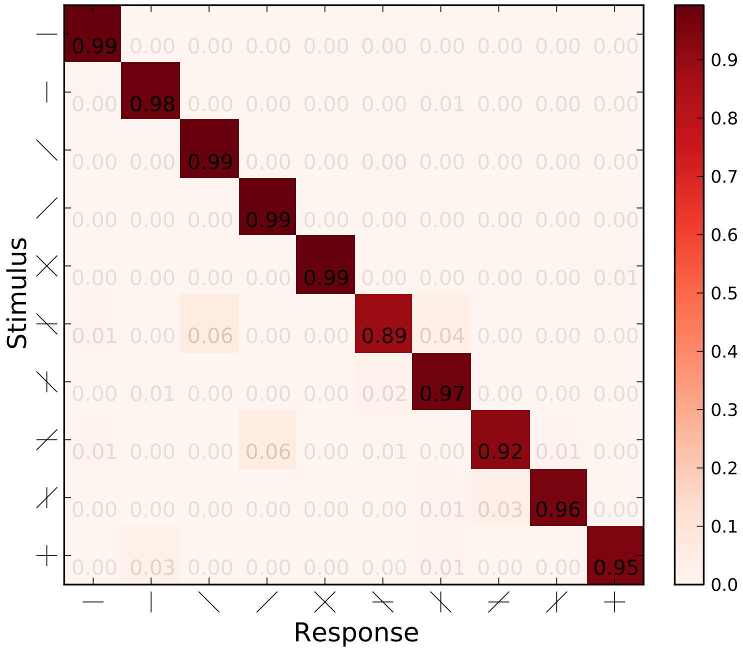
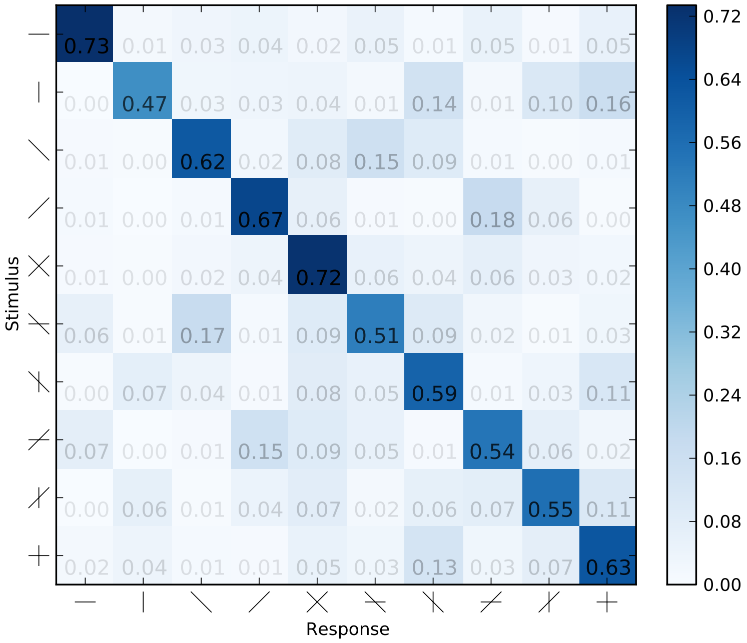


Figure 3.10: Confusion matrix for larger uncrowded symbol (Experiment 3)

which are more evenly distributed, and smaller overall, than in Experiment 2 (Figure 3.6). Besides the different magnitudes, the results between experiments are very similar. There were two small differences between horizontal and vertical scissors in these results, as in Experiment 2. First, the horizontal cardinal was more likely to be lost than the vertical cardinal (0.15 versus 0.04, respectively). Second, vertical scissors were more likely to be confused with + (0.11, versus 0.025 for horizontal). Interestingly, there was no horizontal versus vertical difference in the "oblique lost" error, which was a significant difference when using the smaller symbols (Experiment 2).

#### 3.2.7.3 Mislocation error rates

Mislocation error rates for each of the symbols are shown in Figure 3.13. This plot agrees very well with the results from Experiment 2 in Figure 3.8, albeit with lower rates overall, especially for the singleton flankers. The asymmetry with the left versus right oblique singleton is unique to the larger symbols. The mislocation errors involving × had almost identical magnitudes, while those involving scissors and + had higher mislocation error rates (versus Experiment 2) when a symbol is repeated, and nearly identical mislocation error rates otherwise.



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.73 | 0.01 | 0.03 | 0.04 | 0.02 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 |
| 0.00 | 0.47 | 0.03 | 0.03 | 0.04 | 0.01 | 0.14 | 0.01 | 0.10 | 0.16 |
| 0.01 | 0.00 | 0.62 | 0.02 | 0.08 | 0.15 | 0.09 | 0.01 | 0.00 | 0.01 |
| 0.01 | 0.00 | 0.01 | 0.67 | 0.06 | 0.01 | 0.00 | 0.18 | 0.06 | 0.00 |
| 0.01 | 0.00 | 0.02 | 0.04 | 0.72 | 0.06 | 0.04 | 0.06 | 0.03 | 0.02 |
| 0.06 | 0.01 | 0.17 | 0.01 | 0.09 | 0.51 | 0.09 | 0.02 | 0.01 | 0.03 |
| 0.00 | 0.07 | 0.04 | 0.01 | 0.08 | 0.05 | 0.59 | 0.01 | 0.03 | 0.11 |
| 0.07 | 0.00 | 0.01 | 0.15 | 0.09 | 0.05 | 0.01 | 0.54 | 0.06 | 0.02 |
| 0.00 | 0.06 | 0.01 | 0.04 | 0.07 | 0.02 | 0.06 | 0.07 | 0.55 | 0.11 |
| 0.02 | 0.04 | 0.01 | 0.01 | 0.05 | 0.03 | 0.13 | 0.03 | 0.07 | 0.63 |

Figure 3.11: Confusion matrix for larger crowded symbol (Experiment 4)

#### 3.2.7.4 Left/right symmetry hypothesis

Table 3.9 and Table 3.10 show the left/right symmetry analysis for features and symbols, respectively. There is a substantial amount of asymmetry, for both features and symbols, especially with the $G^2$ measure of computing confusion matrix correspondence. The singletons showed a considerable asymmetry between their effects on the left versus the right.

#### 3.2.7.5 Features versus symbols (as flankers)

As for the previous experiment (Experiment 2), the hypothesis of symbol-level (versus independent feature-level) flanker influence was tested. Table 3.11 shows the analysis for the larger symbols. Even with the more consistent perceivability of the symbols in this experiment, the results in Table 3.11 are remarkably similar to the results from Experiment 2 in Table 3.8. For × and +, there are significant differences between the symbol-based and the feature-based confusion matrices. Using both differing cell count and $G^2$ yields highly significant results. Using a metric based upon Z-score reveals differences in symbol versus feature interactions for ✗, but this special case was not pursued, as the $G^2$ of this symbol was very similar to the other scissors, and its differing cell count was just slightly higher.

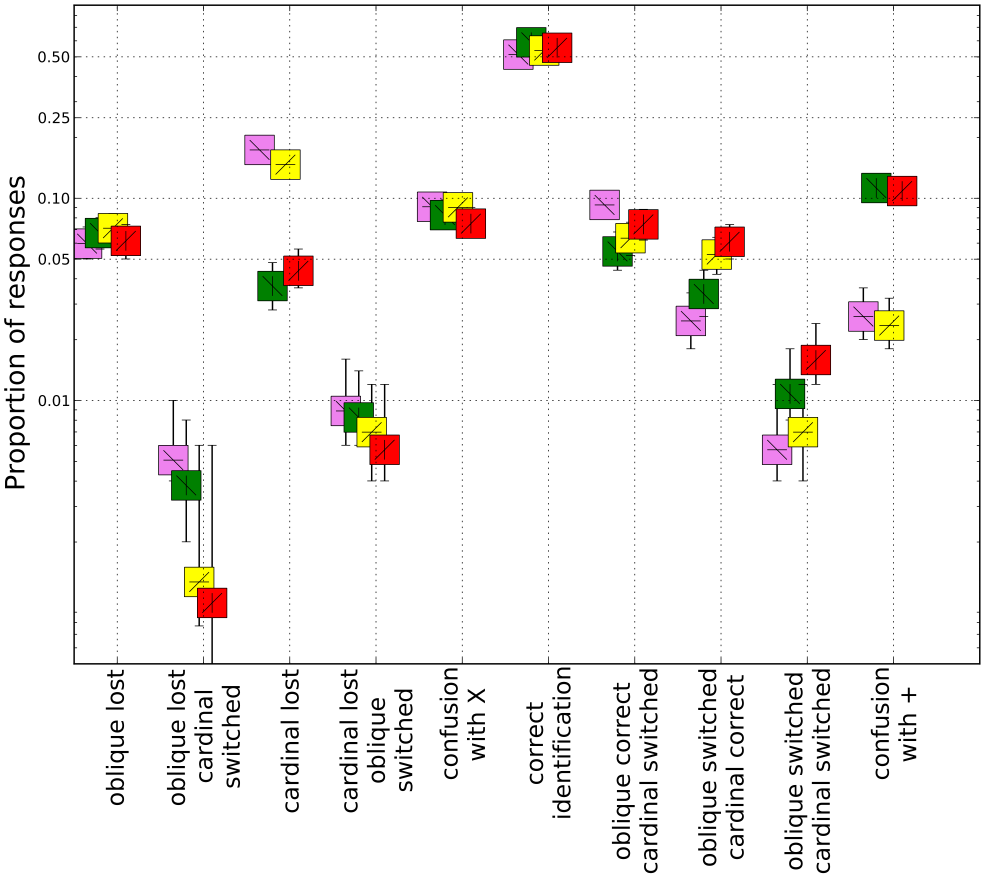


Figure 3.12: Pattern of responses for larger flanked scissors stimuli, grouped over error type and plotted on a logarithmic scale.

| Symbol | # diff | p-value | max Z | p-value | $G^2$ | p-value |
|---|---|---|---|---|---|---|
| – | 8 | 0.170 | 3.47 | **0.030** | 100 | 0.075 |
| \| | 8 | 0.140 | 3.07 | 0.130 | 106 | **0.030** |
| \ | 10 | **0.030** | 2.88 | 0.340 | 106 | **0.032** |
| / | 9 | 0.070 | 3.03 | 0.170 | 114 | **0.008** |

Table 3.9: Experiment 4, left versus right *feature* symmetry hypothesis

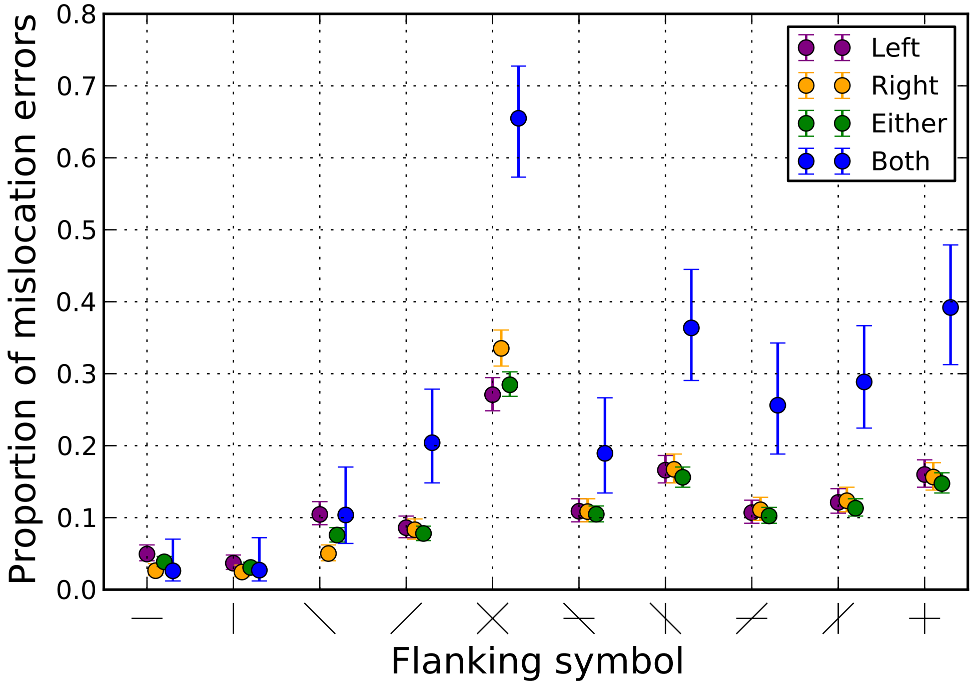


Figure 3.13: Mislocation error rate for each symbol type in Experiment 4

| Symbol | # diff | p-value | max Z | p-value | $G^2$ | p-value |
|---:|:---:|:---:|:---:|:---:|:---:|:---:|
| − | 10 | **0.010** | 4.44 | **0.010** | 156 | <**0.001** |
| \| | 9 | **0.010** | 2.94 | 0.140 | 123 | **0.002** |
| ⟍ | 13 | **0.010** | 4.7 | **0.010** | 164 | <**0.001** |
| ⟋ | 6 | 0.160 | 2.87 | 0.230 | 91.8 | 0.193 |
| × | 9 | **0.020** | 2.65 | 0.290 | 125 | **0.001** |
| ⍀ | 7 | 0.070 | 2.83 | 0.230 | 125 | **0.001** |
| ⫮ | 4 | 0.510 | 2.74 | 0.320 | 92.8 | 0.174 |
| ⌿ | 6 | 0.120 | 2.81 | 0.250 | 131 | <**0.001** |
| ∤ | 7 | 0.080 | 2.91 | 0.190 | 136 | <**0.001** |
| + | 7 | 0.050 | 2.51 | 0.530 | 127 | <**0.001** |

Table 3.10: Experiment 4, left versus right *symbol* symmetry hypothesis

| Symbol | # diff | p-value | max Z | p-value | $G^2$ | p-value |
|---|---|---|---|---|---|---|
| × | 12 | **0.010** | 4.02 | **0.020** | 105 | **0.038** |
| ⧷ | 3 | 0.480 | 2.4 | 0.610 | 49.7 | 0.998 |
| ⫮ | 4 | 0.200 | 2.66 | 0.230 | 48.8 | 0.998 |
| ⌿ | 3 | 0.500 | 2.78 | 0.270 | 56 | 0.985 |
| ∤ | 5 | 0.130 | 3.6 | **0.040** | 49.1 | 0.998 |
| + | 12 | **0.010** | 3.23 | 0.090 | 109 | **0.021** |

Table 3.11: Experiment 4, conditional confusion matrices for features vs. symbols decomposition

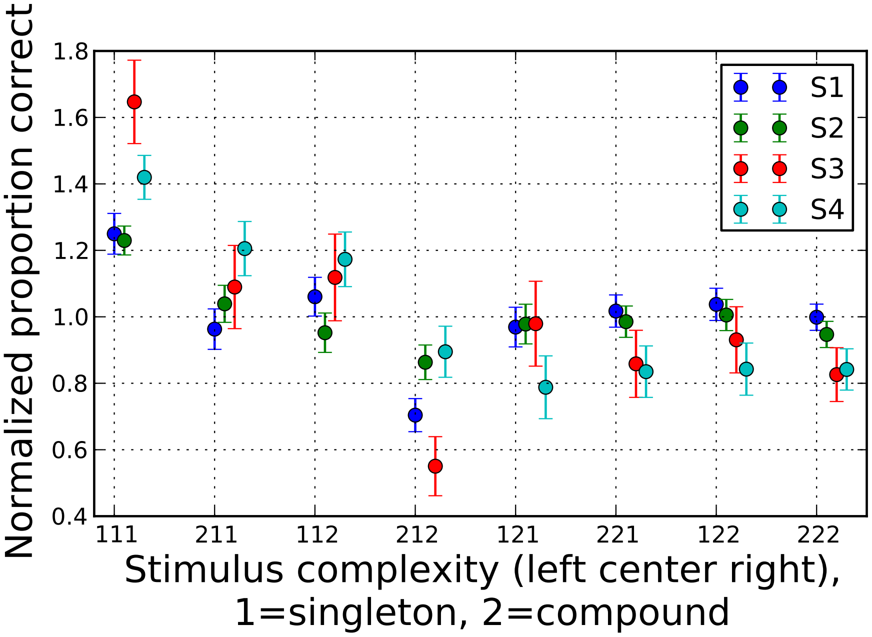


Figure 3.14: Proportion correct as a function of stimulus complexity for each subject, normalized to average proportion correct for each subject. Key on x-axis indicates number of features in left flanker, target, and right flanker symbols, from three singletons (leftmost column) to three compounds (rightmost column).

#### 3.2.7.6 Proportion correct by stimulus complexity

One advantage of our stimuli is that the "complexity" of each stimulus is known, since they are composed of well-defined features. Specifically, the stimulus complexity can be defined by the number of features (1 or 2) in the target and each flanker. Figure 3.14 plots the proportion correct for each type of stimulus, starting with three singletons on the left, to three compounds on the right. The key in the x-axis ticks indicates the complexity of each symbol. Singleton symbols are indicated by a “1” and compound symbols are indicated by a “2.” The order of the three symbols in each tick mark is: left flanker, target symbol, right flanker. For example, “212” means that the left flanker is a compound symbol, the target symbol is a singleton, and the right flanker is a compound symbol. The results are fairly consistent across subjects and are relatively intuitive. Subjects found it easiest to identify the center of three singletons, followed by a singleton flanked by one compound and one singleton, and so on. Surprisingly, results were nearly the same for 121, 221, 122 and 222 stimuli, in terms of normalized proportion correct. This implies that for these stimuli, the number of flanker features did not significantly affect performance with crowded compound targets.

### 3.2.8 Discussion

The goal of these experiments was to study feature interactions in crowding. Because of the well-defined nature of our stimuli, aspects of feature detection and integration could be directly evaluated in a way that overcame assumptions about the putative features of letters. We were particularly interested in acquiring results to help constrain a modeling of crowding based on interactions of symbols and/or features. The asymmetries and inhomogeneities we found are indispensable for such a goal.

To make a satisfiable model for these data would require a scheme that integrated both feature-level and symbol-level interactions. Especially for the × symbol, and slightly less so for the + symbol, symbol-level interactions dominate the observed results. For the scissors stimuli, on the other hand, feature-level interactions could not be rejected. Across all types of stimuli, there are strong dependencies on the entirety of features in the display.

Results of mislocation errors from both experiments revealed interesting trends. First, it was notable that withstanding the difference in percent-correct patterns between the two experiments (Experiment 2 and Experiment 4), relative mislocation error rates agreed well between the two letter sizes. Mislocation error rates for a singleton target decreased as a function of flanker complexity, while mislocation error rates for compound targets were relatively flat—that is, independent of the number of flanker features. The decrease could be a result of the dissimilarity effect in crowding, in which dissimilar flankers crowd less than similar flankers (Bernard and Chung [2011]; Kooi et al. [1994]; Zhang et al. [2009]). The non-dependence of flanker complexity for compound targets could reveal a saturation in terms of features in the display. It is possible that once a sufficient numbers of target and flanker features are present, additional features do not radically affect perception of a crowded stimulus.

In summary, a model capable of capturing all of the effects we observed would be challenging. In particular, the lack of independent usage of target and flanker features challenges the most naive “bag of features” models. The difference in behavior based on complexity class is unsurprising based on modulation of crowding by similarity (previous paragraph). The difference between the horizontal and vertical features could be a function of orientation relative to visual field location. The distinct effects of × and + are harder to explain. It is possible that the symmetry in these shapes conferred a cue unavailable in the scissors. Another possible explanation is their likeness to conventional familiar symbols (an “X”, “x,” or multiplication symbol; and a plus sign or “t,” respectively). More experiments are necessary to explore these possibilities.

## 3.3 Crowded Tumbling-E acuity at varying contrast and eccentricity.

### 3.3.1 Introduction

While the deleterious influence of crowding has been established for some time, it is unclear exactly how different stimulus attributes modulate the amount of crowding. For example, it was shown previously that when the contrast of the stimulus is low, crowding is reduced (Kothe and Regan [1990]; Giaschi et al. [1993]) or absent (Simmers et al. [1999]). However, several of these studies used printed letter charts in a way that confounded acuity limitations and spacing limitations (Kothe and Regan [1990]; Giaschi et al. [1993]), and did not test very close spacings. Specifically, these studies generally measured crowded versus uncrowded acuity. Crowded acuity was determined as the threshold acuity when the target was flanked by bars or letters at a fixed nominal letter spacing. The entire stimulus (with a fixed nominal spacing) could be made larger or smaller.

This method of co-varying the size and spacing was widely used in the past, either by varying viewing distance (Flom, Weymouth, and Kahneman [1963]; Latham and Whitaker [1996]), or with printed cards (Jacobs [1979]), but fell out of favor. There are some recent notable (and relevant) exceptions (Chung [2014]; Song, Levi, and Pelli [2014]; Pelli, Song, and Levi [2011]). A more common procedure nowadays is to use a fixed size stimulus, and test various flanker spacings. The typical result is that flankers located near the target greatly impair performance, while far flankers do not impact performance whatsoever.

Ways to measure the critical spacing include threshold spacing derived from percent correct measurements (Flom, Weymouth, and Kahneman [1963]; Bouma [1970]; Tripathy and Cavanagh [2002]; Siderov, Waugh, and Bedell [2013]), or contrast thresholds (Chung, Levi, and Legge [2001]; Pelli, Palomares, and Majaj [2004]; Strasburger, Harvey, and Rentschler [1991]; Levi, Hariharan, and Klein [2002]). Most prevalent is critical spacing derived from constant stimuli presentations. Psychometric functions are fit to the data, and a variety of options exist for critical spacing quantification. For percent correct measurements, fixed numerical percent correct thresholds can be used (50%-90%), as well as relative metrics, such as decrease of 1/e from unflanked performance (Tripathy and Cavanagh [2002]). Some have determined the farthest flanker location yielding a statistically significant difference in percent correct from unflanked stimuli (Danilova and Bondarko [2007]). For contrast thresholds, several determinations have also been used, including the intersection of a two-line fit to contrast-versus-spacing on log-log plots, mathematically (Chung, Levi, and Legge [2001]) or by eye (Pelli, Palomares, and Majaj [2004]). The spacing causing doubling of unflanked contrast threshold has also been used (Levi, Hariharan, and Klein [2002]).

Technical details may be a reason identification of crowding at low contrast has remained elusive. While the crowding zone for high-contrast, peripherally located stimuli can be easily determined due to its large extent, when contrast is low, stimuli must be larger simply to enable correct identification. The impact of crowding on such large stimuli could thus be minimal. The targets themselves may exceed the critical spacing at a given eccentricity. This is the usual explanation used by contemporary researchers (Siderov, Waugh, and Bedell [2013]) who do find evidence of low-contrast crowding. Indeed, very close spacings must be used to reveal low-contrast crowding, especially at locations close to the fovea.

The effect of contrast on crowded target identification has been a goal of some studies, but none has explored the parameter space systematically. Some studies have looked at differences *between* target and flanker contrasts (Chung, Levi, and Legge [2001]; Pelli, Palomares, and Majaj [2004]; Rashal and Yeshurun [2014]). Another, less well known study (van Nes and Jacobs [1981]) looked at a variety of contrasts and eccentricities, but over a more limited range of values, and did not modulate flanker spacing.

This chapter summarizes an experiment we performed to systematically explore the effects of eccentricity, contrast, and spacing on target identification of Tumbling-E patterns. The first part (Section 3.3.2 to Section 3.3.8) takes a more clinical approach, and shows how usage of traditional letter charts in the periphery may be susceptible to undesirable crowding influences, including in low contrast situations. These sections have been modified from Coates, Chin, and Chung ([2013c]), with Section 3.3.7 and Section 3.3.8 originating in supplemental Appendices A and B, respectively. The final section of this chapter, Section 3.3.9, introduces a quantitative model of the empirical data, which has been presented in talk (Coates, Chin, and Chung [2013a]) and poster (Coates, Chin, and Chung [2013b]) form at scientific meetings.

### 3.3.2 Clinical motivation

Visual acuity measurement is one of the most fundamental methods to assess visual performance in the clinic, and the most common instrument for assessing acuity is the printed letter chart (Bailey [2012]). Letter charts are utilized in a wide variety of situations, so it is important to understand the factors that affect acuity measurements under differing stimulus and observer conditions. The current best letter chart design is likely to be the Bailey-Lovie chart (Bailey and Lovie [1976]), or variants of it (e.g. the ETDRS chart (Ferris et al. [1982]) or the Lea symbol chart (Hyvärinen, Näsänen, and Laurinen [1980])). A characteristic of these charts is that there are five optotypes on each line and the spacing between adjacent optotypes (one optotype width) is designed to minimize the effect of contour interaction. Contour interaction refers to the degrading effect on acuity due to the presence of nearby contours. Although acuity in the presence of this effect can be informative, for example to help detect amblyopia (Hyvärinen, Näsänen, and Laurinen [1980]; McGraw et al. [2000]), the usual desire of clinicians is to avoid the deleterious influence that may introduce undesired variability to measurements of acuity (Bailey and Lovie [1976]). The adoption of a spacing of one full optotype width between adjacent optotypes on letter charts comes from the findings of Flom et al. (Flom, Weymouth, and Kahneman [1963]; Flom, Heath, and Takahashi [1963]). Flom and colleagues measured the accuracy for identifying the orientation of small, high-contrast Landolt-C optotypes in the presence of flanking bars at a range of target-flanker spacings. When expressed in terms of multiples of the size of the gap of the Landolt-C, they found that flanking bars beyond 5 gap widths (equivalently "5 bar widths", which equals one full letter width) had little detrimental effect on identification of the direction of the gap in the Landolt-C. However, these results, and the design of the Bailey-Lovie chart, assume foveal viewing and are based on high-contrast targets.

In the clinic, it is not uncommon to encounter patients who are unable to view a letter chart foveally, as in cases of people with central vision loss or even for patients with mild macular oedema. It is necessary to understand how the measured acuity of these patients might be affected by contour interaction, or "crowding." Flom ([1991]) made a distinction between these two terms, but we use them interchangeably in this study (Chung, Levi, and Legge [2001]). Previous studies have shown that the deleterious effect of crowding on acuity extends over 5 bar widths in the periphery (Jacobs [1979]; Leat, Li, and Epp [1999]), but the maximum spatial extent of the interference, in terms of bar widths at resolution threshold, has not been quantified (Flom [1991]). There has been extensive study of the angular spatial extent of crowding in the periphery (Bouma [1970]; Toet and Levi [1992]; Pelli, Palomares, and Majaj [2004]), and it is well known that isolated letter acuity changes with eccentricity (Jacobs [1979]; Weymouth [1958]). However, since the nominal critical spacing (the letter separation in terms of bar or letter widths necessary to overcome crowding) is dependent on both of these two variables, how it changes with eccentricity at resolution threshold remains an open question. The non-trivial issue of nominal versus angular critical spacing is discussed further in Section 3.3.7. The first goal of this study is to identify the nominal critical spacing in the periphery.

In addition to high contrast acuity, low contrast acuity is routinely assessed for some groups of patients in the clinic (e.g. low vision patients or patients with cataracts or corneal problems), since low contrast acuity may be more sensitive than traditional high contrast acuity in detecting certain abnormal ocular conditions (Regan and Neima [1984]; Woods, Tregear, and Mitchell [1998]; Kleiner et al. [1988]; Regan and Neima [1983]; Schneck et al. [2004]; Haegerstrom-Portnoy [2005]). While it is now established that acuity measurements from low contrast charts viewed foveally will be less affected by crowding than their high contrast counterparts (Kothe and Regan [1990]; Bailey et al. [1993]; Giaschi et al. [1993]; Pascal and Abadi [1995]; Simmers et al. [1999]; Siderov, Waugh, and Bedell [2013]), it is unclear if this same reduction in the influence of crowding occurs peripherally. The second goal of this study is to determine the effect of contrast on the critical spacing in the periphery.

When using low contrast letter charts, how the contrast of the optotype affects the measured acuity, particularly in the presence of flanking letters in the periphery, is also an open question. For single-letter testing, it has been shown that at the fovea, above some critical contrast (the minimum contrast that still yields the maximal acuity) in the range of 20-40% Weber contrast, there is little change in letter or grating acuity with contrast (Herse and Bedell [1989]; van Nes and Jacobs [1981]). In the periphery, Thibos, Still, and Bradley ([1996]) found a critical contrast of approximately 20% for resolution of gratings located at 30 degrees in the nasal visual field. The critical contrast for reading has also been identified, with values ranging from 2-5% (O'Brien, Mansfield, and Legge [2000]), 10% (Chung and Tjan [2009]), to 20% (Legge, Rubin, and Luebker [1987]). It is unknown exactly how the measurement of acuity in the presence of adjacent letters, as in the case of a letter chart, affects the critical contrast, and how the critical contrast for acuity measurement changes from foveal to peripheral viewing of a chart. If the critical contrast for a given condition (e.g. peripheral viewing of a chart) is below the contrast of the printed letters on a low contrast acuity chart, the usefulness and the effectiveness of this chart as a diagnostic aid may be affected.

Given all these considerations, the general goal of this paper is to examine the interplay of letter contrast, viewing eccentricity and letter spacing on acuity measurement. Specifically, the goals are: (1) to quantify the nominal critical spacing for high contrast optotypes in the periphery, (2) to evaluate how the nominal critical spacing changes with contrast in the periphery, and (3) to determine the critical contrast for acuity measurement in the fovea and periphery in the presence of crowding.

## 3.3.3 Methods

### 3.3.3.1 Stimulus Characteristics

Acuity was measured using Tumbling E optotypes adhering to the recommended Sloan dimensions (Sloan [1959]; NAS-NRC Committee on Vision [1980]). The limbs (bars) and gaps of each character were one-fifth of the overall optotype size, which had equal width and height. For testing flanked acuity, four additional Tumbling Es appeared, located above, below, and to the left and right of the target letter. The orientation of the target and each of the four Tumbling E flankers (when present) was completely random, with the limbs of each letter pointing to the left, right, up, or down. The edge-to-edge separation between the target and each of the flankers was specified as a multiple of the size of one limb of the "E", occupying 1, 2, 4, 5, 10, or 20 bar widths. Eight different levels of contrast were evaluated: -2.5%, -3.4%, -6.7%, -12.5%, -22%, -44%, -70%, and -99% Weber contrast. Weber contrast is defined as (L-$L_b$)/$L_b$, where L indicates the luminance of the foreground optotypes and $L_b$ denotes the luminance of the background. The contrast of the flankers was always the same as that of the target. The stimuli appeared in the fovea or one of three eccentricities in the lower visual field: 3°, 5°, or 10°. The stimuli were presented for 150 ms, a duration short enough to avoid voluntary saccadic eye movements to the stimulus once subjects fixated on the fixation target. As soon as the subjects responded the next stimulus appeared. Figure 3.15 presents a cartoon schematic showing the possible stimuli used in the experiment.

### 3.3.3.2 Testing conditions

Testing took place in a dim room with less than 1 cd/m$^2$ of ambient light. A 19' NEC Accusync 120 CRT monitor at a resolution of 1280x1024 pixels was used. The luminance of the white background displayed on the monitor was 75 cd/m$^2$. Luminance measurements were performed using a Minolta LS100 photometer. Subjects viewed the stimuli binocularly with

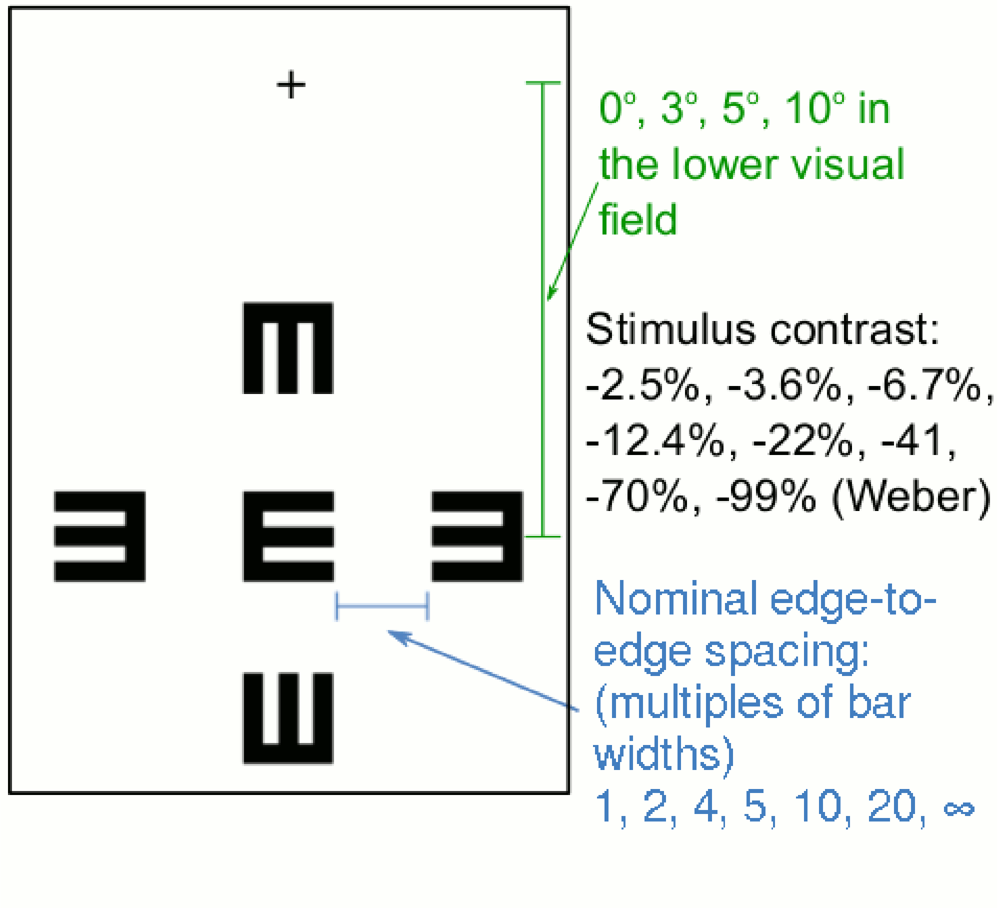

0°, 3°, 5°, 10° in the lower visual field
Stimulus contrast: -2.5%, -3.6%, -6.7%, -12.4%, -22%, -41, -70%, -99% (Weber)
Nominal edge-to-edge spacing: (multiples of bar widths) 1, 2, 4, 5, 10, 20, ∞

Figure 3.15: Stimulus used in this experiment.

their habitual correction. Distance from the monitor depended on the retinal eccentricity being tested. For the foveal condition, subjects were seated 2.4m from the monitor; for the 3° condition, at 1.8m; and for the remaining conditions (5° and 10°), 40cm from the monitor. At the farthest viewing distance (2.4m), one pixel on the monitor subtended 0.43 minutes of arc. For the eccentric conditions, a cross (which was present throughout a trial), served as the fixation target and the target E (flanked or unflanked) appeared at the appropriate eccentricity below the cross. The size of the fixation cross was 3.1mm, so the angular subtense of the cross varied with viewing distance, having a size of approximately 27’ at 40cm and 6’ at 1.8m. To avoid masking effects, the fixation cross disappeared when foveal targets were displayed. Stimuli were rendered and displayed with custom software written in the Python programming language using the PsychoPy psychophysics library (Peirce [2008]).

Table 3.12: Subject demographics

| Subject | Gender | Age | Best corrected visual acuity | Refractive errors |
|---|---|---|---|---|
| S1 | M | 20 | OD: 20/20<br>OS: 20/20 | OD: -5.57<br>OS: -4.75 |
| S2 | M | 27 | OD: 20/16<br>OS: 20/16 | OD: -11.25 -0.50 × 003<br>OS: -11.75 -0.25 × 124 |
| S3 | F | 22 | OD: 20/20<br>OS: 20/16 | OD: -4.00 -0.75 × 165<br>OS: -3.75 -0.75 × 170 |
| S4 | F | 19 | OD: 20/12.6<br>OS: 20/12.6 | OD:-2.00<br>OS:-2.00 |
| S4 | F | 20 | OD: 20/20<br>OS: 20/20 | OD: -2.50 -0.50 × 176<br>OS: -3.00 -0.50 × 013 |

#### 3.3.3.3 Subjects

Five subjects participated in this study. Table 3.12 lists the demographic information of the subjects. Written informed consent was obtained from each subject after the procedures of the experiment were explained, and before the commencement of data collection. The experimental protocol was approved by the Institutional Review Board at the University of California, Berkeley, and conformed to the Declaration of Helsinki.

#### 3.3.3.4 Psychophysical Procedure

Threshold letter size (specified as the minimum angle of resolution, MAR, in units of minutes of arc) for each condition was determined using an adaptive 3-down, 1-up staircase procedure, which identified the threshold for 79% correct performance. For each staircase run, the target Tumbling E, with or without its four flankers, initially appeared at a size significantly above threshold for the selected testing eccentricity: 2° letters at the fovea, 2.5° letters at 3° eccentricity, 3° letters at 5° eccentricity, and 4° letters at 10° eccentricity. The stimulus size was reduced after three consecutive correct trials; and increased after a single incorrect trial. The amount by which the stimulus size changed after each of these reversals became progressively smaller, using the following sequence: three reversals of one log unit, four reversals of 0.2 log units, and five reversals of 0.1 log units. A single staircase ended when all 12 reversals were completed and took on average 78 trials, with 95% of the staircases taking between 50 and 200 trials. There was no systematic effect of retinal eccentricity, stimulus contrast, or letter spacing on the number of trials it took to estimate a threshold. The threshold was determined as the average of the sizes at which the reversals occurred, with the exclusion of the first two reversals, which were not used in the threshold calculation.

#### 3.3.3.5 Testing Sequence

Every condition (eccentricity, contrast, spacing) was tested at least twice, with the order of execution determined as follows. For each subject, a random order of eccentricities was constructed from the set of eight eccentricities (four eccentricities, each appearing twice). To test an eccentricity, a random ordering of contrasts was generated. For each contrast, the order of flanker spacings (including unflanked) was randomized. For each of these conditions (eccentricity, contrast, and spacing), the software first displayed the parameters (eccentricity, contrast, and spacing) about to be tested, then commenced the staircase procedure defined in "Psychophysical Procedure" to determine the threshold. Testing all conditions for one eccentricity (all contrasts and all spacings) took an hour to an hour and a half. This randomization of the sequence of trials was performed in order to minimize the effects of fatigue and practice.

#### 3.3.3.6 Data Analysis: Fitting Individual Acuity vs. Letter Spacing

To evaluate how acuity is affected by the spacing between adjacent letters, and to derive the critical spacing for the different contrasts and eccentricities, thresholds are first analyzed as a function of letter spacing. The primary method of modeling the data is to adopt the formulation used by previous studies (Song [2009]; Song, Levi, and Pelli [2014]; Pelli, Song, and Levi [2011]; Chung [2014]). This method has been used to model crowded acuity in the fovea and periphery of normal subjects, amblyopes and people with age-related macular degeneration. In this model, data for a given condition are fit using a two line function, with acuity plotted as threshold size against nominal spacing, on logarithmic axes. Figure 3.16 shows an example of this model with subject data collected at 3° eccentricity in the lower visual field and high contrast (-99%) stimuli. When flankers are far from the target (or absent), acuity is unaffected by the spacing of the flankers, and the y-values form a horizontal line. When the flankers are in close proximity to the target, acuity is affected by spacing. These data are well described by a line with a slope of negative one, which implies a complete trade-off between acuity and spacing. This complete trade-off between acuity and spacing is a direct consequence of the fixed angular size of the crowding zone at any given eccentricity (Pelli, Palomares, and Majaj [2004]). The intersection of the two lines is the critical spacing, by definition. The basis of this formulation is described in further detail in Section 3.3.8. Although the fit is performed as described above, in this paper results are presented with the units of edge to edge letter spacing (in bar widths) on the abscissa, and MAR (in minutes) on the ordinate to better relate our findings to clinical practice. The effectiveness of this model in fitting the present data is demonstrated in the Results section. While more complex formulations have been used to fit data like ours, such as the rectangular parabola (Gurnsey, Roddy, and Chanab [2011]), the simplicity of the two-line fit, as well as the clear interpretation of its parameters, justify its use. With this fit, the dependent variable increases monotonically as flankers approach the target. Some researchers have identified a facilitation effect, whereby flankers very near the target may actually aid its identification (Flom, Weymouth, and Kah-

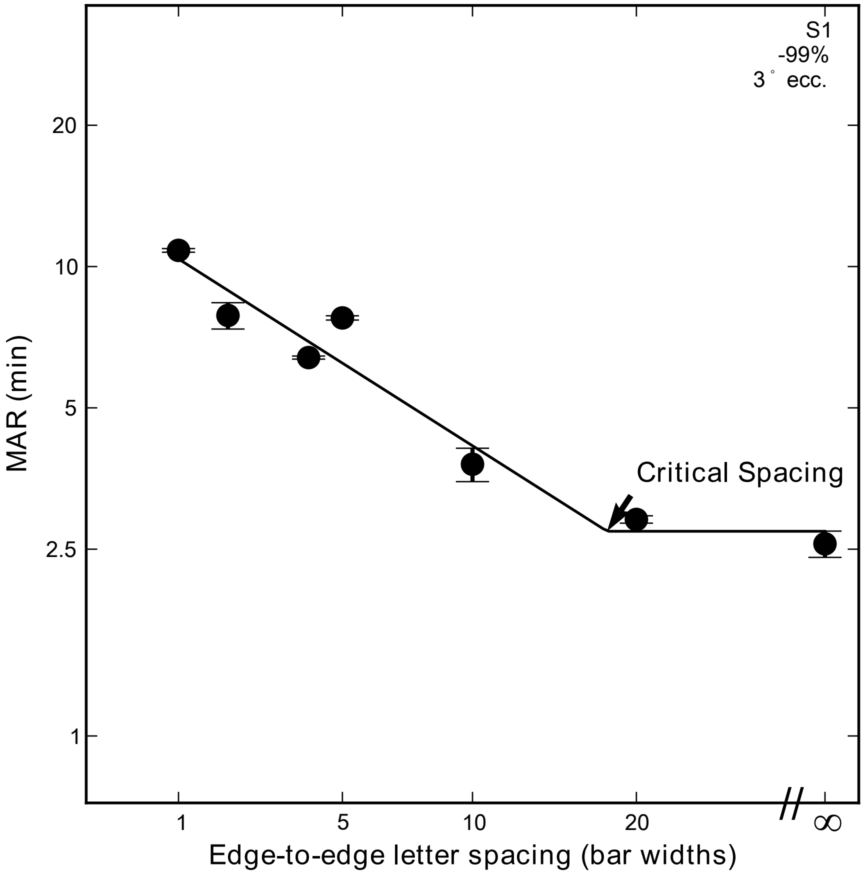


Figure 3.16: Canonical two-line fit. Subject S1's data with high-contrast stimuli at 3° in the lower visual field demonstrating the two-line fit of acuity versus letter spacing. The datum plotted at a letter spacing marked with "∞" represents unflanked acuity. The critical spacing is where the two lines intersect. Error bars indicate the standard deviation between the thresholds from the subjects two separate staircase runs. To the right of the critical spacing acuity is flat, implying that it is unaffected by crowding. To the left of the critical spacing, adjacent characters are within the "crowding zone," and thus, crowding is evident. The slope in this portion is constrained to -1.

neman [1963]; Siderov, Waugh, and Bedell [2013]; Takahashi [1968]; Danilova and Bondarko [2007]), though with percent correct as the dependent variable. It is not clear that the same effect would be apparent when acuity is measured, nor has there been evidence of this effect in the periphery.

#### 3.3.3.7 Data Analysis: Fitting Individual Acuity vs. Contrast Data

To evaluate how acuity is affected by stimulus contrast, the critical contrast for acuity measurements for the different eccentricities and letter spacings is derived. To do so, acuity is plotted as a function of contrast on log-log axes, for each eccentricity and letter spacing. The acuity versus contrast function can also be described by a two-line fit, but unlike the acuity versus spacing fit, the slope of the decreasing portion of the curve is allowed to vary, since there is neither empirical or theoretical justification to constrain it. The critical contrast is defined as the contrast at which threshold begins to worsen from its optimal value, which is achieved at full contrast. This critical contrast is the value on the abscissa where the two lines intersect. A similar fit has been used previously by O'Brien et al. (O'Brien, Mansfield, and Legge [2000]) and Chung and Tjan (Chung and Tjan [2009]), but with reading speed as the ordinate.

#### 3.3.3.8 Curve Fitting

Curve fitting was accomplished using the scipy.opt optimization library in Python. Summed square error was minimized using the L-BFGS algorithm, an iterative fitting procedure capable of non-linear fitting. When the dependent variable was an acuity measurement, errors were minimized on a log axis, as suggested by Westheimer (Westheimer [1979a]).

### 3.3.4 Results

#### 3.3.4.1 Acuity vs. letter spacing

First, threshold size is analyzed as a function of nominal letter spacing. Figure 3.17 shows the individual subject data (S1-S5, separate rows) for all four eccentricities (different curves in each panel) at each stimulus contrast (each contrast in a column). As expected, acuity worsens as eccentricity increases, with the lowest curve in each panel (smallest threshold) representing data obtained at the fovea, and each curve above corresponding to data obtained at the more eccentric target locations. Acuity also worsens as contrast decreases (an upward shift in the family of four curves with the columns going from left to right). The unflanked foveal acuity measured at the highest contrast corresponds to a threshold of approximately one minute of arc for four of the subjects (1.08, 1.09, 1.0, and 0.92 min, respectively, for S1-S4), with much poorer acuity for S5 (3.19 min), who had higher overall variability. We suspect that location uncertainty for the foveal targets and short stimulus duration (150 ms) made the task difficult for our observers (Baron and Westheimer [1973]), which could account for why the high contrast unflanked foveal acuity was not better.

