Review paper

UDK: 655.261
655.262

Legibility based on differentiation of characters: A review of empirical findings fundamental for the type design practice

ABSTRACT

The purpose of the study was to analyse findings in the field of the pattern recognition, visual word recognition and legibility ofgraphemes. The paper focuses on the most significant findings for comprehending the function of the graphemic characteristic during word and letter recognition. The aim of the study, as it is presented in the discussion part, was to analyse the factors which have induced inconsistencies between the findings acquired by confusion matrix, as well as to compare findings which continually lead type scholars to progression.

KEY WORDS

letter recognition, feature distinctiveness, typeface legibility

Uroš Nedeljković, Irma Puškarević, Bojan Banjanin, Ivan Pinćjer

Faculty of Technical Sciences, Graphic Engineering and Design, Novi Sad

Corresponding author:
Uroš Nedeljković
e-mail: urosned@uns.ac.rs

First recieved: 01.04.2013.
Accepted: 29.05.2013.

Introduction

While the subject of weather the serifs contributed to readability was debated among the circle of theoreticians and practitioners, not even the numerous empirical research could provide a precise answer. The results of many studies have pointed out that there is no difference between reading the serif or sans serif typefaces (eg (Paterson & Tinker, 1932; Poulton, 1965; De Lange, Esterhuizen, & Beatty, 1993). The results provided by certain researchers could not be considered externally valid, which some of the researchers themselves concluded (Tinker, 1963; Zachrisson, 1965 prema Lund, 1999) since they noticed great differences in readability within the group of either serif or sans serif typefaces. Ole Land says that the presence or absence of a serif could be an influential factor, but for the process of reading, a completely ephemeral for measurement (Lund, 1999). A myriad of other factors have been noticed as more significant for both readability and legibility, such as: sizes of the typefaces, line width, tracking, paragraph uniformity and the relations of the text color–the background (Paterson & Tinker, 1944; Tinker & Paterson, 1946); x-height, stroke width, counterform (Paterson & Tinker, 1932; Cheetham, Poulton, & Grimbly, 1965; Poulton, 1965; Poulton, 1972).

After a series of empirical findings from the first half of the last century, and a stream of empirical findings in the field of legibility and readability of typefaces, little research has sustained a grounding theory. Since the majority of psychologists do not possess sufficient knowledge of typography, their studies lacked internal validity (Lund, 1999), whereas typography scholars have published numerous ungrounded work in their attempt to base the discourse stylistically and ideologically. Seen from our hindsight, providing the theoretical framework for this field was less probable without the cognitive information processing theory. Numerous research regarding word superiority effect support different hypothesis (Cattell, 1886; Reicher, 1969; Wheeler, 1970; McClelland & Johnston, 1977) but only with the advancement of the eye tracking and the neural network modeling in the '80s scholars were able to obtain precise answers to the question: How we recognize words? (Rayner & Pollatsek, 1989; McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989). The findings that the bottom-up processing in parallel distribution (PDP) starts from feature level indicates that if we want to progress the legibility of the typeface we need to explore the feature level ie. feature distinctiveness. Therefore, this paper analyzes key findings on legibility of letters within type which have been perceived as meaningful for designers and type practitioner.

Letter recognition

In the recent decades, theories have been developed which show that the letters construction the words are used for the recognition of the word. However, cognitive psychologists also assume different approaches for letter recognition that are actually used for research of the pattern recognition. Two main approaches to the problem of letter recognition (or pattern) are the template-matching model and the feature detection model. These two approaches are in accordance with the theory that the mental representation has an array of common characteristics with the object, therefore that the recognition of the pattern rests upon those common features. Still, the differences are displayed within the operationalization or formalization of the process of recognition (Kostić, 2010).

The first approach is derived from the assumption that the representation native within the long term memory matches a real world pattern i.e. that it has the same structural features like the object of observation-recognition (Selfridge & Neisser, 1960; as cited in Kostić, 2010).

However, the chief problem this theory faces is explaining the wide variations of shape, or in this case the letters contained within various typefaces and the handwriting we are able to recognize effortlessly. According to this theory, that would imply that the brain has a special template for the handwritten or the calligraphic letter 'A' in its every shape, as well as the multiple variations of the simple serif or sans serif letters 'A'. That would mean that each new sample is studied and stored every time a letter or a pattern significantly differs from the existing templates [1]. The main deficiency of the template-matching model is stated by Neisser, is that the matching process requires a previous stimulus normalization (adjusting the stimulus to the prototypical position, size and orientation), and the suggestions of such a process in most cases do not have a psychological or neurophysiological validity (Neisser, 1966; as cited in Grainger, et al. 2008). Despite the arguments against this theory as an alternative the model is favored by a few researchers. Larsen i Bundesen (1996) have developed a program called Template-matching pandemonium that reconciled the differences between the two approaches recognizing different handwritten numerals with a high degree of precision.

A decade ago, a consensus has been reached in favor of a more liberally interpreted analogy between the pattern and its representation, therefore, the recent empirical research are based on the structural approach of the feature detection theory. Instead of observing the entire pattern, the essence of feature detection is in the assumption that the brain decodes distinctive features that separate one pattern from the other. Type designers, graphic designers and other practitioners are familiar with the notion that all of the glyphs of the alphabet are defined according to this limited number of strokes (the vertical, the horizontal, the conjoining, the diagonal, the curved). The strokes that compose them represent their distinctive features.

