Evaluation of MILSTD 2525 Glyph Features in a Visual Search Paradigm

Navaneethan Siva M.A., Alex Chaparro, Ph.D., and Evan Palmer, Ph.D.
Department of Psychology, Wichita State University

MILSTD 2525 is a document that outlines the composition and use of a set of standardized symbology by the US Department of Defense to represent vehicles, equipment and personnel on tactical interfaces. These symbols are primarily multivariate glyphs that visualize the status of military units. This study selected a subset of commonly used glyph features in order to investigate their relative efficiency in a search paradigm. Performance across the different features as well as within levels of each feature was examined. Stimuli were tested using an oddball search paradigm with set sizes of 6, 12 and 18. The dependent variables of interest were search efficiency, RT, and accuracy. Results show that search asymmetries occur with MILSTD 2525 glyph features and that these features differ in search efficiency from one another. The authors discuss the relative search performance of these features and implications for glyph design.

INTRODUCTION

Human perception and attention are resource-limited systems (Wickens & Hollands, 2000;Anderson, 2009;Miller, 1956) and these capacity limits can lead to operator error. Nonetheless, many jobs require successful, consistent and wide application of attention for sustained periods to monitor and control complex systems. For example, pilots of unmanned aerial vehicles (UAVs) are external to the aircraft they control and are therefore completely reliant on their interface to perform tasks. Due to the complexity of these systems, the variety of sensor data and the growing expectation for simultaneous control of multiple UAVs, many of the pilot's tasks require sustained and diffuse attention.

One solution to mitigating these capacity limitations is to provide information in a way that optimizes data transfer. A popular choice has been to visualize the data. Since people are unable to effectively process large amounts of raw data in a numerical format, the information is often transformed into a graphic representation. There are a variety of ways in which this can be done, the most familiar being data plotted on a Cartesian plane by frequency, chronology or some other logical dimension based on the data's context. Graphical data visualizations are thought to leverage the highest bandwidth channel from digital information to the brain through the human visual system (Ware, 2004). If done well, visual information can be gathered particularly efficiently and even allow the detection of emergent properties that would not have otherwise been obvious. However, some sets of information contain more variables than can be presented comprehensibly in such a format. These multivariate data are instead visualized in a number of different ways, such as in a '*glyph*'. Glyphs are visually distinct graphical entities that represent a number of data variables via their physical attributes (Ward, 2002).

There are a variety of different glyph designs, although most work under the same concept. Siva, Palmer, & Chaparro (2012) reviewed the available literature on glyph perception and defined an agenda for further investigation to optimize their design. That work identified four categories of glyph-based tasks that pertain to interfaces: (i) searching for a particular glyph or glyph feature, (ii) comparing glyphs, (iii) holistically perceiving glyph displays to ascertain global patterns, and (iv) tracking dynamically changing glyph and glyph features over time. One of the major tasks people engage in with glyphs is visual search, to locate and identify a particular piece of information (as in i) as well as to compare one piece of data to another (as in ii). Glyphs generally display the same data variables about several units but not all of these data variables are equally important. Prior research conducted by Siva, Chaparro, Nguyen, & Palmer (2013) indicates that not all features of glyphs guide search equally well. Therefore, a danger exists in mapping critical information variables onto physical features of a glyph that are hard to identify (e.g. a numeric representation of altitude). The efficiency of visual search for a particular feature in a glyph can impact the overall efficiency of the glyph's design. The Siva et al (2013) study focused on Chernoff faces, which map data onto various features of schematic faces. The work indicated that faces were not particularly effective as glyphs, casting doubt on the underlying motivation of this approach. In this study, the stimuli are based on MILSTD 2525 glyphs as they take advantage of much more generalizable physical features (i.e. color, shape, size, orientation etc.). MILSTD 2525 is a document that outlines the composition and use of a set of standardized symbology by the US Department of Defense to represent vehicles, equipment and personnel on tactical interfaces (Department of Defense, 2008).