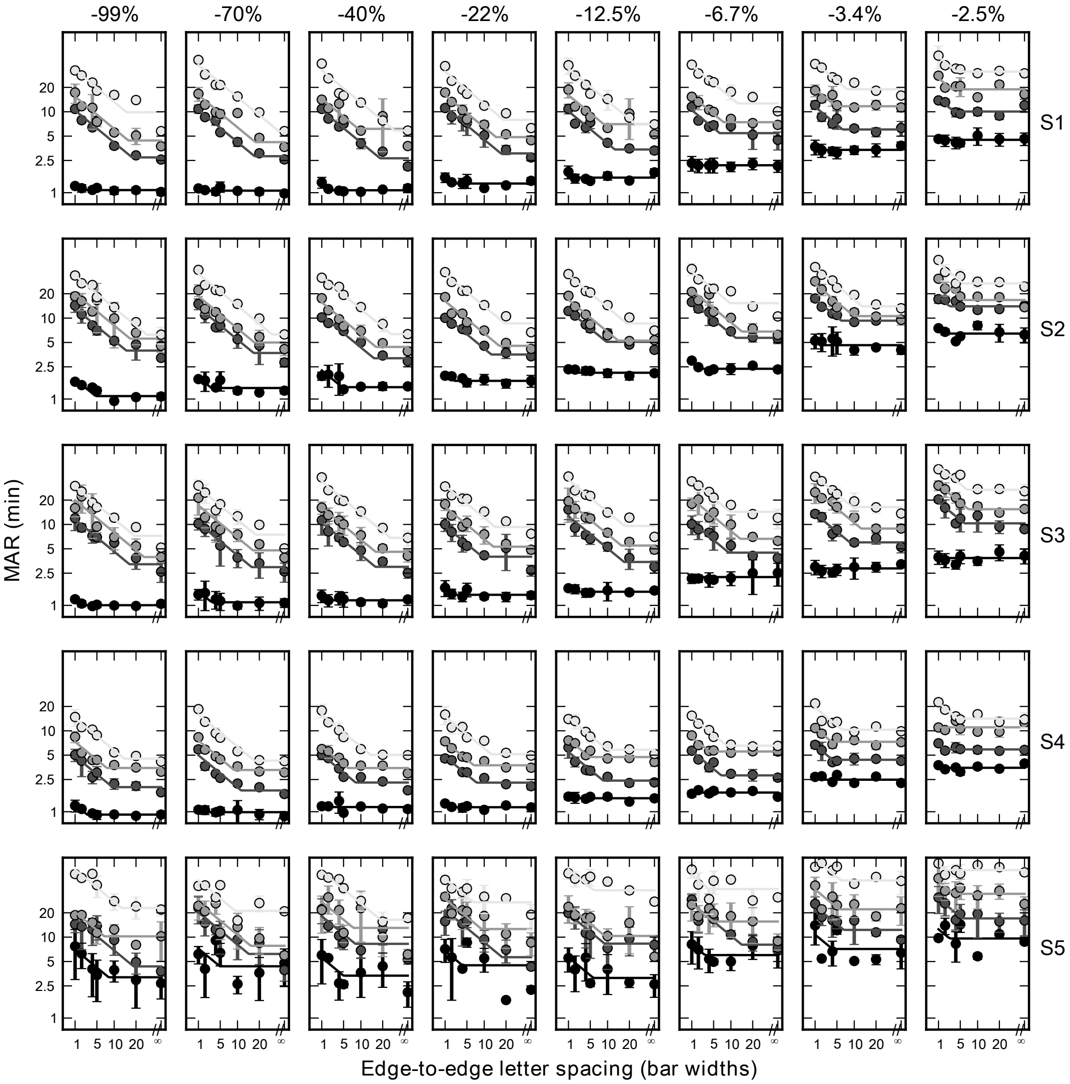


Figure 3.17: Complete individual subject data showing acuity versus letter spacing at the four eccentricities tested (different shaded curves in each panel), at all stimulus contrasts. Each column is a given contrast and each row is a particular subject. Error bars indicate the standard deviation between the thresholds from the subject's two separate staircase runs, and the lines indicate fits. In each plot, the lowest curve comprises the foveal condition, with each successive eccentricity (3°, 5°, and 10°, respectively), stacked above.

Table 3.13: Nominal critical spacing at high contrast (-99%), in bar widths.

| Subject | Fovea | 3° | 5° | 10° |
|---|---|---|---|---|
| S1 | 2 (0.88-3) | 18 (16-20) | 16 (8.4-26) | 15 (9.5-30) |
| S2 | 4.7 (3.6-6.6) | 15 (8.3-23) | 14 (3.2-38) | 27 (17-46) |
| S3 | 2.2 (0.95-3.1) | 16 (4.3-43) | 21 (7.5-41) | 19 (15-33) |
| S4 | 3 (-0.033-4.6) | 9.5 (3.6-15) | 7.7 (5-9.6) | 14 (9.6-17) |
| S5 | 8.2 (-0.95-35) | 17 (2-39) | 6.9 (3.4-51) | 13 (7.9-20) |
| **AVG** | **4.4+/-3** | **15+/-3** | **14+/-5.3** | **19+/-5.9** |

Mean of 1000 Monte Carlo simulations, with 95% confidence intervals in parentheses. Last row, in bold, indicates average (AVG) of the five subjects.

Table 3.14: Nominal critical spacing (bar widths) at all contrasts, mean +/- standard deviation across subjects

| Contrast | Fovea | 3° | 5° | 10° |
|---|---|---|---|---|
| -99.0% | 4.4+/-3 | 15+/-3 | 14+/-5.3 | 19+/-5.9 |
| -70.0% | 3.6+/-2.2 | 16+/-2.7 | 15+/-3.4 | 22+/-9.2 |
| -40.0% | 3.2+/-1.8 | 14+/-4.5 | 12+/-3.9 | 22+/-5.1 |
| -22.0% | 4+/-3.7 | 13+/-4 | 12+/-3.3 | 15+/-5.7 |
| -12.5% | 2.7+/-1.7 | 12+/-3.5 | 11+/-3.6 | 15+/-6.4 |
| -6.7% | 2+/-0.76 | 9.9+/-2.9 | 8.9+/-3.6 | 9.6+/-2.7 |
| -3.4% | 2.8+/-1.7 | 6.2+/-1.5 | 7.4+/-3.1 | 8.4+/-2.5 |
| -2.5% | 1.9+/-0.53 | 4.2+/-1.6 | 4.4+/-2 | 4.5+/-1.6 |

To model the data, we used the constrained two-line fit as described above. $R^2$ statistics for the fit to the peripheral data averaged 0.85 (+/-0.17) across all subjects and contrasts, implying that the two-line fit provides an excellent description of the peripheral data. However, in the fovea, the $R^2$ values for the fit are typically small numbers, yielding an average of 0.33 (+/-0.3) across subjects and contrasts. This is due to the fact that the foveal crowding functions are relatively unaffected by the flankers for the range of spacings and contrasts tested, which is evident in Figure 3.17 by the flatness of the foveal curves. The two-line fit yields very small nominal critical spacings, meaning that a straight line would fit the data almost as well as the model, resulting in a low $R^2$ despite a small sum of squared error. Regardless, it is parsimonious to have a single model that can describe the data accurately across all conditions.

The two-line fit summarizes the acuity at each condition with two parameters: the uncrowded acuity (the ordinate corresponding to the horizontal portion of the curve) and the

Table 3.15: Pairwise significance testing of critical spacing as a function of stimulus contrast values from Table 3.14.

| | -2.5% | -3.4% | -6.7% | -12.5% | -22.0% | -40.0% | -70.0% | -99.0% |
|---|---|---|---|---|---|---|---|---|
| -2.5% | | n.s. | **0.002** | **h.s.** | **h.s.** | **h.s.** | **h.s.** | **h.s.** |
| -3.4% | | | n.s. | **0.04** | **0.003** | **h.s.** | **h.s.** | **h.s.** |
| -6.7% | | | | n.s. | n.s. | **0.01** | **0.002** | **h.s.** |
| -12.5% | | | | | n.s. | n.s. | n.s. | n.s. |
| -22.0% | | | | | | n.s. | n.s. | n.s. |
| -40.0% | | | | | | | n.s. | n.s. |
| -70.0% | | | | | | | | n.s. |

Adjusted p values are from Tukey HSD test. Those in bold are significant at 0.05 level.
**h.s.** = highly significant (adj. p<0.001) n.s. = not significant at 0.05 level

nominal critical spacing (the abscissa corresponding to the intersection of the two lines). Table 3.13 lists the nominal critical spacings at -99% contrast for each subject as a function of eccentricity. To determine confidence intervals, 1000 individual Monte Carlo simulations based on the subject data were generated, and the model was fit for each simulation (Kingdom and Prins [2009]). The reported statistics indicate the mean of the fitted parameter values and the 95% confidence interval range. Table 3.14 shows fits at all contrasts that were tested. The foveal nominal critical spacing (averaged 4.4 bar widths) is generally much smaller than the peripheral values (15 - 20 bar widths), with the three peripheral values being very similar to each other. The average value of the foveal critical spacing (4.4 bar widths) agrees with previous reports (Flom, Weymouth, and Kahneman [1963]). The novel contribution of this study is the finding that the nominal critical spacing in the periphery, known to be greater than 5 bar widths (Flom [1991]; Jacobs [1979]), is 15-20 bar widths at the eccentricities tested.

At -99% contrast, a repeated-measures ANOVA (using the software package R (Ihaka and Gentleman [1996])) revealed a significant effect of eccentricity on critical spacing ($F_{3,18}$ = 7.753, p = 0.002). Posthoc pair-wise comparison using the Tukey HSD test showed that the fovea was different from the nonfoveal eccentricities ($p_{adj} < 0.03$ for the fovea versus each of the three eccentricities), while the non-foveal eccentricities were not different from each other ($p_{adj}$>0.5). As shown in Table 3.14, lower contrasts yielded smaller critical spacings at all eccentricities, with a larger decrease in the periphery. In the fovea, the low contrast critical spacing was half of the high contrast critical spacing, while at 10° eccentricity the low contrast critical spacing was a quarter of the high contrast critical spacing.

A repeated-measures ANOVA showed that contrast indeed had an effect on critical spacing ($F_{7,128}$ = 18.351, p<0.001), although the interaction between contrast and eccentricity was not significant ($F_{21,128}$=1.326, p=0.171). Furthermore, post-hoc pairwise comparison using the Tukey HSD test revealed which contrasts were significantly different from each

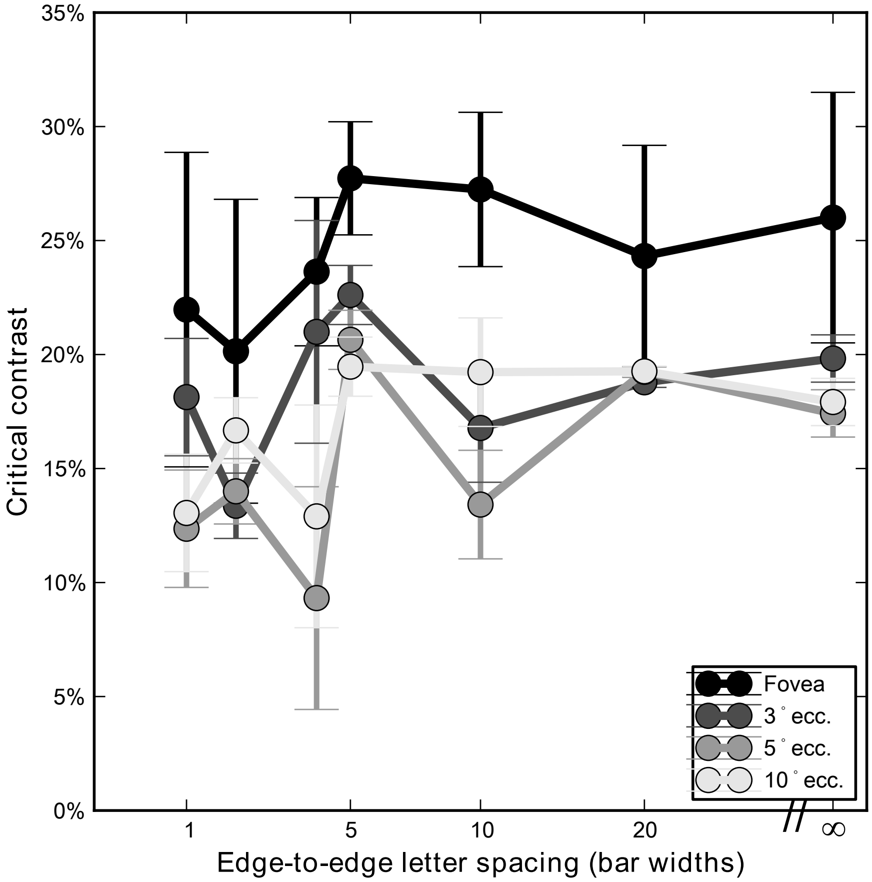


Figure 3.18: Critical contrasts for each eccentricity at the various letter spacings, averaged over all subjects. Error bars represent the standard deviation between the five subjects on the given condition.

other. Adjusted p-values from these comparisons are given in Table 3.15. In general, at 12.5% contrast and above, none of the corresponding critical spacings were significantly different from each other. Particularly at the lowest contrast (-2.5%), the nominal critical spacing was markedly reduced from the high contrast critical spacing, decreasing to less than 5 bar widths (see Table 3.14).

### 3.3.4.2 Acuity vs. contrast

In addition to the effect of letter spacing at each eccentricity and contrast level, the effect of contrast on acuity for a given condition (eccentricity and letter spacing) was analyzed, using the unconstrained two-line fit described earlier. The main parameter of interest in this

Table 3.16: Critical contrast (absolute value %) at all letter spacings, mean +/- standard deviation across subjects.

| Spacing (bar widths) | Fovea | 3 deg | 5 deg | 10 deg |
|---|---|---|---|---|
| 1 | 22+/-6.9 | 18+/-5.3 | 12+/-9.4 | 13+/-7.7 |
| 2 | 20+/-6.7 | 13+/-10 | 14+/-8.3 | 17+/-4.7 |
| 4 | 24+/-3.3 | 21+/-7.5 | 9.3+/-3.8 | 13+/-9.3 |
| 5 | 28+/-2.5 | 23+/-7.6 | 21+/-9.7 | 19+/-8.5 |
| 10 | 27+/-3.4 | 17+/-6.6 | 13+/-5.6 | 19+/-7.8 |
| 20 | 24+/-4.9 | 19+/-7 | 19+/-7.4 | 19+/-7.3 |
| Unflanked | 26+/-5.5 | 20+/-5.2 | 17+/-6 | 18+/-5.6 |
| AVG | 24+/-5.6 | 19+/-7.7 | 15+/-8.3 | 17+/-7.9 |

analysis is the critical contrast, the contrast value at which acuity begins to worsen with decreasing contrast. Figure 3.18 shows a summary of the critical contrasts at each eccentricity for all letter spacings, averaged across subjects, and Table 3.16 lists all the critical contrasts, averaged across subjects. Critical contrasts were on average lower in the periphery (14.5% (flanked) to 18% (unflanked)) than at the fovea (22% (flanked) to 26% (unflanked)), with similar values at the three non-foveal eccentricities. Repeated-measures ANOVA with both eccentricity and spacing as factors revealed a significant effect of eccentricity on critical contrast ($F_{3,112}$) = 9.635, p<0.001), a nearly significant effect of spacing ($F_{6,112}$) = 2.128, p=0.056), and no interaction ($F_{18,112}$) = 0.472, p=0.97). Post-hoc pair-wise comparison using the Tukey HSD test showed that the fovea was different from the non-foveal eccentricities ($p_{adj}$<0.01 for the fovea versus each of the three eccentricities), while the peripheral eccentricities were not different from each other ($p_{adj}$>0.24).

### 3.3.5 Discussion

In his classic 1991 review of contour interaction and crowding (Flom [1991]), Flom noted that the critical spacing value of "5 gap widths" in the fovea had not yet been extended to the retinal periphery. This extent has now been quantified as approximately 15-20 bar widths between 3 and 10 degrees eccentricity, as shown by the black points in Figure 3.19. The critical spacing is relatively invariant to changes in retinal location over this range of eccentricities. For comparison, the horizontal dotted line in Figure 3.19 indicates the character spacing on a modern chart designed with the principles to avoid foveal crowding (Bailey [2012]). Note that the characters on such a chart are outside the critical spacing at the fovea (the dotted line is above the critical spacing we measured), meaning acuity is unaffected by crowding for this letter spacing. Outside the fovea, however, because the critical spacing is much larger, adjacent characters on such a chart are within the critical spacing. Thus, non-foveal acuity measurements using a traditional letter chart may not be optimal as they are

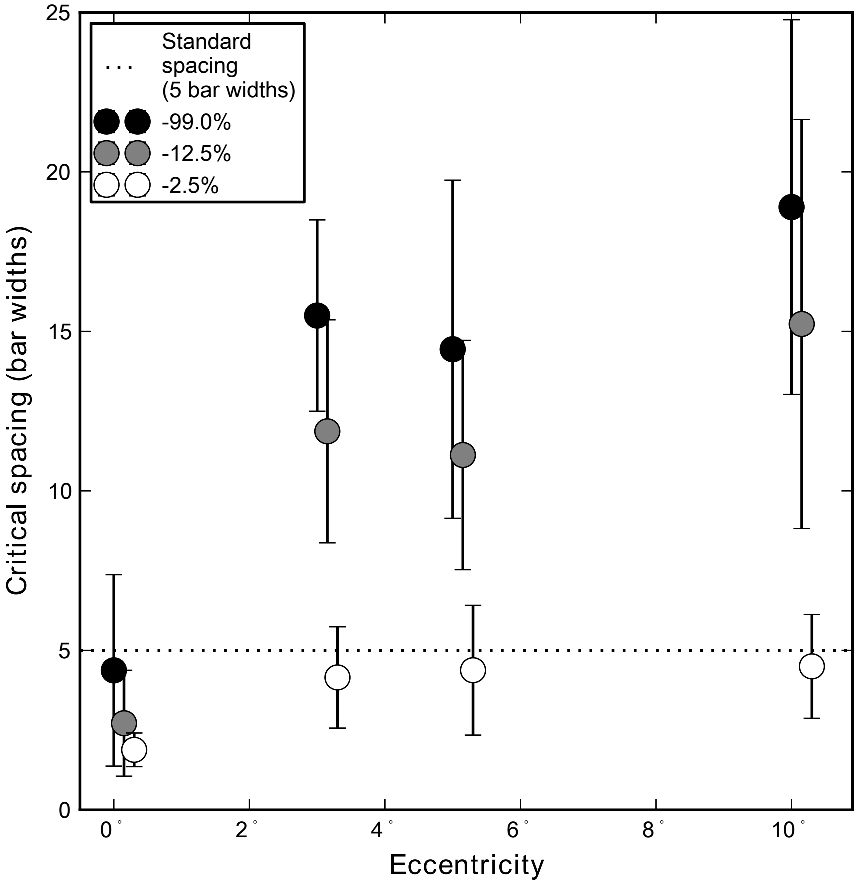


Figure 3.19: Critical spacing plotted as a function of eccentricity for contrasts of -99% (black dots), -12.5% (gray dots), and -2.5% (white dots). Each point represents the average of the five subjects, and error bars indicate the standard deviation. The dotted line shows the spacing of standard chart designs following Bailey-Lovie guidelines, which have 1 letter width (5 bar widths) between each character. Values that fall below the dotted line indicate acuity measurements not limited by crowding based on the letter spacing of a standard letter chart; acuity measurements that fall above the line will be limited by crowding with the letter spacing recommended by Bailey-Lovie chart design. We chose to show the critical spacing for -12.5% contrast to illustrate that for the commercially available low contrast versions of the Bailey-Lovie or ETDRS charts, which have a contrast close to -12.5%, the letter spacing is smaller than the critical spacing in the periphery. Hence, acuity measured using these low contrast charts (as well as high contrast charts) for patients who cannot view foveally may underestimate the peripheral acuity.

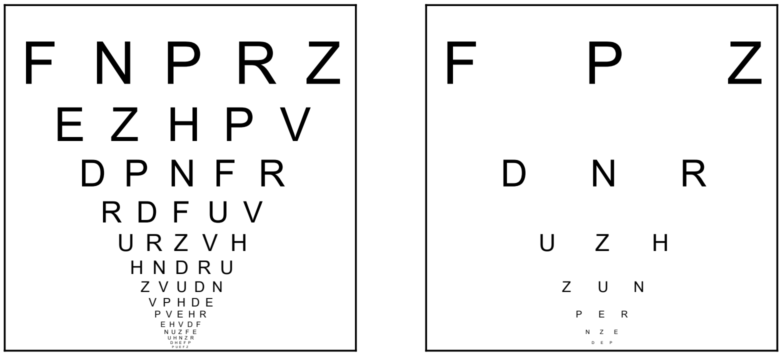


Figure 3.20: Schematic demonstration of a modified Bailey-Lovie/ETDRS chart with 3 letter widths (15 bar widths) critical spacing. Every other line was removed, and every other character of the remaining lines was removed.

limited by crowding. Figure 3.20 illustrates how a traditional letter chart (left side) could be modified to yield optimal acuity for peripheral viewing up to about 10 degrees (right side). Alternately, optotypes may be presented in isolation, if isolated letter cards are available. However, a clinician may be interested in assessing additional information with a letter chart, such as the search ability of patients. This is especially important for patients with central vision loss who often lose their place during reading of text or when viewing letters on an acuity chart. Even if isolated letters are used, it is important to know how much whitespace is necessary to surround a single letter, since any edges in the visual environment may cause lateral interference. Finally, if there is no alternative to using a traditional letter chart to assess peripheral acuity, in Section 3.3.8 we describe a simple way to predict the optimal (isolated letter) acuity based on the crowded acuity. This is possible for two reasons: 1) the crowded thresholds fall on the line with a slope of negative one as described earlier, and 2) the nominal critical spacing is roughly invariant to retinal location within 3-10 degrees eccentricity.

Since the first groundbreaking studies of Flom et al. (Flom, Weymouth, and Kahneman [1963]; Flom, Heath, and Takahashi [1963]), there have been many explorations of crowding, but all with different aims from the present study. For high contrast targets in the periphery, critical spacing has primarily been analyzed in terms of absolute angular distance (Chung, Levi, and Legge [2001]; Jacobs [1979]; Bouma [1970]; Toet and Levi [1992]; Pelli, Palomares, and Majaj [2004]; Gurnsey, Roddy, and Chanab [2011]; Tripathy and Cavanagh [2002]; Strasburger, Harvey, and Rentschler [1991]; Latham and Whitaker [1996]). The now well-established finding that absolute critical spacing changes linearly with eccentricity, and is independent of stimulus variables such as size is useful to researchers, but is of less interest when considering performance on letter charts, for which the character-to-character spacing is fixed physically, and the whole chart scales with distance. Furthermore, previous studies have measured thresholds in various ways that introduce confounding factors. First, some studies have measured threshold as a reduction in percent correct with stimuli of fixed size (Bouma [1970]; Tripathy and Cavanagh [2002]; Hess et al. [2000]), which may not directly translate to results in a threshold acuity paradigm where target and flankers are size-scaled together. Others used threshold contrast for identifying fixed-size stimuli (Chung, Levi, and Legge [2001]; Pelli, Palomares, and Majaj [2004]; Strasburger, Harvey, and Rentschler [1991]; Hariharan, Levi, and Klein [2005]; Levi, Hariharan, and Klein [2002]; Levi, Klein, and Hariharan [2002]) which is potentially a confound for crowding in general (Petrov, Popple, and McKee [2007]), and definitely cannot be utilized if evaluating the effect of contrast on critical spacing. There are several studies that have considered high-contrast, peripheral crowding with flanker spacing measured in terms of bar-widths at resolution threshold (Jacobs [1979]; Leat, Li, and Epp [1999]; Gurnsey, Roddy, and Chanab [2011]; Latham and Whitaker [1996]). Jacobs (Jacobs [1979]) and Leat, et al. (Leat, Li, and Epp [1999]) did measure threshold acuity and showed that the critical spacing for crowding in the periphery exceeded five bar widths and was potentially much greater, but the maximal spatial extent was not identified. Latham and Whitaker (Latham and Whitaker [1996]) and Gurnsey et al. (Gurnsey, Roddy, and Chanab [2011]), scaled target, flankers, and spacing as in the present study, and fit their data with complex mathematical functions, but did not determine the critical spacing in terms of bar widths that would be useful to peripheral letter chart design, nor did they examine the effects of contrast. Lastly, while Tripathy et al. (Tripathy and Cavanagh [2002]), did measure the angular critical spacing in the periphery using low contrast letters, they utilized contrast to equate the effective visibility of stimuli of various sizes, whereas we systematically varied contrast and measured acuity.

We have shown that in the periphery, the nominal critical spacing is smaller when acuity was assessed using low contrast letters than with high contrast letters. The weaker effect of crowding on acuity measurement with low contrast letters has previously been shown in the fovea (Kothe and Regan [1990]; Bailey et al. [1993]; Giaschi et al. [1993]; Pascal and Abadi [1995]; Simmers et al. [1999]; Siderov, Waugh, and Bedell [2013]), and here we report a similar effect in the periphery. The effect of crowding on acuity is even weaker in the periphery, where the low contrast critical spacing is a third of the high contrast critical spacing for the lowest contrast (-2.5%), reducing from 15-20 bar widths down to 4-5 bar widths (see Table 3.14). At this low contrast (-2.5%), nominal edge-to-edge critical spacing in the periphery was as small as the extent of high contrast letters in the periphery (4.4 bar widths). Besides determining the critical spacing required for optimal acuity measurement using letter charts with multiple letters, we were also interested in determining the critical contrast for acuity measurement that would make the assessment of low contrast acuity useful. At the fovea, acuity is independent of contrast above a letter contrast of approximately 24%. In the periphery, this critical contrast is about 17%. These findings imply that if using a letter chart printed in a letter contrast of, for example, 20%, there will be little difference in peripheral acuity between this letter chart and the high contrast version of the chart, whereas the foveal acuity (the condition which the chart may have been designed for), would exhibit a measurable difference in acuity. In other words, the additional information that could be obtained by measuring low contrast acuities will be lost. Low contrast acuity has been shown to be more sensitive in picking up diseases (Regan and Neima [1984]; Woods, Tregear, and Mitchell [1998]; Kleiner et al. [1988]; Regan and Neima [1983]; Haegerstrom-Portnoy [2005]), but to benefit from the measurement, the contrast should be low enough to affect acuity, particularly for the specific condition in which it is utilized, such as in the periphery. Here we show that the letters should be printed at a (Weber) contrast of 17% or lower in order for the chart to be useful in helping the diagnosis of diseases or to evaluate how contrast affects acuity. In sum, greater care should be used when employing tests based on contrast for measuring acuity in the periphery.

### 3.3.6 Conclusions

This study identified the nominal critical spacing for high contrast letters in the periphery, finding a critical spacing of approximately 15-20 bar widths from 3 to 10 degrees eccentricity in the lower vision field. This translates to a required increase in letter spacing from a one character gap (5 bars widths) to a 3-4 character gap (15-20 bars widths) if a chart is intended for use in the periphery such that the acuity measurement will not be affected by crowding. Thus modern letter charts, designed to avoid the effects of high contrast foveal crowding, will exhibit effects of crowding when used in the periphery. Two solutions to this problem were offered: the reduction in acuity due to crowding can be predicted mathematically, or optotypes should be given greater isolation when charts are used peripherally, such as illustrated in the right panel of Figure 3.20.

Decrease in contrast leads to reduced nominal critical spacing (less influence of crowding) for a wide range of contrasts and eccentricities, with a greater reduction in the periphery than in the fovea. Low contrast charts used in the fovea will yield acuity measurements unaffected by crowding, as noted by numerous previous reports. In the periphery, the decrease in critical spacing is more marked (even less crowding), but the low contrast peripheral critical spacing may still exhibit more crowding than the 5 bar width spacing of traditional letter charts. The finding that there is a small (but significant) difference between the critical contrast in the fovea versus the periphery implies that care should be taken when comparing contrast-dependent effects based on peripheral acuity measurements.

### 3.3.7 Flom vs. Bouma, quantifying the critical spacing of crowding.

The “critical spacing” of crowding (maximum distance over which flankers interfere with recognition of a target) was first quantified in the 1960s by Flom, et al. (1963) (Flom, Weymouth, and Kahneman [1963]) in the fovea with small high contrast Landolt Cs (at resolution thresholds: 5 min letters for normal observers). A crucial result was that the maximum spatial extent of the foveal zone of interference scaled with the MAR of the subjects, including amblyopic eyes, which have a much larger MAR than normal eyes. Others have reproduced their finding that foveal crowding extends no further than an edge-to-edge

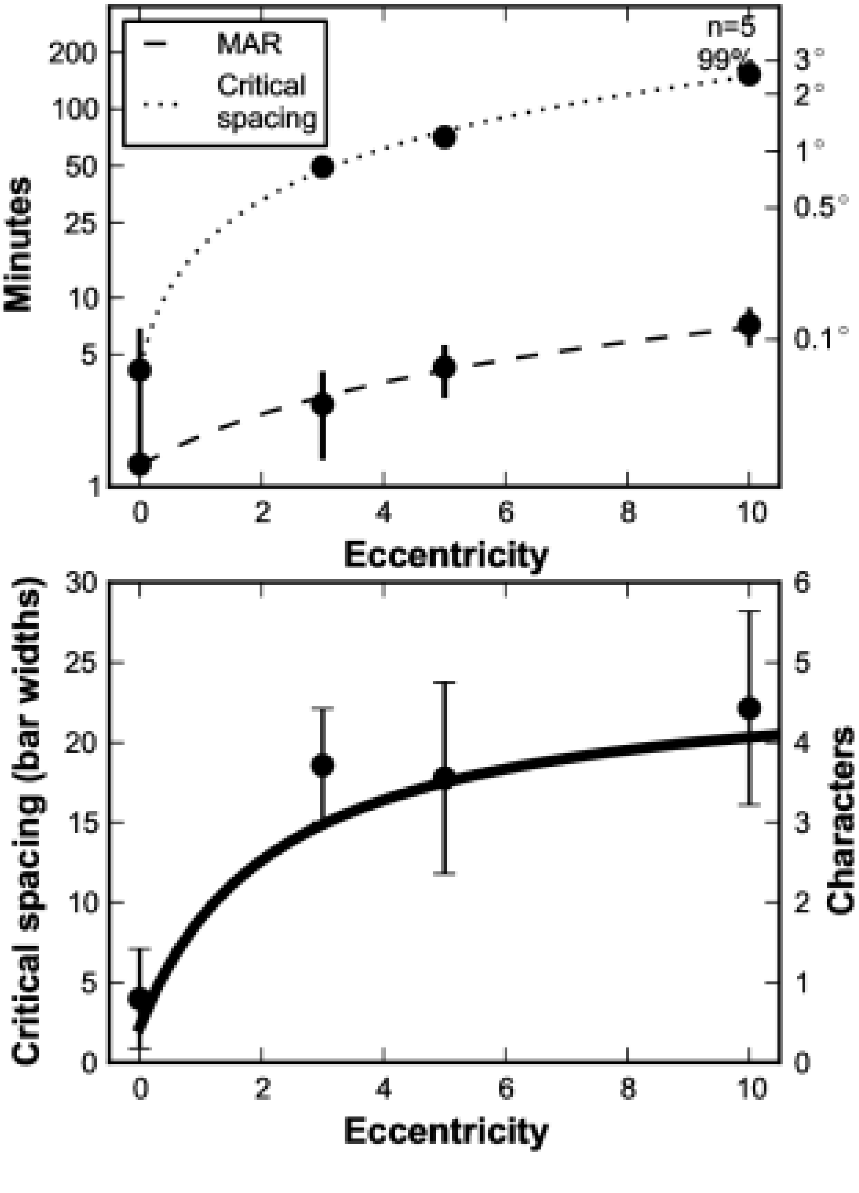


Figure 3.21: Computation of nominal critical spacing from absolute measures. Top panel: dotted line shows threshold MAR: 1/5 the threshold letter size. Dotted line shows absolute edge-to-edge critical spacing. Bottom panel shows nominal edge-to-edge critical spacing.

spacing equal to 1× the size of the target (or 5× the gap/bar width). Levi, et al. (Levi, Hariharan, and Klein [2002]) have shown this size dependence with a wide range of foveal target sizes (4' to greater than 100'), measured as the distance at which contrast thresholds increase instead of percent correct or MAR.

Peripheral studies, on the other hand, typically measure the spatial extent of crowding in terms of the angular separation between adjacent optotypes, beginning with Bouma (Bouma [1970]), who estimated crowding to extend approximately 0.5 times the eccentricity (in degrees). The suitability of this method is bolstered by the finding that the spatial extent of peripheral crowding (unlike foveal) is independent of the target size over a large range (Pelli, Palomares, and Majaj [2004]; Tripathy and Cavanagh [2002]; Levi, Hariharan, and Klein [2002]).

Information about the crowding zone in terms of bar widths at the acuity limit is important in order to identify limitations of peripheral acuity measurements using letter charts. Researchers who have evaluated peripheral crowding in terms of bar widths have found that peripheral crowding extends farther than 5 bar widths (Jacobs [1979]), the spacing of most commercially available charts, and potentially much farther. As we demonstrated earlier, we found in our study that peripheral crowding extends to 15–20 bar widths.

These two ways to specify the critical spacing are not mutually exclusive, and extending Flom's analysis into the periphery is possible. MAR for isolated letters increases linearly with eccentricity (known since Weymouth, 1958), and the extent of crowding also increases linearly with eccentricity, following Bouma's "law" (Bouma [1970]). The Bouma fraction determines the angular spatial extent of crowding, as a multiple of the eccentricity. It may depend on the task (Pelli, Palomares, and Majaj [2004]), typically taking on values from 0.15 to 0.5. The slope of isolated letter MAR vs. eccentricity can vary from 0.15 to 0.8 (measured in minutes; cf. Jacobs ([1979])).

Fundamentally, the nominal edge to edge critical spacing is determined by dividing the angular critical spacing (in min) by the isolated MAR (in min) at each eccentricity, yielding the number of bar widths for the crowding zone at the acuity limit. Figure 3.21 illustrates this procedure on the data from our 5 subjects. The top panel shows the average MAR and the average angular critical spacing (determined using the two line fit described in the previous section). The bottom panel shows the curve that results from dividing the two fitted curves. The points and error bars show the empirical nominal critical spacings determined by averaging the nominal critical spacing of the 5 subjects at each eccentricity.

The parametric line fit for isolated letter MAR is (0.57*E+1.28 min) and for critical spacing is (14.4*E+4.13 min). The corresponding $E_2$ (eccentricity where the foveal value doubles) measures are thus: $E_2$ (MAR)=2.27, $E_2$ (crit. spac.)=0.29, which are comparable with the literature (Latham and Whitaker [1996]; Strasburger, Rentschler, and Jüttner [2011]). The Bouma fraction we determined from our subjects was in the range 0.2-0.3, matching the literature (Pelli, Palomares, and Majaj [2004]).

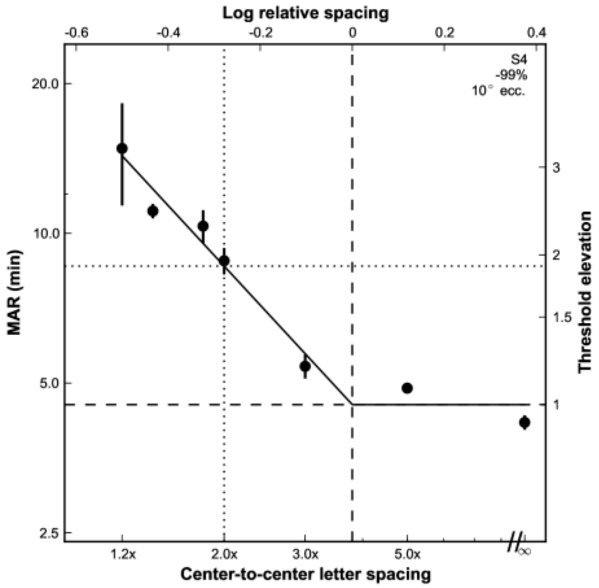


Figure 3.22: Crowded acuity as a function of center-to-center nominal spacing, illustrating -1 slope. See text for usage in estimating crowded acuity.

### 3.3.8 Two line fit justification, and prediction of crowded acuity.

This document summarizes (and motivates) the constrained two line fit of acuity versus center-to-center nominal character spacing used throughout the study. The fit originates from Denis Pelli and colleagues (Song, Levi, and Pelli [2014]; Pelli, Song, and Levi [2011]; Song [2009]) and has been used to model crowded acuity in the periphery and fovea of normal observers, amblyopes, and subjects with AMD (Chung [2014]).

The fit (shown in the Figure 3.22) has two main features. First, a horizontal portion outside the critical spacing, where flankers do not affect recognition of the target, and thus the threshold is a constant function, limited only by isolated acuity for the given condition. Second, a portion where the nearby flankers “crowd” the target, causing acuity to worsen. The crowded region can be represented by a line with a slope of -1. Here size threshold is inversely proportional to letter spacing: smaller spacings yield greater thresholds. The fit for our data is good, especially in the periphery, where the average $R^2$ is 0.85. Furthermore, there is a clear intuitive motivation for the relationship, which is described below.

The success of this model in fitting the data allows prediction of crowded acuity. Since the log spatial extent and the log magnitude of crowding-limited acuity (size threshold) are inversely proportional, knowledge of either can be used to predict performance on any part of the curve. For example, consider the dotted and dashed lines shown in Figure 3.22, which will be used to demonstrate how to estimate flanked acuity. First note that the abscissa employs a measure of spacing that is more commonly used: the spacing is given in center-to-center terms, and it is measured in units of multiples of whole character widths, rather than bar-widths. The dashed lines indicates the critical spacing and unflanked MAR for this condition, at 10 degrees eccentricity. The vertical line indicates the critical center-to-center spacing (in character widths), here approximately 4×, which corresponds to just under 15 bar widths (edge-to-edge). The upper axis shows how each letter spacing relates to this critical spacing (in log units). For example, ∼-0.3 log units halves the spacing: going from 4× to 2×. The horizontal dashed line shows the isolated MAR. The dotted lines demonstrate the consequence of using a standard letter chart spacing (one with a 1 letter (5 bar widths) edge-to-edge separation) under these conditions. First, the vertical line shows that at this center-to-center spacing (2× letter width), the relative spacing is 0.28 log units [$\log_{10}(4)$-$\log_{10}(2)$] closer than the critical spacing. Due to the inverse proportionality relationship, the crowded MAR will be $10^{0.28}$ (=1.9) times worse than the isolated acuity. This is approximately ∼0.3 log units, or 3 lines worse on a logarithmic acuity chart.

In fact, since we found the nominal critical spacing to be roughly invariant over the eccentricities we tested (3-10°, see previous section), the elevations of MAR across subjects were all approximately 2 times for this letter spacing throughout the periphery. This can be compared to the results of Jacobs ([1979]), who found a threshold elevation of approximately 1.5× for his 5 bar width condition from 2-10° eccentricity, which is comparable to the example condition: his flanking bars correspond to the innermost bars of Tumbling Es at 2×center-to-center, though the differing experimental paradigms prevent exact comparison.

The success of the negative one slope has a clear intuitive meaning. The argument depends fundamentally on the observation that the angular size of the crowding zone (the area over which flankers interfere with the target) is dependent solely on the eccentricity, being invariant to the size of the characters and other stimulus manipulations. The dependence on eccentricity has strong support dating back to Bouma ([1970]), and the independence on stimulus conditions (including size) also has some evidence, especially in the periphery (Pelli, Palomares, and Majaj [2004]; Levi, Hariharan, and Klein [2002]; Tripathy and Cavanagh [2002]). In Figure 3.23, the crowding zone is represented by the dashed ellipses. The elliptical shape has been identified previously (Toet and Levi [1992]). The basic idea is that when flankers are within the same ellipse as the target, features from the flankers will be "integrated" with the target and impair its correct identification. Each row represents a different letter size. The variable of interest is the spacing that will cause the two flanking characters to fall outside of the crowding zone. Put another way, at what letter spacing (which depends on the character size) is the critical spacing for crowding achieved? The schematic figure depicts the exact letter spacing at which the flanking optotypes are at the critical spacing for each letter size. Note the center-to-center separation (in letters) and the letter size (in arbitrary units) indicated beside each row. Because of the trade-off between size and spacing, the critical separation and the letter size have a constant product of "8," which is the horizontal radius of the crowding zone oval in this cartoon. So, larger letters permit a smaller spacing whereby flankers are beyond the critical spacing of interference, while smaller letters need to be farther away (in terms of multiples of letters). This exact inverse proportionality motivates the use of the -1 slope to fit the data, which (as stated), also has empirical support from data. The efficacy of this construction dictates the use of

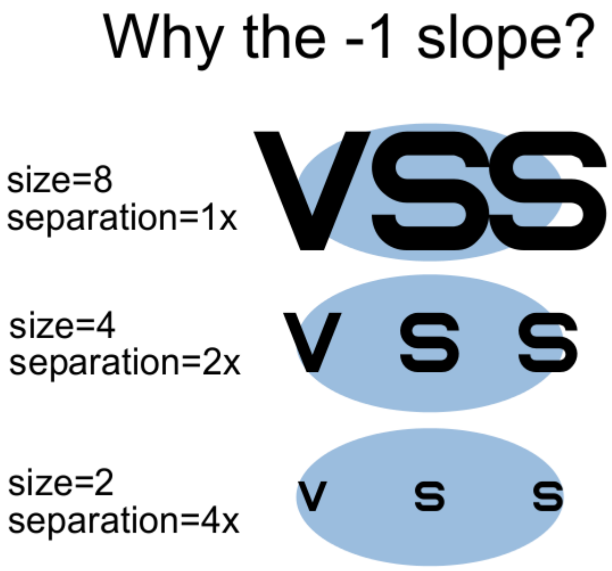


Figure 3.23: Schematic of fixed critical spacing with size-varying stimuli.

center-to-center spacing instead of edge-to-edge spacing, although conversion after fitting is possible, as was described previously.

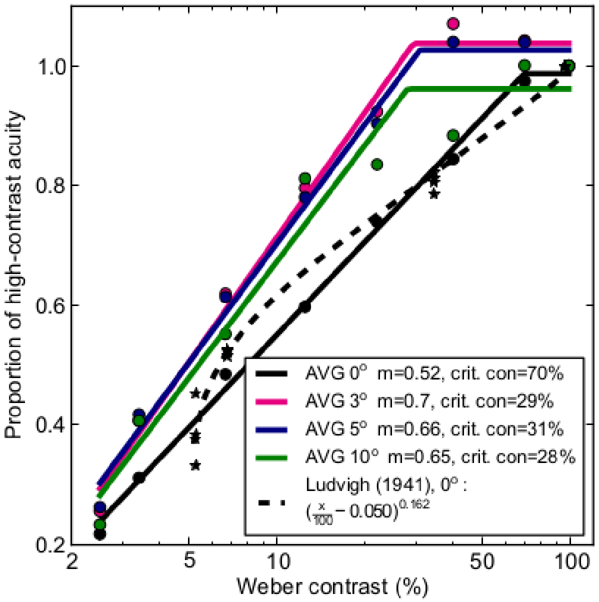


Figure 3.24: Acuity as a function of contrast, relative to high-contrast acuity. Colored dots and lines are empirical data and fits, respectively, from the present study. Stars are individual data and dashed line is fit from Ludvigh ([1941]).

## 3.3.9 A quantitative model of contrast-dependent crowded acuity across the visual field

Using the results of this experiment, it is possible to formalize a quantitative model of crowded acuity across the visual field at varying stimulus contrasts.

### 3.3.9.1 Baseline: foveal contrast-varying acuity

First, acuity in the absence of crowding can be specified as a baseline. What is the proper formulation for contrast-dependent acuity? One previous experiment which directly addressed this question is the classic paper of Ludvigh ([1941]). In this, Ludvigh modeled the *relative* acuity, or acuity normalized to a subject's best (high-contrast) acuity. There are some nuances with using his published data tables. Ludvigh measured acuities as Snellen denominators, and computed the arithmetic average of the values of three subjects, and the threshold he used corresponded to 60% correct identification. Despite these differences, Ludvigh's data are in good agreement with our foveal results, as the black symbols and lines in Figure 3.24 indicate. His fit, shown by the dotted line in Figure 3.24, comprised the function $y = (x - 0.050)^{0.162}$, yielding a relative acuity $y$ for a contrast proportion $x$. There is no justification given for this equation, so a safe assumption is that it simply presented the best mathematical fit. The shape of the function (particularly at low contrasts) is determined by just the two moderately low contrast points (5.3% and 6.8%), which would

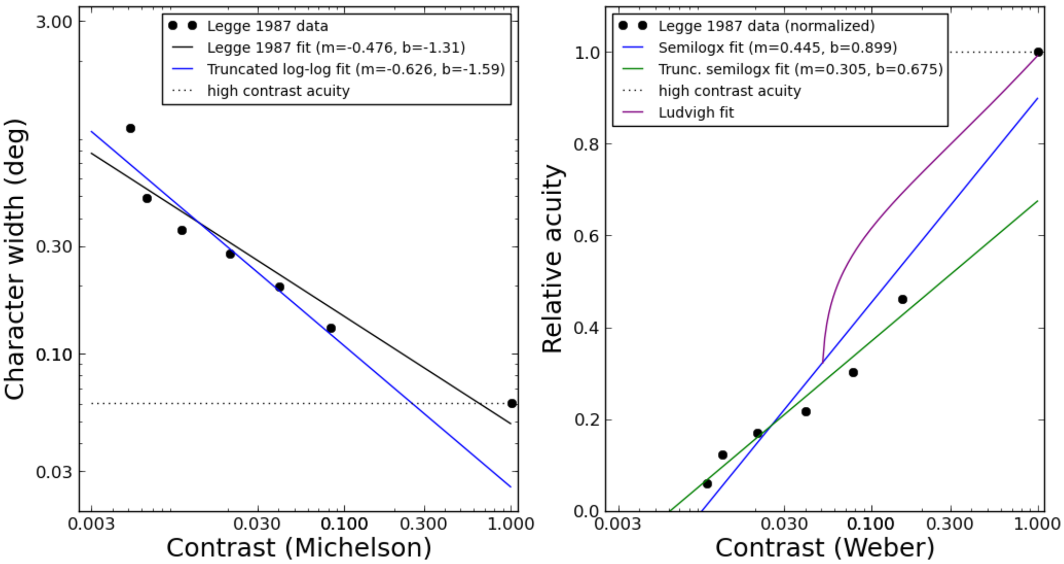


Figure 3.25: Refit of Legge 1987. Left panel shows data in original format and fit, with the addition of a truncated fit, where the high-contrast data point is omitted. Right curve shows data formatted as semilogx, with linear relative acuity on ordinate.

presumably be noisy. Furthermore, because of the form of the equation, contrasts below 0.05 (5% contrast) are undefined, which is problematic. Nevertheless, as shown by Figure 3.24, the data is not inconsistent with our empirical results or fit. Our data, on the other hand, consist of more points with improved coverage of the function. To fit our data, we use a two-line (hinged) function, which could be fit by either minimizing squared error or maximizing likelihood. There was little difference in practice for the two methods on these data, so only least-squares results are reported. For the foveal condition (black line on Figure 3.24), below approximately 70% a straight line fit our empirical data well. All three of Ludvigh's subjects were slightly above the relative acuity predicted by our model at 6.8% contrast. His other points (at 5.3%, 34.4%, and 100%) fall almost exactly on our line—the last of these by design, of course. Note the parameters of this line as given on the legend: a two-line function with slope 0.52 that saturates at 70% contrast. The saturation point is defined as the "critical contrast," abbreivated "crit. con" on the figure legend.

Another relevant result from the literature comes from the lab of Gordon Legge. Legge et al. (Legge, Rubin, and Luebker [1987]) measured contrast thresholds using Sloan letters and a method of limits to identify the minimum contrast for reliable (6 out of 8 correct) identification of letters of various sizes. Instead of plotting the data with the target size on the abscissa and contrast threshold on the ordinate, as typically done in a contrast-sensivity (CSF) plot, they plotted log contrast on the abscissa, and the corresponding size on the ordinate, much like Figure 3.24. Curiously, they defined contrast in terms of the Michelson measure ($\frac{L_{max}-L_{min}}{L_{max}+L_{min}}$), which is often used for gratings, instead of the Weber definition ($\frac{L_{max}-L_{min}}{L_{max}}$), which is more common with letter optotypes. Their data are plotted on Figure 3.25, with the left panel a simple re-plot of their data and fit as described in their paper. An additional line is fit, where the high contrast data point is excluded. Since this point is the only one constraining the fit at high contrasts, its exclusion permits determining the lower portion of a hinged-like fit. The high-contrast acuity is indicated by the horizontal dotted line. Thus the hinge would occur where the blue and dotted lines intersect, which is 0.3 for these data. However, since they used a Michelson contrast measure, this would correspond to a higher Weber value (0.46), but still less than the 0.7 that we found. The slope for the truncated fit is -0.63, much steeper than the slope of -0.5 for the full data.

To more clearly compare, the right panel transforms the data in two ways. First, the contrast values are converted to their Weber equivalents before further computation. Second, the ordinate is normalized to the best (high-contrast) acuity, which includes a reciprocal operation. The blue curve fits the data as-is (including all points), and has a slope of 0.445. The truncated curve (excluding the top point), becomes shallower, with a slope of 0.3, and would imply a "hinge" in the wrong direction, since extending the line does not cross the high-contrast ordinate value of 1.0 until much farther than the 1.0 contrast value. The unmodified fit seems more reasonable on these axes. Lastly, the purple curve plots the Ludvigh fit, which does not correspond well with either the empirical data or any of the fits.

It is also useful to place our findings in the context of typical results from other experiments which focused on measuring the contrast sensitivity function (CSF). There is a connection here that is often missed. It is well known that high-contrast acuity should correspond to the high-frequency cut-off the curve on the CSF as Legge and Rubin remind us (Legge and Rubin [1986]). As Legge and Rubin note, Virsu, Lehtio, and Rovamo ([1981]) measured a correlation of 0.84 between CSF cutoff and Snellen acuity. But, just as Legge, Rubin, and Luebker ([1987]) plotted their data as acuity versus log contrast, there are other results which may allow similar re-analysis. An obvious candidate is the study by Herse and Bedell ([1989]), which reported contrast thresholds for letters and gratings under various relevant conditions, such as at non-foveal eccentricities. We will return to their results after a discussion of general principles governing peripheral acuity.

#### 3.3.9.2 Acuity versus eccentricity: $E_2$, etc.

The variation of acuity with retinal location has also been well studied, with many examples, dating back at least to 1891 (Wertheim [1980]). The aforementioned Herse and Bedell ([1989]) is one representative example, as well as results from Jacobs ([1979]) (which summarizes many of the previous studies), Virsu and Rovamo ([1979]); Virsu, Näsänen, and Osmoviita ([1987]), Phelps ([1984]), and many others. Expressing how thresholds vary with eccentricity is a common operation, and has been quantified in several ways. Cortical magnification factor (M), or its inverse ($M^{-1}$), popularized by Virsu and Rovamo ([1979]), links performance in the visual field with physiological parameters, particularly ganglion cell and cortical cell density. "M-scaling" can be used to equalize performance across the visual field, such as increasing the size of letter optotypes to make them equally legible, as in the demo chart of Anstis ([1974]). Thresholds increase approximately linearly with most spatial tasks out to 20° or further (Weymouth [1958]). For brevity, we will call plots that express threshold-versus-eccentricity "TvE" plots, in parallel to the ubiquitous "TvC" (threshold-versus-contrast), or "TvN" (threshold-versus-noise) plots. Weymouth observed that the TvE "curves" (straight lines, actually) for different tasks had divergent slopes.

Quantification of these differences typically employs the "$E_2$" measure, popularized by Levi and Klein (Levi, Klein, and Aitsebaomo [1985]). One way to conceptualize $E_2$ is that this number reflects the eccentricity at which the foveal thresholds *double*. It is likewise the location where the TvE curve crosses the X-axis (in the negative direction). Algebra shows that these quantities are equal (except for an opposite sign), for the TvE line defined by y=mx+b.

$$mx + b = 2b \quad \text{(For what “x” does the foveal value double?)}$$
$$mx = b$$
$$x = \frac{b}{m}$$

$$mx + b = 0 \quad \text{(Where does the line intersect the X-axis?)}$$
$$mx = -b$$
$$x = -\frac{b}{m}$$

As shown above, mathematically, $E_2$ is equal to $\frac{b}{m}$, or the intercept divided by the slope. It is therefore sensitive to the foveal value, which must be accurately measured. A comprehensive history and usage of $E_2$ is given in Strasburger's exhaustive review of peripheral vision (Strasburger, Rentschler, and Jüttner [2011]). High-contrast grating or letter acuity, measured extensively by many clinicians and researchers, typically have an $E_2$ of 1.5° , which corresponds well with ganglion cell density (Levi, Klein, and Aitsebaomo [1985]), a well-known result that has been confirmed with newer technologies such as adaptive optics (Rossi and Roorda [2010]). Tasks like vernier and other hyperacuity tasks, or crowded identification, have a smaller $E_2$ of 0.6-0.8°, which more closely matches the density in cortex (Levi, Klein, and Aitsebaomo [1985]).

Strasburger's review (Strasburger, Rentschler, and Jüttner [2011]) presents a table summarizing $E_2$ values from the literature, but unfortunately its haphazard ordering makes it difficult to see the correspondence between task-dependent acuities as measured by different researchers. To ameliorate this, I have created a new table of relevant $E_2$ values in Table 3.17, ordered by $E_2$ magnitude. Note the large variation in $E_2$ values, although most values are in the range 1.2-3.0. Importantly, these values are larger than those associated

Table 3.17: $E_2$ values for acuity

| $E_2$ value | Visual task | Source, transcribed from primary reference except where noted |
|---|---|---|
| 0.88 | Landolt C acuity | Virsu 1987[1] |
| 0.977 | Landolt C acuity (mean of 20)* | (Weymouth [1958]) |
| 1.21 | Tumbling-bars, inferior (KL)+ | (Latham and Whitaker [1996]) |
| 1.275 | Tumbling-E MAR w/AO | (Rossi and Roorda [2010]) |
| 1.4 | Critical print size @ 80% | (Chung, Mansfield, and Legge [1998]) |
| 1.5 | Single-letter acuity | Herse and Bedell (1989)[2] |
| 1.512 | MAR; ave. 4 meridians | Weymouth 1928[3] |
| 1.62 | Tumbling-bars, inferior (DW)+ | (Latham and Whitaker [1996]) |
| 1.790 | MAR; horizontal (EC) | Ludvigh 1941[3] |
| 1.716 | MAR; horizontal (ave.) | Ludvigh 1941[3] |
| 1.8 | Landolt C acuity (Subj. M)* | (Weymouth [1958]) |
| 1.959 | MAR horizontal | Fick[3] |
| 2.2 | Grating acuity (PA) | (Levi, Klein, and Aitsebaomo [1985]) |
| 2.3 | Letter acuity | Anstis (1974)[4] |
| 2.41 | Grating acuity | Harvey (1985)[5] |
| 2.461 | MAR temporal | Wertheim (Weymouth [1958]) |
| 2.6 | Grating acuity (ave.) | Westheimer[6] |
| 2.676 | Landolt C acuity (Subj. K)* | (Weymouth [1958]) |
| 2.7 | Grating acuity | Virsu 1987[1] |
| 3.0 | Grating acuity (JM) | (Levi, Klein, and Aitsebaomo [1985]) |

[1] Referenced in (Drasdo [1991])
[2] Referenced in (Chung, Mansfield, and Legge [1998])
[3] Referenced in (Weymouth [1958])
[4] Referenced in (Strasburger, Rentschler, and Jüttner [2011])
[5] Referenced in (Herse and Bedell [1989])
[6] Referenced in (Levi, Klein, and Aitsebaomo [1985])

*: low luminance and short exposure (leading to difficulty in the task) are offered as possible explanations for the small average number. However, the two individual subjects shown have more reasonable values. Perhaps the method of averaging was inappropriate?
+: For both subjects, the superior visual field $E_2$ was smaller, while the two horizontal fields yielded ∼33% larger $E_2$s.

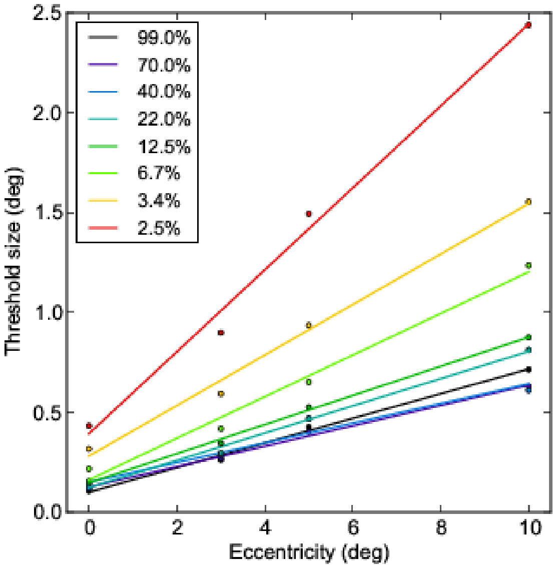


Figure 3.26: Threshold letter size as a function of contrast across eccentricities for subject S4.

with hyperacuity tasks, which are in the range 0.6–0.8 (Levi, Klein, and Aitsebaomo [1985]; Strasburger, Rentschler, and Jüttner [2011]).

For our Tumbling-E task, Figure 3.26 plots representative subject TvE data, showing threshold letter size (in degrees) versus eccentricity for the different contrasts. Note the spectrally-ordered color coding of contrast, which persists for the next two plots. The TvE plots seem to steepen as the contrast reduces, but can the numerical progression be parsimoniously captured? Figure 3.27 plots the intercept versus slope of these data, which hints toward a linear relationship between the two variables. Indeed, as Figure 3.28 shows, $E_2$ (or b/m), is constant across contrasts, with an average value of 2.5, in agreement with Table 3.17. The $E_2$s of our subjects (given in tabular format later, in Table 3.18), ranged from 1.73 to 3.2.

#### 3.3.9.3 Peripheral contrast-varying acuity

The previous section evaluated how acuity changes with eccentricity using the $E_2$ measure, and showed how this relationship holds across contrasts. Although the effect of contrast on acuity in the fovea has been thoroughly evaluated, fewer studies have explored the effects of contrast reduction on peripheral acuity. The aforementioned study of Herse and Bedell ([1989]) is one example. Although their goal was to capture the letter CSF (LCSF), the axes can be switched as described above to yield acuity versus contrast curves. Similarly, the results of Virsu and Rovamo ([1979]) for peripheral grating CSF can be transformed in this way. Lastly, Strasburger, Harvey, and Rentschler ([1991]) measured contrast thresholds for

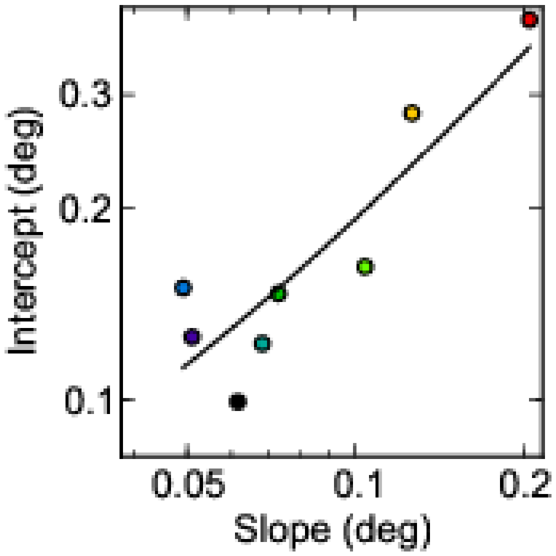


Figure 3.27: Intercept versus slope of threshold letter size at various contrast levels. Contrast colors match those of Figure 3.26.

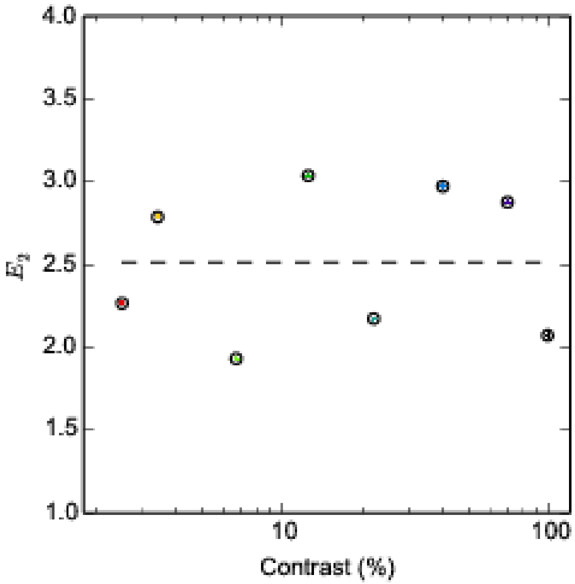


Figure 3.28: $E_2$ derived from the data of Figure 3.27 across contrasts.

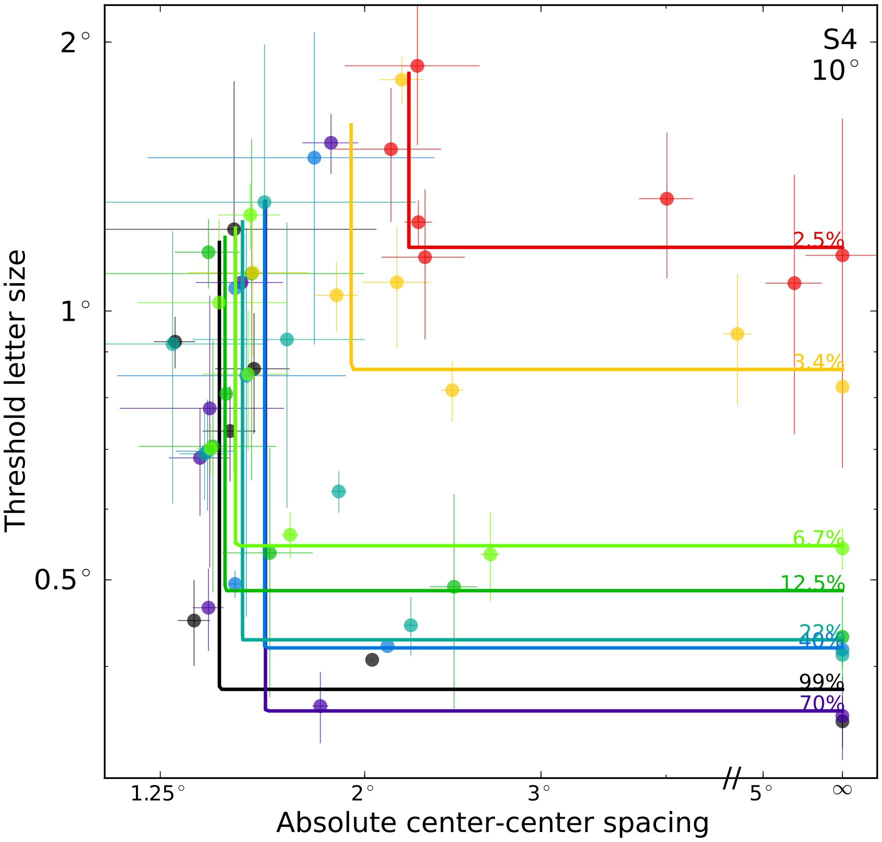


Figure 3.29: Representative data plot, with abscissa transformed to absolute units of visual angle. S4, 10°, all contrasts.

fixed letter sizes much like Herse and Bedell ([1989]). However, when these authors plotted their data as TvE curves for each contrasts (much like we do in our Figure 3.26), they rejected a linear relationship of threshold size with eccentricity for all contrasts–they found a steepening of TvE with lower contrasts, and linearity with high contrast only out to 6° of retinal eccentricity. Nevertheless, for our range of conditions, we found linear variations of threshold size with eccentricity at all contrasts, so we will move on to a discussion of acuity versus eccentricity.

#### 3.3.9.4 Contrast-varying critical spacing

Critical spacing is the minimal flanker spacing at which crowding occurs, and is typically measured using the center-to-center distance between two objects. It can be determined from plots such as Figure 3.17 by noting the value on the x-axis where the “kink-point” between the two lines occurs. Note that the units for the abscissa in Figure 3.17 are nominal, however. Since peripheral crowding is thought to be mostly invariant to target size (Levi, Hariharan, and Klein [2002]; Tripathy and Cavanagh [2002]; Pelli, Palomares, and Majaj [2004]), depending only on eccentricity, these nominal values must be converted to absolute degrees of visual angle. Furthermore, the nominal value must be converted from edge-to-edge spacing to center-to-center spacing. This simply requires adding “1” to the nominal spacing. Figure 3.29 is a representative plot of the data from Figure 3.17 converted to absolute units on the abscissa. To perform this operation, the nominal units (plus one, to yield center-to-center spacing) on the abscissa are simply multiplied by the ordinate, which represents the letter size at that threshold. The resultant product is the corresponding absolute center-to-center spacing. It is worth a reminder that our experimental paradigm fixed the nominal spacing, and varied the whole stimulus size. Thus the resultant threshold measures are not located at evenly distributed absolute spacings, such as those experiments that traverse absolute flanker spacing with the method of constant stimuli.

This plot is illuminating in several ways. First, note that the constrained two-line plot becomes the intersection of a vertical and horizontal line with this coordinate system. This is by construction, since a line with a slope of -1 (on log-log coordinates) implies a reciprocal relationship that gets canceled to a constant with the multiplication described in the previous paragraph. Second, note the systematic change in baseline threshold size (ordinate of the horizontal part of the curve) which elevates with decreasing contrast. Finally, and most importantly, the absolute critical spacing is evident from the ordinate corresponding to the vertical section of each curve. Many of these spacings all lined up at the same ordinate, specifically those of higher contrasts (cooler colors). The lower contrasts, however, represented by the hotter colors, exhibit an increase in the critical spacing. This increase of the crowding zone for low-contrast stimuli is a new finding. It has not been identified before because those experiments that have used low contrast often covaried contrast and size to keep a constant performance level (Siderov, Waugh, and Bedell [2013]). The contrasts we used (down to 2.5%) are also lower than those typically employed by crowding researchers.

Figure 3.30 plots the difference in measured critical spacing on the Y axis (in degrees), versus the difference in letter size on the x-axis (in degrees). Both these changes are in relation to the fitted high-contrast value (critical spacing and unflanked acuity, respectively) at each eccentricity. The symbols indicate eccentricity, as shown in the legend. Contrast is colored as in Figure 3.26, from hot (low contrast) to cold (high contrast) spectral colors. As already shown, lower contrasts yield larger threshold letter sizes at each eccentricity (see Section 3.3.9.3). Figure 3.30 shows that there is also a concomitant increase in the extent of the crowding zone (versus the high contrast crowding zone). Remarkably, these values change in a linear fashion.

The absolute critical spacings should vary systemically across eccentricity. Figure 3.31 plots the “Bouma ratio,” or the ratio of critical spacing divided by eccentricity. As expected, this ratio is relatively constant across the visual field, with a slightly different value for each subject. Individual differences for this quantity are not unexpected (Pelli et al. [2007]).

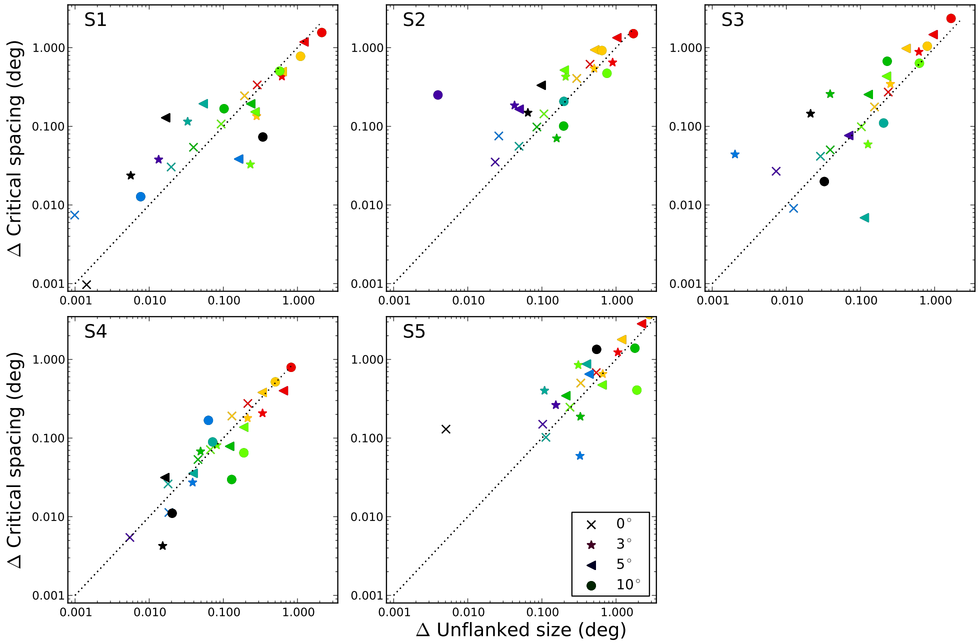


Figure 3.30: Linear change in critical spacing relative to high-contrast critical spacing as a function of linear change in acuity (relative to high-contrast acuity), across eccentricity (indicated by marker) and contrasts (colored as Figure 3.26: hotter colors denote lower contrasts.) Identity (1-to-1) line is indicated.

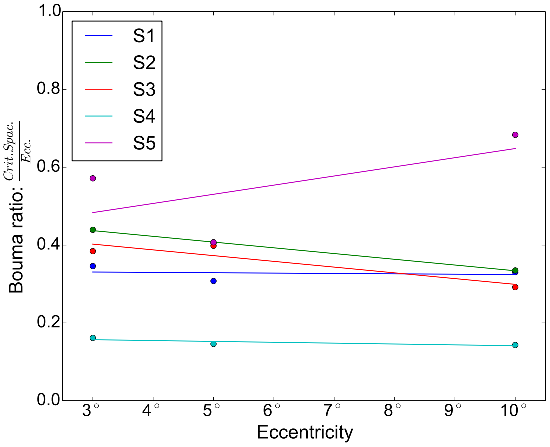


Figure 3.31: Absolute critical spacing size divided by eccentricity. (“Bouma ratio”)

Table 3.18: Fitted model parameters for the five subjects: high-contrast foveal acuity, $E_2$ (acuity changes with eccentricity), and high-contrast Bouma ratio for size of absolute critical spacing.

| Subject | Threshold( -100%, Fovea) (min) | $E_2$ | $b_{HC}$ |
|---|---|---|---|
| S1 | 1.06 | 2.4 | 0.33 |
| S2 | 1.25 | 2.9 | 0.35 |
| S3 | 1.1 | 3.2 | 0.32 |
| S4 | 0.96 | 2.9 | 0.14 |
| S5 | 2.8 | 1.73 | 0.65 |

Table 3.19: General constants used for all subjects. Foveal and peripheral critical contrast and slope from two-line fit of Figure 3.24, for normalized acuity versus contrast.

| Fovea | | Periphery | |
|---|---|---|---|
| Crit. contrast | m | Crit. contrast | m |
| 0.70 | 0.52 | 0.30 | 0.67 |

The various values are shown in the last column of Table 3.18, reasonably within the range expected from the literature (Pelli, Palomares, and Majaj [2004]).

#### 3.3.9.5 Full quantitative model

Now, given the components described in the previous sections, it is possible to create a single equation to model the critical spacing under the combination of all stimulus manipulations described.

**Terms:** *thresh(C,E)* = Threshold at contrast C and eccentricity E. *HC*=high contrast

**High contrast** threshold at eccentric locations can be determined using subject's $E_2$ and foveal acuity. Specifically, each subject's high-contrast acuity grows with eccentricity at a rate dictated by their particular $E_2$ (see Table 3.18 for values).

$$thresh(HC,E) = thresh(HC,0) \times \frac{E}{E_2} + thresh(HC,0) \tag{3.4}$$

**Contrast-dependent** threshold across the visual field can be determined using two-line fit:

$$\text{(con >= critical contrast(C,E))} : thresh(C,E) = thresh(HC,E) \tag{3.5}$$

$$\text{(con < crit. con.(C,E))} : thresh(C,E) = \frac{thresh(HC,E)}{1.0 - m_E \times \log_{10} \frac{crit.con.(E)}{con}} \tag{3.6}$$

In words, Eq. 3.5 and Eq. 3.6 mathematically define the two-line fit from Figure 3.24. If the contrast exceeds the critical contrast for that eccentricity (see Table 3.19), the threshold size will equal that of a high-contrast stimulus, which was determined in Eq. 3.4. If the contrast is less than the critical contrast, the threshold size will be elevated from the high-contrast threshold, as determined by the appropriate slope for the eccentricity ($\mathrm{m}_E$) from Table 3.19. Since it is actually the *proportion* of acuity that is linear, the scaling term appears in the denominator of Eq. 3.6.

**Crowding-limited** threshold is determined by Bouma ratio, low contrast (large size) factor, and the constrained two-line fit:

$$\text{critical spacing(C,E)} = b \times E + (thresh(C, E) - thresh(HC, E)) \tag{3.7}$$

$$\text{(spacing >= crit. spac.(C,E))} : thresh(C, E) = thresh(HC, E) \tag{3.8}$$

$$\text{(spacing < crit. spac.(C,E))} : thresh(C, E) = (\log_{10} \frac{crit.spac.(C, E)}{spacing}) \times thresh(HC, E) \tag{3.9}$$

First, critical spacing is determined in the usual way as a multiple of the eccentricity (Eq. 3.7). The correction factor from Section 3.3.9.4 is added to account for the low contrast/larger size effect. Lastly, the size versus spacing two-line fit is used. If the center-to-center flanker spacing exceeds the critical spacing, the size threshold is simply the uncrowded threshold—performance is unaffected by crowding. If the flankers are within the critical spacing, threshold elevation occurs. The exact amount of elevation defined in Eq. 3.9 is determined by the constrained two-line fit, as described in Section 3.3.8. In words, the threshold elevation is directly proportional to the difference between $\log_{10}$(nominal critical spacing) and $\log_{10}$(nominal spacing), or, equivalently, the $\log_{10}$ of their ratio.

Finally, **crowding-limited**, **contrast-dependent**, threshold is the *worse* of the two thresholds:

$$thresh(C, E) = maximum(thresh_{crowding}(C, E), thresh_{contrast}(C, E)) \tag{3.10}$$

For a statistical validation that the parameters of Table 3.18 and Table 3.19 are appropriate, a rigorous model checking procedure would iterate through the possible model space. For example, the fovea vs. periphery distinction for the contrast-dependent two-line fit parameters appears justified based on Figure 3.24, but a principled validation would allow: (1) all parameters free, (2) all parameters yoked together, or (3) fovea vs. periphery (as we have done), and compare the chi-squared value (or AIC/BIC) for each of these possibilities. In a similar vein, an illuminating diagnostic procedure is to look at the residual thresholds versus the maximum-likelihood estimates. Ideally the residual plots will be absent of any systematic patterns of errors, as systematic error patterns indicate missing terms in the model (Mosteller and Tukey [1977]).

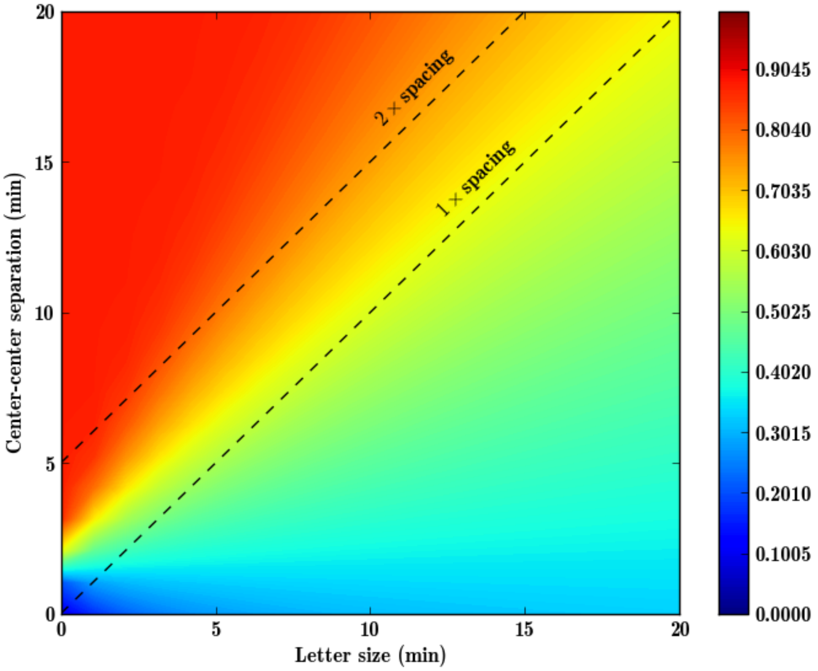


Figure 3.32: Hypothetical performance levels for crowded letter identification across size and spacing, at the fovea.

#### 3.3.9.6 Methodological questions: procedure

As discussed previously, there are several advantages to measuring crowding using the procedure described. Mainly, since size and absolute spacing are covaried, the threshold measured will always yield a true limit of discrimination, whether it is determined by crowding, or by acuity limitations. Other paradigms require stimulus titration to identify appropriate size/spacing combinations for a given eccentricity. The downside of the crowded acuity method is that it is not always clear what the limiting factor is. Typical experiments to identify the critical spacing vary the absolute spacing with fixed size stimuli (Bouma [1970]; Toet and Levi [1992]; Pelli, Palomares, and Majaj [2004]). To test size invariance, some researchers have varied the size with fixed (absolute) spacings (Tripathy and Cavanagh [2002]; Pelli, Palomares, and Majaj [2004]), and more recently one group has varied the *relative* spacing (Siderov, Waugh, and Bedell [2013]) at different sizes.

The general procedure can be made precise by considering Figure 3.32. This figure shows a simulation of how stimuli of different sizes and spacings may yield different performance levels for identification of a crowded letter at the fovea. The exact experimental details are not crucial, but rather the parameter space indicated by Figure 3.32. Most experiments determining critical spacing correspond to a vertical slice of this graph, with the critical spacing (or equivalently, the size of the crowding zone) determined by the ordinate where a certain threshold performance is achieved. For the data of Figure 3.32, it is clear that performance (and critical spacing) roughly scales with target size and separation in tandem. Note that this is particular to foveal targets and does not hold in the periphery. In the periphery, performance for a fixed separation is roughly independent of target size. Figure 3.32 would have the appearance of horizontal bands in that case.

What is missing from this figure is performance degradation based on acuity limitations. There would also be a minimum size defined by the acuity limit. Performance below this limit would be compromised. A plot like Figure 3.32 demonstrating this effect would be made up of vertical stripes, with a sharp transition indicating the acuity threshold. Horizontal slices would exhibit a sigmoidal shape, due to the known shape of performance versus size.

The result of these two limitations (crowding and acuity) would then be the product of these two two-dimensional plots, if these two factors are truly independent. It is not inconceivable there there is an interaction between acuity and crowding, making the compound effects more complicated. Is it possible to isolate the effects with our data? Unlike traditional methods, the crowded acuity staircase procedure finds thresholds at *diagonal* lines on the size/spacing grid of Figure 3.32–the dashed lines indicating the iso-nominal spacing lines. Instead of using the size threshold from the staircase as we have done, it may be possible to place each raw data point on the two dimensional grid, and perform two-dimensional smoothing (such as kernel density estimation) to reveal the true iso-performance levels at each location on Figure 3.32. These results could potentially shed light on the complex interplay of size and spacing above and beyond the method we have employed, which (in effect) loses information. The addition of the third independent variable, contrast, complicates isolation of the true interactions even further, since contrast and threshold size are (necessarily) confounded.

#### 3.3.9.7 Chapter summary

In this chapter, we have presented the results of an experiment that systematically explored the effects of stimulus contrast, visual field location, stimulus size and flanker spacing of Tumbling-E stimuli across the visual field. The ultimate goal was a quantitative model capable of describing crowded acuity at different values of each stimulus parameter. Along the way, it was necessary to model the various contributing factors to threshold *uncrowded* acuity, such as the effects of contrast, eccentricity, and their combinations, which have not been conclusively quantified previously. This model is a useful tool in describing expected baseline performance as a normative measure. Furthermore, the various coefficients presumably yield insight about the underlying anatomical sources contributing to performance, and may help advance our understanding of the locus of phenomena such as crowding.

## 3.4 Crowded Tumbling-E acuity in an S-cone isolation paradigm

### 3.4.1 Introduction

The physiological locus of crowding remains elusive. Early work successfully identified that crowding was not retinal, based on evidence from contralateral effects (Flom, Heath, and Takahashi [1963]). In these experiments, Flom, Heath, and Takahashi ([1963]) showed that eye of origin was not a factor in crowding effects. Specifically, when the target is presented to one eye, and the flanker(s) to the other eye, crowding is just as strong as when the two stimuli are presented to the same eye (see also Tripathy and Levi ([1994])). Due to the anatomical architecture of the visual system, this demonstrated that the interactions in crowding happen *after* the signals from the two eyes are combined—in primary visual cortex (V1) or later. Specifically, signals from the two eyes remain separate until V1.

Since then, however, progress in pinpointing a neural locus has been slow and inconclusive. Using psychophysics and reasoning about cortical distance, Liu et al. ([2009]) claimed than crowding was unlikely to originate in V1, as this would require signal propagation across the vertical meridian. Due to the anatomy, however, the neural locations of cross-meridian objects would be quite distant. Thus these authors posited that crowding occurs beyond V3, instead proposing hV4 or LOC (lateral occipital cortex). Using EEG, Chicherov, Plomp, and Herzog ([2014]) showed that the N1 signal is associated with crowding. This implicates high-level areas, since this signal originates in LOC. Others who champion a locus beyond V1 for crowding include Freeman and Simoncelli ([2011]) and Freeman, Donner, and Heeger ([2011]), who identify V2, or inter-area correlations, from psychophysics and modeling. Finally, Motter ([2009]), identifies V4 based on electrophysiology.

Using functional magnetic resonance imaging, however, Millin et al. ([2014]), did see a suppression of signals in V1 that could be attributed to crowding. Other neuroimaging results from the same lab confirmed this (Kwon et al. [2014]). With EEG and fMRI, Chen et al. ([2014]) confirmed modulation of signals in V1, believing early suppression to be a large contributor to crowding. Also using fMRI, Anderson et al. ([2012]) found signal changes across all of visual cortex, from V1 to V4, with the largest magnitudes in later areas.

Our approach, using only psychophysics, was to target a visual pathway with known anatomical segregation, the short wavelength sensitive (SWS) cone, or "S-cone," system. Due to its unique anatomical structure, study of this pathway has unique theoretical implications, and also practical benefits. For the latter issue, it has been shown than measuring visual function while isolating the S-cone pathway may yield diagnostic information about certain retinal diseases, including diabetic retinopathy, glaucoma, and retinitis pigmentosa (Johnson et al. [1993]; Adams et al. [1987]; Sandberg and Berson [1977]; Greenstein et al. [1989]). Theoretically, the koniocellular pathway, driven primarily by S-cones, features distinct loci in V1 (Hendry and Reid [2000]). It is conceivable that any phenomenon dependent on lateral interactions could exhibit identifiable properties that reflect the layout of

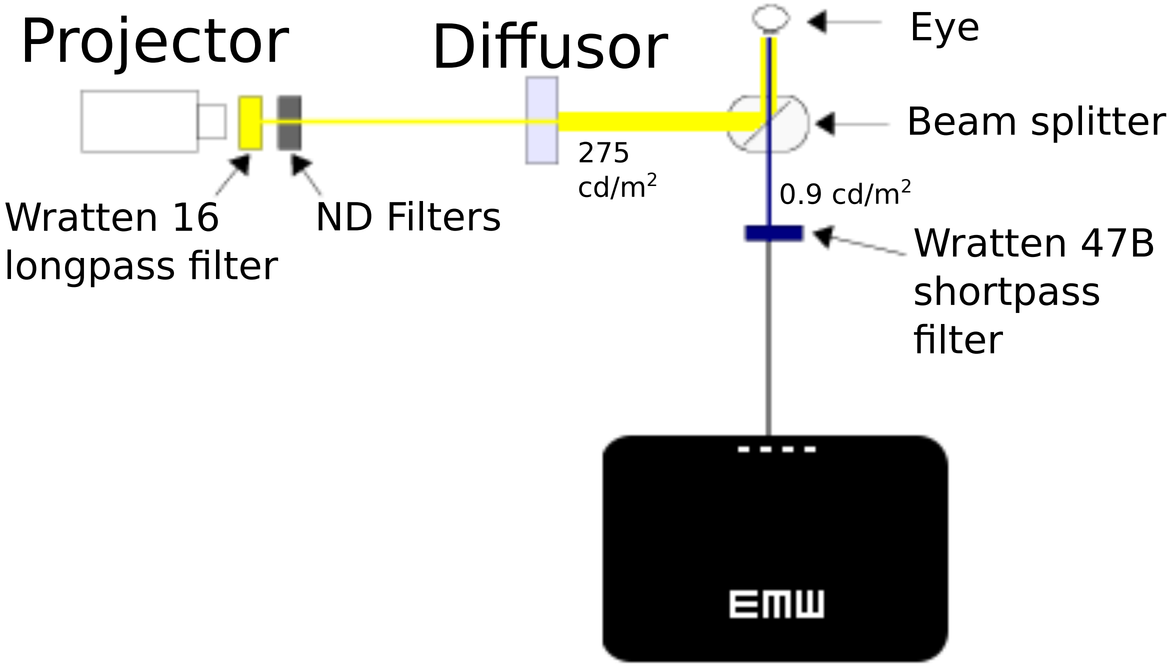


Figure 3.33: Experimental setup for S-cone isolation experiments.

the underlying pathway. A behavioral difference versus experiments that target traditional visual channels may yield insights about the nature of the interactions. How S-cone acuity changes with peripheral viewing has been explored by Anderson, Zlatkova, and Demirel ([2002]), who identified that peripheral acuity was limited by sampling from small bistratified ganglion cells. Other researchers have studied a variety of phenomena in the context of the S-cone channel, including surround suppression (Xiao and Wade [2010]), binocular rivalry (O'Shea and Williams [1996]), and oculomotor distractor effects (Sumner, Adamjee, and Mollon [2002]). As far as we know, there has been no previous measurement of crowding using an S-cone isolation paradigm.

## 3.4.2 Methods

### 3.4.2.1 Experimental apparatus

The experimental setup comprised a projector, computer screen, and optical bench with a beam splitter and various optical filters to yield the desired experimental conditions for S-cone isolation. A schematic diagram is shown in Figure 3.33. In our design of the apparatus we followed the examples of several classic papers from the literature (Rabin and Adams

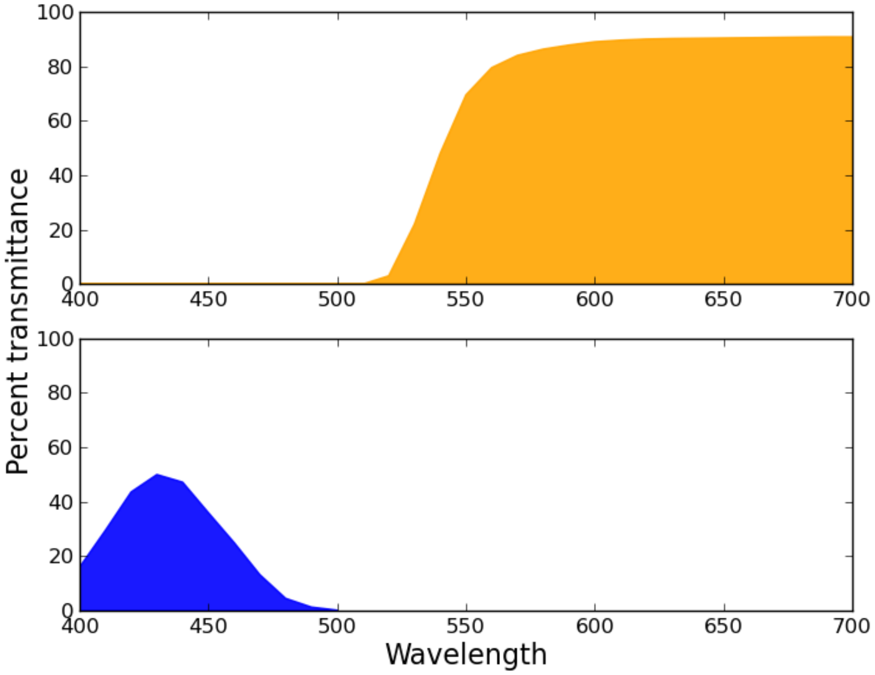


Figure 3.34: Transmittance for Wratten filters used in experiment. Top plot: Wratten 16 for adapting yellow/orange background. Bottom plot: Wratten 47b for short-wavelength-limited stimulus.

[1990]; Swanson [1989]; Anderson, Zlatkova, and Demirel [2002]; Anderson et al. [2003]). Basically, a blue stimulus display was overlaid with a bright orange/yellow background whose function was to adapt the L and M cones. To achieve this, the output from a Kodak Ektagraphic III A Projector was passed through a Wratten 16 long-pass filter and onto a diffusing screen made from tracing paper mounted on a clear glass plate. The luminance of the diffused orange/yellow light was 275 cd/m$^2$, as measured with a Minolta LS100 photometer. A beamsplitter, located in front of the subject's testing eye, combined the adapting background in a 50/50 ratio with a 15" CRT computer screen. Directly in front of the beam splitter a Wratten 47B short-pass filter was used to limit the stimuli to short wavelength light. The transmittance of the two chromatic filters used is shown graphically in Figure 3.34. The computer displayed high contrast white stimuli on a black background. The luminance of the target, after passing through the filter, was measured at 0.9 cd/m$^2$. The image of the adapting screen completely covered the monitor on the beam splitter. The experiment was conducted in a completely dark room.

A fixation line appeared near the top of the CRT display, while the stimuli were centered at 3, 5, or 10 degrees in the lower visual field. Additionally, for foveal conditions, both the fixation line and stimuli were centered in the display. Stimuli comprised Tumbling-E

Table 3.20: Summary of conditions

| | | | Stimulus | | Background | |
|---|---|---|---|---|---|---|
| | Condition | Flankers | Filter | Luminance | Filter | Luminance |
| S2 | S-cone$_2$ | 2 | Wratten 47B | 0.9 cd/m$^2$ | Wratten 16 | 275 cd/m$^2$ |
| A2 | achromatic$_2$ | 2 | Wratten 47B | 0.9 cd/m$^2$ | (off) | |
| S4 | S-cone$_4$ | 4 | Wratten 47B | 0.9 cd/m$^2$ | Wratten 16 | 275 cd/m$^2$ |
| A4 | achromatic$_4$ | 4 | Wratten 47B | 0.9 cd/m$^2$ | (off) | |
| LC4 | low contrast$_4$ | 4 | 1.8ND | 0.9 cd/m$^2$ | 2ND | 100 cd/m$^2$ |

patterns satisfying Sloan specifications (NAS-NRC Committee on Vision [1980]), rendered and presented using custom software written in Python with the PsychoPy psychophysics library (Peirce [2008]). When flanked, the target E was surrounded by either four flankers (above, below, and to each side), or by flankers positioned to the left and right of the target. Flanker spacing was specified in nominal units, relative to the stimulus size, and measured from the center of the target to the center of the flanker. For the condition with two flankers, 5 spacings were tested: 1.5×, 1.75×, 2×, 3×, and 6×. For the conditions with four flankers, an additional spacing of 1.25× was tested. Stimuli were presented for 150 milliseconds, after which subjects responded with the arrow keys on a traditional computer keyboard. Immediately after the subjects responded for a trial, the next stimulus appeared.