There are a number of basic principles from the visual search literature that can inform research with glyphs. First, there is research suggesting that different physical properties support varying levels of search efficiency (Wolfe & Horowitz, 2004). For instance, searching for a red square among green squares is fast and efficient, while searching for a digital 2 among digital 5s is slow and inefficient (e.g., Palmer, Horowitz, Torralba, & Wolfe, 2011). Based on the average composition of glyphs, and specifically that of MILSTD 2525 glyphs, some separation in search efficiency should exist for different features. Another well-established phenomenon are search asymmetries, where it is easier to find feature A among feature B than it is to find B among A (Treisman & Gormican, 1988). Consider Figure 1, which demonstrates this principle. An object with a unique feature among a group lacking that feature are easy to find but the reverse case is not true. Similarly, as seen in Figure 2, objects with more of a given feature will be easier to search for among a group with less of that feature than vice versa (Wolfe, 2001). This is important because each feature in a glyph can vary physically relative to the magnitude of the data it represents. Thus, if one unit has less of a feature than the units around it, search to find this unit will be difficult. This could lead to trouble in cases where the data is coded so that lower magnitudes of a feature signify a critical situation (e.g., dangerously slow air speeds).

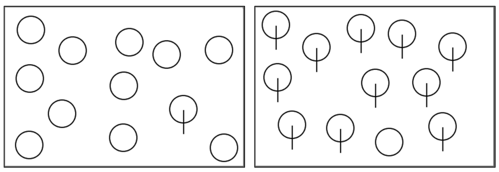

Figure 1. It's easier to search for a Q among O's than it is to search for an O among Q's.

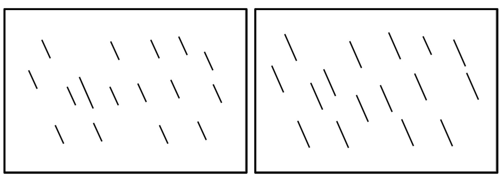

Figure 2. It's easier to search for short among long than it is to search for long among short.

Since most glyphs, including MILSTD 2525 glyphs, are made up of several alterable combinations of simple features, it is reasonable to expect that each of these glyph features will yield varying levels of search efficiency. In fact, independent of glyphs, many of these features have been tested using a visual search paradigm as described above (e.g. colors, geometric shapes, line segments). In this study the authors investigate whether glyphs, specialized multi-feature objects, also share these principles.

MILSTD 2525 glyphs are made up of a set of defined features that can be optionally added to a base icon. Figure 3 demonstrates the layout of some of the possible features on a glyph. The color fill is used to indicate a unit's *standard identity*, or allegiance. For example, red for hostile, blue for friendly and green for neutral. The frame is double coded and can refer to both a unit's standard identity and *battle dimension* (e.g., air, land, sea). For example, a pointed frame defines a hostile unit and an open frame on the bottom, as shown in Figure 3, defines an air unit. Several features were selected for evaluation in the study and all others were held constant.

The goal of this study was to investigate whether disparities exist in terms of visual search efficiency and performance between the chosen MILSTD 2525 glyph features as well as between the different levels of each feature. Based on research from the visual search literature, if search asymmetries exist in glyphs, stimuli with lower magnitudes should be difficult to locate among stimuli with moderate magnitudes while stimuli with higher magnitudes should be easier to locate. Also, based on the difference in physical properties of the chosen features, there should be a noticeable difference in search performance between them.

METHODS

Participants. A sample of 32 undergraduate psychology students from Wichita State University, ages 18 - 35 (M = 19.94, SD = 3.14) participated in the study. Eight participants did not complete the study in the allotted period and their data was not included in the analysis. We extended the study duration to account for this issue. All gave informed consent and were remunerated with course credit for their participation.

Materials. MILSTD 2525 glyphs were created using Adobe Illustrator, mirroring images from the military standards documentation For the study, *standard identity* and *battle dimension*, reflected by the overall fill and frame of the glyph, were held constant as a friendly (blue fill) air unit (frame with curved top and an open base). Four features were manipulated for the study. As seen in Figure 4, the direction indicator is a line that identifies both air speed (through its relative length) and direction (through its angle) of the unit. The icon in the centre represents the type of air unit and the text field provides additional information about the unit.