There were five luminance conditions in the main experiment of the study, summarized in Table 3.20. Two conditions (S2 and S4) used the S-cone isolating setup described above, with either two or four flankers. For two other conditions (A2 and A4) the projector was turned off, yielding a blue-on-black stimulus. Both two and four flankers were tested with this setup. The remaining condition (LC4) used four flankers with a highly reduced achromatic stimulus contrast. Contrast was reduced in two ways: the projector was again used, but with an 2.0 ND filter in place of the Wratten 16 filter, and the CRT screen luminance was reduced by placing a 1.8 ND filter in the optical path in place of the Wratten 47B filter. This ND filter matched the luminance of the blue-filtered stimulus in the main experiment: 0.9 cd/m$^2$. The luminance of the white background from the projector was 100 cd/m$^2$.

Finally, there was a separate control experiment to validate S-cone isolation conditions, at 8° in the lower vision field. For this experiment, neutral density filters were placed between the Wratten 16 filter and the diffusing screen, to test various luminances of the yellow adapting background. The stimulus display incorporated the Wratten 47B filter. The details of this experiment, and its motivation, are described in the subsequent Results section.

Five young subjects with normal vision participated in the experiment. All wore their normal correction and viewed the stimulus with their right eye. Their left eye was occluded with an eye patch. Subjects were seated, with their head placed in a chin rest. At the beginning of each condition subjects dark adapted for a few minutes. Subject YTY had extensive experience with peripheral vision experiments, while the other four had relatively little experience with experiments in peripheral vision. All subjects practiced each condition at each eccentricity at least once before data collection.

#### 3.4.2.2 Psychophysical procedure

In the first phase of the experiment, both of the two-flanker conditions were tested for each subject (in a random order). Next, the four-flanker conditions were tested, with the order randomized. Each block for a condition comprised all flanker spacings at a randomly chosen eccentricity. For a given set of stimulus parameters (luminance condition, eccentricity, spacing), a three-down, one-up staircase determined the smallest stimulus size that could be identified at 79% correct. Each staircase began with a stimulus that was well above threshold. To determine the final threshold, the last eight out of ten reversals were averaged. For each stimulus condition, two staircases were run on different days, and the average of the two runs is reported.

### 3.4.3 Results

#### 3.4.3.1 Validation of S-cone isolation

To confirm that our experimental paradigm was indeed isolating S-cones, we used similar logic as Rabin and Adams ([1990]) and Anderson, Zlatkova, and Demirel ([2002]). As stated by  Anderson, Zlatkova, and Demirel ([2002]), when plotting acuity (in terms of threshold letter size) versus background luminance, the curve will exhibit a steeply rising portion where acuity is mediated by L and M cones, and then a plateau which indicates the S-cone regime. The curve flattens since S-cone acuity is unaffected by the background. Rabin and Adams ([1990]) note that this shape reflects underlying Stiles $\pi$ mechanisms. Anderson, Zlatkova, and Demirel ([2002]) confirmed this effect at a variety of locations in the visual field, from the fovea to 35° in the nasal and temporal retina.

Figure 3.35 shows our results from this control experiment for each subject. The target stimuli was unchanged from the main experiment, located at 8° in the lower visual field. Neutral density filters placed between the projector and the diffusing screen modulated the luminance of the adapting background. A range of luminances were tested, from very dim backgrounds to several that were brighter than that used in the main experiment. The red dashed line in Figure 3.35 indicates the luminance of the background used in the main experiment (275 cd/m$^2$), well within the plateau region corresponding to the S-cone regime.

#### 3.4.3.2 Main experimental results

Figure 3.36 plots representative results for all four eccentricities of one subject, for the achromatic two-flanker condition. As expected, acuity worsens with eccentricity, and worsens when flankers are located near the target (left side of plot). The good fit of the "constrained -1 slope" two-line fit described earlier (Section 3.3.8) is evident in this plot. One fact that

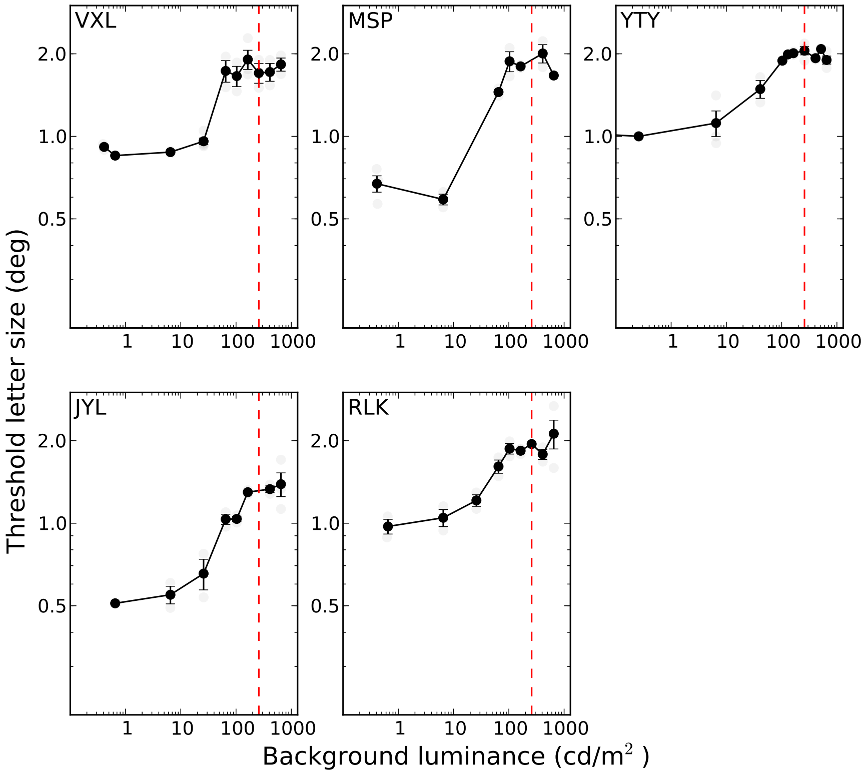


Figure 3.35: Threshold letter size as a function of yellow background luminance for S-cone isolation stimuli. Average and standard deviation of two staircases. Red dashed line indicates background luminance of main experiment.

is also clear is that the foveal condition has little crowding at the smallest spacing tested (1.5×), motivating the 1.25× spacing that was added for the four flanker conditions.

The complete set of raw empirical results for this experiment are shown in Figure 3.37 and Figure 3.38. The only difference in these two plots is whether the abscissa (flanker spacing) is specified in nominal units, or instead in absolute units of visual angle. To convert from the nominal units of Figure 3.37 to the absolute units of Figure 3.38, the nominal spacing (dependent variable) is multiplied by the ordinate (letter size), with the resultant value specifying a center-to-center spacing in degrees of visual angle. If the model completely describes the data, the plot will have the appearance of an "L"–a hinged vertical and horizontal line, shown by the dashed lines in Figure 3.38. The horizontal section, like the plateau in the nominal plots, indicates that the widely separated flankers are not affecting target identification. The vertical section corresponds to the crowded regime. In this regime there is a trade-off between spacing and size, as described in the earlier chapters. Although

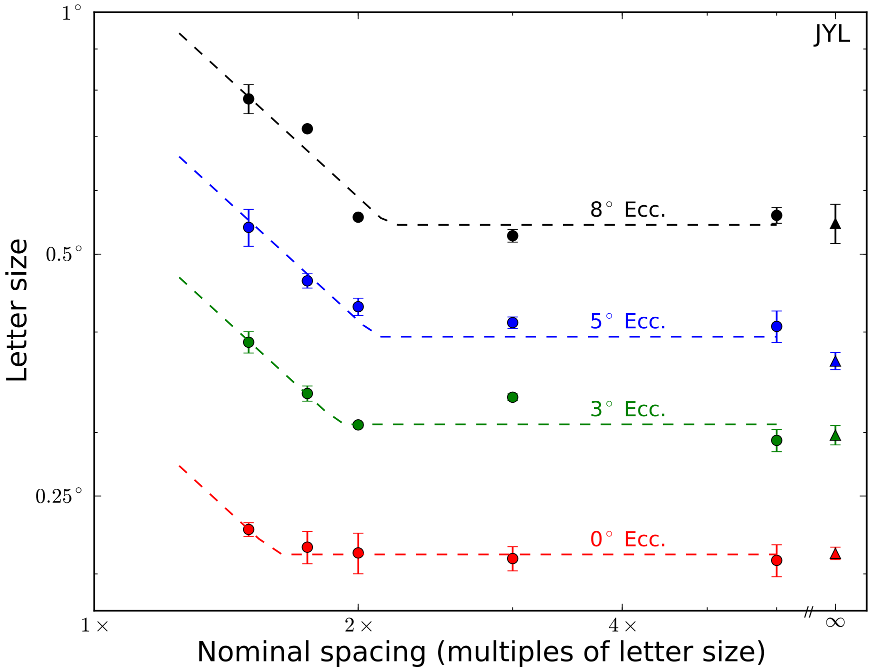


Figure 3.36: Representative results for all eccentricities and spacings for one condition. Points indicate mean of two staircases. Dashed lines indicate constrained two-line fit. Error bars indicate standard deviation of two staircases.

the fit is not perfect, it summarizes the data reasonably well overall, and quite well in some cases. For certain subjects and conditions, such as the achromatic four-flanker condition (third columns), it is clear that there is sometimes a curved knee in the function rather than an abrupt angle. This observation hints that a more complex function such as a rectangular parabola (Gurnsey, Roddy, and Chanab [2011]) could also fit the data, at the expense of additional parameters.

#### 3.4.3.3 Fitted summary results

Figure 3.39 summarizes how the fitted two-line parameters change across eccentricity. Each row is a different variable or analysis method, and each column is a subject (or the subject averages, last column). The five luminance conditions are presented as sets of points/curves on each panel. The X-axis is always degrees of eccentricity, and the Y-axis depends on the analysis. For each row (from top) ordinate is: nominal critical spacing (top row), absolute critical spacing (second row), Bouma ratio ($\frac{\text{absolute spacing}}{\text{Eccentricity}}$): (third row) or, threshold unflanked letter size (bottom row). The blue lines/points are the condition with the bright yellow background, for S-cone isolation. The black lines are the achromatic condition, with-

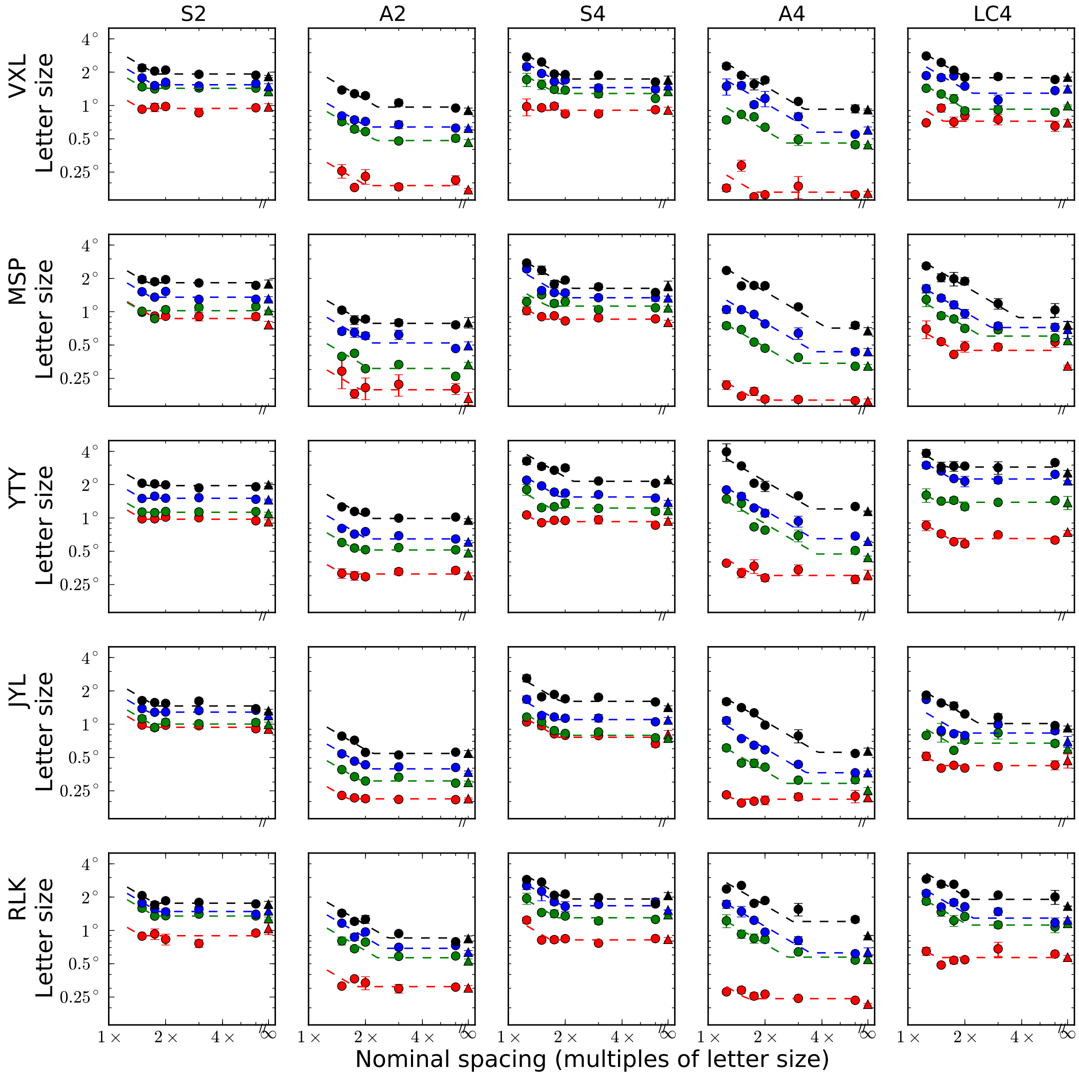


Figure 3.37: Threshold letter size as a function of nominal flanker spacing. Rows indicate subjects, columns indicate luminance condition, from left: S-cone$_2$, achromatic$_2$, S-cone$_4$, achromatic$_4$, low contrast$_4$. Eccentricities are colored as in Figure 3.36

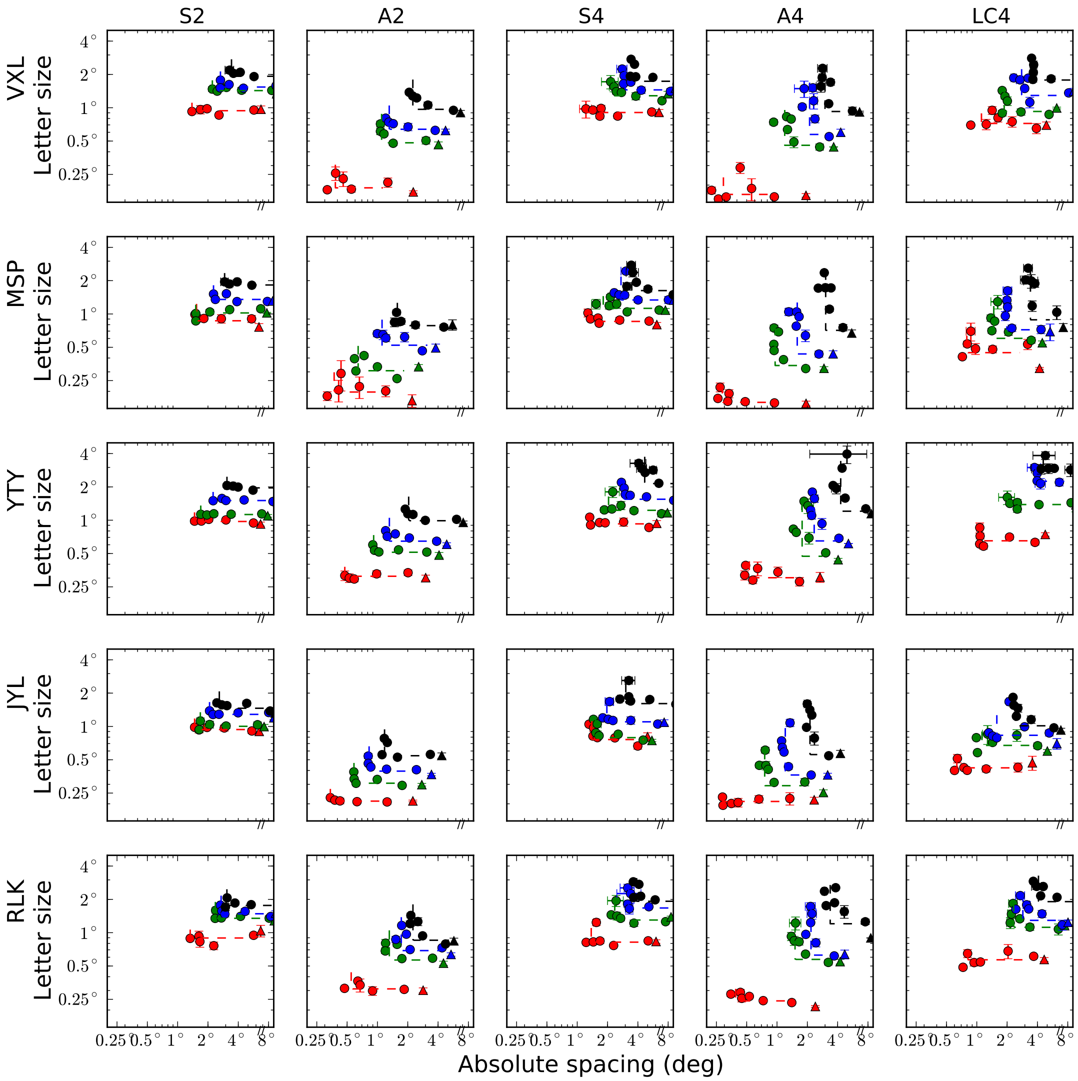


Figure 3.38: Threshold letter size as a function of absolute flanker spacing. Same data as Figure 3.37, but with converted abscissa. Eccentricities are colored as in Figure 3.36.

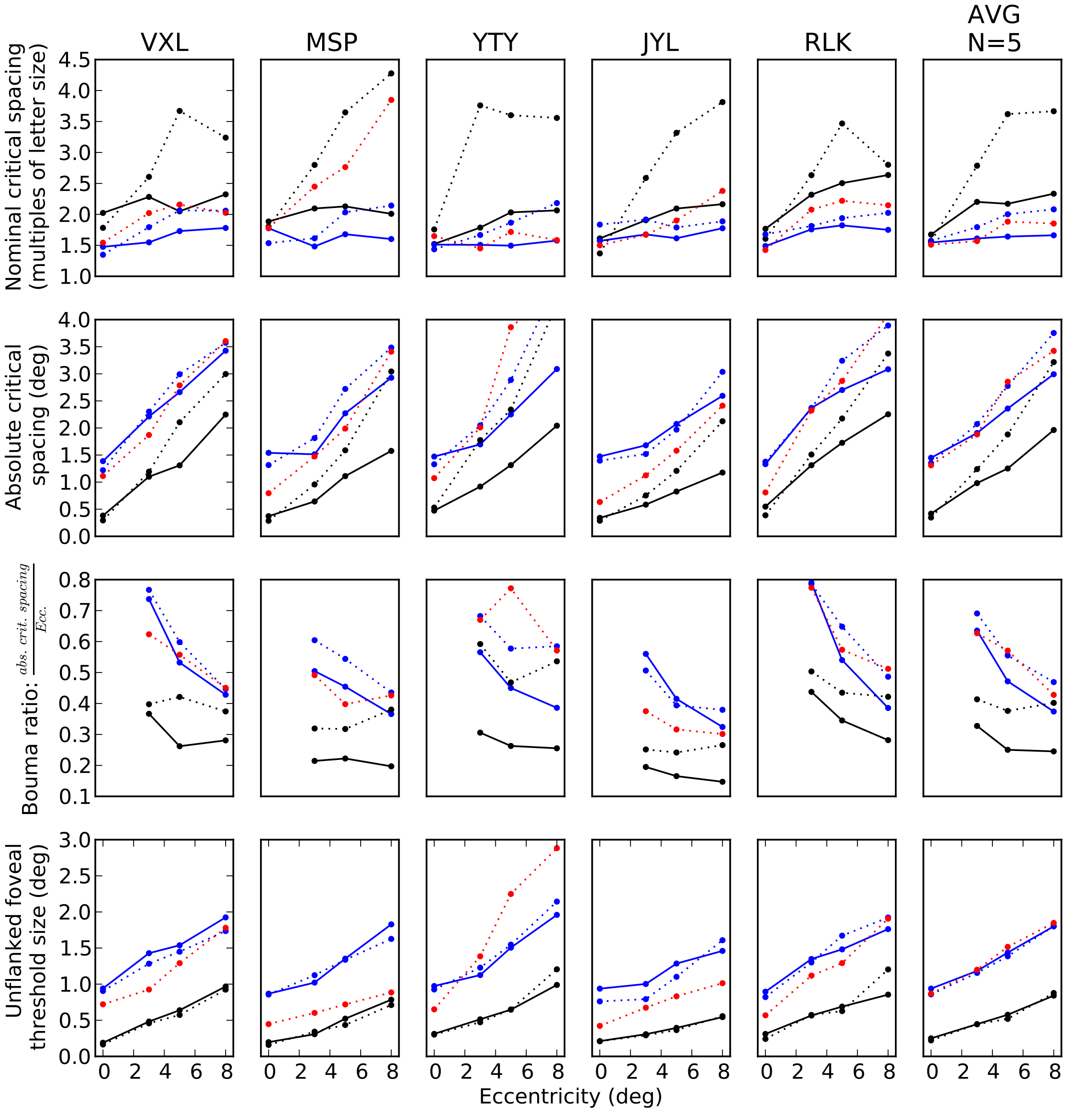


Figure 3.39: Parameters of two-line fits from Figure 3.38, as a function of eccentricity. Units of ordinate differ for each row, see text for details. Subjects are in columns. For each panel, blue curves indicate the S-cone conditions, black curves indicate achromatic conditions, and red curve indicates low contrast condition. Solid lines indicate two flankers, and dotted lines indicate four flankers.

out the adapting background. Solid lines indicate two radial flankers, while dotted lines indicate four flankers. Lastly, the red curve/points indicates the low contrast condition.

The bottom row depicts the threshold letter size in the absence of flankers (the ordinate of the flat section on the constrained two-line fit). The top row indicates the critical spacing (kink-point on two-line fit) in terms of multiples of the unflanked size (shown in the bottom row). It is rare to plot the data in this way, since the critical spacing in terms of multiples of the stimulus size is not terribly useful. This is because critical spacing varies predominantly with eccentricity, independent of stimulus size (Tripathy and Cavanagh [2002]; Pelli, Palomares, and Majaj [2004]). The shape is consistent with Figure 3.19, showing an increase in nominal spacing with eccentricity, then a plateau, for most subjects.

The critical spacing in absolute units of visual angle (second row), is much more useful, as the curves now show a natural grouping, and there is also a linear trend with eccentricity. It appears that the lines corresponding to the same flanker condition (i.e., same line style of solid or dashed) are parallel, implying similar slopes. The intercepts, on the other hand, seem to vary based on luminance condition (line color). We shall explore this further in the next section.

The remaining row shows the Bouma ratio—the absolute critical spacing divided by the eccentricity. The point at the fovea is undefined, since the eccentricity is zero. Interestingly, this plot has less regularity than expected. Generally the Bouma ratio should be the same at all eccentricities, such as our Figure 3.31. While the two high contrast conditions (black lines) may exhibit the expected straight line, the other curves show a decrease with greater eccentricity. Note that although there are individual counterexamples given in Pelli et al. ([2007]), for all subjects their trend goes the opposite way, with increasing ratio versus eccentricity. Our result calls into question the validity of a single ratio capable of describing critical spacing across eccentricities for these data.

#### 3.4.3.4 Slope and intercept of critical spacing versus eccentricity

As shown in the previous section, the ratio of the critical spacing divided by eccentricity (Bouma ratio) appears uninformative for our empirical data. The measure of $E_2$, described in Section 3.3.9.2, may be useful since it has a dependency also on the intercept term (rather than just the slope). Figure 3.40 plots the variables for lines fitted to each of the subjects/conditions from the third row of Figure 3.39. For each panel, the condition is indicated by the abscissa. The conditions are ordered as follows: S-cone 2-flanker, achromatic 2-flanker, S-cone 4-flanker, achromatic 4-flanker, low contrast 4-flanker.

$E_2$, shown in the rightmost panel, resists any clear interpretation. The intercept and slope, however, exhibit the qualitative patterns described in the previous section—the intercept is associated with the chromatic condition (achromatic or S-cone isolating), while the slope is associated with the crowding condition (two or four). This observation challenges the generality of the Bouma ratio as sufficient for capturing all changes in critical spacing versus eccentricity. Similarly, $E_2$ failed to capture important patterns in the data.

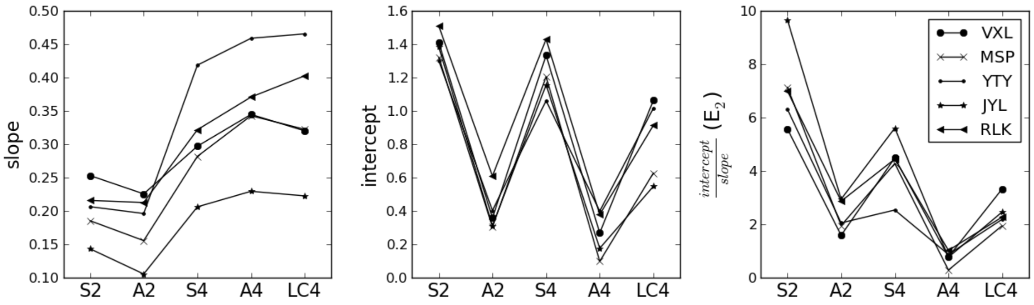


Figure 3.40: Slope, intercept, and E$_2$ ($\frac{intercept}{slope}$) of absolute critical spacing versus eccentricity for each subject, across conditions.

It could be hypothesized that the spatial extent of crowding is determined by the unflanked foveal threshold for each subject. Figure 3.41 plots the unflanked foveal threshold against both the intercept and slope of the linear fit of critical spacing versus eccentricity for each subject and condition. The good correlation of the intercept with the foveal threshold (left panel) is not surprising. At the fovea, the limiting factor in identification is the size of the target stimulus, since overlap masking will be the dominant detrimental effect. The regression line fitted to all data (green dashed line) has slope 0.64 and intercept 0.035. The relationship of the slope with the unflanked foveal threshold (right panel) is more interesting. The regression lines indicate that within each condition, larger foveal values are associated with higher (steeper) slopes. For the two shallowest conditions, involving two flankers (black and blue symbols/lines), this relationship is weakest.

### 3.4.4 Summary

This section presented experiments measuring foveal and peripheral crowded acuity under an S-cone isolation paradigm. First, we found that the constrained two-line model, used to determine the critical spacing from crowded acuity, fit our data well. The presentation of the results in absolute units (Figure 3.38) demonstrates this actuality especially well. A complicated model may capture more of the rounded shape visible in some conditions, but the two-line fit is parsimonious, provides an excellent fit for most conditions, and its defining parameters are easily interpretable—the kink point is the critical spacing, and the ordinate of the flat portion is the unflanked size. The success of this model under our stimulus conditions is important, since it validates the generality of the spacing versus size trade-off concept (see Section 3.3.8).

The control condition of a very low contrast white-on-white stimulus (red curves in Figure 3.39) indicate that the extent of crowding is correlated with per-subject resolution parameters. Since the low contrast stimuli had similar critical spacings as the S-cone stimuli, this suggests that the lateral interactions can be modeled as a function of the threshold unflanked letter size and eccentricity. Therefore, the interactions we observed in the S-cone

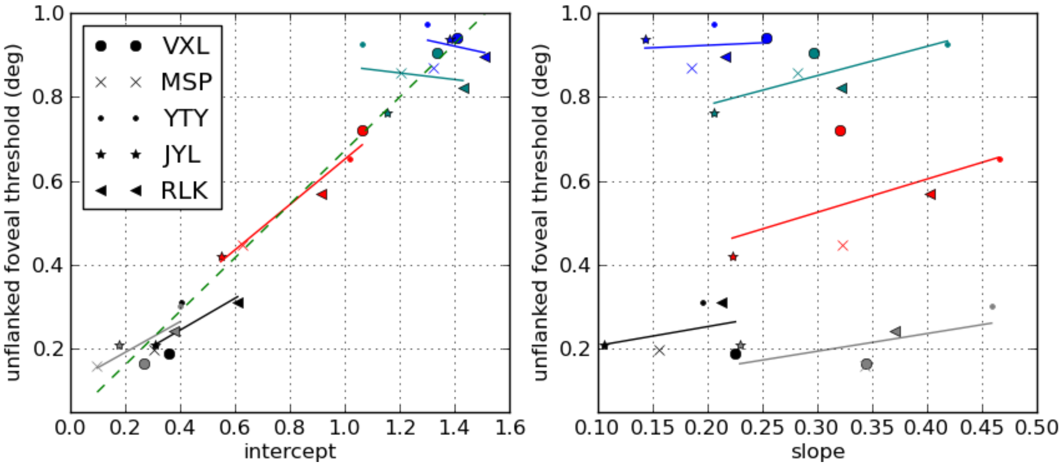


Figure 3.41: Unflanked foveal threshold versus intercept (left panel) and slope (right panel) of absolute critical spacing versus eccentricity. Symbol indicates subject as shown in key, and color indicates condition as follows: blue: S-cone$_2$, black: achromatic$_2$, teal: S-cone$_4$, gray: achromatic$_4$, red: low contrast$_4$. Lines are per-condition linear regressions. Green dashed line (left panel only) is regression for aggregated data.

isolation paradigm can be explained by the low-contrast model described in the previous section.

This observation could suggest one of two things. If crowding does have a single neural locus, it must influence a channel that is insensitive to the chromatic content of an image. This could include either the luminance channel in the standard model of the color system (Boynton [1979]), or mechanisms after the pathways are combined. Alternatively, it has been proposed that crowding acts at multiple levels of the visual system (Whitney and Levi [2011]), representing a fundamental aspect of visual processing (Chaney, Fischer, and Whitney [2014]). If so, our results could simply be interpreted as showing that the spatial extent of interference in the S-cone pathway is similar to the extent of interference in the traditional luminance-based channel. Performing a quantitative comparison, such as using cortical distance (Levi, Klein, and Aitsebaomo [1985]) is an intriguing possibility.

# Chapter 4

# Dissertation summary and conclusion

The two main themes that have I explored in this dissertation are letter recognition and crowding. To pursue the topic of letter recognition, I presented two approaches: the statistical analysis of confusion matrices, and modeling letter recognition using a convolutional neural network. To pursue the topic of crowding, I presented experiments with artificial letter-like symbols, as well as experiments using Tumbling-E stimuli under various conditions of contrast and luminance.

Consideration of confusion matrices was commonplace in the past, but has become less popular in recent time. The generality of letter recognition results is surely one reason for their fall from favor. It seems that at one time (1960s-1980s) there was hope to identify a single "objective" and "true" set of letter confusions, particularly for the capital letters. Obtaining such a definitive result would shed much light on the neural components involved in letter processing. Instead, like in so many areas of science, knowledge has made the picture fuzzier, not clearer. I believe it is both the multidimensional structure of letters, as well as the flexible and dynamic nature of our processing of visual form, that constitute the difficulties. Nevertheless, I believe our results are a solid step towards identifying real patterns in letter recognition results that may reveal condition-specific differences, detectable in whole confusion matrices.

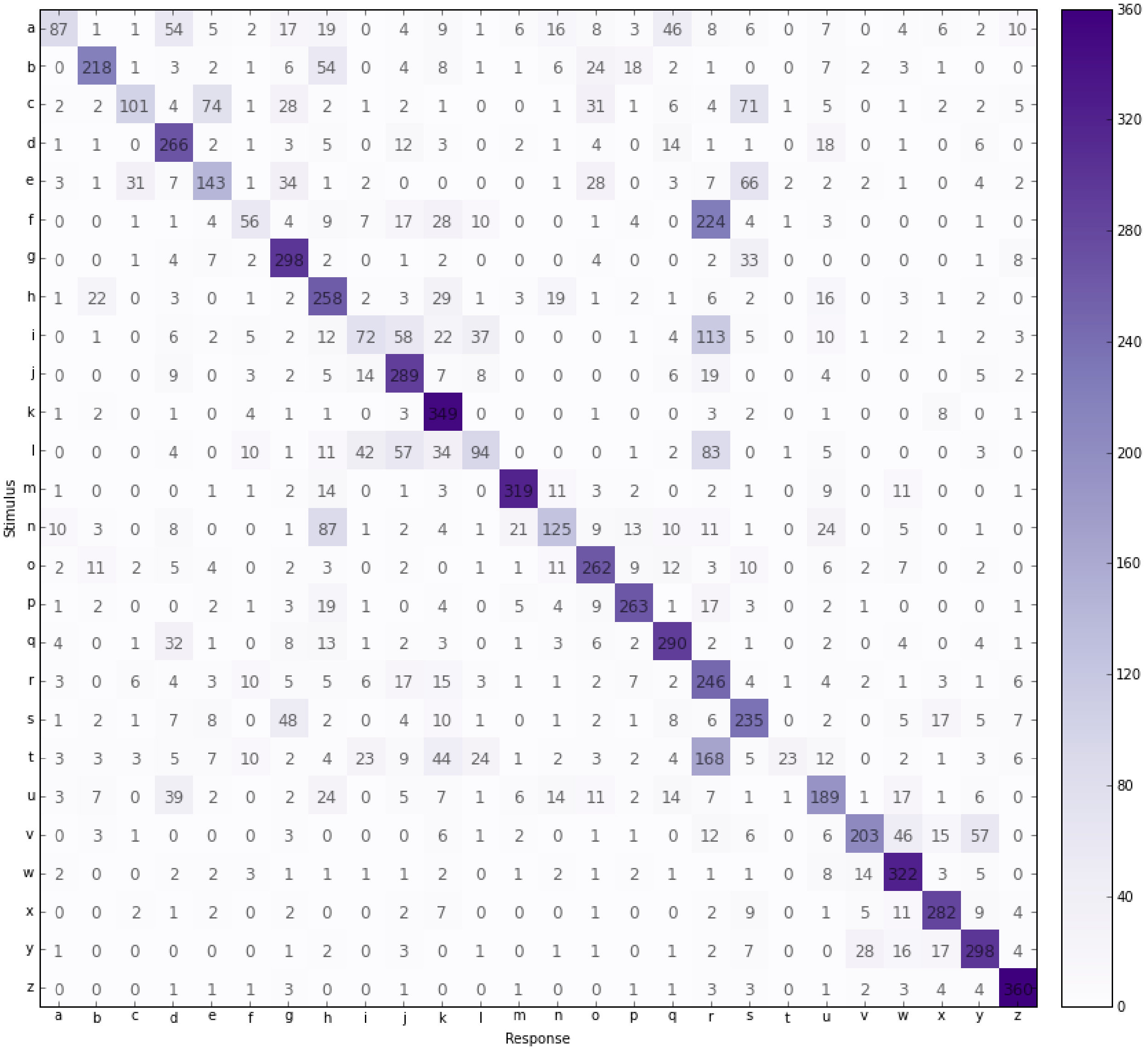
Our evaluation of a convolutional neural network is a first step towards modeling aspects of the visual system in a more biologically-inspired way. While other approaches, such as ideal-observer models, form a useful baseline for many tasks, there remains a gap in usage of biologically realistic models. Neurally plausible models may be effective at reflecting known phenomena from psychophysics, but more work is needed to test these networks with empirical results. For example, a myriad of models have been proposed for hyperacuity, but none has "stuck." I believe applying a convolutional network to a benchmark task such as this is promising. Then, a logical next step is to extend the network in various ways, such as to exhibit crowding, and evaluate its performance.

There is more opportunity for cross-pollination of the ideas in this thesis. The convolutional neural network could be applied to the line segments stimuli, and additive clustering could be used to analyze the results. For a crowding-limited neural network, the manipulations described for the Tumbling-E patterns, such as low contrast of the target and flankers, could be studied. Analysis of the spread of activation in the network may yield insights into underlying neural mechanisms. “Digital lesioning” is another effective method that could be employed along these lines.

# Appendix A

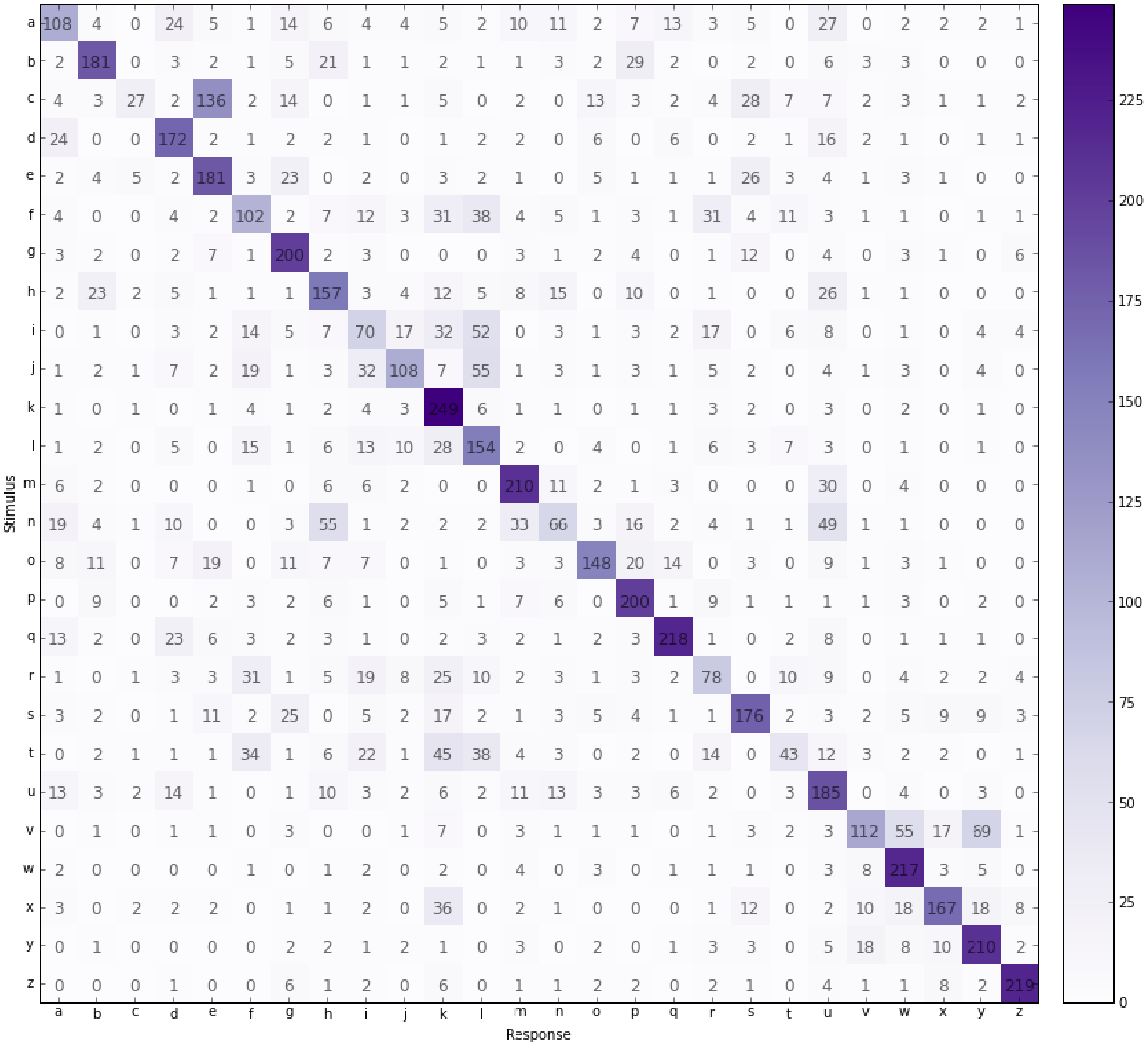
# Raw data

## A.1 Confusion matrices



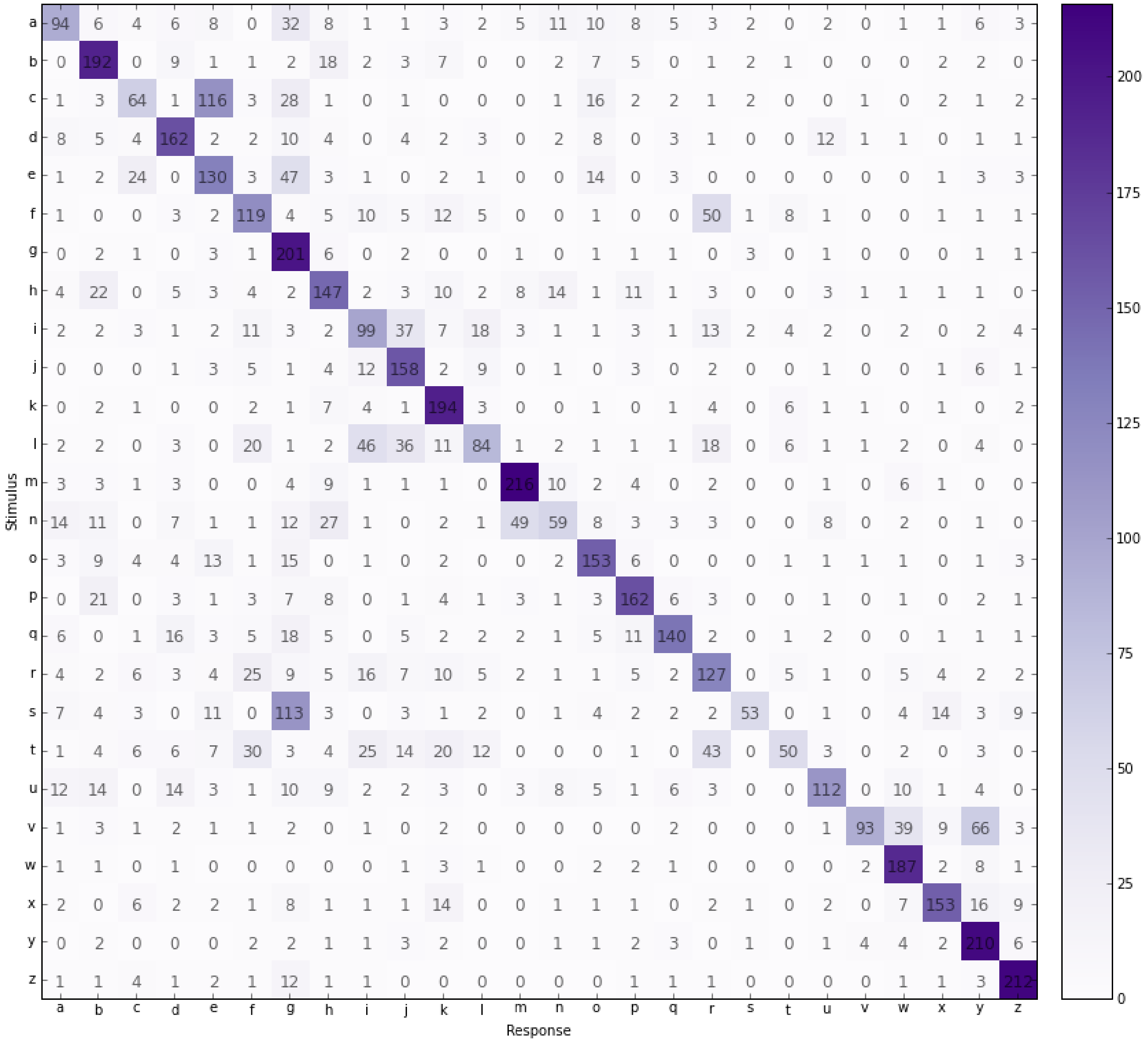
| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 87 | 1 | 1 | 54 | 5 | 2 | 17 | 19 | 0 | 4 | 9 | 1 | 6 | 16 | 8 | 3 | 46 | 8 | 6 | 0 | 7 | 0 | 4 | 6 | 2 | 10 |
| b | 0 | 218 | 1 | 3 | 2 | 1 | 6 | 54 | 0 | 4 | 8 | 1 | 1 | 6 | 24 | 18 | 2 | 1 | 0 | 0 | 7 | 2 | 3 | 1 | 0 | 0 |
| c | 2 | 2 | 101 | 4 | 74 | 1 | 28 | 2 | 1 | 2 | 1 | 0 | 0 | 1 | 31 | 1 | 6 | 4 | 71 | 1 | 5 | 0 | 1 | 2 | 2 | 5 |
| d | 1 | 1 | 0 | 266 | 2 | 1 | 3 | 5 | 0 | 12 | 3 | 0 | 2 | 1 | 4 | 0 | 14 | 1 | 1 | 0 | 18 | 0 | 1 | 0 | 6 | 0 |
| e | 3 | 1 | 31 | 7 | 143 | 1 | 34 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 28 | 0 | 3 | 7 | 66 | 2 | 2 | 2 | 1 | 0 | 4 | 2 |
| f | 0 | 0 | 1 | 1 | 4 | 56 | 4 | 9 | 7 | 17 | 28 | 10 | 0 | 0 | 1 | 4 | 0 | 224 | 4 | 1 | 3 | 0 | 0 | 0 | 1 | 0 |
| g | 0 | 0 | 1 | 4 | 7 | 2 | 298 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 33 | 0 | 0 | 0 | 0 | 0 | 1 | 8 |
| h | 1 | 22 | 0 | 3 | 0 | 1 | 2 | 258 | 2 | 3 | 29 | 1 | 3 | 19 | 1 | 2 | 1 | 6 | 2 | 0 | 16 | 0 | 3 | 1 | 2 | 0 |
| i | 0 | 1 | 0 | 6 | 2 | 5 | 2 | 12 | 72 | 58 | 22 | 37 | 0 | 0 | 0 | 1 | 4 | 113 | 5 | 0 | 10 | 1 | 2 | 1 | 2 | 3 |
| j | 0 | 0 | 0 | 9 | 0 | 3 | 2 | 5 | 14 | 289 | 7 | 8 | 0 | 0 | 0 | 0 | 6 | 19 | 0 | 0 | 4 | 0 | 0 | 0 | 5 | 2 |
| k | 1 | 2 | 0 | 1 | 0 | 4 | 1 | 1 | 0 | 3 | 349 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 2 | 0 | 1 | 0 | 0 | 8 | 0 | 1 |
| l | 0 | 0 | 0 | 4 | 0 | 10 | 1 | 11 | 42 | 57 | 34 | 94 | 0 | 0 | 0 | 1 | 2 | 83 | 0 | 1 | 5 | 0 | 0 | 0 | 3 | 0 |
| m | 1 | 0 | 0 | 0 | 1 | 1 | 2 | 14 | 0 | 1 | 3 | 0 | 319 | 11 | 3 | 2 | 0 | 2 | 1 | 0 | 9 | 0 | 11 | 0 | 0 | 1 |
| n | 10 | 3 | 0 | 8 | 0 | 0 | 1 | 87 | 1 | 2 | 4 | 1 | 21 | 125 | 9 | 13 | 10 | 11 | 1 | 0 | 24 | 0 | 5 | 0 | 1 | 0 |
| o | 2 | 11 | 2 | 5 | 4 | 0 | 2 | 3 | 0 | 2 | 0 | 1 | 1 | 11 | 262 | 9 | 12 | 3 | 10 | 0 | 6 | 2 | 7 | 0 | 2 | 0 |
| p | 1 | 2 | 0 | 0 | 2 | 1 | 3 | 19 | 1 | 0 | 4 | 0 | 5 | 4 | 9 | 263 | 1 | 17 | 3 | 0 | 2 | 1 | 0 | 0 | 0 | 1 |
| q | 4 | 0 | 1 | 32 | 1 | 0 | 8 | 13 | 1 | 2 | 3 | 0 | 1 | 3 | 6 | 2 | 290 | 2 | 1 | 0 | 2 | 0 | 4 | 0 | 4 | 1 |
| r | 3 | 0 | 6 | 4 | 3 | 10 | 5 | 5 | 6 | 17 | 15 | 3 | 1 | 1 | 2 | 7 | 2 | 246 | 4 | 1 | 4 | 2 | 1 | 3 | 1 | 6 |
| s | 1 | 2 | 1 | 7 | 8 | 0 | 48 | 2 | 0 | 4 | 10 | 1 | 0 | 1 | 2 | 1 | 8 | 6 | 235 | 0 | 2 | 0 | 5 | 17 | 5 | 7 |
| t | 3 | 3 | 3 | 5 | 7 | 10 | 2 | 4 | 23 | 9 | 44 | 24 | 1 | 2 | 3 | 2 | 4 | 168 | 5 | 23 | 12 | 0 | 2 | 1 | 3 | 6 |
| u | 3 | 7 | 0 | 39 | 2 | 0 | 2 | 24 | 0 | 5 | 7 | 1 | 6 | 14 | 11 | 2 | 14 | 7 | 1 | 1 | 189 | 1 | 17 | 1 | 6 | 0 |
| v | 0 | 3 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 6 | 1 | 2 | 0 | 1 | 1 | 0 | 12 | 6 | 0 | 6 | 203 | 46 | 15 | 57 | 0 |
| w | 2 | 0 | 0 | 2 | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 0 | 8 | 14 | 322 | 3 | 5 | 0 |
| x | 0 | 0 | 2 | 1 | 2 | 0 | 2 | 0 | 0 | 2 | 7 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 9 | 0 | 1 | 5 | 11 | 282 | 9 | 4 |
| y | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 3 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 7 | 0 | 0 | 28 | 16 | 17 | 298 | 4 |
| z | 0 | 0 | 0 | 1 | 1 | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 3 | 3 | 0 | 1 | 2 | 3 | 4 | 4 | 360 |

Figure A.1: AXL1



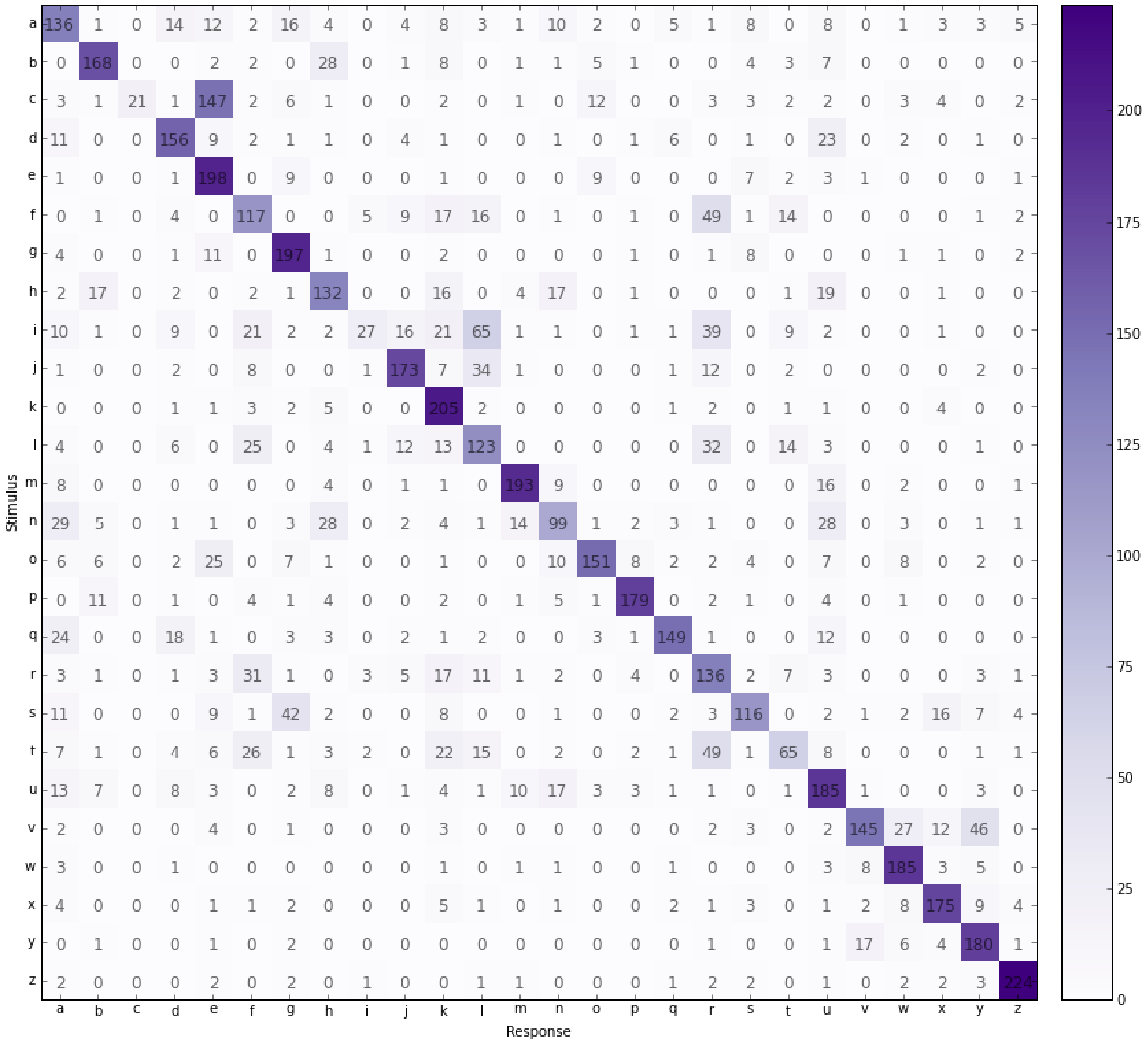
| Stimulus \ Response | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 108 | 4 | 0 | 24 | 5 | 1 | 14 | 6 | 4 | 4 | 5 | 2 | 10 | 11 | 2 | 7 | 13 | 3 | 5 | 0 | 27 | 0 | 2 | 2 | 2 | 1 |
| b | 2 | 181 | 0 | 3 | 2 | 1 | 5 | 21 | 1 | 1 | 2 | 1 | 1 | 3 | 2 | 29 | 2 | 0 | 2 | 0 | 6 | 3 | 3 | 0 | 0 | 0 |
| c | 4 | 3 | 27 | 2 | 136 | 2 | 14 | 0 | 1 | 1 | 5 | 0 | 2 | 0 | 13 | 3 | 2 | 4 | 28 | 7 | 7 | 2 | 3 | 1 | 1 | 2 |
| d | 24 | 0 | 0 | 172 | 2 | 1 | 2 | 2 | 1 | 0 | 1 | 2 | 2 | 0 | 6 | 0 | 6 | 0 | 2 | 1 | 16 | 2 | 1 | 0 | 1 | 1 |
| e | 2 | 4 | 5 | 2 | 181 | 3 | 23 | 0 | 2 | 0 | 3 | 2 | 1 | 0 | 5 | 1 | 1 | 1 | 26 | 3 | 4 | 1 | 3 | 1 | 0 | 0 |
| f | 4 | 0 | 0 | 4 | 2 | 102 | 2 | 7 | 12 | 3 | 31 | 38 | 4 | 5 | 1 | 3 | 1 | 31 | 4 | 11 | 3 | 1 | 1 | 0 | 1 | 1 |
| g | 3 | 2 | 0 | 2 | 7 | 1 | 200 | 2 | 3 | 0 | 0 | 0 | 3 | 1 | 2 | 4 | 0 | 1 | 12 | 0 | 4 | 0 | 3 | 1 | 0 | 6 |
| h | 2 | 23 | 2 | 5 | 1 | 1 | 1 | 157 | 3 | 4 | 12 | 5 | 8 | 15 | 0 | 10 | 0 | 1 | 0 | 0 | 26 | 1 | 1 | 0 | 0 | 0 |
| i | 0 | 1 | 0 | 3 | 2 | 14 | 5 | 7 | 70 | 17 | 32 | 52 | 0 | 3 | 1 | 3 | 2 | 17 | 0 | 6 | 8 | 0 | 1 | 0 | 4 | 4 |
| j | 1 | 2 | 1 | 7 | 2 | 19 | 1 | 3 | 32 | 108 | 7 | 55 | 1 | 3 | 1 | 3 | 1 | 5 | 2 | 0 | 4 | 1 | 3 | 0 | 4 | 0 |
| k | 1 | 0 | 1 | 0 | 1 | 4 | 1 | 2 | 4 | 3 | 249 | 6 | 1 | 1 | 0 | 1 | 1 | 3 | 2 | 0 | 3 | 0 | 2 | 0 | 1 | 0 |
| l | 1 | 2 | 0 | 5 | 0 | 15 | 1 | 6 | 13 | 10 | 28 | 154 | 2 | 0 | 4 | 0 | 1 | 6 | 3 | 7 | 3 | 0 | 1 | 0 | 1 | 0 |
| m | 6 | 2 | 0 | 0 | 0 | 1 | 0 | 6 | 6 | 2 | 0 | 0 | 210 | 11 | 2 | 1 | 3 | 0 | 0 | 0 | 30 | 0 | 4 | 0 | 0 | 0 |
| n | 19 | 4 | 1 | 10 | 0 | 0 | 3 | 55 | 1 | 2 | 2 | 2 | 33 | 66 | 3 | 16 | 2 | 4 | 1 | 1 | 49 | 1 | 1 | 0 | 0 | 0 |
| o | 8 | 11 | 0 | 7 | 19 | 0 | 11 | 7 | 7 | 0 | 1 | 0 | 3 | 3 | 148 | 20 | 14 | 0 | 3 | 0 | 9 | 1 | 3 | 1 | 0 | 0 |
| p | 0 | 9 | 0 | 0 | 2 | 3 | 2 | 6 | 1 | 0 | 5 | 1 | 7 | 6 | 0 | 200 | 1 | 9 | 1 | 1 | 1 | 1 | 3 | 0 | 2 | 0 |
| q | 13 | 2 | 0 | 23 | 6 | 3 | 2 | 3 | 1 | 0 | 2 | 3 | 2 | 1 | 2 | 3 | 218 | 1 | 0 | 2 | 8 | 0 | 1 | 1 | 1 | 0 |
| r | 1 | 0 | 1 | 3 | 3 | 31 | 1 | 5 | 19 | 8 | 25 | 10 | 2 | 3 | 1 | 3 | 2 | 78 | 0 | 10 | 9 | 0 | 4 | 2 | 2 | 4 |
| s | 3 | 2 | 0 | 1 | 11 | 2 | 25 | 0 | 5 | 2 | 17 | 2 | 1 | 3 | 5 | 4 | 1 | 1 | 176 | 2 | 3 | 2 | 5 | 9 | 9 | 3 |
| t | 0 | 2 | 1 | 1 | 1 | 34 | 1 | 6 | 22 | 1 | 45 | 38 | 4 | 3 | 0 | 2 | 0 | 14 | 0 | 43 | 12 | 3 | 2 | 2 | 0 | 1 |
| u | 13 | 3 | 2 | 14 | 1 | 0 | 1 | 10 | 3 | 2 | 6 | 2 | 11 | 13 | 3 | 3 | 6 | 2 | 0 | 3 | 185 | 0 | 4 | 0 | 3 | 0 |
| v | 0 | 1 | 0 | 1 | 1 | 0 | 3 | 0 | 0 | 1 | 7 | 0 | 3 | 1 | 1 | 1 | 0 | 1 | 3 | 2 | 3 | 112 | 55 | 17 | 69 | 1 |
| w | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 2 | 0 | 4 | 0 | 3 | 0 | 1 | 1 | 1 | 0 | 3 | 8 | 217 | 3 | 5 | 0 |
| x | 3 | 0 | 2 | 2 | 2 | 0 | 1 | 1 | 2 | 0 | 36 | 0 | 2 | 1 | 0 | 0 | 0 | 1 | 12 | 0 | 2 | 10 | 18 | 167 | 18 | 8 |
| y | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 2 | 1 | 2 | 1 | 0 | 3 | 0 | 2 | 0 | 1 | 3 | 3 | 0 | 5 | 18 | 8 | 10 | 210 | 2 |
| z | 0 | 0 | 0 | 1 | 0 | 0 | 6 | 1 | 2 | 0 | 6 | 0 | 1 | 1 | 2 | 2 | 0 | 2 | 1 | 0 | 4 | 1 | 1 | 8 | 2 | 219 |

Figure A.2: KMA1



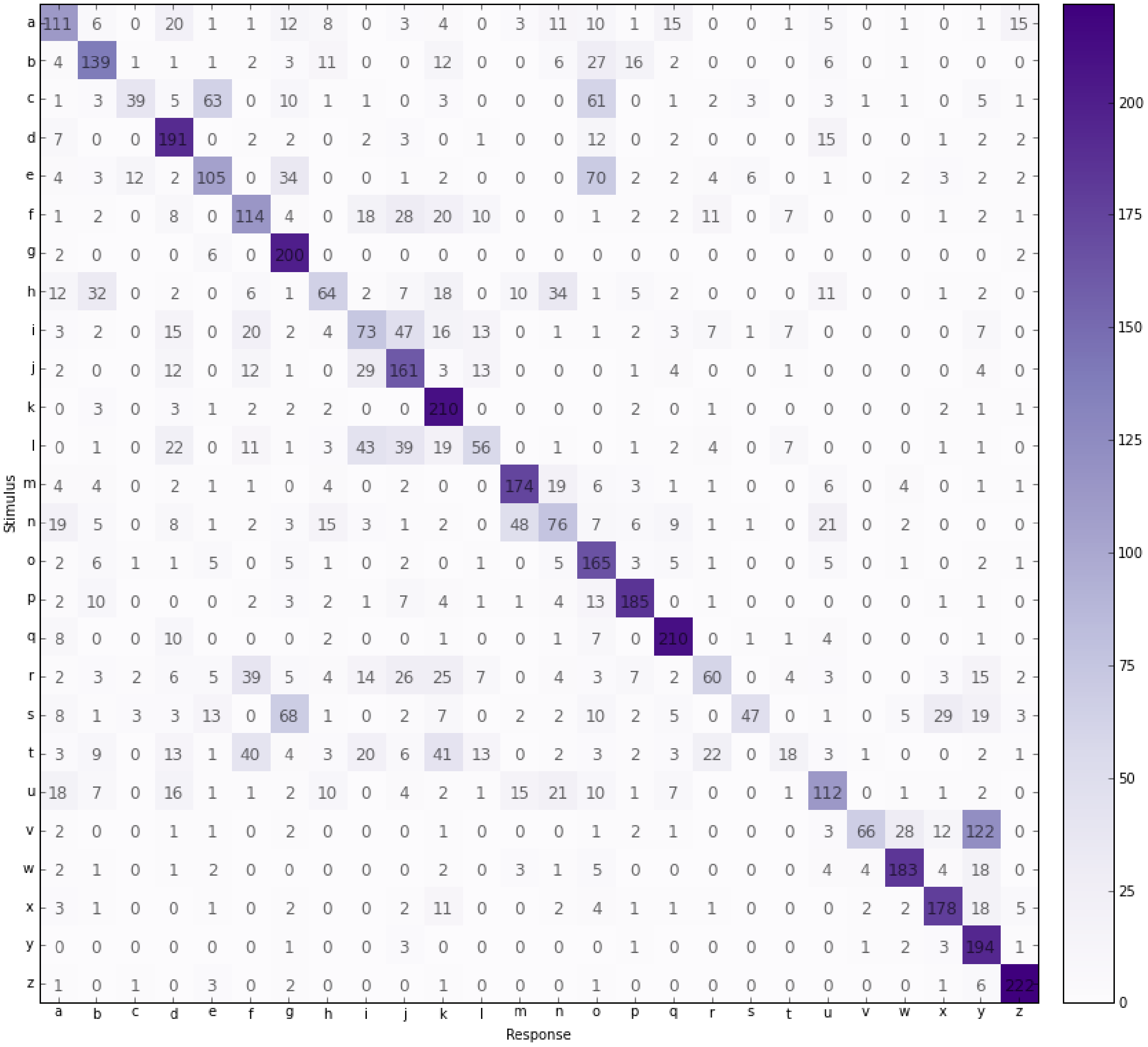
| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 94 | 6 | 4 | 6 | 8 | 0 | 32 | 8 | 1 | 1 | 3 | 2 | 5 | 11 | 10 | 8 | 5 | 3 | 2 | 0 | 2 | 0 | 1 | 1 | 6 | 3 |
| b | 0 | 192 | 0 | 9 | 1 | 1 | 2 | 18 | 2 | 3 | 7 | 0 | 0 | 2 | 7 | 5 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 2 | 2 | 0 |
| c | 1 | 3 | 64 | 1 | 116 | 3 | 28 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 16 | 2 | 2 | 1 | 2 | 0 | 0 | 1 | 0 | 2 | 1 | 2 |
| d | 8 | 5 | 4 | 162 | 2 | 2 | 10 | 4 | 0 | 4 | 2 | 3 | 0 | 2 | 8 | 0 | 3 | 1 | 0 | 0 | 12 | 1 | 1 | 0 | 1 | 1 |
| e | 1 | 2 | 24 | 0 | 130 | 3 | 47 | 3 | 1 | 0 | 2 | 1 | 0 | 0 | 14 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 3 |
| f | 1 | 0 | 0 | 3 | 2 | 119 | 4 | 5 | 10 | 5 | 12 | 5 | 0 | 0 | 1 | 0 | 0 | 50 | 1 | 8 | 1 | 0 | 0 | 1 | 1 | 1 |
| g | 0 | 2 | 1 | 0 | 3 | 1 | 201 | 6 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| h | 4 | 22 | 0 | 5 | 3 | 4 | 2 | 147 | 2 | 3 | 10 | 2 | 8 | 14 | 1 | 11 | 1 | 3 | 0 | 0 | 3 | 1 | 1 | 1 | 1 | 0 |
| i | 2 | 2 | 3 | 1 | 2 | 11 | 3 | 2 | 99 | 37 | 7 | 18 | 3 | 1 | 1 | 3 | 1 | 13 | 2 | 4 | 2 | 0 | 2 | 0 | 2 | 4 |
| j | 0 | 0 | 0 | 1 | 3 | 5 | 1 | 4 | 12 | 158 | 2 | 9 | 0 | 1 | 0 | 3 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 1 | 6 | 1 |
| k | 0 | 2 | 1 | 0 | 0 | 2 | 1 | 7 | 4 | 1 | 194 | 3 | 0 | 0 | 1 | 0 | 1 | 4 | 0 | 6 | 1 | 1 | 0 | 1 | 0 | 2 |
| l | 2 | 2 | 0 | 3 | 0 | 20 | 1 | 2 | 46 | 36 | 11 | 84 | 1 | 2 | 1 | 1 | 1 | 18 | 0 | 6 | 1 | 1 | 2 | 0 | 4 | 0 |
| m | 3 | 3 | 1 | 3 | 0 | 0 | 4 | 9 | 1 | 1 | 1 | 0 | 216 | 10 | 2 | 4 | 0 | 2 | 0 | 0 | 1 | 0 | 6 | 1 | 0 | 0 |
| n | 14 | 11 | 0 | 7 | 1 | 1 | 12 | 27 | 1 | 0 | 2 | 1 | 49 | 59 | 8 | 3 | 3 | 3 | 0 | 0 | 8 | 0 | 2 | 0 | 1 | 0 |
| o | 3 | 9 | 4 | 4 | 13 | 1 | 15 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 153 | 6 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 3 |
| p | 0 | 21 | 0 | 3 | 1 | 3 | 7 | 8 | 0 | 1 | 4 | 1 | 3 | 1 | 3 | 162 | 6 | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 1 |
| q | 6 | 0 | 1 | 16 | 3 | 5 | 18 | 5 | 0 | 5 | 2 | 2 | 2 | 1 | 5 | 11 | 140 | 2 | 0 | 1 | 2 | 0 | 0 | 1 | 1 | 1 |
| r | 4 | 2 | 6 | 3 | 4 | 25 | 9 | 5 | 16 | 7 | 10 | 5 | 2 | 1 | 1 | 5 | 2 | 127 | 0 | 5 | 1 | 0 | 5 | 4 | 2 | 2 |
| s | 7 | 4 | 3 | 0 | 11 | 0 | 113 | 3 | 0 | 3 | 1 | 2 | 0 | 1 | 4 | 2 | 2 | 2 | 53 | 0 | 1 | 0 | 4 | 14 | 3 | 9 |
| t | 1 | 4 | 6 | 6 | 7 | 30 | 3 | 4 | 25 | 14 | 20 | 12 | 0 | 0 | 0 | 1 | 0 | 43 | 0 | 50 | 3 | 0 | 2 | 0 | 3 | 0 |
| u | 12 | 14 | 0 | 14 | 3 | 1 | 10 | 9 | 2 | 2 | 3 | 0 | 3 | 8 | 5 | 1 | 6 | 3 | 0 | 0 | 112 | 0 | 10 | 1 | 4 | 0 |
| v | 1 | 3 | 1 | 2 | 1 | 1 | 2 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 93 | 39 | 9 | 66 | 3 |
| w | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 1 | 0 | 0 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 2 | 187 | 2 | 8 | 1 |
| x | 2 | 0 | 6 | 2 | 2 | 1 | 8 | 1 | 1 | 1 | 14 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 1 | 0 | 2 | 0 | 7 | 153 | 16 | 9 |
| y | 0 | 2 | 0 | 0 | 0 | 2 | 2 | 1 | 1 | 3 | 2 | 0 | 0 | 1 | 1 | 2 | 3 | 0 | 1 | 0 | 1 | 4 | 4 | 2 | 210 | 6 |
| z | 1 | 1 | 4 | 1 | 2 | 1 | 12 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 212 |

Figure A.3: RXK



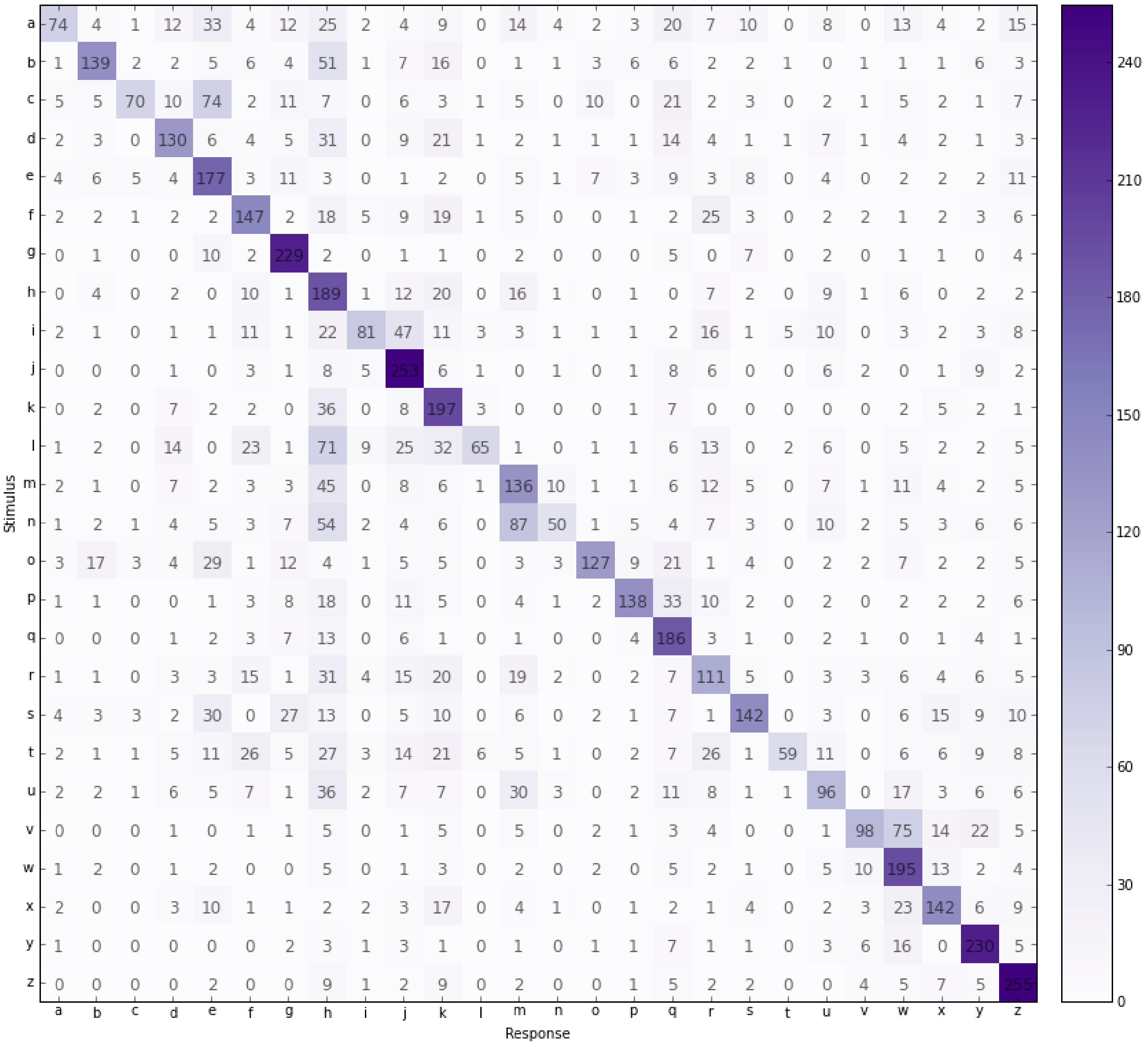
| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 136 | 1 | 0 | 14 | 12 | 2 | 16 | 4 | 0 | 4 | 8 | 3 | 1 | 10 | 2 | 0 | 5 | 1 | 8 | 0 | 8 | 0 | 1 | 3 | 3 | 5 |
| b | 0 | 168 | 0 | 0 | 2 | 2 | 0 | 28 | 0 | 1 | 8 | 0 | 1 | 1 | 5 | 1 | 0 | 0 | 4 | 3 | 7 | 0 | 0 | 0 | 0 | 0 |
| c | 3 | 1 | 21 | 1 | 147 | 2 | 6 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 12 | 0 | 0 | 3 | 3 | 2 | 2 | 0 | 3 | 4 | 0 | 2 |
| d | 11 | 0 | 0 | 156 | 9 | 2 | 1 | 1 | 0 | 4 | 1 | 0 | 0 | 1 | 0 | 1 | 6 | 0 | 1 | 0 | 23 | 0 | 2 | 0 | 1 | 0 |
| e | 1 | 0 | 0 | 1 | 198 | 0 | 9 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 7 | 2 | 3 | 1 | 0 | 0 | 0 | 1 |
| f | 0 | 1 | 0 | 4 | 0 | 117 | 0 | 0 | 5 | 9 | 17 | 16 | 0 | 1 | 0 | 1 | 0 | 49 | 1 | 14 | 0 | 0 | 0 | 0 | 1 | 2 |
| g | 4 | 0 | 0 | 1 | 11 | 0 | 197 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 8 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |
| h | 2 | 17 | 0 | 2 | 0 | 2 | 1 | 132 | 0 | 0 | 16 | 0 | 4 | 17 | 0 | 1 | 0 | 0 | 0 | 1 | 19 | 0 | 0 | 1 | 0 | 0 |
| i | 10 | 1 | 0 | 9 | 0 | 21 | 2 | 2 | 27 | 16 | 21 | 65 | 1 | 1 | 0 | 1 | 1 | 39 | 0 | 9 | 2 | 0 | 0 | 1 | 0 | 0 |
| j | 1 | 0 | 0 | 2 | 0 | 8 | 0 | 0 | 1 | 173 | 7 | 34 | 1 | 0 | 0 | 0 | 1 | 12 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 |
| k | 0 | 0 | 0 | 1 | 1 | 3 | 2 | 5 | 0 | 0 | 205 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 4 | 0 | 0 |
| l | 4 | 0 | 0 | 6 | 0 | 25 | 0 | 4 | 1 | 12 | 13 | 123 | 0 | 0 | 0 | 0 | 0 | 32 | 0 | 14 | 3 | 0 | 0 | 0 | 1 | 0 |
| m | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 1 | 0 | 193 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 2 | 0 | 0 | 1 |
| n | 29 | 5 | 0 | 1 | 1 | 0 | 3 | 28 | 0 | 2 | 4 | 1 | 14 | 99 | 1 | 2 | 3 | 1 | 0 | 0 | 28 | 0 | 3 | 0 | 1 | 1 |
| o | 6 | 6 | 0 | 2 | 25 | 0 | 7 | 1 | 0 | 0 | 1 | 0 | 0 | 10 | 151 | 8 | 2 | 2 | 4 | 0 | 7 | 0 | 8 | 0 | 2 | 0 |
| p | 0 | 11 | 0 | 1 | 0 | 4 | 1 | 4 | 0 | 0 | 2 | 0 | 1 | 5 | 1 | 179 | 0 | 2 | 1 | 0 | 4 | 0 | 1 | 0 | 0 | 0 |
| q | 24 | 0 | 0 | 18 | 1 | 0 | 3 | 3 | 0 | 2 | 1 | 2 | 0 | 0 | 3 | 1 | 149 | 1 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 |
| r | 3 | 1 | 0 | 1 | 3 | 31 | 1 | 0 | 3 | 5 | 17 | 11 | 1 | 2 | 0 | 4 | 0 | 136 | 2 | 7 | 3 | 0 | 0 | 0 | 3 | 1 |
| s | 11 | 0 | 0 | 0 | 9 | 1 | 42 | 2 | 0 | 0 | 8 | 0 | 0 | 1 | 0 | 0 | 2 | 3 | 116 | 0 | 2 | 1 | 2 | 16 | 7 | 4 |
| t | 7 | 1 | 0 | 4 | 6 | 26 | 1 | 3 | 2 | 0 | 22 | 15 | 0 | 2 | 0 | 2 | 1 | 49 | 1 | 65 | 8 | 0 | 0 | 0 | 1 | 1 |
| u | 13 | 7 | 0 | 8 | 3 | 0 | 2 | 8 | 0 | 1 | 4 | 1 | 10 | 17 | 3 | 3 | 1 | 1 | 0 | 1 | 185 | 1 | 0 | 0 | 3 | 0 |
| v | 2 | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 0 | 2 | 145 | 27 | 12 | 46 | 0 |
| w | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 8 | 185 | 3 | 5 | 0 |
| x | 4 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 5 | 1 | 0 | 1 | 0 | 0 | 2 | 1 | 3 | 0 | 1 | 2 | 8 | 175 | 9 | 4 |
| y | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 17 | 6 | 4 | 180 | 1 |
| z | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 2 | 2 | 0 | 1 | 0 | 2 | 2 | 3 | 224 |

Figure A.4: RXL1



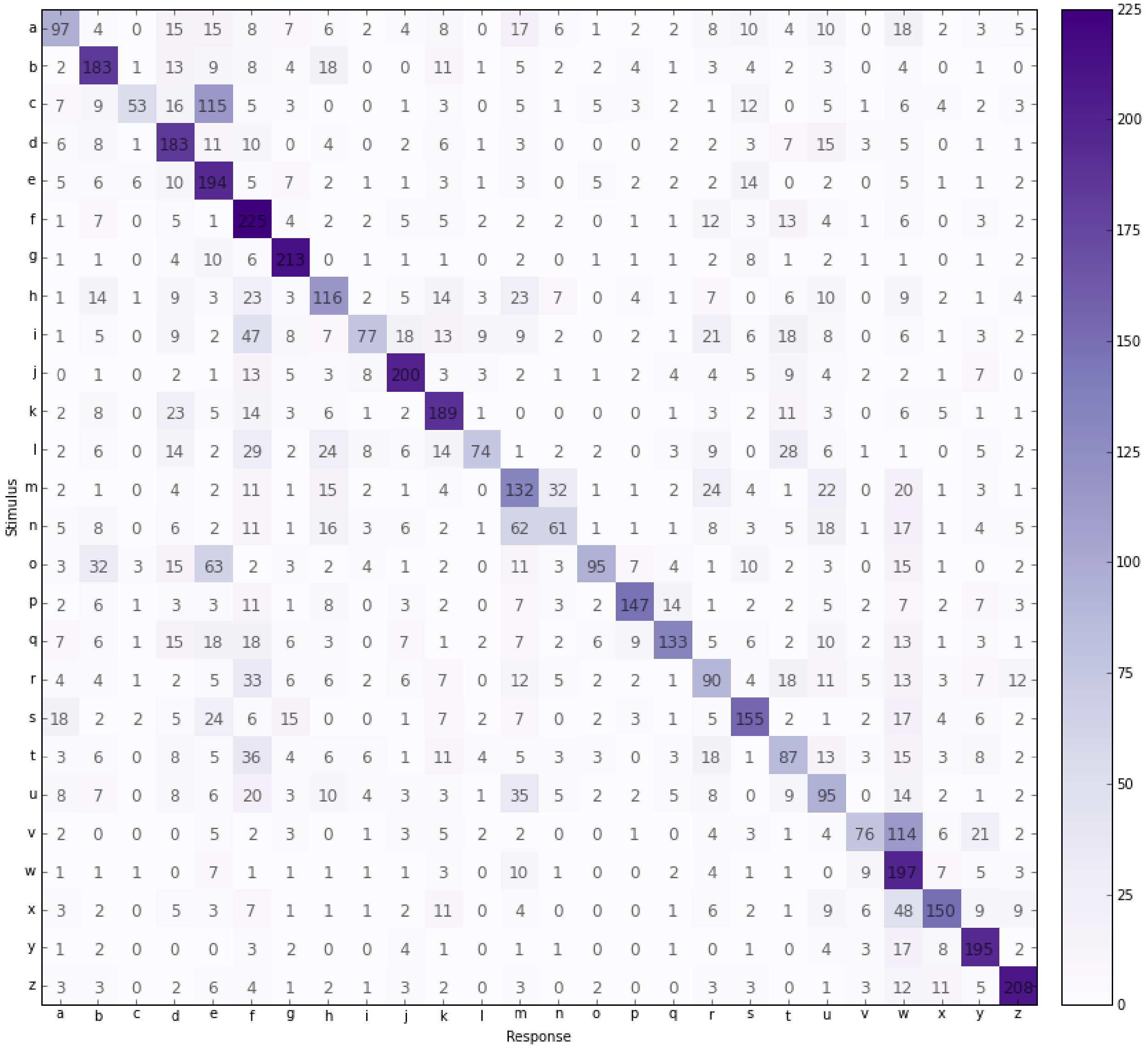
|   | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 111 | 6 | 0 | 20 | 1 | 1 | 12 | 8 | 0 | 3 | 4 | 0 | 3 | 11 | 10 | 1 | 15 | 0 | 0 | 1 | 5 | 0 | 1 | 0 | 1 | 15 |
| b | 4 | 139 | 1 | 1 | 1 | 2 | 3 | 11 | 0 | 0 | 12 | 0 | 0 | 6 | 27 | 16 | 2 | 0 | 0 | 0 | 6 | 0 | 1 | 0 | 0 | 0 |
| c | 1 | 3 | 39 | 5 | 63 | 0 | 10 | 1 | 1 | 0 | 3 | 0 | 0 | 0 | 61 | 0 | 1 | 2 | 3 | 0 | 3 | 1 | 1 | 0 | 5 | 1 |
| d | 7 | 0 | 0 | 191 | 0 | 2 | 2 | 0 | 2 | 3 | 0 | 1 | 0 | 0 | 12 | 0 | 2 | 0 | 0 | 0 | 15 | 0 | 0 | 1 | 2 | 2 |
| e | 4 | 3 | 12 | 2 | 105 | 0 | 34 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 70 | 2 | 2 | 4 | 6 | 0 | 1 | 0 | 2 | 3 | 2 | 2 |
| f | 1 | 2 | 0 | 8 | 0 | 114 | 4 | 0 | 18 | 28 | 20 | 10 | 0 | 0 | 1 | 2 | 2 | 11 | 0 | 7 | 0 | 0 | 0 | 1 | 2 | 1 |
| g | 2 | 0 | 0 | 0 | 6 | 0 | 200 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| h | 12 | 32 | 0 | 2 | 0 | 6 | 1 | 64 | 2 | 7 | 18 | 0 | 10 | 34 | 1 | 5 | 2 | 0 | 0 | 0 | 11 | 0 | 0 | 1 | 2 | 0 |
| i | 3 | 2 | 0 | 15 | 0 | 20 | 2 | 4 | 73 | 47 | 16 | 13 | 0 | 1 | 1 | 2 | 3 | 7 | 1 | 7 | 0 | 0 | 0 | 0 | 7 | 0 |
| j | 2 | 0 | 0 | 12 | 0 | 12 | 1 | 0 | 29 | 161 | 3 | 13 | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 4 | 0 |
| k | 0 | 3 | 0 | 3 | 1 | 2 | 2 | 2 | 0 | 0 | 210 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 |
| l | 0 | 1 | 0 | 22 | 0 | 11 | 1 | 3 | 43 | 39 | 19 | 56 | 0 | 1 | 0 | 1 | 2 | 4 | 0 | 7 | 0 | 0 | 0 | 1 | 1 | 0 |
| m | 4 | 4 | 0 | 2 | 1 | 1 | 0 | 4 | 0 | 2 | 0 | 0 | 174 | 19 | 6 | 3 | 1 | 1 | 0 | 0 | 6 | 0 | 4 | 0 | 1 | 1 |
| n | 19 | 5 | 0 | 8 | 1 | 2 | 3 | 15 | 3 | 1 | 2 | 0 | 48 | 76 | 7 | 6 | 9 | 1 | 1 | 0 | 21 | 0 | 2 | 0 | 0 | 0 |
| o | 2 | 6 | 1 | 1 | 5 | 0 | 5 | 1 | 0 | 2 | 0 | 1 | 0 | 5 | 165 | 3 | 5 | 1 | 0 | 0 | 5 | 0 | 1 | 0 | 2 | 1 |
| p | 2 | 10 | 0 | 0 | 0 | 2 | 3 | 2 | 1 | 7 | 4 | 1 | 1 | 4 | 13 | 185 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| q | 8 | 0 | 0 | 10 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 1 | 7 | 0 | 210 | 0 | 1 | 1 | 4 | 0 | 0 | 0 | 1 | 0 |
| r | 2 | 3 | 2 | 6 | 5 | 39 | 5 | 4 | 14 | 26 | 25 | 7 | 0 | 4 | 3 | 7 | 2 | 60 | 0 | 4 | 3 | 0 | 0 | 3 | 15 | 2 |
| s | 8 | 1 | 3 | 3 | 13 | 0 | 68 | 1 | 0 | 2 | 7 | 0 | 2 | 2 | 10 | 2 | 5 | 0 | 47 | 0 | 1 | 0 | 5 | 29 | 19 | 3 |
| t | 3 | 9 | 0 | 13 | 1 | 40 | 4 | 3 | 20 | 6 | 41 | 13 | 0 | 2 | 3 | 2 | 3 | 22 | 0 | 18 | 3 | 1 | 0 | 0 | 2 | 1 |
| u | 18 | 7 | 0 | 16 | 1 | 1 | 2 | 10 | 0 | 4 | 2 | 1 | 15 | 21 | 10 | 1 | 7 | 0 | 0 | 1 | 112 | 0 | 1 | 1 | 2 | 0 |
| v | 2 | 0 | 0 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 3 | 66 | 28 | 12 | 122 | 0 |
| w | 2 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 3 | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | 183 | 4 | 18 | 0 |
| x | 3 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 11 | 0 | 0 | 2 | 4 | 1 | 1 | 1 | 0 | 0 | 0 | 2 | 2 | 178 | 18 | 5 |
| y | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 194 | 1 |
| z | 1 | 0 | 1 | 0 | 3 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 222 |