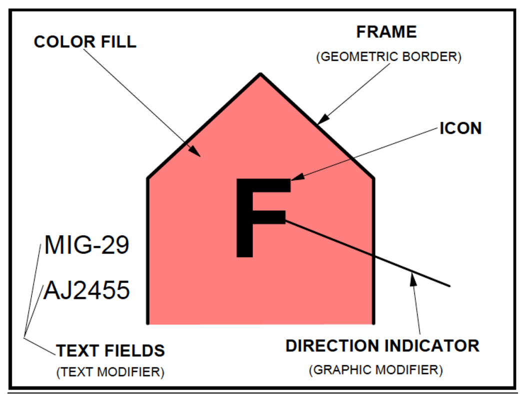

Figure 3. Illustration of the basic features of a MILSTD 2525 glyph from the DOD Interfacs Standards Document (Department of Defense, 2008).

A default setting for each of the attributes was selected to be used as a distractor and each was then manipulated to represent two levels of distinctiveness: minor and major. In total, this made eight target stimuli: four features broken into two levels of distinctiveness. Speed was split into a slower speed (a short line) and a faster speed (a longer line), direction was split into a shallow turn (a 20 degree counter clockwise rotation) and sharp turn (a 90 degree clockwise turn), the text identifier was broken into an easy (alphabetic character change) and difficult (numerical character change), and unit type was split into a rotary air (similar position and size icon) and rotary drone (split icon with different sizing). These manipulations were meant to highlight how physical variation in these glyphs can be minor (e.g. slow speed, shallow turn, rotary air, easy identifier) or major (fast, sharp turn, rotary drone, difficult identifier). The two text identifier stimuli were omitted from the analysis (see below). Figure 4 illustrates each of the pairs for speed, direction and unit type. When presented on the testing computers at a viewing distance of 45 cm, a 6 x 6 stimulus array subtended a 39 deg x 33 deg square centered on the screen. Each item fit within a 4 deg x 4 deg area and was displayed as seen in Figure 4 against a white screen background. Each presentation introduced jitter to every stimulus location. Target position in the grid was randomized and target type was randomly selected without replacement from a trial list.

Design and Procedure. The experiment was an "oddball" search in which participants were instructed to determine which side of the screen a target glyph was shown among a group of neutral distractors by pressing a "left side" key, a "right side" key or a "no target present" key. Before the start of the experiment, participants were shown all the possible target glyphs. Upon a correct response from the participant, the target was highlighted green and a "correct" tone was played. Subsequent to an incorrect response from the participant, the target was highlighted red and an "incorrect" tone was played. If a participant indicated seeing a target on a non-target trial, a message was presented in red stating "Incorrect: Target is absent" and if they correctly identify a non target trial the message read "Correct: Target is absent". To prevent participant fatigue, participants were offered breaks at the end of each block. Additionally, experiment-running times were kept to seventy-five minutes or less.

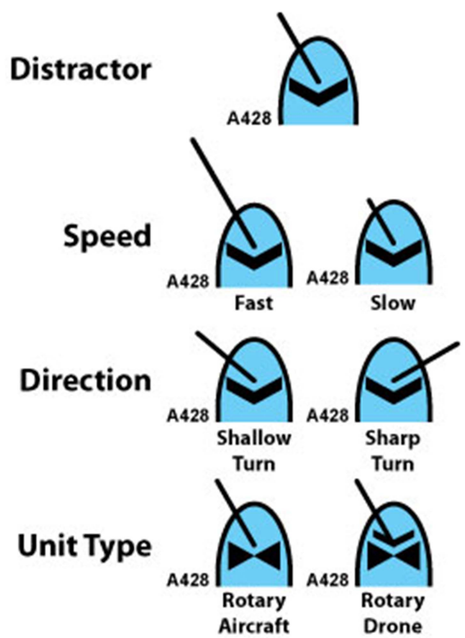

Figure 4. The six target glyphs and the neutral distractor. The six target glyphs represent small or large variation for the variables being represented.