Figure A.5: YTY1



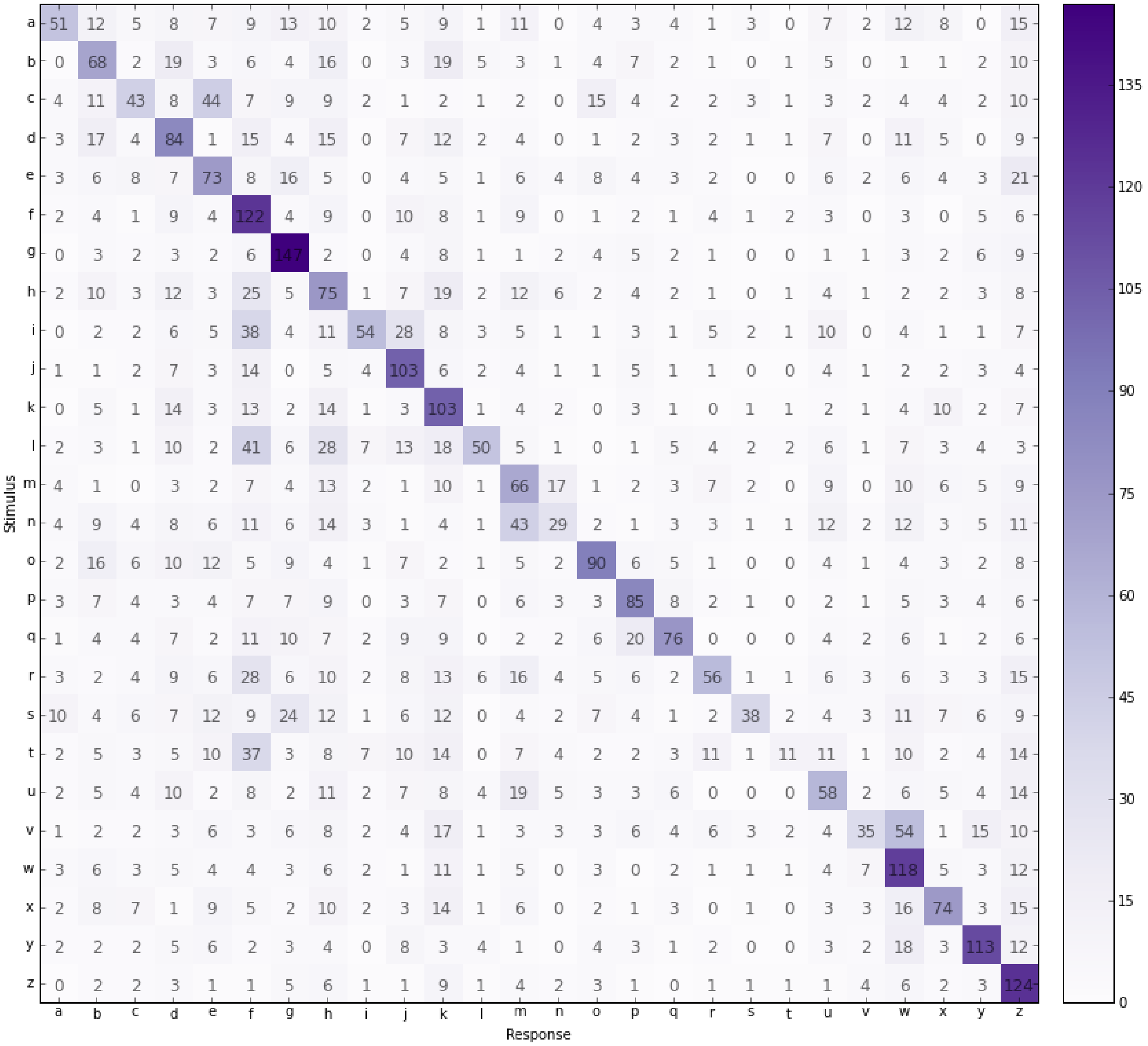
| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 74 | 4 | 1 | 12 | 33 | 4 | 12 | 25 | 2 | 4 | 9 | 0 | 14 | 4 | 2 | 3 | 20 | 7 | 10 | 0 | 8 | 0 | 13 | 4 | 2 | 15 |
| b | 1 | 139 | 2 | 2 | 5 | 6 | 4 | 51 | 1 | 7 | 16 | 0 | 1 | 1 | 3 | 6 | 6 | 2 | 2 | 1 | 0 | 1 | 1 | 1 | 6 | 3 |
| c | 5 | 5 | 70 | 10 | 74 | 2 | 11 | 7 | 0 | 6 | 3 | 1 | 5 | 0 | 10 | 0 | 21 | 2 | 3 | 0 | 2 | 1 | 5 | 2 | 1 | 7 |
| d | 2 | 3 | 0 | 130 | 6 | 4 | 5 | 31 | 0 | 9 | 21 | 1 | 2 | 1 | 1 | 1 | 14 | 4 | 1 | 1 | 7 | 1 | 4 | 2 | 1 | 3 |
| e | 4 | 6 | 5 | 4 | 177 | 3 | 11 | 3 | 0 | 1 | 2 | 0 | 5 | 1 | 7 | 3 | 9 | 3 | 8 | 0 | 4 | 0 | 2 | 2 | 2 | 11 |
| f | 2 | 2 | 1 | 2 | 2 | 147 | 2 | 18 | 5 | 9 | 19 | 1 | 5 | 0 | 0 | 1 | 2 | 25 | 3 | 0 | 2 | 2 | 1 | 2 | 3 | 6 |
| g | 0 | 1 | 0 | 0 | 10 | 2 | 229 | 2 | 0 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 5 | 0 | 7 | 0 | 2 | 0 | 1 | 1 | 0 | 4 |
| h | 0 | 4 | 0 | 2 | 0 | 10 | 1 | 189 | 1 | 12 | 20 | 0 | 16 | 1 | 0 | 1 | 0 | 7 | 2 | 0 | 9 | 1 | 6 | 0 | 2 | 2 |
| i | 2 | 1 | 0 | 1 | 1 | 11 | 1 | 22 | 81 | 47 | 11 | 3 | 3 | 1 | 1 | 1 | 2 | 16 | 1 | 5 | 10 | 0 | 3 | 2 | 3 | 8 |
| j | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 8 | 5 | 253 | 6 | 1 | 0 | 1 | 0 | 1 | 8 | 6 | 0 | 0 | 6 | 2 | 0 | 1 | 9 | 2 |
| k | 0 | 2 | 0 | 7 | 2 | 2 | 0 | 36 | 0 | 8 | 197 | 3 | 0 | 0 | 0 | 1 | 7 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 2 | 1 |
| l | 1 | 2 | 0 | 14 | 0 | 23 | 1 | 71 | 9 | 25 | 32 | 65 | 1 | 0 | 1 | 1 | 6 | 13 | 0 | 2 | 6 | 0 | 5 | 2 | 2 | 5 |
| m | 2 | 1 | 0 | 7 | 2 | 3 | 3 | 45 | 0 | 8 | 6 | 1 | 136 | 10 | 1 | 1 | 6 | 12 | 5 | 0 | 7 | 1 | 11 | 4 | 2 | 5 |
| n | 1 | 2 | 1 | 4 | 5 | 3 | 7 | 54 | 2 | 4 | 6 | 0 | 87 | 50 | 1 | 5 | 4 | 7 | 3 | 0 | 10 | 2 | 5 | 3 | 6 | 6 |
| o | 3 | 17 | 3 | 4 | 29 | 1 | 12 | 4 | 1 | 5 | 5 | 0 | 3 | 3 | 127 | 9 | 21 | 1 | 4 | 0 | 2 | 2 | 7 | 2 | 2 | 5 |
| p | 1 | 1 | 0 | 0 | 1 | 3 | 8 | 18 | 0 | 11 | 5 | 0 | 4 | 1 | 2 | 138 | 33 | 10 | 2 | 0 | 2 | 0 | 2 | 2 | 2 | 6 |
| q | 0 | 0 | 0 | 1 | 2 | 3 | 7 | 13 | 0 | 6 | 1 | 0 | 1 | 0 | 0 | 4 | 186 | 3 | 1 | 0 | 2 | 1 | 0 | 1 | 4 | 1 |
| r | 1 | 1 | 0 | 3 | 3 | 15 | 1 | 31 | 4 | 15 | 20 | 0 | 19 | 2 | 0 | 2 | 7 | 111 | 5 | 0 | 3 | 3 | 6 | 4 | 6 | 5 |
| s | 4 | 3 | 3 | 2 | 30 | 0 | 27 | 13 | 0 | 5 | 10 | 0 | 6 | 0 | 2 | 1 | 7 | 1 | 142 | 0 | 3 | 0 | 6 | 15 | 9 | 10 |
| t | 2 | 1 | 1 | 5 | 11 | 26 | 5 | 27 | 3 | 14 | 21 | 6 | 5 | 1 | 0 | 2 | 7 | 26 | 1 | 59 | 11 | 0 | 6 | 6 | 9 | 8 |
| u | 2 | 2 | 1 | 6 | 5 | 7 | 1 | 36 | 2 | 7 | 7 | 0 | 30 | 3 | 0 | 2 | 11 | 8 | 1 | 1 | 96 | 0 | 17 | 3 | 6 | 6 |
| v | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 5 | 0 | 1 | 5 | 0 | 5 | 0 | 2 | 1 | 3 | 4 | 0 | 0 | 1 | 98 | 75 | 14 | 22 | 5 |
| w | 1 | 2 | 0 | 1 | 2 | 0 | 0 | 5 | 0 | 1 | 3 | 0 | 2 | 0 | 2 | 0 | 5 | 2 | 1 | 0 | 5 | 10 | 195 | 13 | 2 | 4 |
| x | 2 | 0 | 0 | 3 | 10 | 1 | 1 | 2 | 2 | 3 | 17 | 0 | 4 | 1 | 0 | 1 | 2 | 1 | 4 | 0 | 2 | 3 | 23 | 142 | 6 | 9 |
| y | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 1 | 3 | 1 | 0 | 1 | 0 | 1 | 1 | 7 | 1 | 1 | 0 | 3 | 6 | 16 | 0 | 230 | 5 |
| z | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 9 | 1 | 2 | 9 | 0 | 2 | 0 | 0 | 1 | 5 | 2 | 2 | 0 | 0 | 4 | 5 | 7 | 5 | 255 |

Figure A.6: AXL3



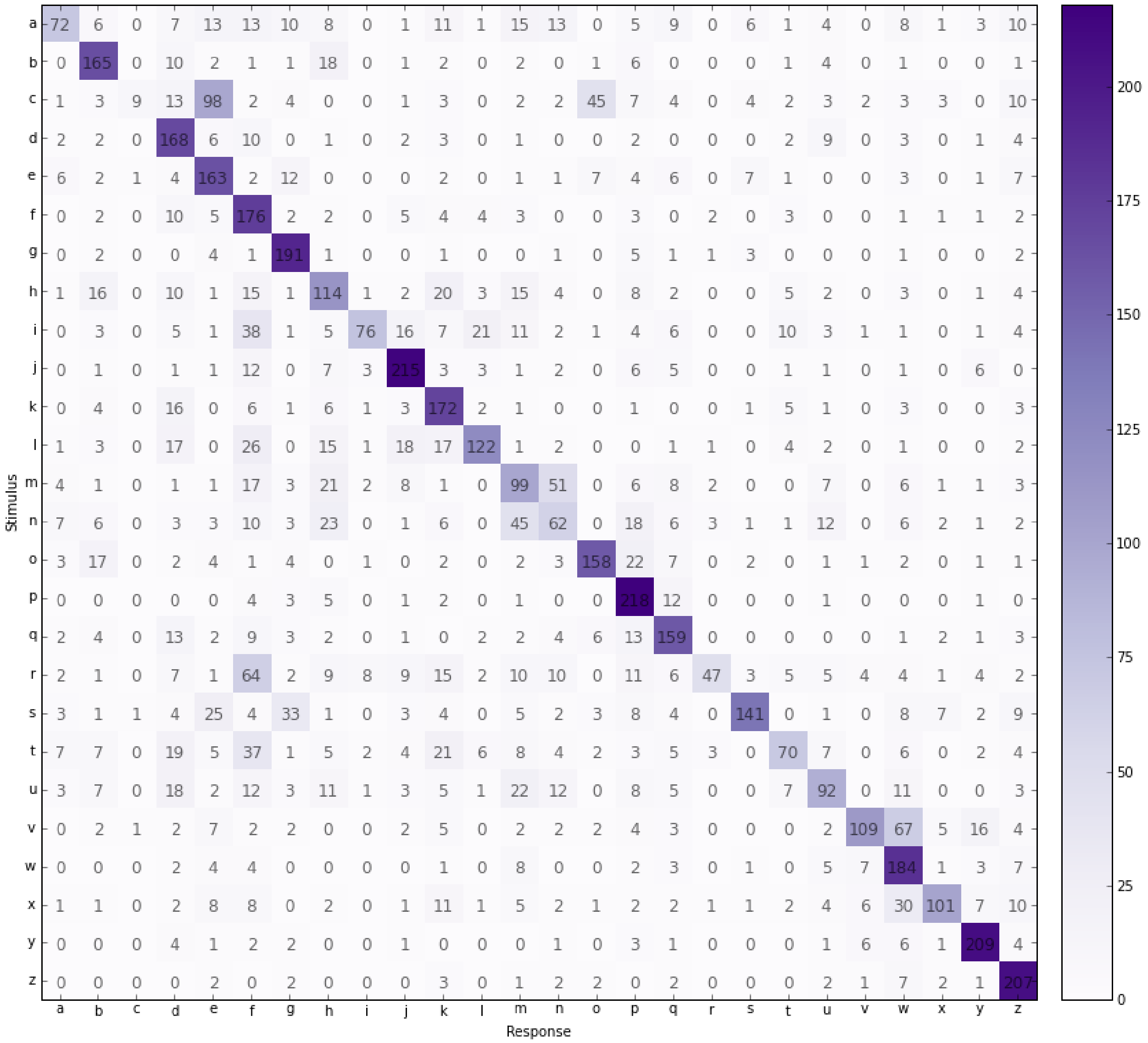
| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 97 | 4 | 0 | 15 | 15 | 8 | 7 | 6 | 2 | 4 | 8 | 0 | 17 | 6 | 1 | 2 | 2 | 8 | 10 | 4 | 10 | 0 | 18 | 2 | 3 | 5 |
| b | 2 | 183 | 1 | 13 | 9 | 8 | 4 | 18 | 0 | 0 | 11 | 1 | 5 | 2 | 2 | 4 | 1 | 3 | 4 | 2 | 3 | 0 | 4 | 0 | 1 | 0 |
| c | 7 | 9 | 53 | 16 | 115 | 5 | 3 | 0 | 0 | 1 | 3 | 0 | 5 | 1 | 5 | 3 | 2 | 1 | 12 | 0 | 5 | 1 | 6 | 4 | 2 | 3 |
| d | 6 | 8 | 1 | 183 | 11 | 10 | 0 | 4 | 0 | 2 | 6 | 1 | 3 | 0 | 0 | 0 | 2 | 2 | 3 | 7 | 15 | 3 | 5 | 0 | 1 | 1 |
| e | 5 | 6 | 6 | 10 | 194 | 5 | 7 | 2 | 1 | 1 | 3 | 1 | 3 | 0 | 5 | 2 | 2 | 2 | 14 | 0 | 2 | 0 | 5 | 1 | 1 | 2 |
| f | 1 | 7 | 0 | 5 | 1 | 225 | 4 | 2 | 2 | 5 | 5 | 2 | 2 | 2 | 0 | 1 | 1 | 12 | 3 | 13 | 4 | 1 | 6 | 0 | 3 | 2 |
| g | 1 | 1 | 0 | 4 | 10 | 6 | 213 | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 1 | 1 | 1 | 2 | 8 | 1 | 2 | 1 | 1 | 0 | 1 | 2 |
| h | 1 | 14 | 1 | 9 | 3 | 23 | 3 | 116 | 2 | 5 | 14 | 3 | 23 | 7 | 0 | 4 | 1 | 7 | 0 | 6 | 10 | 0 | 9 | 2 | 1 | 4 |
| i | 1 | 5 | 0 | 9 | 2 | 47 | 8 | 7 | 77 | 18 | 13 | 9 | 9 | 2 | 0 | 2 | 1 | 21 | 6 | 18 | 8 | 0 | 6 | 1 | 3 | 2 |
| j | 0 | 1 | 0 | 2 | 1 | 13 | 5 | 3 | 8 | 200 | 3 | 3 | 2 | 1 | 1 | 2 | 4 | 4 | 5 | 9 | 4 | 2 | 2 | 1 | 7 | 0 |
| k | 2 | 8 | 0 | 23 | 5 | 14 | 3 | 6 | 1 | 2 | 189 | 1 | 0 | 0 | 0 | 0 | 1 | 3 | 2 | 11 | 3 | 0 | 6 | 5 | 1 | 1 |
| l | 2 | 6 | 0 | 14 | 2 | 29 | 2 | 24 | 8 | 6 | 14 | 74 | 1 | 2 | 2 | 0 | 3 | 9 | 0 | 28 | 6 | 1 | 1 | 0 | 5 | 2 |
| m | 2 | 1 | 0 | 4 | 2 | 11 | 1 | 15 | 2 | 1 | 4 | 0 | 132 | 32 | 1 | 1 | 2 | 24 | 4 | 1 | 22 | 0 | 20 | 1 | 3 | 1 |
| n | 5 | 8 | 0 | 6 | 2 | 11 | 1 | 16 | 3 | 6 | 2 | 1 | 62 | 61 | 1 | 1 | 1 | 8 | 3 | 5 | 18 | 1 | 17 | 1 | 4 | 5 |
| o | 3 | 32 | 3 | 15 | 63 | 2 | 3 | 2 | 4 | 1 | 2 | 0 | 11 | 3 | 95 | 7 | 4 | 1 | 10 | 2 | 3 | 0 | 15 | 1 | 0 | 2 |
| p | 2 | 6 | 1 | 3 | 3 | 11 | 1 | 8 | 0 | 3 | 2 | 0 | 7 | 3 | 2 | 147 | 14 | 1 | 2 | 2 | 5 | 2 | 7 | 2 | 7 | 3 |
| q | 7 | 6 | 1 | 15 | 18 | 18 | 6 | 3 | 0 | 7 | 1 | 2 | 7 | 2 | 6 | 9 | 133 | 5 | 6 | 2 | 10 | 2 | 13 | 1 | 3 | 1 |
| r | 4 | 4 | 1 | 2 | 5 | 33 | 6 | 6 | 2 | 6 | 7 | 0 | 12 | 5 | 2 | 2 | 1 | 90 | 4 | 18 | 11 | 5 | 13 | 3 | 7 | 12 |
| s | 18 | 2 | 2 | 5 | 24 | 6 | 15 | 0 | 0 | 1 | 7 | 2 | 7 | 0 | 2 | 3 | 1 | 5 | 155 | 2 | 1 | 2 | 17 | 4 | 6 | 2 |
| t | 3 | 6 | 0 | 8 | 5 | 36 | 4 | 6 | 6 | 1 | 11 | 4 | 5 | 3 | 3 | 0 | 3 | 18 | 1 | 87 | 13 | 3 | 15 | 3 | 8 | 2 |
| u | 8 | 7 | 0 | 8 | 6 | 20 | 3 | 10 | 4 | 3 | 3 | 1 | 35 | 5 | 2 | 2 | 5 | 8 | 0 | 9 | 95 | 0 | 14 | 2 | 1 | 2 |
| v | 2 | 0 | 0 | 0 | 5 | 2 | 3 | 0 | 1 | 3 | 5 | 2 | 2 | 0 | 0 | 1 | 0 | 4 | 3 | 1 | 4 | 76 | 114 | 6 | 21 | 2 |
| w | 1 | 1 | 1 | 0 | 7 | 1 | 1 | 1 | 1 | 1 | 3 | 0 | 10 | 1 | 0 | 0 | 2 | 4 | 1 | 1 | 0 | 9 | 197 | 7 | 5 | 3 |
| x | 3 | 2 | 0 | 5 | 3 | 7 | 1 | 1 | 1 | 2 | 11 | 0 | 4 | 0 | 0 | 0 | 1 | 6 | 2 | 1 | 9 | 6 | 48 | 150 | 9 | 9 |
| y | 1 | 2 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 4 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 4 | 3 | 17 | 8 | 195 | 2 |
| z | 3 | 3 | 0 | 2 | 6 | 4 | 1 | 2 | 1 | 3 | 2 | 0 | 3 | 0 | 2 | 0 | 0 | 3 | 3 | 0 | 1 | 3 | 12 | 11 | 5 | 208 |

Figure A.7: KMA3



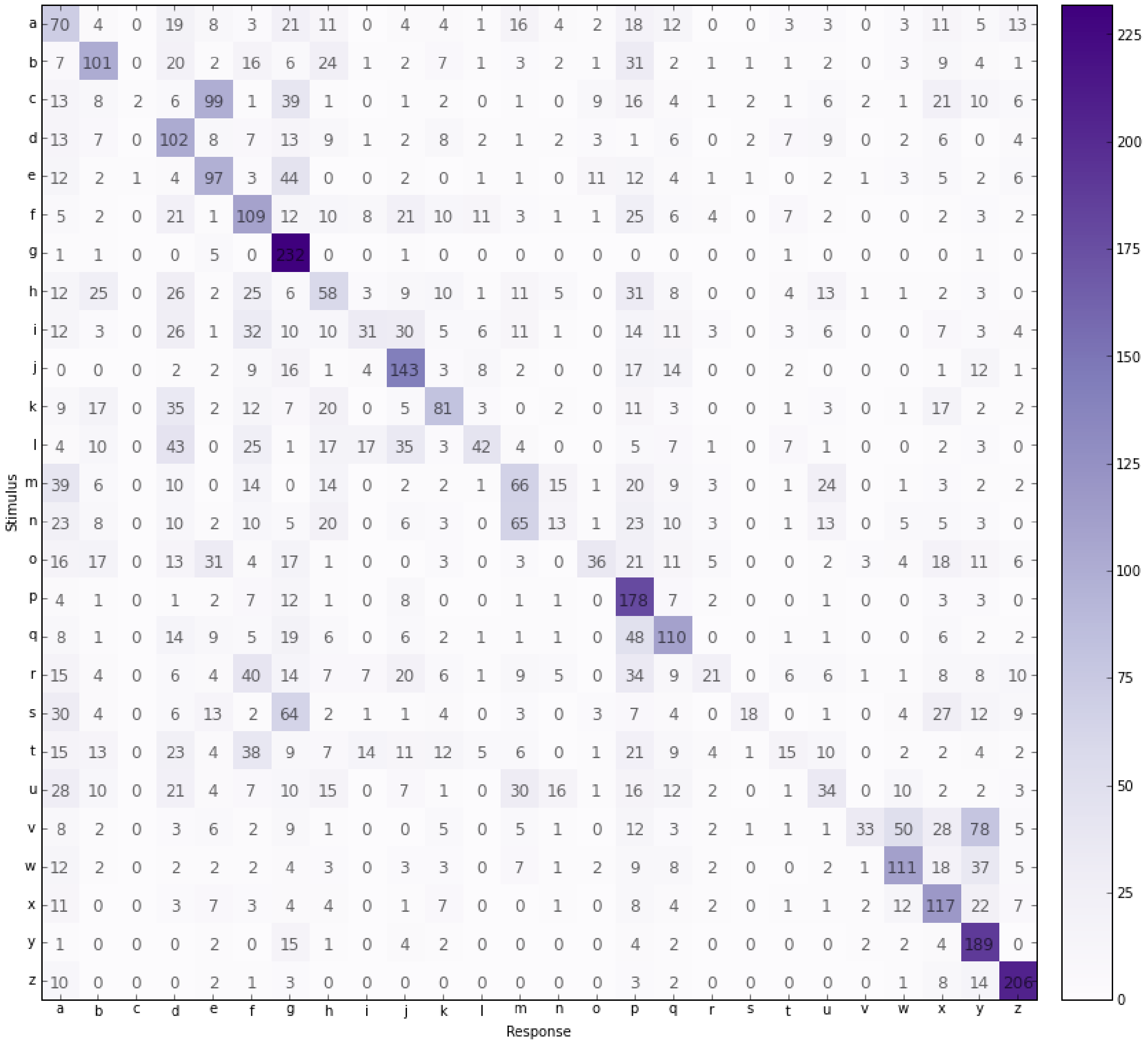
|   | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 51 | 12 | 5 | 8 | 7 | 9 | 13 | 10 | 2 | 5 | 9 | 1 | 11 | 0 | 4 | 3 | 4 | 1 | 3 | 0 | 7 | 2 | 12 | 8 | 0 | 15 |
| b | 0 | 68 | 2 | 19 | 3 | 6 | 4 | 16 | 0 | 3 | 19 | 5 | 3 | 1 | 4 | 7 | 2 | 1 | 0 | 1 | 5 | 0 | 1 | 1 | 2 | 10 |
| c | 4 | 11 | 43 | 8 | 44 | 7 | 9 | 9 | 2 | 1 | 2 | 1 | 2 | 0 | 15 | 4 | 2 | 2 | 3 | 1 | 3 | 2 | 4 | 4 | 2 | 10 |
| d | 3 | 17 | 4 | 84 | 1 | 15 | 4 | 15 | 0 | 7 | 12 | 2 | 4 | 0 | 1 | 2 | 3 | 2 | 1 | 1 | 7 | 0 | 11 | 5 | 0 | 9 |
| e | 3 | 6 | 8 | 7 | 73 | 8 | 16 | 5 | 0 | 4 | 5 | 1 | 6 | 4 | 8 | 4 | 3 | 2 | 0 | 0 | 6 | 2 | 6 | 4 | 3 | 21 |
| f | 2 | 4 | 1 | 9 | 4 | 122 | 4 | 9 | 0 | 10 | 8 | 1 | 9 | 0 | 1 | 2 | 1 | 4 | 1 | 2 | 3 | 0 | 3 | 0 | 5 | 6 |
| g | 0 | 3 | 2 | 3 | 2 | 6 | 147 | 2 | 0 | 4 | 8 | 1 | 1 | 2 | 4 | 5 | 2 | 1 | 0 | 0 | 1 | 1 | 3 | 2 | 6 | 9 |
| h | 2 | 10 | 3 | 12 | 3 | 25 | 5 | 75 | 1 | 7 | 19 | 2 | 12 | 6 | 2 | 4 | 2 | 1 | 0 | 1 | 4 | 1 | 2 | 2 | 3 | 8 |
| i | 0 | 2 | 2 | 6 | 5 | 38 | 4 | 11 | 54 | 28 | 8 | 3 | 5 | 1 | 1 | 3 | 1 | 5 | 2 | 1 | 10 | 0 | 4 | 1 | 1 | 7 |
| j | 1 | 1 | 2 | 7 | 3 | 14 | 0 | 5 | 4 | 103 | 6 | 2 | 4 | 1 | 1 | 5 | 1 | 1 | 0 | 0 | 4 | 1 | 2 | 2 | 3 | 4 |
| k | 0 | 5 | 1 | 14 | 3 | 13 | 2 | 14 | 1 | 3 | 103 | 1 | 4 | 2 | 0 | 3 | 1 | 0 | 1 | 1 | 2 | 1 | 4 | 10 | 2 | 7 |
| l | 2 | 3 | 1 | 10 | 2 | 41 | 6 | 28 | 7 | 13 | 18 | 50 | 5 | 1 | 0 | 1 | 5 | 4 | 2 | 2 | 6 | 1 | 7 | 3 | 4 | 3 |
| m | 4 | 1 | 0 | 3 | 2 | 7 | 4 | 13 | 2 | 1 | 10 | 1 | 66 | 17 | 1 | 2 | 3 | 7 | 2 | 0 | 9 | 0 | 10 | 6 | 5 | 9 |
| n | 4 | 9 | 4 | 8 | 6 | 11 | 6 | 14 | 3 | 1 | 4 | 1 | 43 | 29 | 2 | 1 | 3 | 3 | 1 | 1 | 12 | 2 | 12 | 3 | 5 | 11 |
| o | 2 | 16 | 6 | 10 | 12 | 5 | 9 | 4 | 1 | 7 | 2 | 1 | 5 | 2 | 90 | 6 | 5 | 1 | 0 | 0 | 4 | 1 | 4 | 3 | 2 | 8 |
| p | 3 | 7 | 4 | 3 | 4 | 7 | 7 | 9 | 0 | 3 | 7 | 0 | 6 | 3 | 3 | 85 | 8 | 2 | 1 | 0 | 2 | 1 | 5 | 3 | 4 | 6 |
| q | 1 | 4 | 4 | 7 | 2 | 11 | 10 | 7 | 2 | 9 | 9 | 0 | 2 | 2 | 6 | 20 | 76 | 0 | 0 | 0 | 4 | 2 | 6 | 1 | 2 | 6 |
| r | 3 | 2 | 4 | 9 | 6 | 28 | 6 | 10 | 2 | 8 | 13 | 6 | 16 | 4 | 5 | 6 | 2 | 56 | 1 | 1 | 6 | 3 | 6 | 3 | 3 | 15 |
| s | 10 | 4 | 6 | 7 | 12 | 9 | 24 | 12 | 1 | 6 | 12 | 0 | 4 | 2 | 7 | 4 | 1 | 2 | 38 | 2 | 4 | 3 | 11 | 7 | 6 | 9 |
| t | 2 | 5 | 3 | 5 | 10 | 37 | 3 | 8 | 7 | 10 | 14 | 0 | 7 | 4 | 2 | 2 | 3 | 11 | 1 | 11 | 11 | 1 | 10 | 2 | 4 | 14 |
| u | 2 | 5 | 4 | 10 | 2 | 8 | 2 | 11 | 2 | 7 | 8 | 4 | 19 | 5 | 3 | 3 | 6 | 0 | 0 | 0 | 58 | 2 | 6 | 5 | 4 | 14 |
| v | 1 | 2 | 2 | 3 | 6 | 3 | 6 | 8 | 2 | 4 | 17 | 1 | 3 | 3 | 3 | 6 | 4 | 6 | 3 | 2 | 4 | 35 | 54 | 1 | 15 | 10 |
| w | 3 | 6 | 3 | 5 | 4 | 4 | 3 | 6 | 2 | 1 | 11 | 1 | 5 | 0 | 3 | 0 | 2 | 1 | 1 | 1 | 4 | 7 | 118 | 5 | 3 | 12 |
| x | 2 | 8 | 7 | 1 | 9 | 5 | 2 | 10 | 2 | 3 | 14 | 1 | 6 | 0 | 2 | 1 | 3 | 0 | 1 | 0 | 3 | 3 | 16 | 74 | 3 | 15 |
| y | 2 | 2 | 2 | 5 | 6 | 2 | 3 | 4 | 0 | 8 | 3 | 4 | 1 | 0 | 4 | 3 | 1 | 2 | 0 | 0 | 3 | 2 | 18 | 3 | 113 | 12 |
| z | 0 | 2 | 2 | 3 | 1 | 1 | 5 | 6 | 1 | 1 | 9 | 1 | 4 | 2 | 3 | 1 | 0 | 1 | 1 | 1 | 1 | 4 | 6 | 2 | 3 | 124 |

Figure A.8: RXK3



| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 72 | 6 | 0 | 7 | 13 | 13 | 10 | 8 | 0 | 1 | 11 | 1 | 15 | 13 | 0 | 5 | 9 | 0 | 6 | 1 | 4 | 0 | 8 | 1 | 3 | 10 |
| b | 0 | 165 | 0 | 10 | 2 | 1 | 1 | 18 | 0 | 1 | 2 | 0 | 2 | 0 | 1 | 6 | 0 | 0 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 1 |
| c | 1 | 3 | 9 | 13 | 98 | 2 | 4 | 0 | 0 | 1 | 3 | 0 | 2 | 2 | 45 | 7 | 4 | 0 | 4 | 2 | 3 | 2 | 3 | 3 | 0 | 10 |
| d | 2 | 2 | 0 | 168 | 6 | 10 | 0 | 1 | 0 | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 9 | 0 | 3 | 0 | 1 | 4 |
| e | 6 | 2 | 1 | 4 | 163 | 2 | 12 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 7 | 4 | 6 | 0 | 7 | 1 | 0 | 0 | 3 | 0 | 1 | 7 |
| f | 0 | 2 | 0 | 10 | 5 | 176 | 2 | 2 | 0 | 5 | 4 | 4 | 3 | 0 | 0 | 3 | 0 | 2 | 0 | 3 | 0 | 0 | 1 | 1 | 1 | 2 |
| g | 0 | 2 | 0 | 0 | 4 | 1 | 191 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 5 | 1 | 1 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 2 |
| h | 1 | 16 | 0 | 10 | 1 | 15 | 1 | 114 | 1 | 2 | 20 | 3 | 15 | 4 | 0 | 8 | 2 | 0 | 0 | 5 | 2 | 0 | 3 | 0 | 1 | 4 |
| i | 0 | 3 | 0 | 5 | 1 | 38 | 1 | 5 | 76 | 16 | 7 | 21 | 11 | 2 | 1 | 4 | 6 | 0 | 0 | 10 | 3 | 1 | 1 | 0 | 1 | 4 |
| j | 0 | 1 | 0 | 1 | 1 | 12 | 0 | 7 | 3 | 215 | 3 | 3 | 1 | 2 | 0 | 6 | 5 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 6 | 0 |
| k | 0 | 4 | 0 | 16 | 0 | 6 | 1 | 6 | 1 | 3 | 172 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 5 | 1 | 0 | 3 | 0 | 0 | 3 |
| l | 1 | 3 | 0 | 17 | 0 | 26 | 0 | 15 | 1 | 18 | 17 | 122 | 1 | 2 | 0 | 0 | 1 | 1 | 0 | 4 | 2 | 0 | 1 | 0 | 0 | 2 |
| m | 4 | 1 | 0 | 1 | 1 | 17 | 3 | 21 | 2 | 8 | 1 | 0 | 99 | 51 | 0 | 6 | 8 | 2 | 0 | 0 | 7 | 0 | 6 | 1 | 1 | 3 |
| n | 7 | 6 | 0 | 3 | 3 | 10 | 3 | 23 | 0 | 1 | 6 | 0 | 45 | 62 | 0 | 18 | 6 | 3 | 1 | 1 | 12 | 0 | 6 | 2 | 1 | 2 |
| o | 3 | 17 | 0 | 2 | 4 | 1 | 4 | 0 | 1 | 0 | 2 | 0 | 2 | 3 | 158 | 22 | 7 | 0 | 2 | 0 | 1 | 1 | 2 | 0 | 1 | 1 |
| p | 0 | 0 | 0 | 0 | 0 | 4 | 3 | 5 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 218 | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| q | 2 | 4 | 0 | 13 | 2 | 9 | 3 | 2 | 0 | 1 | 0 | 2 | 2 | 4 | 6 | 13 | 159 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 3 |
| r | 2 | 1 | 0 | 7 | 1 | 64 | 2 | 9 | 8 | 9 | 15 | 2 | 10 | 10 | 0 | 11 | 6 | 47 | 3 | 5 | 5 | 4 | 4 | 1 | 4 | 2 |
| s | 3 | 1 | 1 | 4 | 25 | 4 | 33 | 1 | 0 | 3 | 4 | 0 | 5 | 2 | 3 | 8 | 4 | 0 | 141 | 0 | 1 | 0 | 8 | 7 | 2 | 9 |
| t | 7 | 7 | 0 | 19 | 5 | 37 | 1 | 5 | 2 | 4 | 21 | 6 | 8 | 4 | 2 | 3 | 5 | 3 | 0 | 70 | 7 | 0 | 6 | 0 | 2 | 4 |
| u | 3 | 7 | 0 | 18 | 2 | 12 | 3 | 11 | 1 | 3 | 5 | 1 | 22 | 12 | 0 | 8 | 5 | 0 | 0 | 7 | 92 | 0 | 11 | 0 | 0 | 3 |
| v | 0 | 2 | 1 | 2 | 7 | 2 | 2 | 0 | 0 | 2 | 5 | 0 | 2 | 2 | 2 | 4 | 3 | 0 | 0 | 0 | 2 | 109 | 67 | 5 | 16 | 4 |
| w | 0 | 0 | 0 | 2 | 4 | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 8 | 0 | 0 | 2 | 3 | 0 | 1 | 0 | 5 | 7 | 184 | 1 | 3 | 7 |
| x | 1 | 1 | 0 | 2 | 8 | 8 | 0 | 2 | 0 | 1 | 11 | 1 | 5 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 4 | 6 | 30 | 101 | 7 | 10 |
| y | 0 | 0 | 0 | 4 | 1 | 2 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 0 | 0 | 0 | 1 | 6 | 6 | 1 | 209 | 4 |
| z | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 3 | 0 | 1 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 7 | 2 | 1 | 207 |

Figure A.9: RXL3



| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 70 | 4 | 0 | 19 | 8 | 3 | 21 | 11 | 0 | 4 | 4 | 1 | 16 | 4 | 2 | 18 | 12 | 0 | 0 | 3 | 3 | 0 | 3 | 11 | 5 | 13 |
| b | 7 | 101 | 0 | 20 | 2 | 16 | 6 | 24 | 1 | 2 | 7 | 1 | 3 | 2 | 1 | 31 | 2 | 1 | 1 | 1 | 2 | 0 | 3 | 9 | 4 | 1 |
| c | 13 | 8 | 2 | 6 | 99 | 1 | 39 | 1 | 0 | 1 | 2 | 0 | 1 | 0 | 9 | 16 | 4 | 1 | 2 | 1 | 6 | 2 | 1 | 21 | 10 | 6 |
| d | 13 | 7 | 0 | 102 | 8 | 7 | 13 | 9 | 1 | 2 | 8 | 2 | 1 | 2 | 3 | 1 | 6 | 0 | 2 | 7 | 9 | 0 | 2 | 6 | 0 | 4 |
| e | 12 | 2 | 1 | 4 | 97 | 3 | 44 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 11 | 12 | 4 | 1 | 1 | 0 | 2 | 1 | 3 | 5 | 2 | 6 |
| f | 5 | 2 | 0 | 21 | 1 | 109 | 12 | 10 | 8 | 21 | 10 | 11 | 3 | 1 | 1 | 25 | 6 | 4 | 0 | 7 | 2 | 0 | 0 | 2 | 3 | 2 |
| g | 1 | 1 | 0 | 0 | 5 | 0 | 232 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| h | 12 | 25 | 0 | 26 | 2 | 25 | 6 | 58 | 3 | 9 | 10 | 1 | 11 | 5 | 0 | 31 | 8 | 0 | 0 | 4 | 13 | 1 | 1 | 2 | 3 | 0 |
| i | 12 | 3 | 0 | 26 | 1 | 32 | 10 | 10 | 31 | 30 | 5 | 6 | 11 | 1 | 0 | 14 | 11 | 3 | 0 | 3 | 6 | 0 | 0 | 7 | 3 | 4 |
| j | 0 | 0 | 0 | 2 | 2 | 9 | 16 | 1 | 4 | 143 | 3 | 8 | 2 | 0 | 0 | 17 | 14 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 12 | 1 |
| k | 9 | 17 | 0 | 35 | 2 | 12 | 7 | 20 | 0 | 5 | 81 | 3 | 0 | 2 | 0 | 11 | 3 | 0 | 0 | 1 | 3 | 0 | 1 | 17 | 2 | 2 |
| l | 4 | 10 | 0 | 43 | 0 | 25 | 1 | 17 | 17 | 35 | 3 | 42 | 4 | 0 | 0 | 5 | 7 | 1 | 0 | 7 | 1 | 0 | 0 | 2 | 3 | 0 |
| m | 39 | 6 | 0 | 10 | 0 | 14 | 0 | 14 | 0 | 2 | 2 | 1 | 66 | 15 | 1 | 20 | 9 | 3 | 0 | 1 | 24 | 0 | 1 | 3 | 2 | 2 |
| n | 23 | 8 | 0 | 10 | 2 | 10 | 5 | 20 | 0 | 6 | 3 | 0 | 65 | 13 | 1 | 23 | 10 | 3 | 0 | 1 | 13 | 0 | 5 | 5 | 3 | 0 |
| o | 16 | 17 | 0 | 13 | 31 | 4 | 17 | 1 | 0 | 0 | 3 | 0 | 3 | 0 | 36 | 21 | 11 | 5 | 0 | 0 | 2 | 3 | 4 | 18 | 11 | 6 |
| p | 4 | 1 | 0 | 1 | 2 | 7 | 12 | 1 | 0 | 8 | 0 | 0 | 1 | 1 | 0 | 178 | 7 | 2 | 0 | 0 | 1 | 0 | 0 | 3 | 3 | 0 |
| q | 8 | 1 | 0 | 14 | 9 | 5 | 19 | 6 | 0 | 6 | 2 | 1 | 1 | 1 | 0 | 48 | 110 | 0 | 0 | 1 | 1 | 0 | 0 | 6 | 2 | 2 |
| r | 15 | 4 | 0 | 6 | 4 | 40 | 14 | 7 | 7 | 20 | 6 | 1 | 9 | 5 | 0 | 34 | 9 | 21 | 0 | 6 | 6 | 1 | 1 | 8 | 8 | 10 |
| s | 30 | 4 | 0 | 6 | 13 | 2 | 64 | 2 | 1 | 1 | 4 | 0 | 3 | 0 | 3 | 7 | 4 | 0 | 18 | 0 | 1 | 0 | 4 | 27 | 12 | 9 |
| t | 15 | 13 | 0 | 23 | 4 | 38 | 9 | 7 | 14 | 11 | 12 | 5 | 6 | 0 | 1 | 21 | 9 | 4 | 1 | 15 | 10 | 0 | 2 | 2 | 4 | 2 |
| u | 28 | 10 | 0 | 21 | 4 | 7 | 10 | 15 | 0 | 7 | 1 | 0 | 30 | 16 | 1 | 16 | 12 | 2 | 0 | 1 | 34 | 0 | 10 | 2 | 2 | 3 |
| v | 8 | 2 | 0 | 3 | 6 | 2 | 9 | 1 | 0 | 0 | 5 | 0 | 5 | 1 | 0 | 12 | 3 | 2 | 1 | 1 | 1 | 33 | 50 | 28 | 78 | 5 |
| w | 12 | 2 | 0 | 2 | 2 | 2 | 4 | 3 | 0 | 3 | 3 | 0 | 7 | 1 | 2 | 9 | 8 | 2 | 0 | 0 | 2 | 1 | 111 | 18 | 37 | 5 |
| x | 11 | 0 | 0 | 3 | 7 | 3 | 4 | 4 | 0 | 1 | 7 | 0 | 0 | 1 | 0 | 8 | 4 | 2 | 0 | 1 | 1 | 2 | 12 | 117 | 22 | 7 |
| y | 1 | 0 | 0 | 0 | 2 | 0 | 15 | 1 | 0 | 4 | 2 | 0 | 0 | 0 | 0 | 4 | 2 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 189 | 0 |
| z | 10 | 0 | 0 | 0 | 2 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 8 | 14 | 206 |

Figure A.10: YTY3

## A.2 Luce choice model parameters

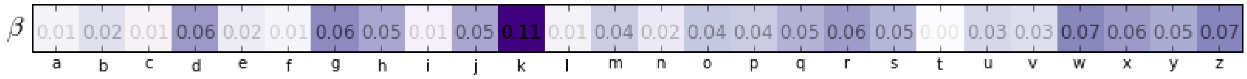


Figure A.11: AXL1: Luce choice $\beta$s

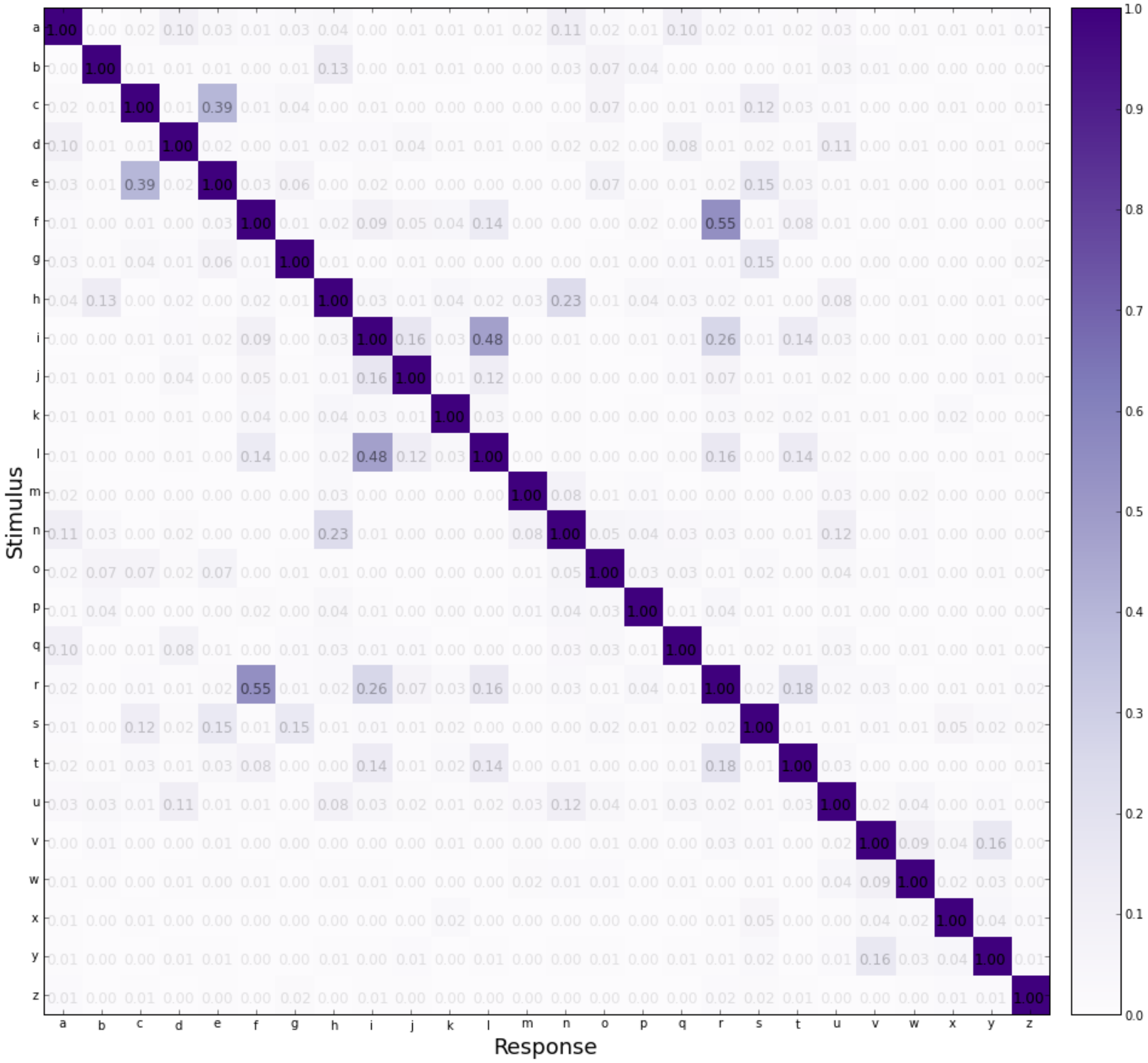


Figure A.12: AXL1: Luce choice $\eta$s

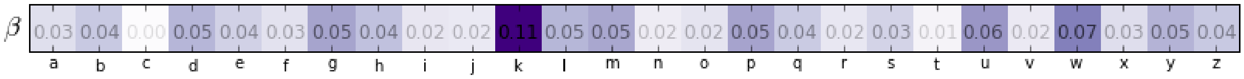


Figure A.13: KMA1: Luce choice $\beta$s

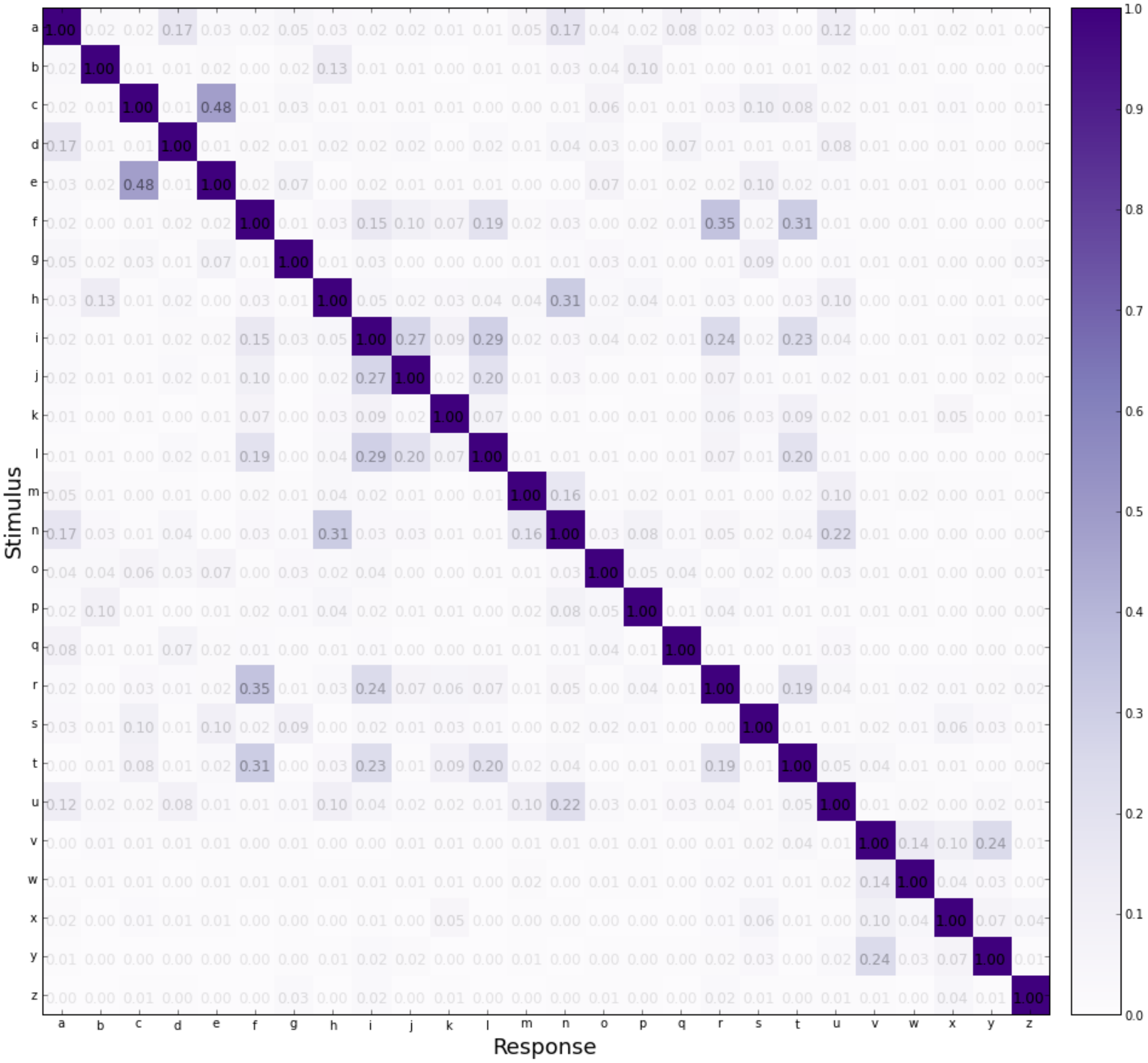
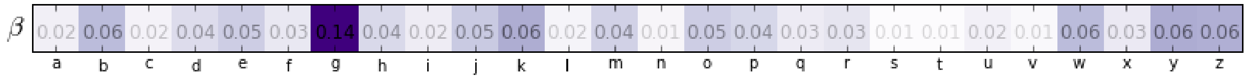
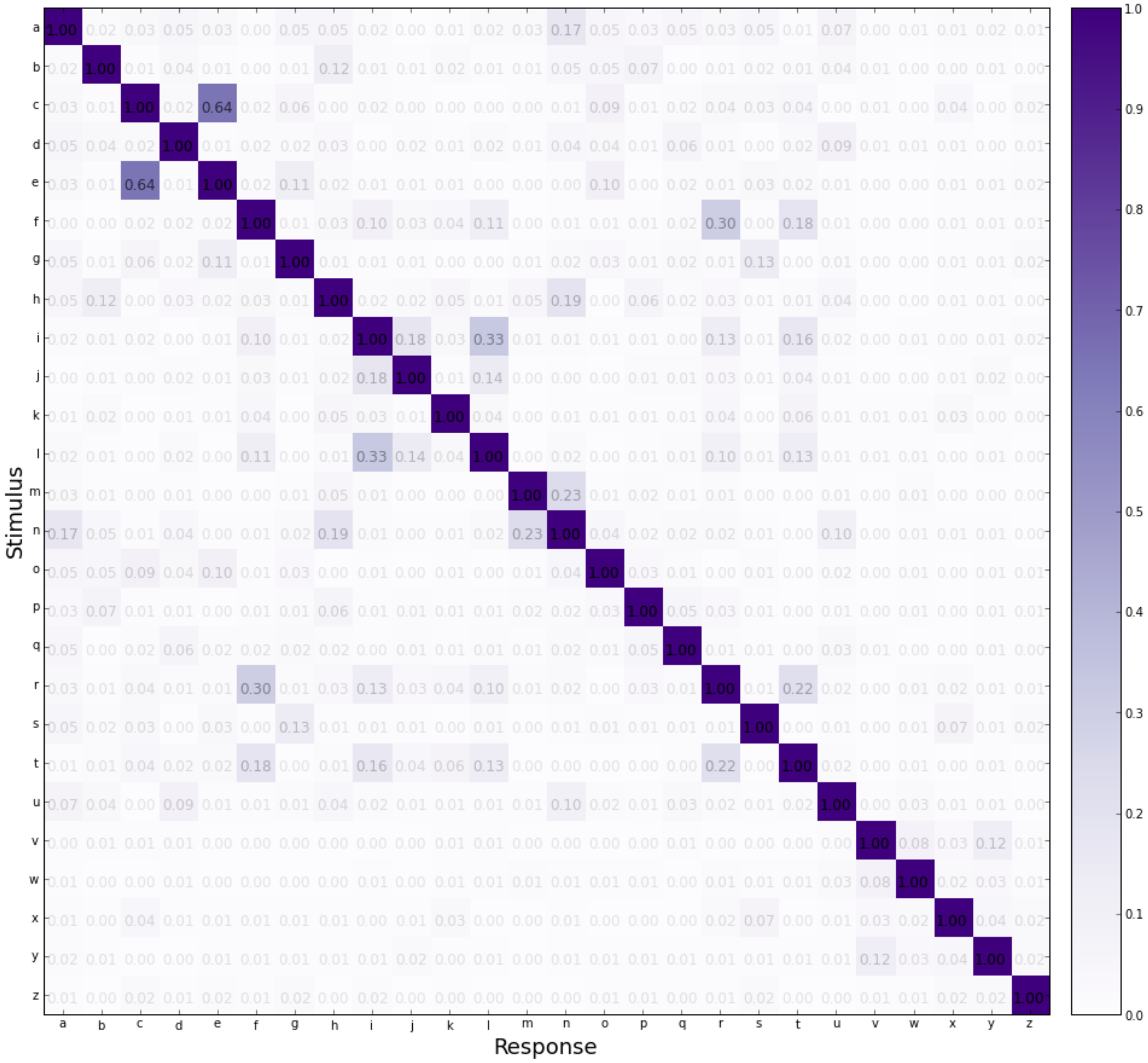