The experiment consisted of a practice block and 10 experimental blocks. Each block contained 81 trials: 3 presentations of each target stimulus (3 x 8 = 24) and 3 presentations of a non-target scenario (24 + 3 = 27) at set sizes of 6, 12 and 18 (27 x 3 = 81). This resulted in a total of 891 trials consisting of 30 presentations of each stimulus at each set size. Trials were randomly intermixed during each block. At the end of the experiment, participants were debriefed and asked what strategies they used during the experiment and whether they found any of the features to be particularly easy or difficult to detect.

Dependent Measures and Data Analysis. Response times (RTs) and proportion correct data were collected. Our primary data analyses examined the slopes of the function relating response time to set size as a measure of search efficiency (Wolfe, 1998). This was done to determine whether search efficiency differed across the different features. In addition, similar analyses were done on response time and proportion correct data to determine if differences existed in these factors for search performance. Unfortunately, the text field feature needed to be filtered from the analysis because the majority of participants were not aware of its presence as a target, a variable unaccounted for in the studies design. Homogeneity of variance was tested and unless explicitly declared there was a non-significant result.

RESULTS

Feature and Distinctiveness in Search Efficiency. Search efficiency slopes were calculated using RT and set size data and then submitted to a 3 x 2 (Feature x Distinctiveness) repeated measures design analysis of variance (ANOVA). Mauchly's test for sphericity was significant, $\chi^2(2) = 16.49$, $p < .001$ for feature, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .67$). The analysis detected a main effect of feature, $F(1.31, 30.12) = 5.50$, $p < .05$, $\eta_p^2 = .19$. Bonferroni-corrected post-hoc tests showed that speed (93.15 ms/item) had a significantly higher slope than unit type (47.93 ms/item) and was therefore less efficient. The speed feature did not have a significantly greater slope than the direction feature (57.28 ms/item), however the relationship approached significance ($p = .06$). There was also a significant effect of distinctiveness, $F(1, 23) = 20.53$, $p < .001$, $\eta_p^2 = .47$. Bonferroni-corrected post-hoc tests showed that minor distinctions (91.16 ms/item) had significantly higher search efficiency slopes than major distinctions (41.08 ms/item), meaning that search for smaller distinctions was less efficient than major distinctions. There were no other significant effects. These data are plotted in Figure 5.

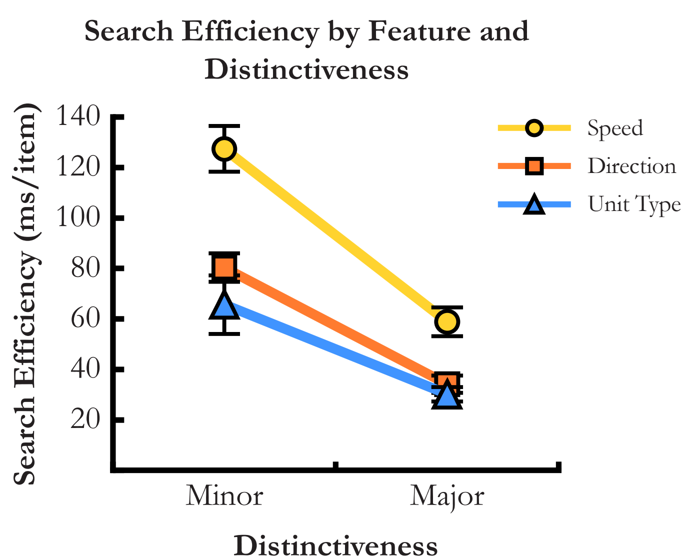

Figure 5. Search efficiency by feature and distinctiveness. The analyses showed speed was significantly different from unit type but not direction. All distinctiveness differences were significant except for rotary aircraft and rotary drone. Error bars show standard error of the mean.