Figure A.14: KMA1: Luce choice $\eta$s. Symmetric around diagonal.



|  | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.02 | 0.06 | 0.02 | 0.04 | 0.05 | 0.03 | 0.14 | 0.04 | 0.02 | 0.05 | 0.06 | 0.02 | 0.04 | 0.01 | 0.05 | 0.04 | 0.03 | 0.03 | 0.01 | 0.01 | 0.02 | 0.01 | 0.06 | 0.03 | 0.06 | 0.06 |

Figure A.15: RXK1: Luce choice $\beta$s



|  | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 1.00 | 0.02 | 0.03 | 0.05 | 0.03 | 0.00 | 0.05 | 0.05 | 0.02 | 0.00 | 0.01 | 0.02 | 0.03 | 0.17 | 0.05 | 0.03 | 0.05 | 0.03 | 0.05 | 0.01 | 0.07 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 |
| b | 0.02 | 1.00 | 0.01 | 0.04 | 0.01 | 0.00 | 0.01 | 0.12 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.05 | 0.05 | 0.07 | 0.00 | 0.01 | 0.02 | 0.01 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 |
| c | 0.03 | 0.01 | 1.00 | 0.02 | 0.64 | 0.02 | 0.06 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.09 | 0.01 | 0.02 | 0.04 | 0.03 | 0.04 | 0.00 | 0.01 | 0.00 | 0.04 | 0.00 | 0.02 |
| d | 0.05 | 0.04 | 0.02 | 1.00 | 0.01 | 0.02 | 0.02 | 0.03 | 0.00 | 0.02 | 0.01 | 0.02 | 0.01 | 0.04 | 0.04 | 0.01 | 0.06 | 0.01 | 0.00 | 0.02 | 0.09 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 |
| e | 0.03 | 0.01 | 0.64 | 0.01 | 1.00 | 0.02 | 0.11 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.02 | 0.01 | 0.03 | 0.02 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| f | 0.00 | 0.00 | 0.02 | 0.02 | 0.02 | 1.00 | 0.01 | 0.03 | 0.10 | 0.03 | 0.04 | 0.11 | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 | 0.30 | 0.00 | 0.18 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 |
| g | 0.05 | 0.01 | 0.06 | 0.02 | 0.11 | 0.01 | 1.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.02 | 0.03 | 0.01 | 0.02 | 0.01 | 0.13 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| h | 0.05 | 0.12 | 0.00 | 0.03 | 0.02 | 0.03 | 0.01 | 1.00 | 0.02 | 0.02 | 0.05 | 0.01 | 0.05 | 0.19 | 0.00 | 0.06 | 0.02 | 0.03 | 0.01 | 0.01 | 0.04 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 |
| i | 0.02 | 0.01 | 0.02 | 0.00 | 0.01 | 0.10 | 0.01 | 0.02 | 1.00 | 0.18 | 0.03 | 0.33 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.13 | 0.01 | 0.16 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.02 |
| j | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 0.03 | 0.01 | 0.02 | 0.18 | 1.00 | 0.01 | 0.14 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 | 0.01 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.02 | 0.00 |
| k | 0.01 | 0.02 | 0.00 | 0.01 | 0.01 | 0.04 | 0.00 | 0.05 | 0.03 | 0.01 | 1.00 | 0.04 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.00 | 0.06 | 0.01 | 0.01 | 0.01 | 0.03 | 0.00 | 0.00 |
| l | 0.02 | 0.01 | 0.00 | 0.02 | 0.00 | 0.11 | 0.00 | 0.01 | 0.33 | 0.14 | 0.04 | 1.00 | 0.00 | 0.02 | 0.00 | 0.01 | 0.01 | 0.10 | 0.01 | 0.13 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 |
| m | 0.03 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.05 | 0.01 | 0.00 | 0.00 | 0.00 | 1.00 | 0.23 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| n | 0.17 | 0.05 | 0.01 | 0.04 | 0.00 | 0.01 | 0.02 | 0.19 | 0.01 | 0.00 | 0.01 | 0.02 | 0.23 | 1.00 | 0.04 | 0.02 | 0.02 | 0.02 | 0.01 | 0.00 | 0.10 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 |
| o | 0.05 | 0.05 | 0.09 | 0.04 | 0.10 | 0.01 | 0.03 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.04 | 1.00 | 0.03 | 0.01 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 |
| p | 0.03 | 0.07 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.06 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 1.00 | 0.05 | 0.03 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 |
| q | 0.05 | 0.00 | 0.02 | 0.06 | 0.02 | 0.02 | 0.02 | 0.02 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.05 | 1.00 | 0.01 | 0.01 | 0.00 | 0.03 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 |
| r | 0.03 | 0.01 | 0.04 | 0.01 | 0.01 | 0.30 | 0.01 | 0.03 | 0.13 | 0.03 | 0.04 | 0.10 | 0.01 | 0.02 | 0.00 | 0.03 | 0.01 | 1.00 | 0.01 | 0.22 | 0.02 | 0.00 | 0.01 | 0.02 | 0.01 | 0.01 |
| s | 0.05 | 0.02 | 0.03 | 0.00 | 0.03 | 0.00 | 0.13 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 1.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.07 | 0.01 | 0.02 |
| t | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0.18 | 0.00 | 0.01 | 0.16 | 0.04 | 0.06 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.22 | 0.00 | 1.00 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 |
| u | 0.07 | 0.04 | 0.00 | 0.09 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.10 | 0.02 | 0.01 | 0.03 | 0.02 | 0.01 | 0.02 | 1.00 | 0.00 | 0.03 | 0.01 | 0.01 | 0.00 |
| v | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.08 | 0.03 | 0.12 | 0.01 |
| w | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.03 | 0.08 | 1.00 | 0.02 | 0.03 | 0.01 |
| x | 0.01 | 0.00 | 0.04 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.07 | 0.00 | 0.01 | 0.03 | 0.02 | 1.00 | 0.04 | 0.02 |
| y | 0.02 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.12 | 0.03 | 0.04 | 1.00 | 0.02 |
| z | 0.01 | 0.00 | 0.02 | 0.01 | 0.02 | 0.01 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.02 | 1.00 |

Figure A.16: RXK1: Luce choice $\eta$s. Symmetric around diagonal.

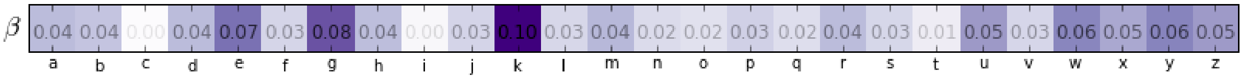


Figure A.17: RXL1: Luce choice $\beta$s

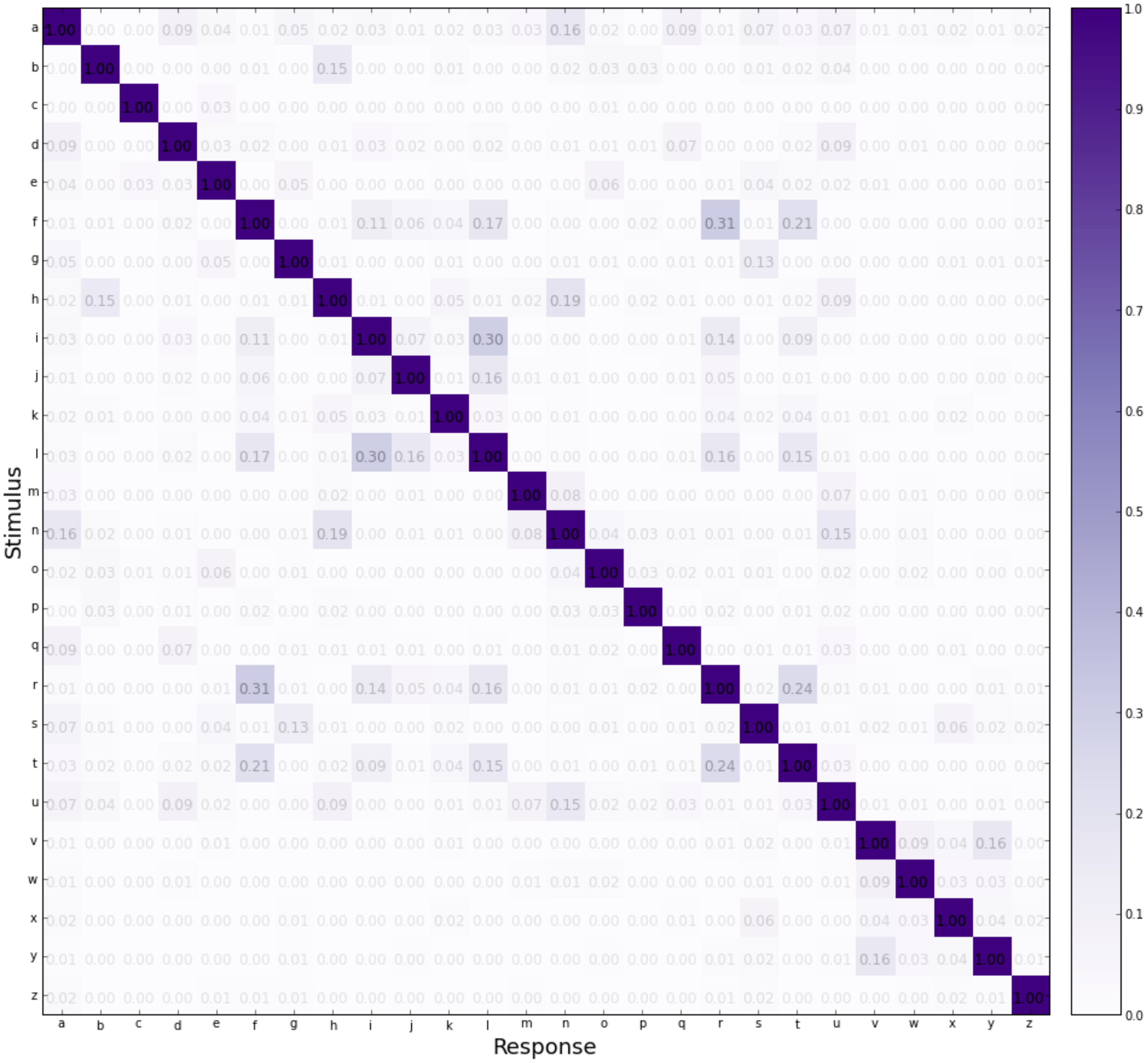


Figure A.18: RXL1: Luce choice $\eta$s. Symmetric around diagonal.

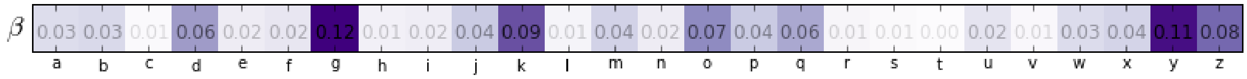


Figure A.19: YTY1: Luce choice $\beta$s

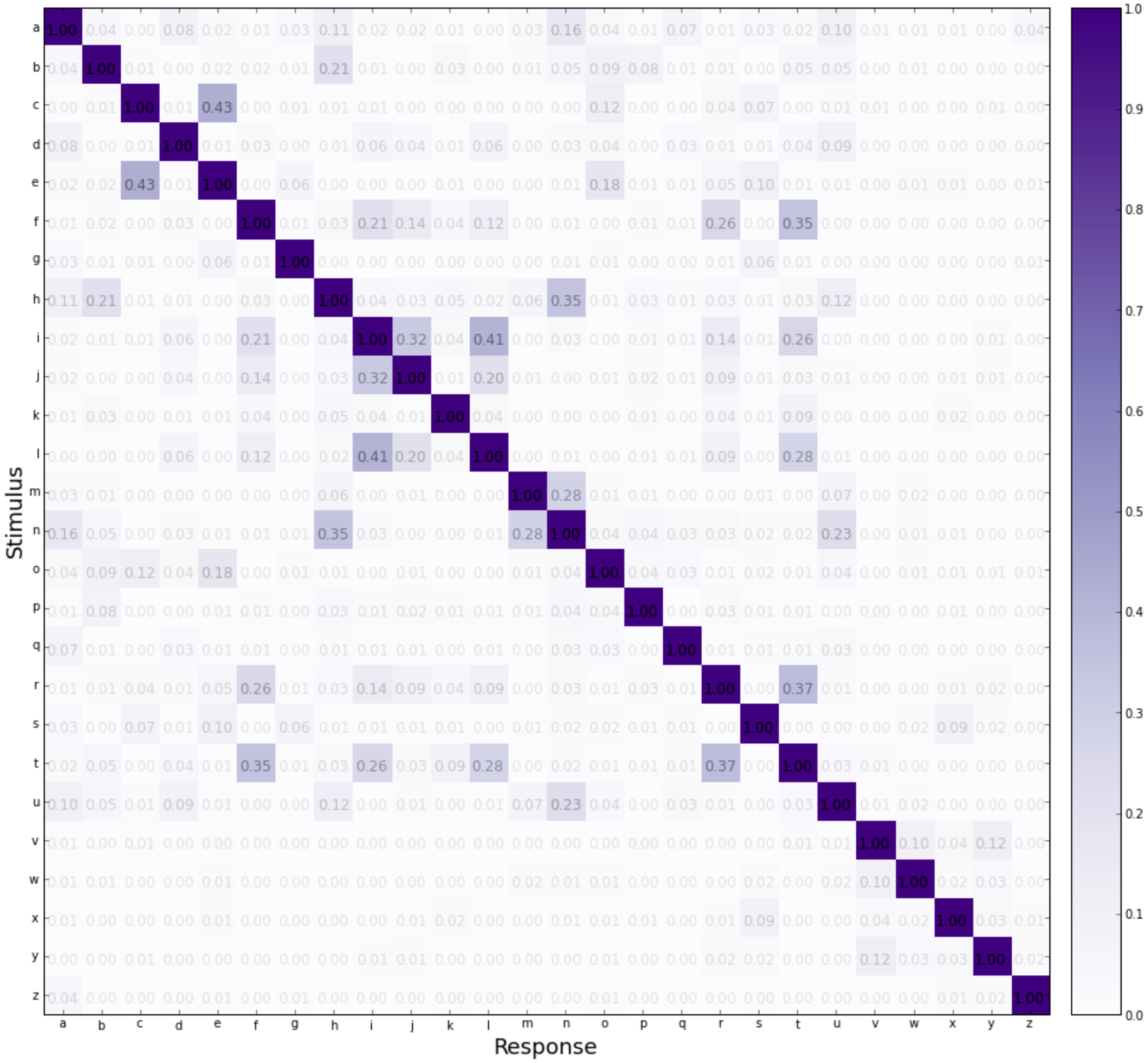


Figure A.20: YTY1: Luce choice $\eta$s. Symmetric around diagonal.

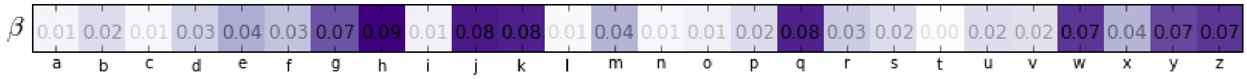


Figure A.21: AXL3: Luce choice $\beta$s

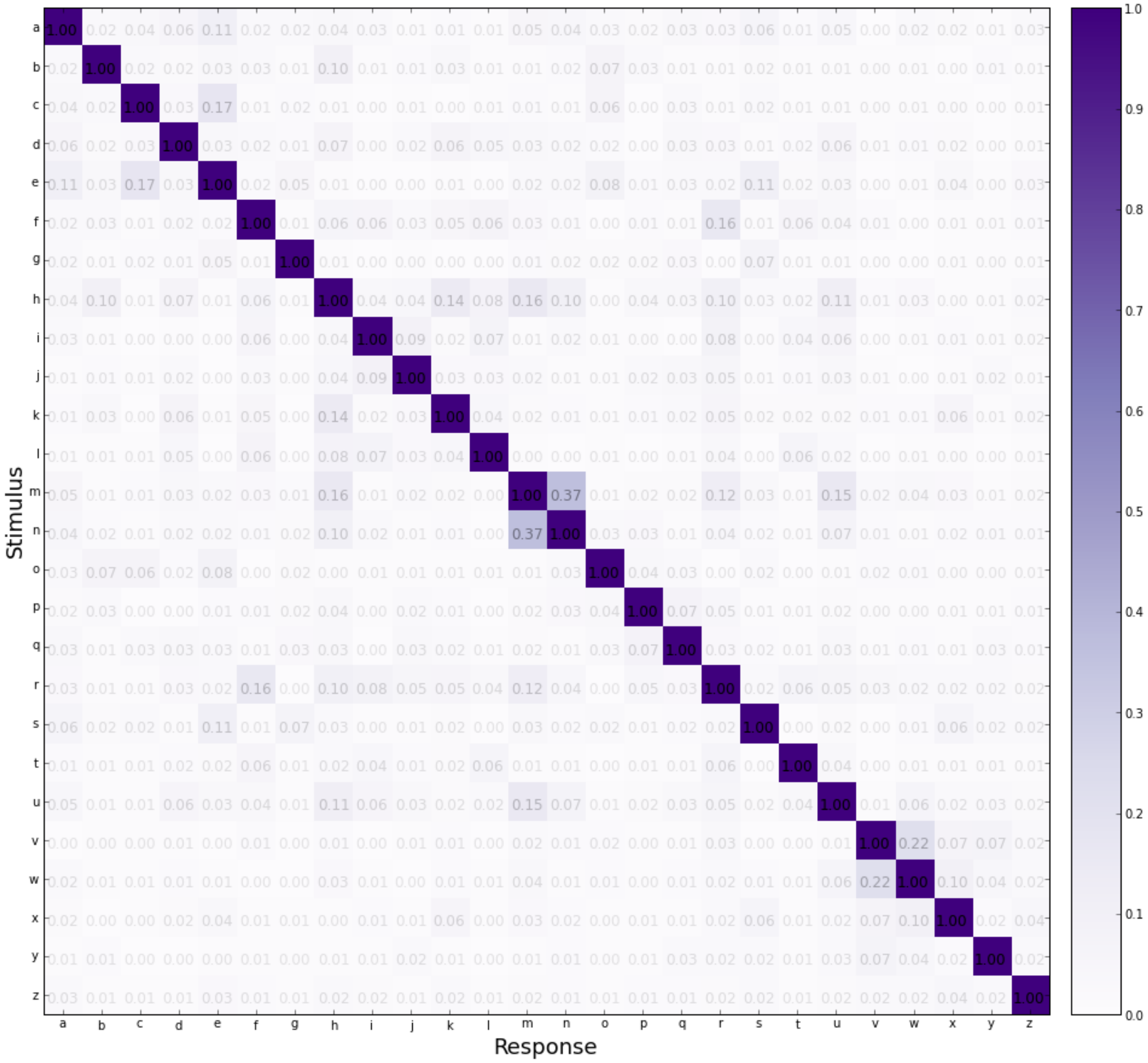
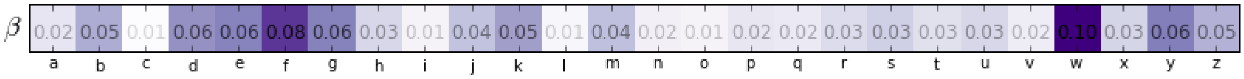
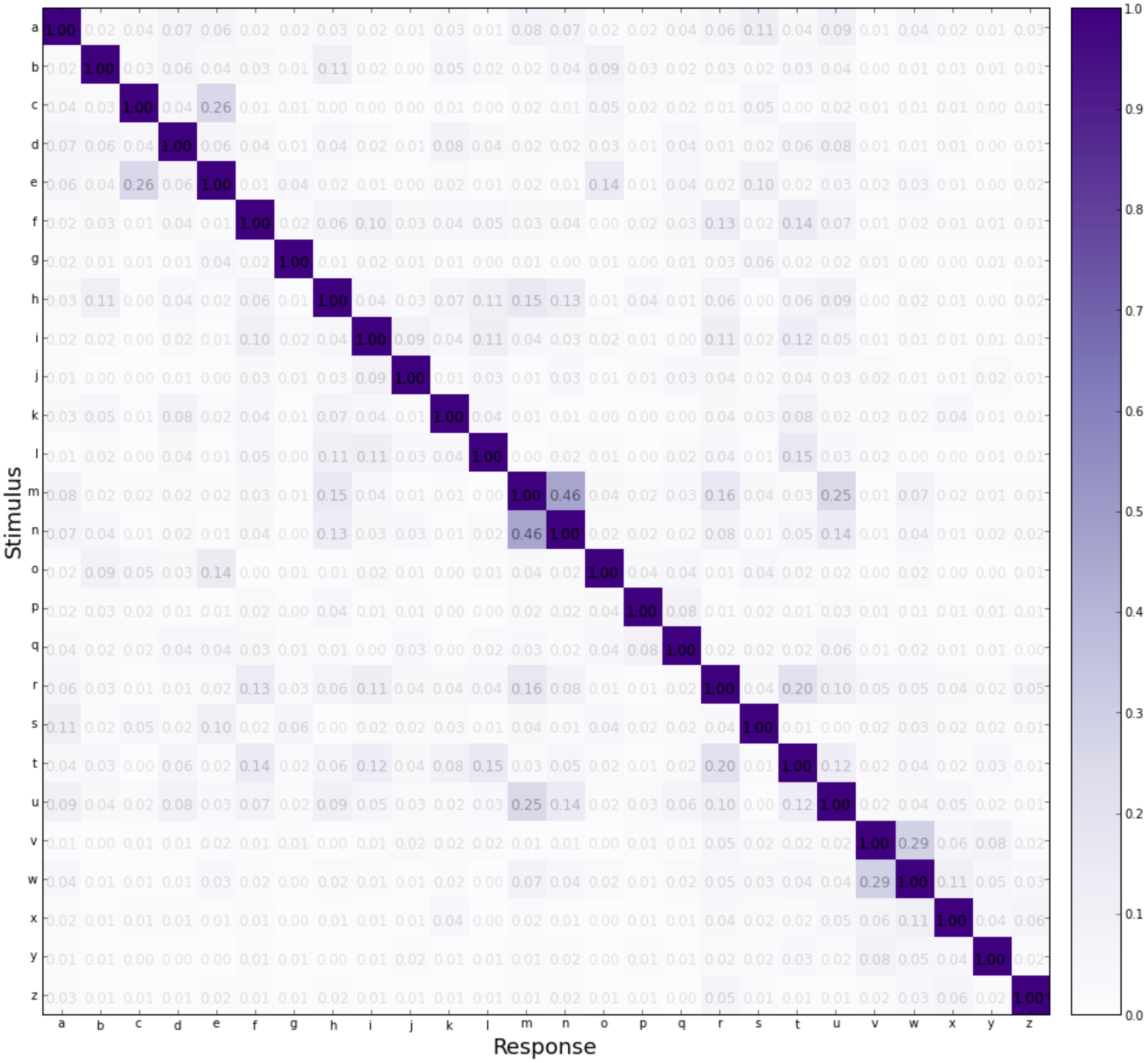


Figure A.22: AXL3: Luce choice $\eta$s. Symmetric around diagonal.



| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.02 | 0.05 | 0.01 | 0.06 | 0.06 | 0.08 | 0.06 | 0.03 | 0.01 | 0.04 | 0.05 | 0.01 | 0.04 | 0.02 | 0.01 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.10 | 0.03 | 0.06 | 0.05 |

Figure A.23: KMA3: Luce choice $\beta$s



| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 1.00 | 0.02 | 0.04 | 0.07 | 0.06 | 0.02 | 0.02 | 0.03 | 0.02 | 0.01 | 0.03 | 0.01 | 0.08 | 0.07 | 0.02 | 0.02 | 0.04 | 0.06 | 0.11 | 0.04 | 0.09 | 0.01 | 0.04 | 0.02 | 0.01 | 0.03 |
| b | 0.02 | 1.00 | 0.03 | 0.06 | 0.04 | 0.03 | 0.01 | 0.11 | 0.02 | 0.00 | 0.05 | 0.02 | 0.02 | 0.04 | 0.09 | 0.03 | 0.02 | 0.03 | 0.02 | 0.03 | 0.04 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 |
| c | 0.04 | 0.03 | 1.00 | 0.04 | 0.26 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 | 0.01 | 0.05 | 0.02 | 0.02 | 0.01 | 0.05 | 0.00 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 |
| d | 0.07 | 0.06 | 0.04 | 1.00 | 0.06 | 0.04 | 0.01 | 0.04 | 0.02 | 0.01 | 0.08 | 0.04 | 0.02 | 0.02 | 0.03 | 0.01 | 0.04 | 0.01 | 0.02 | 0.06 | 0.08 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 |
| e | 0.06 | 0.04 | 0.26 | 0.06 | 1.00 | 0.01 | 0.04 | 0.02 | 0.01 | 0.00 | 0.02 | 0.01 | 0.02 | 0.01 | 0.14 | 0.01 | 0.04 | 0.02 | 0.10 | 0.02 | 0.03 | 0.02 | 0.03 | 0.01 | 0.00 | 0.02 |
| f | 0.02 | 0.03 | 0.01 | 0.04 | 0.01 | 1.00 | 0.02 | 0.06 | 0.10 | 0.03 | 0.04 | 0.05 | 0.03 | 0.04 | 0.00 | 0.02 | 0.03 | 0.13 | 0.02 | 0.14 | 0.07 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 |
| g | 0.02 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 1.00 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.03 | 0.06 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 |
| h | 0.03 | 0.11 | 0.00 | 0.04 | 0.02 | 0.06 | 0.01 | 1.00 | 0.04 | 0.03 | 0.07 | 0.11 | 0.15 | 0.13 | 0.01 | 0.04 | 0.01 | 0.06 | 0.00 | 0.06 | 0.09 | 0.00 | 0.02 | 0.01 | 0.00 | 0.02 |
| i | 0.02 | 0.02 | 0.00 | 0.02 | 0.01 | 0.10 | 0.02 | 0.04 | 1.00 | 0.09 | 0.04 | 0.11 | 0.04 | 0.03 | 0.02 | 0.01 | 0.00 | 0.11 | 0.02 | 0.12 | 0.05 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| j | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.03 | 0.01 | 0.03 | 0.09 | 1.00 | 0.01 | 0.03 | 0.01 | 0.03 | 0.01 | 0.01 | 0.03 | 0.04 | 0.02 | 0.04 | 0.03 | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 |
| k | 0.03 | 0.05 | 0.01 | 0.08 | 0.02 | 0.04 | 0.01 | 0.07 | 0.04 | 0.01 | 1.00 | 0.04 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.04 | 0.03 | 0.08 | 0.02 | 0.02 | 0.02 | 0.04 | 0.01 | 0.01 |
| l | 0.01 | 0.02 | 0.00 | 0.04 | 0.01 | 0.05 | 0.00 | 0.11 | 0.11 | 0.03 | 0.04 | 1.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.02 | 0.04 | 0.01 | 0.15 | 0.03 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 |
| m | 0.08 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.01 | 0.15 | 0.04 | 0.01 | 0.01 | 0.00 | 1.00 | 0.46 | 0.04 | 0.02 | 0.03 | 0.16 | 0.04 | 0.03 | 0.25 | 0.01 | 0.07 | 0.02 | 0.01 | 0.01 |
| n | 0.07 | 0.04 | 0.01 | 0.02 | 0.01 | 0.04 | 0.00 | 0.13 | 0.03 | 0.03 | 0.01 | 0.02 | 0.46 | 1.00 | 0.02 | 0.02 | 0.02 | 0.08 | 0.01 | 0.05 | 0.14 | 0.01 | 0.04 | 0.01 | 0.02 | 0.02 |
| o | 0.02 | 0.09 | 0.05 | 0.03 | 0.14 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 | 0.00 | 0.01 | 0.04 | 0.02 | 1.00 | 0.04 | 0.04 | 0.01 | 0.04 | 0.02 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 |
| p | 0.02 | 0.03 | 0.02 | 0.01 | 0.01 | 0.02 | 0.00 | 0.04 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 0.02 | 0.04 | 1.00 | 0.08 | 0.01 | 0.02 | 0.01 | 0.03 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| q | 0.04 | 0.02 | 0.02 | 0.04 | 0.04 | 0.03 | 0.01 | 0.01 | 0.00 | 0.03 | 0.00 | 0.02 | 0.03 | 0.02 | 0.04 | 0.08 | 1.00 | 0.02 | 0.02 | 0.02 | 0.06 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 |
| r | 0.06 | 0.03 | 0.01 | 0.01 | 0.02 | 0.13 | 0.03 | 0.06 | 0.11 | 0.04 | 0.04 | 0.04 | 0.16 | 0.08 | 0.01 | 0.01 | 0.02 | 1.00 | 0.04 | 0.20 | 0.10 | 0.05 | 0.05 | 0.04 | 0.02 | 0.05 |
| s | 0.11 | 0.02 | 0.05 | 0.02 | 0.10 | 0.02 | 0.06 | 0.00 | 0.02 | 0.02 | 0.03 | 0.01 | 0.04 | 0.01 | 0.04 | 0.02 | 0.02 | 0.04 | 1.00 | 0.01 | 0.00 | 0.02 | 0.03 | 0.02 | 0.02 | 0.01 |
| t | 0.04 | 0.03 | 0.00 | 0.06 | 0.02 | 0.14 | 0.02 | 0.06 | 0.12 | 0.04 | 0.08 | 0.15 | 0.03 | 0.05 | 0.02 | 0.01 | 0.02 | 0.20 | 0.01 | 1.00 | 0.12 | 0.02 | 0.04 | 0.02 | 0.03 | 0.01 |
| u | 0.09 | 0.04 | 0.02 | 0.08 | 0.03 | 0.07 | 0.02 | 0.09 | 0.05 | 0.03 | 0.02 | 0.03 | 0.25 | 0.14 | 0.02 | 0.03 | 0.06 | 0.10 | 0.00 | 0.12 | 1.00 | 0.02 | 0.04 | 0.05 | 0.02 | 0.01 |
| v | 0.01 | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.01 | 0.02 | 0.02 | 0.02 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.05 | 0.02 | 0.02 | 0.02 | 1.00 | 0.29 | 0.06 | 0.08 | 0.02 |
| w | 0.04 | 0.01 | 0.01 | 0.01 | 0.03 | 0.02 | 0.00 | 0.02 | 0.01 | 0.01 | 0.02 | 0.00 | 0.07 | 0.04 | 0.02 | 0.01 | 0.02 | 0.05 | 0.03 | 0.04 | 0.04 | 0.29 | 1.00 | 0.11 | 0.05 | 0.03 |
| x | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.04 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0.05 | 0.06 | 0.11 | 1.00 | 0.04 | 0.06 |
| y | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.02 | 0.08 | 0.05 | 0.04 | 1.00 | 0.02 |
| z | 0.03 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.05 | 0.01 | 0.01 | 0.01 | 0.02 | 0.03 | 0.06 | 0.02 | 1.00 |

Figure A.24: KMA3: Luce choice $\eta$s. Symmetric around diagonal.

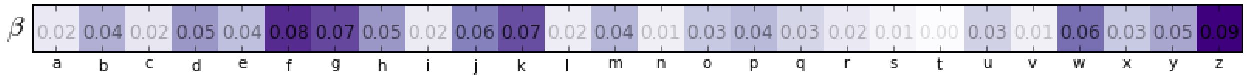


Figure A.25: RXK3: Luce choice $\beta$s

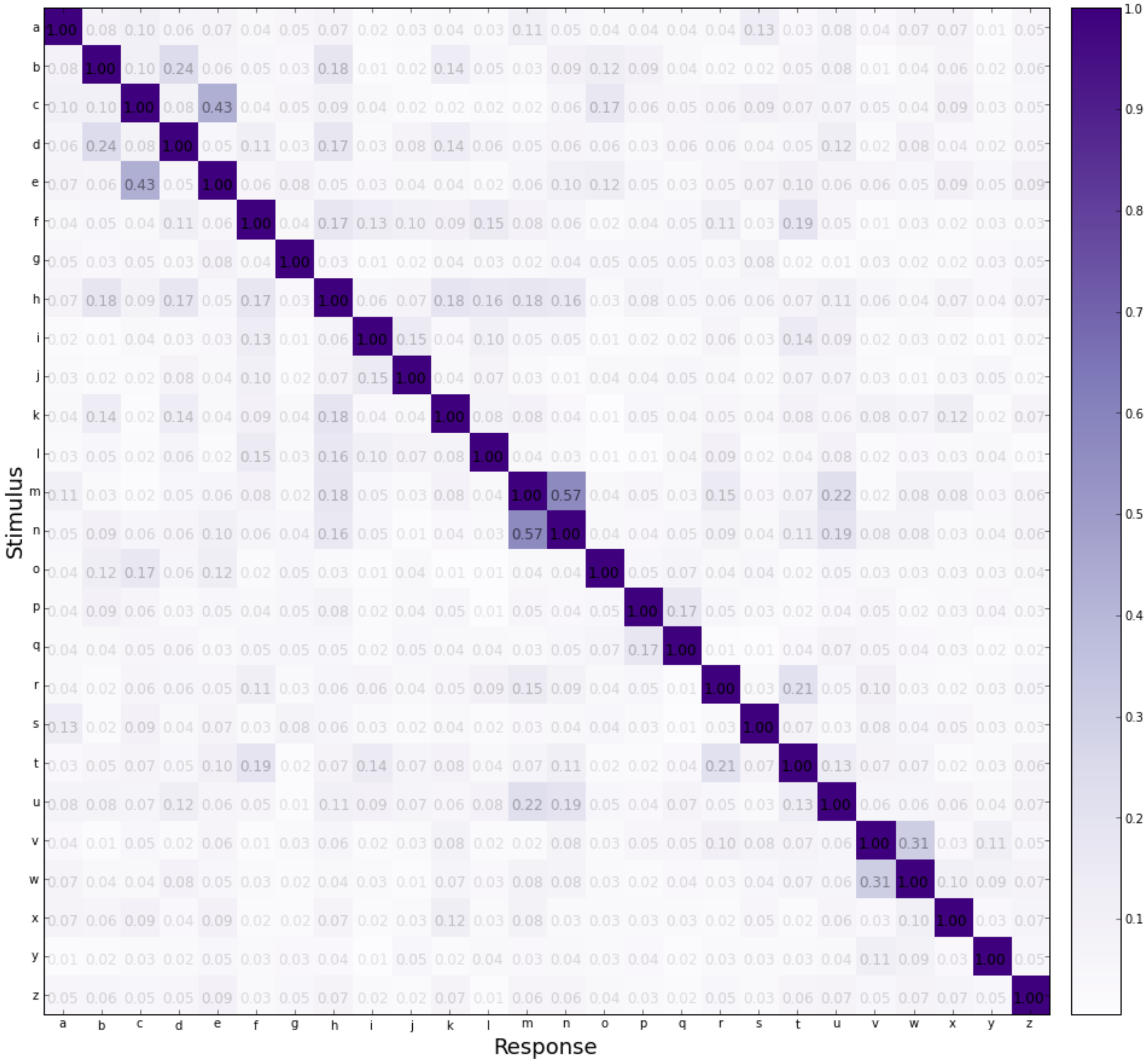


Figure A.26: RXK3: Luce choice $\eta$s

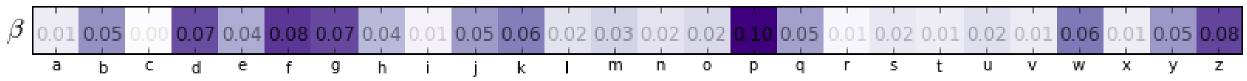


Figure A.27: RXL3: Luce choice $\beta$s

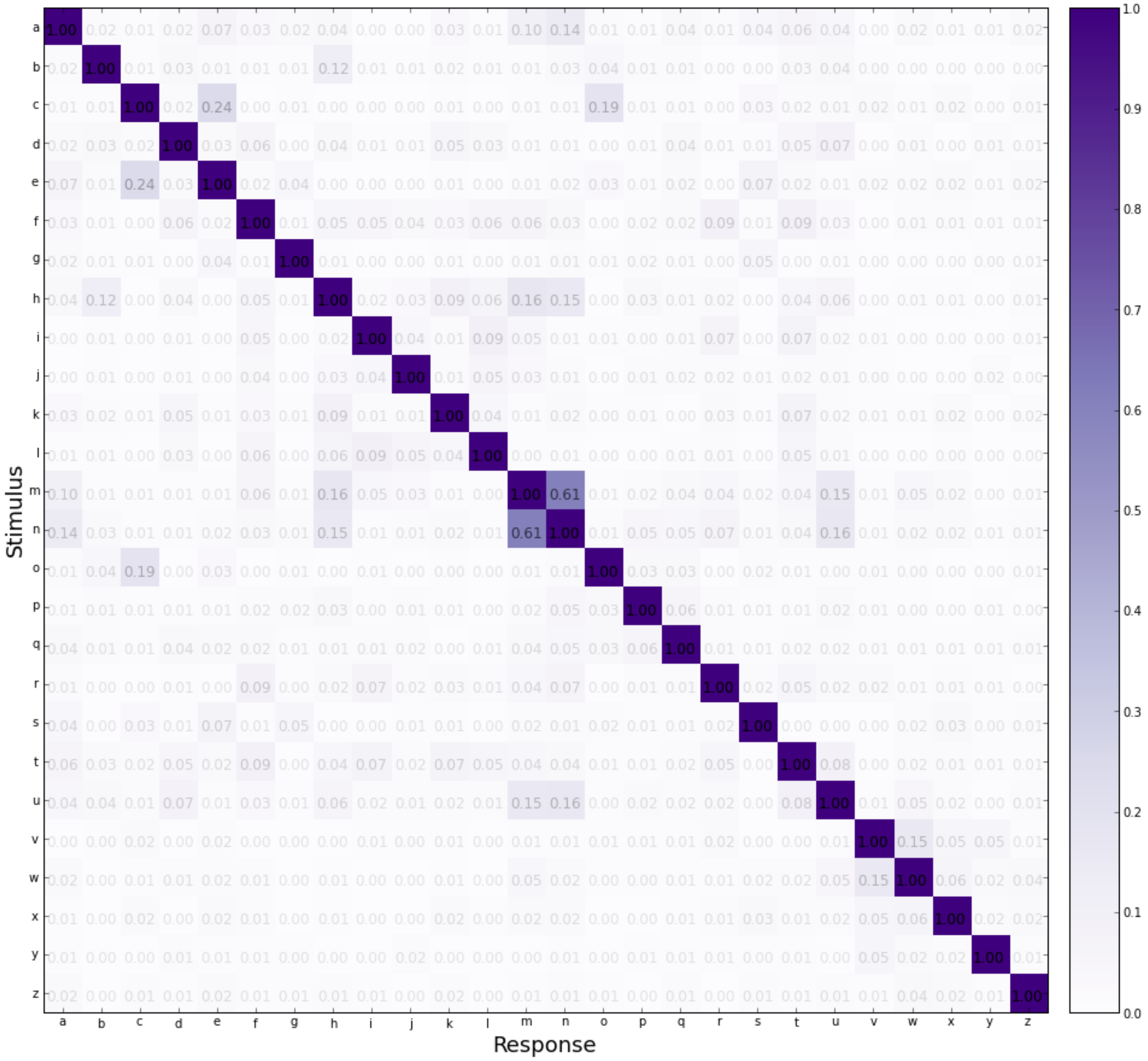


Figure A.28: RXL3: Luce choice $\eta$s. Symmetric around diagonal.