Feature and Distinctiveness in Response Time. RT data were submitted to a 3 x 2 x 3 (Feature x Distinctiveness x Set

Size) repeated measures design analysis of variance (ANOVA). Mauchly's test for sphericity was significant, for feature $\chi^2(2) = 16.83$, $p < .001$, for set size $\chi^2(2) = 10.02$, as well as for each of the corresponding interactions (all $p < .05$) except feature by distinctiveness ($p = .1$). Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (feature, $\varepsilon = .67$; set size, $\varepsilon = .73$). The analysis detected a main effect of feature, $F(1.30, 29.97) = 50.40$, $p < .001$, $\eta_p^2 = .69$. Bonferroni-corrected post-hoc tests showed that all relationships were significant between the three features with unit type (1.41 s) having the quickest average RT, followed by direction (1.82 s) and lastly speed (2.31 s) with the slowest RT (all $p < .001$). There was also a main effect of distinctiveness, $F(1, 23) = 144.61$, $p < .001$, $\eta_p^2 = .86$ and a main effect of set size, $F(1.46, 33.68) = 92.25$, $p < .001$, $\eta_p^2 = .80$. Bonferroni-corrected post-hoc tests confirmed the same results as the previous analysis for distinctiveness with minor distinctions (2.30 s) being slower to identify than major distinctions (1.39 s) and the expected result for set size, with each successively larger set size having a significantly slower RT (Size 6: 14.56 s, Size 12: 18.25 s, Size 18: 22.49 s; all $p < .001$). There were several two way interactions: feature by distinctiveness $F(2, 46) = 14.93$, $p < .001$, $\eta_p^2 = .39$, feature by set size $F(1.67, 38.51) = 4.00$, $p < .05$, $\eta_p^2 = .15$, and distinctiveness by set size, $F(1.40, 32.14) = 16.34$, $p < .001$, $\eta_p^2 = .42$. This indicates that the RTs to find some features were modulated by distinctiveness and set size more than others. There were no other interactions or effects. A set of post hoc t-tests were run to explore the feature by distinctiveness interaction. They found that the only non-significant component of the interaction was between the minor level of unit type (rotary air) and the major type of speed (fast), all other relationships were significant ($p < .05$). These data are plotted in Figure 6.

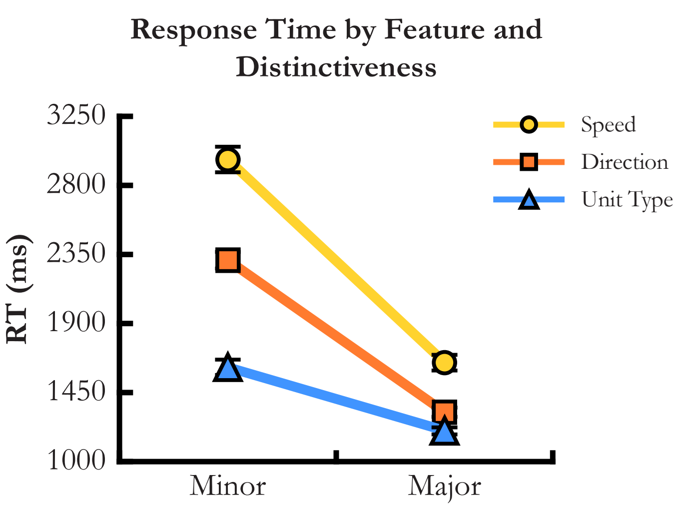

Figure 6. Response time by feature and distinctiveness. The analyses indicate significant differences between all points except fast and rotary air. Error bars show standard error of the mean.

Feature and Distinctiveness in Proportion Correct. Accuracy data were submitted to a 3 x 2 x 3 (Feature x Distinctiveness x Set Size) repeated measures design ANOVA. Mauchly's test for sphericity was significant, for feature $\chi^2(2) = 7.04$, $p < .05$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity for feature ($\varepsilon = .79$). The analysis detected a main effect of feature, $F(1.57, 36.11) = 78.25$, $p < .001$, $\eta_p^2 = .77$. Bonferroni-corrected post-hoc tests showed that all relationships were significant between the three features with unit type (95.4%) having the greatest proportion correct, followed by direction (88.3%) and lastly speed (79.5%) with the least proportion correct of the set (all $p < .001$). There was also a main effect of distinctiveness, $F(1, 23) = 102.62$, $p < .001$, $\eta_p^2 = .82$ and a main effect of set size, $F(2, 46) = 45.18$, $p < .001$, $\eta_p^2 = .66$. Bonferroni-corrected post-hoc tests confirmed similar results as the previous analyses for distinctiveness, with minor distinctions having less accuracy (81.5%) than major distinctions (93.9%). Similarly with set size, post hoc tests found that with each successively larger set size there was significantly lower proportion correct (Size 6: 91.0%, Size 12: 87.3%, Size 18: 84.9%; all $p < .001$). There were several two way interactions: feature by distinctiveness $F(2, 46) = 53.81$, $p < .001$, $\eta_p^2 = .70$, feature by set size $F(4, 92) = 12.05$, $p < .001$, $\eta_p^2 = .34$ and distinctiveness by set size, $F(2, 46) = 24.11$, $p < .001$, $\eta_p^2 = .51$. Accuracy of identification varied as a function of feature distinctiveness and set size. Finally, there was a three-way interaction of feature by distinctiveness by set size, $F(4, 92) = 4.57$, $p < .01$, $\eta_p^2 = .17$. A set of post hoc t-test was run to explore the feature by distinctiveness interaction. They found that the only non-significant component of the interaction was between the minor level of unit type (rotary air) and the major type of direction (sharp turn), all other relationships were significant ($p < .05$). These data are depicted in Figure 7.

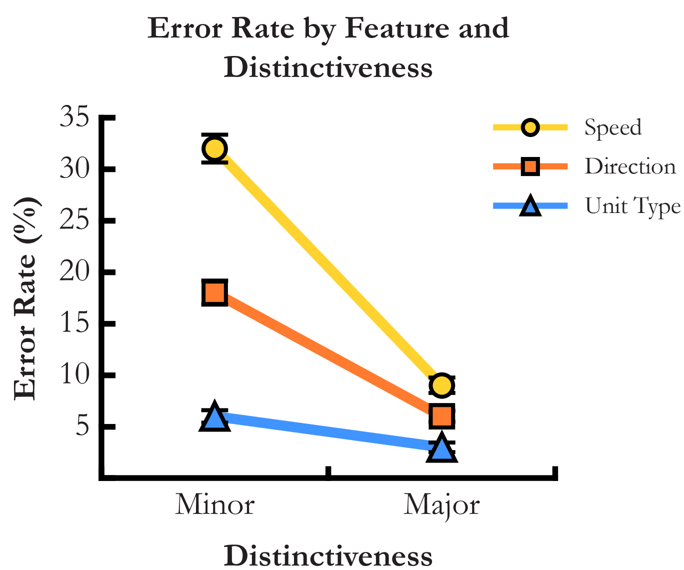

Figure 7. Error rate by feature and distinctiveness. Analyses indicate significant differences between all points except sharp turn and rotary aircraft. Error bars show standard error of the mean.

DISCUSSION

The goal of this study was to 1) determine whether search asymmetries occur for MILSTD 2525 glyphs and 2) see if the features themselves show different levels of search efficiency and performance based on their properties. The results indicate that there are indeed search asymmetries at work and that each of the selected features differs significantly from one another.

There was a significant difference in search efficiency between the speed feature and unit type feature. The difference between the speed and direction features approached significance (p = .06) but generally the features for direction and unit type had statistically indistinguishable efficiencies. The RT data provides a clearer picture with a distinct separation of the three features; speed (2.31 s) is the slowest search feature with unit type (1.41 s) being the fastest and direction (1.82 s) falling in between. This relationship can be seen in Figure 6. The proportion correct data also reinforces this relationship with unit type (95.4%) being the most accurately identified feature and speed (79.5%) being the least accurately identified feature with direction falling between the two (88.3%). Efficiency, RT and proportion correct also indicate that the minor and major distinctions were clearly different from one another with the minor changes being less efficient (91.16 ms/item vs. 41.08 ms/item), slower to be detected (2.3 s vs. 1.39 s) and harder to accurately identify (81.5% vs. 93.9%), as predicted.