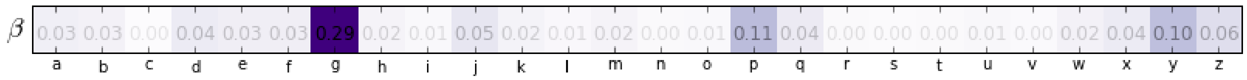


Figure A.29: YTY3: Luce choice $\beta$s

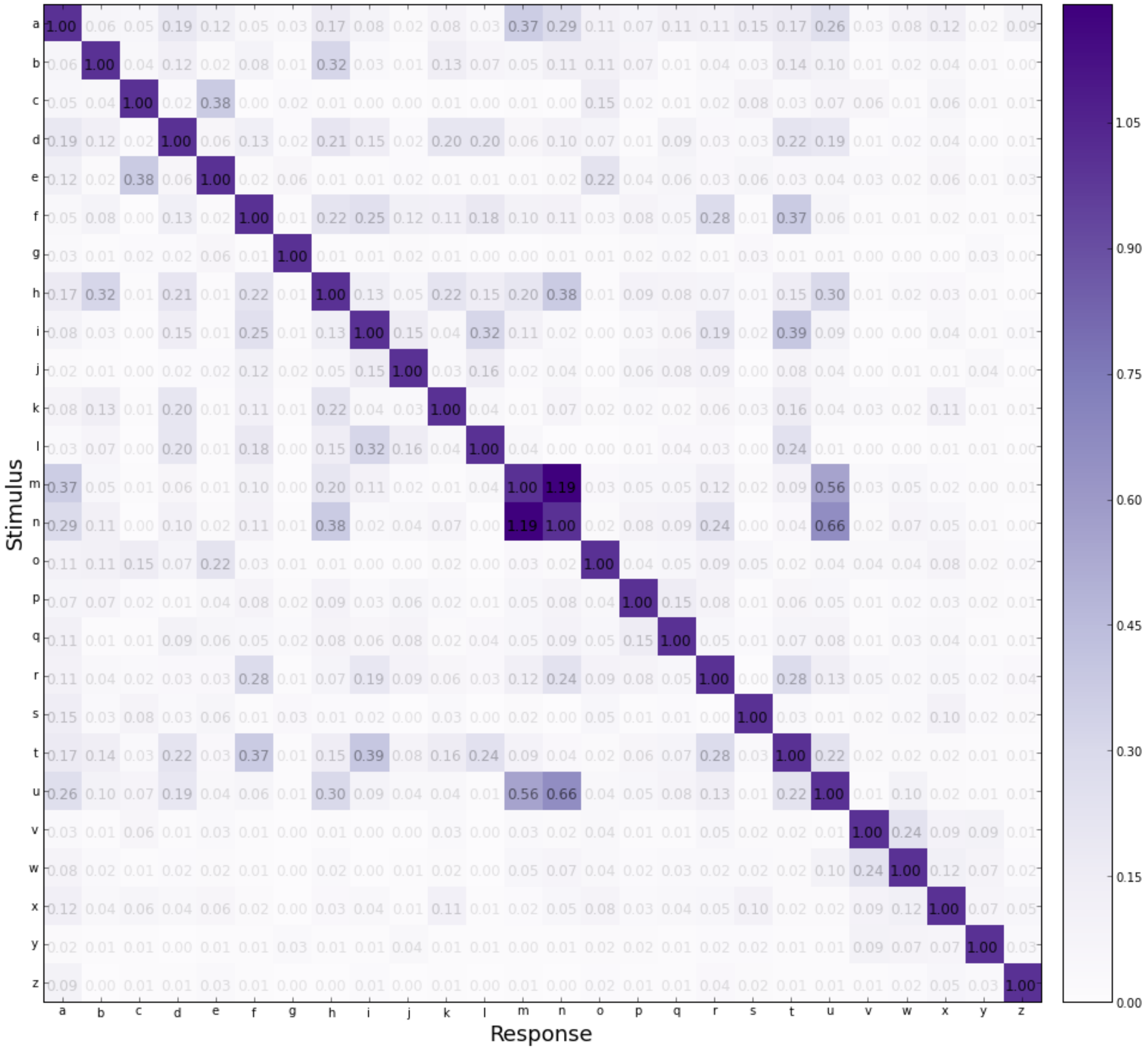


Figure A.30: YTY3: Luce choice $\eta$s. Symmetric around diagonal.

# Bibliography

[1] A. J. Adams, F. Zisman, E. Ai, and G. Bresnick. "Macular edema reduces B cone sensitivity in diabetics". *Applied optics* 26:8 (1987), pp. 1455–1457.

[2] Y. Amit, U. Grenander, and M. Piccioni. "Structural image restoration through deformable templates". *Journal of the American Statistical Association* 86:414 (1991), pp. 376–387.

[3] E. J. Anderson, S. C. Dakin, D. S. Schwarzkopf, G. Rees, and J. A. Greenwood. "The neural correlates of crowding-induced changes in appearance". *Current Biology* 22:13 (2012), pp. 1199–1206.

[4] R. S. Anderson, M. B. Zlatkova, and S. Demirel. "What limits detection and resolution of short-wavelength sinusoidal gratings across the retina?" *Vision research* 42:8 (2002), pp. 981–990.

[5] R. S. Anderson, E. Coulter, M. B. Zlatkova, and S. Demirel. "Short-wavelength acuity: optical factors affecting detection and resolution of blue–yellow sinusoidal gratings in foveal and peripheral vision". *Vision research* 43:1 (2003), pp. 101–107.

[6] S. M. Anstis. "A chart demonstrating variations in acuity with retinal position". *Vision research* 14:7 (1974), pp. 589–592.

[7] F. G. Ashby and J. T. Townsend. "Varieties of perceptual independence". *Psychological review* 93:2 (1986), p. 154.

[8] B. B. Averbeck, P. E. Latham, and A. Pouget. "Neural correlations, population coding and computation". *Nature Reviews Neuroscience* 7:5 (2006), pp. 358–366.

[9] I. L. Bailey. "Perspective: Visual Acuity–Keeping It Clear". *Optometry & Vision Science* 89:9 (2012), pp. 1247–1248.

[10] I. L. Bailey and J. E. Lovie. "New design principles for visual acuity letter charts." *American journal of optometry and physiological optics* 53:11 (1976), pp. 740–745.

[11] I. L. Bailey, T. W. Raasch, P Koh, M Hetland, and A Park. "Contour interaction with high- and low-contrast charts". *Non-invasive Assessment of the Visual System.* Vol. 3. OSA Technical Digest Series. Washington, DC: Optical Society of America, 1993, pp. 228–231.

[12] B. Balas, L. Nakano, and R. Rosenholtz. "A summary-statistic representation in peripheral vision explains visual crowding." *Journal of Vision* 9:12 (Jan. 2009), pp. 13.1–18.

[13] W. S. Baron and G. Westheimer. "Visual acuity as a function of exposure duration". *JOSA* 63:2 (1973), pp. 212–219.

[14] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. J. Goodfellow, A. Bergeron, N. Bouchard, and Y. Bengio. *Theano: new features and speed improvements.* Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop. 2012.

[15] P. J. Beckmann and G. E. Legge. "Preneural limitations on letter identification in central and peripheral vision." *Journal of the Optical Society of America. A, Optics, image science, and vision* 19:12 (Dec. 2002), pp. 2349–62.

[16] H. E. Bedell, J. Siderov, S. J. Waugh, R. Zemanová, F. Pluháček, and L. Musilová. "Contour interaction for foveal acuity targets at different luminances". *Vision research* 89 (2013), pp. 90–95.

[17] P. Bennett and M. Banks. "Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision." *Nature* 326:6116 (1986), pp. 873–876.

[18] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. "Theano: a CPU and GPU Math Expression Compiler". *Proceedings of the Python for Scientific Computing Conference (SciPy).* Oral Presentation. Austin, TX, June 2010.

[19] J.-B. Bernard, C. Aguilar, and E. Castet. "A new font to reduce crowding". *Journal of Vision* 14:10 (2014), pp. 784–784.

[20] J.-B. Bernard and S. T. L. Chung. "The dependence of crowding on flanker complexity and target–flanker similarity". *Journal of vision* 11:8 (2011), p. 1.

[21] I. Biederman. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94:2 (1987), p. 115.

[22] Y. Bishop, S. Fienberg, and P. Holland. "Discrete multivariate analysis: theory and practice" (1975).

[23] D. B. Boles and J. E. Clifford. "An upper-and lowercase alphabetic similarity matrix, with derived generation similarity values". *Behavior Research Methods, Instruments, & Computers* 21:6 (1989), pp. 579–586.

[24] H. Bouma. "Interaction effects in parafoveal letter recognition". *Nature* 226 (1970), pp. 177–178.

[25] H. Bouma. "Visual recognition of isolated lower-case letters." *Vision research* 11:5 (May 1971), pp. 459–74.

[26] R. M. Boynton. *Human color vision.* Holt Rinehart and Winston, 1979.

[27] D. H. Brainard. "The psychophysics toolbox". *Spatial vision* 10 (1997), pp. 433–436.

[28] R. Briggs and D. J. Hocevar. “A new distinctive feature theory for upper case letters.” *The Journal of general psychology* 93 (1975), pp. 87–93.

[29] M. J. Brusco and J. D. Cradit. “ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses”. *Journal of Mathematical Psychology* 49:2 (2005), pp. 142–154.

[30] M. J. Brusco and D. Steinley. “Clustering, seriation, and subset extraction of confusion data.” *Psychological methods* 11:3 (2006), p. 271.

[31] F. W. Campbell and J. Robson. “Application of Fourier analysis to the visibility of gratings”. *The Journal of physiology* 197:3 (1968), pp. 551–566.

[32] W. Chaney, J. Fischer, and D. Whitney. “The hierarchical sparse selection model of visual crowding”. *Frontiers in integrative neuroscience* 8 (2014).

[33] J. Chen, Y. He, Z. Zhu, T. Zhou, Y. Peng, X. Zhang, and F. Fang. “Attention-dependent early cortical suppression contributes to crowding”. *The Journal of Neuroscience* 34:32 (2014), pp. 10465–10474.

[34] V. Chicherov, G. Plomp, and M. H. Herzog. “Neural correlates of visual crowding”. *Neuroimage* 93 (2014), pp. 23–31.

[35] S. T. L. Chung. “Detection and identification of crowded mirror-image letters in normal peripheral vision”. *Vision research* 50:3 (2010), pp. 337–345.

[36] S. T. L. Chung. “Size or spacing: Which limits letter recognition in people with age-related macular degeneration?” *Vision research* 101 (2014), pp. 167–176.

[37] S. T. L. Chung, G. Kumar, and D. R. Coates. “Coarse-to-fine spatial analysis for identifying multiple letters?” *Journal of Vision* 13:9 (2013), pp. 1302–1302.

[38] S. T. L. Chung and G. E. Legge. “Precision of position signals for letters”. *Vision research* 49:15 (2009), pp. 1948–1960.

[39] S. T. L. Chung, G. E. Legge, and B. S. Tjan. “Spatial-frequency characteristics of letter identification in central and peripheral vision”. *Vision research* 42:18 (2002), pp. 2137–2152.

[40] S. T. L. Chung, D. M. Levi, and G. E. Legge. “Spatial-frequency and contrast properties of crowding”. *Vision research* 41:14 (2001), pp. 1833–1850.

[41] S. T. L. Chung, D. M. Levi, and R. W. Li. “Learning to identify contrast-defined letters in peripheral vision”. *Vision research* 46:6 (2006), pp. 1038–1047.

[42] S. T. L. Chung, J. S. Mansfield, and G. E. Legge. “Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision”. *Vision research* 38:19 (1998), pp. 2949–2962.

[43] S. T. L. Chung and B. S. Tjan. “Spatial-frequency and contrast properties of reading in central and peripheral vision”. *Journal of Vision* 9:9 (2009), p. 16.

[44] D. R. Coates, J. M. Chin, and S. T. L. Chung. *Acuity, contrast, eccentricity, and crowding.* Optom Vis Sci 2013; 90 E-Abstract 130461. 2013.

[45] D. R. Coates, J. M. Chin, and S. T. L. Chung. "Acuity, contrast, eccentricity, and crowding". *Journal of Vision* 13:9 (2013), pp. 567–567.

[46] D. R. Coates, J. M. Chin, and S. T. L. Chung. "Factors affecting crowded acuity: Eccentricity and contrast". *Optometry and vision science: official publication of the American Academy of Optometry* 90:7 (2013).

[47] S. Coffin. "Spatial frequency analysis of block letters does not predict experimental confusions". *Perception & Psychophysics* 23:1 (1978), pp. 69–74.

[48] T. F. Cox and M. A. Cox. *Multidimensional scaling.* CRC Press, 2000.

[49] S. C. Dakin, J. Cass, J. A. Greenwood, and P. J. Bex. "Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli". *Journal of Vision* 10:10 (2010), p. 14.

[50] M. V. Danilova and V. M. Bondarko. "Foveal contour interactions and crowding effects at the resolution limit of the visual system". *Journal of Vision* 7:2 (2007), pp. 1–18.

[51] J. G. Daugman. "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression". *Acoustics, Speech and Signal Processing, IEEE Transactions on* 36:7 (1988), pp. 1169–1179.

[52] J. G. Daugman. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters". *JOSA A* 2:7 (1985), pp. 1160–1169.

[53] R. L. DeValois and K. K. DeValois. *Spatial vision.* 14. Oxford University Press, 1988.

[54] N. Drasdo. "Neural substrates and threshold gradients of peripheral vision." *Limits of Vision.* Ed. by J. J. Kulikowski, V Walsh, and M. I. J. Vol. 5. London: Macmillan Press, 1991, pp. 250–264.

[55] J. Duncan and G. W. Humphreys. "Visual search and stimulus similarity." *Psychological review* 96:3 (1989), p. 433.

[56] M. P. Eckstein and A. J. Ahumada. "Classification images: A tool to analyze visual strategies". *Journal of Vision* 2:1 (2002), p. i.

[57] B. A. Eriksen and C. W. Eriksen. "Effects of noise letters upon the identification of a target letter in a nonsearch task". *Perception & psychophysics* 16:1 (1974), pp. 143–149.

[58] E. F. Ester, D. Klee, and E. Awh. "Visual crowding cannot be wholly explained by feature pooling." *Journal of Experimental Psychology: Human Perception and Performance* 40:3 (2014), p. 1022.

[59] E. F. Ester, E. Zilber, and J. T. Serences. “Substitution and pooling in visual crowding induced by similar and dissimilar distractors”. *Journal of vision* 15:1 (2015), p. 4.

[60] D. J. Felleman and D. C. Van Essen. “Distributed hierarchical processing in the primate cerebral cortex”. *Cerebral cortex* 1:1 (1991), pp. 1–47.

[61] F. L. Ferris, A. Kassoff, G. H. Bresnick, and I. Bailey. “New visual acuity charts for clinical research”. *American journal of ophthalmology* 94:1 (1982), pp. 91–96.

[62] D. Fiset, C. Blais, C. Ethier-Majcher, M. Arguin, D. Bub, and F. Gosselin. “Features for identification of uppercase and lowercase letters”. *Psychological Science* 19:11 (2008), pp. 1161–1168.

[63] D. F. Fisher, R. A. Monty, and S Glucksberg. “Visual confusion matrices: fact or artifact?” *The Journal of psychology* 71:1 (Jan. 1969), pp. 111–25.

[64] M. C. Flom. “Contour interaction and the crowding effect”. *Problems in Optometry.* Ed. by R. P. Rubinstein. Vol. 3. Philadelphia, PA: Lippincot, 1991, pp. 237–257.

[65] M. C. Flom, G. G. Heath, and E. Takahashi. “Contour interaction and visual resolution: Contralateral effects”. *Science* 142:3594 (1963), pp. 979–980.

[66] M. C. Flom, F. W. Weymouth, and D. Kahneman. “Visual resolution and contour interaction”. *JOSA* 53:9 (1963), pp. 1026–1032.

[67] J. Freeman, R. Chakravarthi, and D. G. Pelli. “Substitution and pooling in crowding”. *Attention, Perception, & Psychophysics* 74:2 (2012), pp. 379–396.

[68] J. Freeman, T. H. Donner, and D. J. Heeger. “Inter-area correlations in the ventral visual pathway reflect feature integration”. *Journal of vision* 11:4 (2011), p. 15.

[69] J. Freeman and E. P. Simoncelli. “Metamers of the ventral stream”. *Nature neuroscience* 14:9 (2011), pp. 1195–1201.

[70] K. Fukushima. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. *Biological cybernetics* 36:4 (1980), pp. 193–202.

[71] M. J. Gervais, L. O. Harvey, and J. O. Roberts. “Identification confusions among letters of the alphabet.” *Journal of experimental psychology: Human perception and performance* 10:5 (1984), p. 655.

[72] L. H. Geyer and C. G. Dewald. “Feature lists and confusion matrices”. *Perception & Psychophysics* 14 (1973), pp. 471–482.

[73] D. E. Giaschi, D. Regan, S. P. Kraft, and A. C. Kothe. “Crowding and contrast in amblyopia.” *Optometry & Vision Science* 70:3 (1993), pp. 192–197.

[74] E. J. Gibson, J. J. Gibson, A. D. Pick, and H Osser. “A developmental study of the discrimination of letter-like forms.” *Journal of comparative and physiological psychology* 55:6 (Dec. 1962), pp. 897–906.

[75] J. J. Gibson. “The Ecological approach to visual perception” (1986).

[76] G. Gilmore, H Hersh, A Caramazza, and J Griffin. “Multidimensional letter similarity derived from recognition errors”. *Perception & Psychophysics* 25:5 (1979), pp. 425–431.

[77] G. C. Gilmore. “Letters are visual stimuli: A comment on the use of confusion matrices”. *Attention, Perception, & Psychophysics* 37:4 (1985), pp. 389–390.

[78] J Gold, P. J. Bennett, and a. B. Sekuler. “Identification of band-pass filtered letters and faces by human and ideal observers.” *Vision research* 39:21 (Oct. 1999), pp. 3537–60.

[79] F. Gosselin and P. G. Schyns. “Bubbles: a technique to reveal the use of information in recognition tasks”. *Vision research* 41:17 (2001), pp. 2261–2271.

[80] J. Grainger, A. Rey, and S. Dufau. “Letter perception: from pixels to pandemonium”. *Trends in cognitive sciences* 12:10 (2008), pp. 381–387.

[81] V. C. Greenstein, D. C. Hood, R. Ritch, D. Steinberger, and R. E. Carr. “S (blue) cone pathway vulnerability in retinitis pigmentosa, diabetes and glaucoma.” *Investigative ophthalmology & visual science* 30:8 (1989), pp. 1732–1737.

[82] J. A. Greenwood, P. J. Bex, and S. C. Dakin. “Positional averaging explains crowding with letter-like stimuli.” *Proceedings of the National Academy of Sciences of the United States of America* 106:31 (Aug. 2009), pp. 13130–5.

[83] S. Grossberg. “Competitive learning: From interactive activation to adaptive resonance”. *Cognitive science* 11:1 (1987), pp. 23–63.

[84] R. Guillery and S. M. Sherman. “Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system”. *Neuron* 33:2 (2002), pp. 163–175.

[85] R. Gurnsey, G. Roddy, and W. Chanab. “Crowding is size and eccentricity dependent”. *Journal of Vision* 11:7 (15 2011), pp. 1–17.

[86] G. Haegerstrom-Portnoy. “The Glenn A. Fry Award Lecture 2003: vision in elderssummary of findings of the SKI study”. *Optometry & Vision Science* 82:2 (2005), pp. 87–93.

[87] D. Hanus and E. Vul. “Quantifying error distributions in crowding”. *Journal of Vision* 13:4 (2013), p. 17.

[88] S. Hariharan, D. M. Levi, and S. A. Klein. ““Crowding” in normal and amblyopic vision assessed with Gaussian and Gabor Cs”. *Vision research* 45:5 (2005), pp. 617–633.

[89] J. Hegdé. “Time course of visual perception: coarse-to-fine processing and beyond”. *Progress in neurobiology* 84:4 (2008), pp. 405–439.

[90] J. Hegde and D. J. Felleman. “Reappraising the Functional Implications of the Primate Visual Anatomical Hierarchy”. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 13:5 (Oct. 2007), pp. 416–421.

[91] S. H. Hendry and R. C. Reid. "The koniocellular pathway in primate vision". *Annual review of neuroscience* 23:1 (2000), pp. 127–153.

[92] P. R. Herse and H. E. Bedell. "Contrast sensitivity for letter and grating targets under various stimulus conditions." *Optometry & Vision Science* 66:11 (1989), pp. 774–781.

[93] R. F. Hess and D. Field. "Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disarray?" *Vision research* 33:18 (1993), pp. 2663–2670.

[94] R. F. Hess, S. C. Dakin, N. Kapoor, and M. Tewfik. "Contour interaction in fovea and periphery". *JOSA A* 17:9 (2000), pp. 1516–1524.

[95] K. E. Higgins, A. Arditi, and K. Knoblauch. "Detection and identification of mirror-image letter pairs in central and peripheral vision." *Vision research* 36:2 (Jan. 1996), pp. 331–7.

[96] J. J. Hopfield. "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the national academy of sciences* 79:8 (1982), pp. 2554–2558.

[97] D. H. Hubel and T. N. Wiesel. "Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor." *The Journal of comparative neurology* 158:3 (Dec. 1974), pp. 295–305.

[98] D. H. Hubel and T. N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". *The Journal of physiology* 160:1 (1962), pp. 106–154.

[99] A. Hyvärinen and P. O. Hoyer. "Topographic independent component analysis as a model of V1 organization and receptive fields". *Neurocomputing* 38 (2001), pp. 1307–1315.

[100] L. Hyvärinen, R. Näsänen, and P. Laurinen. "New visual acuity test for pre-school children". *Acta ophthalmologica* 58:4 (1980), pp. 507–511.

[101] R. Ihaka and R. Gentleman. "R: a language for data analysis and graphics". *Journal of computational and graphical statistics* 5:3 (1996), pp. 299–314.

[102] J Intriligator and P. Cavanagh. "The spatial resolution of visual attention." *Cognitive psychology* 43:3 (Nov. 2001), pp. 171–216.

[103] A. M. Jacobs, T. A. Nazir, and O. Heller. "Perception of lowercase letters in peripheral vision: A discrimination matrix based on saccade latencies". *Perception & Psychophysics* 46:1 (1989), pp. 95–102.

[104] R. Jacobs. "Visual resolution and contour interaction in the fovea and periphery". *Vision research* 19:11 (1979), pp. 1187–1195.

[105] A. K. Jain, Y. Zhong, and S. Lakshmanan. "Object matching using deformable templates". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 18:3 (1996), pp. 267–278.

[106] C. A. Johnson, A. J. Adams, E. J. Casson, and J. D. Brandt. “Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss”. *Archives of ophthalmology* 111:5 (1993), pp. 645–650.

[107] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. Le-Cun. “Learning invariant features through topographic filter maps”. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE. 2009, pp. 1605–1612.

[108] G Keren and S Baggen. “Recognition models of alphanumeric characters.” *Perception & Psychophysics* 29:3 (Mar. 1981), pp. 234–46.

[109] T. Keunnapas and A.-J. Janson. “Multidimensional Similarity of Letters”. *Perceptual and Motor Skills* 28 (1969), pp. 3–12.

[110] F. A. Kingdom and N. Prins. *Psychophysics: A Practical Introduction.* London: Elsevier/Academic Press, 2009.

[111] S. A. Klein and D. M. Levi. “Position sense of the peripheral retina”. *JOSA A* 4:8 (1987), pp. 1543–1553.

[112] R. C. Kleiner, C. Enger, M. F. Alexander, and S. L. Fine. “Contrast sensitivity in age-related macular degeneration”. *Archives of ophthalmology* 106:1 (1988), pp. 55–57.

[113] F. L. Kooi, A. Toet, S. P. Tripathy, and D. M. Levi. “The effect of similarity and duration on spatial interaction in peripheral vision.” *Spatial vision* 8:2 (Jan. 1994), pp. 255–79.

[114] W Korte. “Uber die Gestaltauffassung im indirekten Sehen”. *Zeitschrift fur Psychologie* 93 (1923), pp. 17–82.

[115] A. C. Kothe and D. Regan. “Crowding depends on contrast.” *Optometry & Vision Science* 67:4 (1990), pp. 283–286.

[116] C. L. Krumhansl. “Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density.” *Psychological Review* 85:5 (1978), pp. 445–463.

[117] M. Kwon, C. Ramachandra, P. Satgunam, B. W. Mel, E. Peli, and B. S. Tjan. “Contour enhancement benefits older adults with simulated central field loss”. *Optometry and vision science: official publication of the American Academy of Optometry* 89:9 (2012), p. 1374.

[118] M. Kwon, P. Bao, R. Millin, and B. S. Tjan. “Radial-tangential anisotropy of crowding in the early visual areas”. *Journal of neurophysiology* 112:10 (2014), pp. 2413–2422.

[119] S. N. Lanthier, E. F. Risko, J. A. Stolz, and D. Besner. “Not all visual features are created equal: Early processing in letter and word recognition”. *Psychonomic bulletin & review* 16:1 (2009), pp. 67–73.

[120] K. Latham and D. Whitaker. “Relative roles of resolution and spatial interference in foveal and peripheral vision”. *Ophthalmic and Physiological Optics* 16:1 (1996), pp. 49–57.

[121] S. J. Leat, W. Li, and K. Epp. “Crowding in central and eccentric vision: the effects of contour interaction and attention.” *Investigative Ophthalmology & Visual Science* 40:2 (1999), pp. 504–512.

[122] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition”. *Proceedings of the IEEE* 86:11 (1998), pp. 2278–2324.

[123] M. D. Lee. “A simple method for generating additive clustering models with limited complexity”. *Machine Learning* 49:1 (2002), pp. 39–58.

[124] M. D. Lee. “On the complexity of additive clustering models”. *Journal of Mathematical Psychology* 45:1 (2001), pp. 131–148.

[125] G. E. Legge and G. S. Rubin. “Contrast sensitivity function as a screening test: a critique.” *American journal of optometry and physiological optics* 63:4 (1986), pp. 265–270.

[126] G. E. Legge, G. S. Rubin, and A. Luebker. “Psychophysics of reading–V. The role of contrast in normal vision”. *Vision research* 27:7 (1987), pp. 1165–1177.

[127] D. M. Levi. “Crowding An essential bottleneck for object recognition : A mini-review”. *Vision Research* 48 (2008), pp. 635–654.

[128] D. M. Levi, S. Hariharan, and S. A. Klein. “Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking”. *Journal of Vision* 2:2 (2002), p. 3.

[129] D. M. Levi, S. A. Klein, and A. Aitsebaomo. “Vernier acuity, crowding and cortical magnification”. *Vision research* 25:7 (1985), pp. 963–977.

[130] D. M. Levi, S. A. Klein, and S. Hariharan. “Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking”. *Journal of Vision* 2:2 (2002), p. 2.

[131] D. M. Levi, S. Song, and D. G. Pelli. “Amblyopic reading is crowded”. *Journal of Vision* 7:2 (2007), p. 21.

[132] L. Liu and A. Arditi. “How crowding affects letter confusion.” *Optometry and vision science : official publication of the American Academy of Optometry* 78:1 (Jan. 2001), pp. 50–5.

[133] T. Liu, Y. Jiang, X. Sun, and S. He. “Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli”. *Current Biology* 19:2 (2009), pp. 127–132.

[134] J. M. Loomis. “A model of character recognition and legibility.” *Journal of Experimental Psychology: Human Perception and Performance* 16:1 (1990), p. 106.

[135] R. D. Luce. “Detection and recognition”. *Handbook of mathematical psychology*. Ed. by R. D. Luce, R. R. Bush, and E. Galanter. Vol. 1. New York: Wiley, 1963, pp. 103–189.

[136] R. D. Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2005.

[137] E. Ludvigh. “Effect of reduced contrast on visual acuity as measured with Snellen test letters”. *Archives of Ophthalmology* 25:3 (1941), pp. 469–474.

[138] S. J. Lupker. “On the nature of perceptual information during letter perception.” *Perception & psychophysics* 25:4 (Apr. 1979), pp. 303–12.

[139] N. J. Majaj, D. G. Pelli, P. Kurshan, and M. C. Palomares. “The role of spatial frequency channels in letter identification.” *Vision research* 42:9 (Apr. 2002), pp. 1165–84.

[140] S Marčelja. “Mathematical description of the responses of simple cortical cells*”. *JOSA* 70:11 (1980), pp. 1297–1300.

[141] J. L. McClelland and D. E. Rumelhart. “An interactive activation model of context effects in letter perception: I. An account of basic findings.” *Psychological Review* 88:5 (1981), pp. 375–407.

[142] P. V. McGraw, B. Winn, L. S. Gray, and D. B. Elliott. “Improving the reliability of visual acuity measures in young children”. *Ophthalmic and Physiological Optics* 20:3 (2000), pp. 173–184.

[143] R. Millin, A. C. Arman, S. T. L. Chung, and B. S. Tjan. “Visual Crowding in V1”. *Cerebral Cortex* 24:12 (2014), pp. 3107–3115.

[144] M. Mitchell. *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1998. ISBN: 0262631857.

[145] F. Mosteller and J. W. Tukey. “Data analysis and regression: a second course in statistics.” *Addison-Wesley Series in Behavioral Science: Quantitative Methods* (1977).

[146] B. C. Motter. “Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface”. *The Journal of Neuroscience* 29:18 (2009), pp. 5749–5757.

[147] S. T. Mueller. *Letter Similarity Data Set Archive*. Sept. 2014. URL: http://obereed.net/lettersim/.

[148] S. T. Mueller and C. T. Weidemann. “Alphabetic letter identification: Effects of perceivability, similarity, and bias”. *Acta psychologica* 139:1 (2012), pp. 19–37.

[149] R. F. Murray and J. M. Gold. “Troubles with bubbles”. *Vision research* 44:5 (2004), pp. 461–470.

[150] A. S. Nandy and B. S. Tjan. “The nature of letter crowding as revealed by first-and second-order classification images”. *Journal of Vision* 7:2 (5 2007), pp. 1–26.

[151] NAS-NRC Committee on Vision. “Recommended stardard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. Committee on vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, D.C”. *Adv Ophthalmol* 41 (1980), pp. 103–148.

[152] D. J. Navarro and T. L. Griffiths. “Latent features in similarity judgments: A nonparametric Bayesian approach”. *Neural computation* 20:11 (2008), pp. 2597–2628.

[153] U. Neisser. *Cognitive psychology.* Appleton-Century-Crofts, 1967.

[154] P. Neri and D. M. Levi. “Spatial resolution for feature binding is impaired in peripheral and amblyopic vision”. *Journal of Neurophysiology* 96:1 (2006), pp. 142–153.

[155] J. Ngiam, Z. Chen, D. Chia, P. W. Koh, Q. V. Le, and A. Y. Ng. “Tiled convolutional neural networks”. *Advances in Neural Information Processing Systems.* 2010, pp. 1279–1287.

[156] R. Nosofsky. “Overall similarity and the identification of separable-dimension stimuli: A choice model analysis”. *Perception & Psychophysics* 38:5 (1985), pp. 415–432.

[157] B. A. O’Brien, J. S. Mansfield, and G. E. Legge. “The effect of contrast on reading speed in dyslexia”. *Vision research* 40:14 (2000), pp. 1921–1935.

[158] T. E. Oliphant. “Python for scientific computing”. *Computing in Science & Engineering* 9:3 (2007), pp. 10–20.

[159] B. A. Olshausen and D. J. Field. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”. *Nature* 381:6583 (1996), pp. 607–609.

[160] R. P. O’Shea and D. R. Williams. “Binocular rivalry with isoluminant stimuli visible only via short-wavelength-sensitive cones”. *Vision research* 36:11 (1996), pp. 1561–1571.

[161] E. Põder. “Spatial-frequency spectra of printed characters and human visual perception”. *Vision Research* 43:14 (June 2003), pp. 1507–1511.

[162] E. Põder and J. Wagemans. “Crowding with conjunctions of simple features.” *Journal of Vision* 7:2 (Jan. 2007), pp. 23.1–12.

[163] D. H. Parish and G. Sperling. “Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination”. *Vision research* 31:7 (1991), pp. 1399–1415.

[164] E. Pascal and R. V. Abadi. “Contour interaction in the presence of congenital nystagmus”. *Vision research* 35:12 (1995), pp. 1785–1789.

[165] J. W. Peirce. “Generating stimuli for neuroscience using PsychoPy”. *Frontiers in neuroinformatics* 2 (2008).

[166] D. G. Pelli, B. Farell, and D. C. Moore. “The remarkable inefficiency of word recognition”. *Nature* 423:6941 (2003), pp. 752–756.

[167] D. G. Pelli, M. Palomares, and N. J. Majaj. “Crowding is unlike ordinary masking: Distinguishing feature integration from detection”. *Journal of Vision* 4:12 (2004), p. 12.

[168] D. G. Pelli, S. Song, and D. M. Levi. “Improving the screening of children for amblyopia”. *Journal of Vision* 11:11 (2011), pp. 411–411.

[169] D. G. Pelli and K. A. Tillman. “The uncrowded window of object recognition”. *Nature Neuroscience* 11:10 (Sept. 2008), pp. 1129–1135.

[170] D. G. Pelli and L. Zhang. “Accurate control of contrast on microcomputer displays”. *Vision research* 31:7 (1991), pp. 1337–1350.

[171] D. G. Pelli, K. A. Tillman, J. Freeman, M. Su, T. D. Berger, and N. J. Majaj. “Crowding and eccentricity determine reading rate”. *Journal of vision* 7:2 (2007), p. 20.

[172] D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore-Page. “Feature detection and letter identification”. *Vision research* 46:28 (2006), pp. 4646–4674.

[173] Y. Petrov and A. V. Popple. “Crowding is directed to the fovea and preserves only feature contrast.” *Journal of Vision* 7:2 (Jan. 2007), pp. 8.1–9.

[174] Y. Petrov, A. V. Popple, and S. P. McKee. “Crowding and surround suppression: Not to be confused”. *Journal of Vision* 7:2 (2007), p. 12.

[175] C. D. Phelps. “Acuity perimetry and glaucoma.” *Transactions of the American Ophthalmological Society* 82 (1984), p. 753.

[176] P. Podgorny and W. R. Garner. “Reaction time as a measure of inter- and intraobject visual similarity : Letters of the alphabet”. *Perception & Psychophysics* 26:1 (1979), pp. 37–52.

[177] J. R. Pomerantz and E. A. Pristach. “Emergent features, attention, and perceptual glue in visual form perception.” *Journal of Experimental Psychology: Human Perception and Performance* 15:4 (1989), p. 635.

[178] J. R. Pomerantz, L. C. Sager, and R. J. Stoever. “Perception of wholes and of their component parts: some configural superiority effects.” *Journal of Experimental Psychology: Human Perception and Performance* 3:3 (1977), p. 422.

[179] J. Portilla and E. P. Simoncelli. “A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients”. *International Journal of Computer Vision* 40:1 (2000), pp. 49–70.

[180] W. Prinzmetal. “Principles of feature integration in visual perception.” *Perception & psychophysics* 30:4 (Oct. 1981), pp. 330–40.

[181] J. Rabin and A. J. Adams. “Visual acuity and contrast sensitivity of the S cone pathway: preliminary measures with letter charts.” *Optometry & Vision Science* 67:11 (1990), pp. 799–802.

[182] E. Rashal and Y. Yeshurun. “Contrast dissimilarity effects on crowding are not simply another case of target saliency”. *Journal of vision* 14:6 (2014), p. 9.

[183] D Regan and D Neima. “Low-contrast letter charts in early diabetic retinopathy, ocular hypertension, glaucoma, and Parkinson’s disease.” *British journal of ophthalmology* 68:12 (1984), pp. 885–889.

[184] D. Regan and D. Neima. “Low-contrast letter charts as a test of visual function”. *Ophthalmology* 90:10 (1983), pp. 1192–1200.

[185] L. N. Reich and H. E. Bedell. “Relative legibility and confusions of letter acuity targets in the peripheral and central retina.” *Optometry and vision science : official publication of the American Academy of Optometry* 77:5 (May 2000), pp. 270–5.

[186] G. M. Reicher. “Perceptual recognition as a function of meaningfulness of stimulus material.” *Journal of experimental psychology* 81:2 (1969), p. 275.

[187] M. Riesenhuber and T. Poggio. “Hierarchical models of object recognition in cortex”. *Nature neuroscience* 2:11 (1999), pp. 1019–1025.

[188] E. A. Rossi and A. Roorda. “The relationship between visual resolution and cone spacing in the human fovea”. *Nature neuroscience* 13:2 (2010), pp. 156–157.

[189] D. E. Rumelhart and P. Siple. “Process of recognizing tachistoscopically presented words.” *Psychological review* 81:2 (1974), p. 99.

[190] W. Ruml. “Constructing distributed representations using additive clustering”. *Advances in Neural Information Processing Systems 14: Proceedings of the 2001 Conference.* Vol. 1. 243. MIT Press. 2002, p. 107.

[191] M. A. Sandberg and E. L. Berson. “Blue and green cone mechanisms in retinitis pigmentosa.” *Investigative ophthalmology & visual science* 16:2 (1977), pp. 149–157.

[192] E. C. Sanford. “The relative legibility of the small letters”. *The American Journal of Psychology* 1:3 (1888), pp. 402–435.

[193] M. E. Schneck, G. Haegerstrom-Portnoy, L. A. Lott, J. A. Brabyn, and G. Gildengorin. “Low contrast vision function predicts subsequent acuity loss in an aged population: the SKI study”. *Vision research* 44:20 (2004), pp. 2317–2325.

[194] O. G. Selfridge. “Pandemonium: a paradigm for learning in Mechanisation of Thought Processes”. *Proceedings of a Symposium Held at the National Physical Laboratory.* London: HMSO, Nov. 1958, pp. 513–526.

[195] O. G. Selfridge and U. Neisser. “Pattern recognition by machine”. *Scientific American* 203 (1960), pp. 60–68.

[196] R. N. Shepard. “Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space”. *Psychometrika* 22:4 (1957), pp. 325–345.

[197] R. N. Shepard. “Stimulus and response generalization: tests of a model relating generalization to distance in psychological space.” *Journal of experimental psychology* 55:6 (June 1958), pp. 509–23.

[198] R. N. Shepard. “The analysis of proximities: Multidimensional scaling with an unknown distance function. II”. *Psychometrika* 27:3 (Sept. 1962), pp. 219–246.

[199] R. N. Shepard and P. Arabie. “Additive clustering: Representation of similarities as combinations of discrete overlapping properties.” *Psychological Review* 86:2 (1979), pp. 87–123.

[200] J. Siderov, S. J. Waugh, and H. E. Bedell. “Foveal contour interaction for low contrast acuity targets”. *Vision research* 77 (2013), pp. 10–13.

[201] A. J. Simmers, L. S. Gray, P. V. McGraw, and B. Winn. “Contour interaction for high and low contrast optotypes in normal and amblyopic observers”. *Ophthalmic and Physiological Optics* 19:3 (1999), pp. 253–260.

[202] L. L. Sloan. “New test charts for the measurement of visual acuity at far and near distances”. *American journal of ophthalmology* 48:6 (1959), pp. 807–813.

[203] J. E. K. Smith. “Alternative biased choice models”. *Mathematical Social Sciences* 23:2 (1992), pp. 199–219.

[204] J. E. K. Smith. “Recognition models evaluated: a commentary on Keren and Baggen.” *Perception & psychophysics* 31:2 (Feb. 1982), pp. 183–9.

[205] J. A. Solomon and D. G. Pelli. “The visual filter mediating letter identification”. *Nature* 369:6479 (1994), pp. 395–397.

[206] S. Song. “Acuity, crowding, feature detection, and fixation in normal and amblyopic vision.” PhD thesis. University of California, Berkeley, 2009.

[207] S. Song, D. M. Levi, and D. G. Pelli. “A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening”. *Journal of Vision* 14:5 (2014), p. 3.

[208] H. Strasburger, L. O. Harvey, and I. Rentschler. “Contrast thresholds for identification of numeric characters in direct and eccentric view”. *Perception & Psychophysics* 49:6 (1991), pp. 495–508.

[209] H. Strasburger and M. Malania. “Source confusion is a major cause of crowding”. *Journal of vision* 13:1 (2013), p. 24.

[210] H. Strasburger, I. Rentschler, and M. Jüttner. “Peripheral vision and pattern recognition: A review”. *Journal of Vision* 11:5 (2011), p. 13.

[211] P. Sumner, T. Adamjee, and J. Mollon. “Signals invisible to the collicular and magnocellular pathways can capture visual attention”. *Current Biology* 12:15 (2002), pp. 1312–1316.

[212] W. H. Swanson. "Short wavelength sensitive cone acuity: individual differences and clinical use". *Applied Optics* 28:6 (1989), pp. 1151–1157.

[213] E. Takahashi. "Effects of flanking contours on visual resolution at foveal and near-foveal loci." PhD thesis. University of California, Berkeley, 1968.

[214] J. B. Tenenbaum. "Learning the Structure of Similarity". *Advances in Neural Information Processing Systems.* 1996, pp. 3–9.

[215] L. N. Thibos, D. L. Still, and A. Bradley. "Characterization of spatial aliasing and contrast sensitivity in peripheral vision". *Vision research* 36:2 (1996), pp. 249–258.

[216] M. A. Tinker. "The relative legibility of the letters, the digits, and of certain mathematical signs". *The Journal of General Psychology* 1:3-4 (1928), pp. 472–496.

[217] A. Toet and D. M. Levi. "The two-dimensional shape of spatial interaction zones in the parafovea". *Vision research* 32:7 (1992), pp. 1349–1357.

[218] J. T. Townsend, G. G. Hu, and R. J. Evans. "Modeling feature perception in brief displays with evidence for positive interdependencies." *Perception & psychophysics* 36:1 (July 1984), pp. 35–49.

[219] J. T. Townsend, G. G. Hu, and F. G. Ashby. "A test of visual feature sampling independence with orthogonal straight lines". *Bulletin of the Psychonomic Society* 15:3 (1980), pp. 163–166.

[220] J. T. Townsend, G. G. Hu, and F. G. Ashby. "Perceptual sampling of orthogonal straight line features". *Psychological Research* 43:3 (1981), pp. 259–275.

[221] J. T. Townsend, G. G. Hu, and H. Kadlec. "Feature sensitivity, bias, and interdependencies as a function of energy and payoffs." *Perception & psychophysics* 43:6 (June 1988), pp. 575–91.

[222] J. T. Townsend. "Theoretical analysis of an alphabetic confusion matrix". *Perception & Psychophysics* 9:1 (1971), pp. 40–50.

[223] J. Townsend. "Alphabetic confusion: A test of models for individuals". *Perception & Psychophysics* 9:6 (1971), pp. 449–454.

[224] A. Treisman and R. Paterson. "Emergent features, attention, and object perception." *Journal of Experimental Psychology: Human Perception and Performance* 10:1 (1984), p. 12.

[225] A. M. Treisman and G. Gelade. "A feature-integration theory of attention". *Cognitive psychology* 12:1 (1980), pp. 97–136.

[226] A. M. Treisman and H. Schmidt. "Illusory conjunctions in the perception of objects". *Cognitive psychology* 14:1 (1982), pp. 107–141.

[227] S. P. Tripathy and P. Cavanagh. "The extent of crowding in peripheral vision does not scale with target size". *Vision research* 42:20 (2002), pp. 2357–2369.

[228] S. P. Tripathy and D. M. Levi. “Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot”. *Vision research* 34:9 (1994), pp. 1127–1138.

[229] A. Tversky. “Features of similarity.” *Psychological Review* 84:4 (1977), p. 327.

[230] A. Tversky and I. Gati. “Similarity, separability, and the triangle inequality.” *Psychological review* 89:2 (1982), p. 123.

[231] C. W. Tyler and L. T. Likova. “Crowding: a neuroanalytic approach.” *Journal of Vision* 7:2 (Jan. 2007), pp. 16.1–9.

[232] W. R. Uttal. “Masking of alphabetic character recognition by dynamic visual noise (DVN)”. *Perception & Psychophysics* 6:2 (1969), pp. 121–128.

[233] F. L. van Nes and J. C. Jacobs. “The effect of contrast on letter and word recognition”. *IPO Annual Progress Report*. Vol. 16. 1981, pp. 72–80.

[234] A. H. van der Heijden, M. S. Malhas, and B. Van Den Roovaart. “An empirical interletter confusion matrix for continuous-line capitals”. *Attention, Perception, & Psychophysics* 35:1 (1984), pp. 85–88.

[235] V. Virsu, P. Lehtio, and J. Rovamo. “Contrast sensitivity in normal and pathological vision.” *Docum Opthal Proc Ser* 39 (1981), pp. 263–272.

[236] V. Virsu and J. Rovamo. “Visual resolution, contrast sensitivity, and the cortical magnification factor”. *Experimental Brain Research* 37:3 (1979), pp. 475–494.

[237] V. Virsu, R. Näsänen, and K. Osmoviita. “Cortical magnification and peripheral vision”. *JOSA A* 4:8 (1987), pp. 1568–1578.

[238] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt. “A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization.” *Psychological bulletin* 138:6 (2012), p. 1172.

[239] J. Wandmacher. “Multicomponent theory of perception”. *Psychological Research* 39:1 (1976), pp. 17–37.

[240] J. Wandmacher. “S-multiplicativity of a stochastic matrix and applications to visual identification”. *Journal of Mathematical Psychology* 16:3 (Dec. 1977), pp. 219–232.

[241] H. Wang, X. He, and G. E. Legge. “Effect of pattern complexity on the visual span for Chinese and alphabet characters”. *Journal of vision* 14:8 (2014), p. 6.

[242] A. B. Watson and A. J. Ahumada. “Modeling acuity for optotypes varying in complexity”. *Journal of Vision* 12:10 (2012), p. 19.

[243] A. B. Watson and A. E. Fitzhugh. “Modelling character legibility”. *SID Digest* 10 (1989), pp. 360–363.

[244] T. Wertheim. “Peripheral visual acuity: Th. Wertheim.” *American journal of optometry and physiological optics* 57:12 (1980), pp. 915–924.

[245] G. Westheimer. “Scaling of visual acuity measurements”. *Archives of ophthalmology* 97:2 (1979), pp. 327–330.

[246] G. Westheimer. “The spatial grain of the perifoveal visual field”. *Vision research* 22:1 (1982), pp. 157–162.

[247] G. Westheimer. “The spatial sense of the eye. Proctor lecture.” *Investigative Ophthalmology & Visual Science* 18:9 (1979), pp. 893–912.

[248] F. W. Weymouth. “Visual sensory units and the minimal angle of resolution”. *American journal of ophthalmology* 46:1 (1958), pp. 102–113.

[249] D. Whitney and D. M. Levi. “Visual crowding: a fundamental limit on conscious perception and object recognition.” *Trends in cognitive sciences* (Mar. 2011), pp. 1–9.

[250] T. D. Wickens. *Multiway contingency table analysis for the social sciences.* Lawrence Erlbaum Associates, Inc, 1989.

[251] R. L. Woods, S. J. Tregear, and R. A. Mitchell. “Screening for ophthalmic disease in older subjects using visual acuity and contrast sensitivity”. *Ophthalmology* 105:12 (1998), pp. 2318–2326.

[252] B. Xiao and A. R. Wade. “Measurements of long-range suppression in human opponent S-cone and achromatic luminance channels”. *Journal of vision* 10:13 (2010), p. 10.

[253] J.-Y. Zhang, T. Zhang, F. Xue, L. Liu, and C. Yu. “Legibility of Chinese characters in peripheral vision and the top-down influences on crowding.” *Vision research* 49:1 (Jan. 2009), pp. 44–53.

[254] J.-Y. Zhang, G.-L. Zhang, L. Liu, and C. Yu. “Whole report uncovers correctly identified but incorrectly placed target information under visual crowding”. *Journal of vision* 12:7 (2012), p. 5.