Together these two main effects suggest a general separation across the three different glyph features studied as well as across the two levels of range explored for each feature. Analysis of the interaction of these factors with both the RT and performance correct measures revealed that, for the most part, each combination of glyph feature and level were distinct in terms of how effectively they guided search. The general order with RT seems to be, from most to least efficient: 1) rotary drone, 2) sharp turn, 3) fast and rotary air, 4) shallow turn, 5) slow. In terms of search accuracy, the relationships are almost identical: 1) rotary drone, 2) sharp turn and rotary air, 3) fast, 4) shallow turn, 5) slow. It is important to note that, although significant, some of these differences are larger than others. Specifically, slow and shallow turn are both much less effective in terms of RT and performance. The rest of the features seem to group near the same relative magnitude of performance.

Treisman & Gormican (1988) demonstrated that short lines among longer lines are difficult to see and that the relative orientation of nearby lines affects the visibility of a target line. The results of this study support those findings and extend them to MILSTD 2525 glyphs. When the slow feature was seen alongside distractors with relatively longer lines and the fast stimulus appeared among distractors that had relatively shorter lines, we observed such a search asymmetry. Similarly, the shallow turn appeared among relatively similar neighbors and the sharp turn was highly distinct, running perpendicular to the distractors' feature. Another consideration is that the line has two spatial codings (size and orientation), while unit type only has one (shape). Since conjunction searches are generally less efficient than single feature searches this may explain the reason why the shape-based feature turned out to be so efficient.

Glyphs are meant to visualize large chunks of information in relatively small areas while allowing an observer to easily see and interpret this information. The more difficult a feature is to spot and the more effort it takes to locate, the more inefficient the glyph becomes as a whole. Our study indicates that there are a variety of differences in visual search for MILSTD 2525 glyph features across and within a small subset of features. Additionally, there seem to be underlying relationships between these features that if stable could prove to be a valuable base for a set of guidelines in feature selection for glyphs. We conclude that given sufficient research into combinations of features with a large enough data set, it will be possible to identify an underlying ordinal relationship between ease and efficiency of search for glyph attributes and thereby construct a guideline toward the selection of such attributes in glyph design.

Limitations and Future Work. Due to insufficient prompting of each target type, many participants did not notice the text identifier feature. As a result search data for this feature was too poor to be used in the analysis.

In applied scenarios, cases where distractors are consistently heterogeneous would be rare. It will be interesting to investigate whether the relationships between features remains stable with the introduction of variability in competing features. If the relationship is modulated by the presence of other features, will it do so in a predictable manner? An additional question that this research raises is the shape of this relationship, it is apparent from the data that not all levels of a single feature are equally distinguishable across glyphs. This work is meant as a starting point for the systematic evaluation of features commonly used in glyphs to help establish a set of guidelines for the design and use of glyphs.

REFERENCES

Anderson, J. R. (2009). *Cognitive psychology and its implications.* (7th Ed.). New York, NY: Worth Publishers.

Department of Defense. (2008). *MIL-STD-2525C: Common Warfighting Symbology*.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*(2), 81–97.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71.

Siva, N., Chaparro, A., Nguyen, D., & Palmer, E. M. (2013). Visual search efficiency for features in Chernoff faces. *Journal of Vision*, *13*(9).

Siva, N., Palmer, E.M., & Chaparro, A. (2012). Human factors principles underlying glyph design: A review of the literature and an agenda for future research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1659–1663.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Ward, M. (2002). A taxonomy of glyph placement strategies for multidimentional data visualization. *Information Visualization*, *1*(3), 194–210.

Ware, C. (2004). *Information visualization: Perception for design* (2nd Ed.). San Francisco, CA: Morgan Kaufmann.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance*, 3rd Ed., 17–118. Prentice-Hall Inc.

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

Wolfe, J. M. (2001). Symposium on asymmetries in visual search: An introduction. *Perception & Psychophysics*, *63*(3), 381–389.

Wolfe, J. M., & Horowitz, T. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501.