The computational model poses the central hypothesis that the letters are recognized according to their features was devised by Oliver Selfridge, in 1959 under the name Pandemonium [2] (Selfridge, 1959). It was one of the first computer models for pattern recognition that represented a few levels of parallel processing. Though it wasn't perfect, the program was very influential for the development of other computationalmodels and the development of artificial intelligence.

The support for the Feature detection theory came from neurophysiological studies. Hubel and Wiesel,

[1] Explaining the deficiencies of this approach professor Kostić states the following: "The approach that is based on template-matching faces a problem of infinite regression of the representation of the presented object."

[2] The Pandemonium model can process patterns in four levels. Every level of findings procession obtains detectors of different graphemic patern. Selfridge metaphorically calls these levels ''the demons''.The first image demon receives sensory stimuli. The processing starts at feature demons level, the activation is produced at the cognitive demons level that ''scream'' when they receive certain feature combinations. The assemblies with the largest number of activations received from the feature detection level, and also make cognitive demons scream the loudest, generate up to the decision demon level. The decision making level is comprised of memorised assembled strokes that make up letter signs and some other signs.

[3] Neisser determined capital letter 'Z' as target letter within two matrix of letters: the matrix of letters with similar distinctive feature (I, V, M, X, E, W), and the matrix of letters with different distinctive features in comparison to the target letter (O, D, U, G, Q,R, C).

determined that certain neurons in the visual cortex respond only to specific stimuli presented on the part of the retina that these neurons cover. Generally, retinal neurons don't respond to a simple illumination, though they do respond to 'specifically oriented line segments"(Hubel & Wiesel, 1979, as cited in Sternberg, et al. 2008). The authors received the Nobel Prize for their breakthrough findings.

In the study of Urlich Neisser (1964) the results support the theory of feature detection.Assuming that recognition starts with identification of distinctive features, the recognition will be more difficult if the targeted pattern (targeted letter) is amongst other patterns that share large number of distinctive features [3]. The subject that participated in this study found it less difficult to point target letter in the matrix of letterswith different distinctive features. In other words, identifying target letter in the matrix of letters with similar distinctive features took longer and was followed by numerous mistakes (Neisser, 1964).

The pattern recognition approaches described here imply the bottom up processing. In other words, they move from the basic units towards the more complex. However, if the patterns that we have observed are put into context i.e. we use letters to compose words and sentences that convey meaning, we are faced with an actual observation. Like any other pattern in nature, the letters are rarely seen individually within actual observation, but rather in a context that induces their meaning. The researchers and typographers are faced with a key question and problem; In what way do we recognize words? Are the properties of the letters themselves a key factor for word recognition or are the words patterns we recognize by their distinct features, in this case the graphemes, or are the word and letter recognition within the word happening on a much higher level – the top-down processing?

Visual Word Recognition

In the attempt to explain the reading process, the cognitive psychologists have assumed a number of models of visual word recognition, start from the whole word recognition (Huey, 1908; Reicher, 1969; Haber & Schindler, 1981), bigrames or supraletter features (Wheeler, 1970; Monk & Hulme, 1983), Serial Letter Recognition to Parallel Letter Recognition (Wheeler, 1970; Monk & Hulme, 1983).

A famous researcher in the field of psycholinguistics, James Cattel was the first to discover the intricate effect known today as the Word Superiority Effect (WSE). The basis of the experiment was to display words to the participants in a very short time interval (5-10ms), thereby discovering a more accurate recognition of letters. From this experiment Cattel deduced that words are better recognized since they present whole units that we can discern. The same effect was repeated in Reicher's(1969) and Wheeler's (1970) research. The results showed that letters were more easily and more accurately recognized if they are within a context rather than within a pseudo word or isolated. Based on theresults a hypothesis was established that word recognition includes long-term memory and word pattern recognition, as well as the word shape or individual parts (familiar letter units – supraletter feature).

Empirical results consistent with the word shape recognition method have been reached through experiments in which the participants performed proofreading task (Haber & Schindler, 1981; Monk & Hulme, 1983). The experiments with the texts were conceived in such a manner that all texts contain an equal number of: errors within words that possessed a word shape [4] ,and the errors within words that did not contain a word shape [5]. The results have shown a doubled probability to overlook the errorsin words consistent by shape than errors in words not consistent by shape but with a correctly spelled word (Haber & Schindler, 1981). This backs up the notion that we will miss these errors because we are used to the shape a word is characterized by. However, these results had an alternative interpretation. In the attempt to prove that the error is being missed on account of a confusion on a letter level and not on a whole word shape level, a similar experiment (Experiment 1) has been conducted in a study entitled Word Shape's in Poor Shape for the Race to the Lexicon, Paap et al. (1984). Their results show that the participants made the most mistakes with words that had the same word and letter shape as the original word (than-tban; 15% of mistakes) and with the words that had a simi-

[4] Improperly spelled words which contained the word shape had the same schematic of the descenders and the ascender (eg.'test' is a properly spelled word, while 'tesf' is an error consistent with the shape of a properly spelled word)

[5] Improperly spelled words which did not maintain the word shape have a different schematic of the descenders and ascenders (the properly spelled word is 'test', while the improperly spelled word is 'tesc').

lar letter shape but different word shape (than-tnan; 19% of mistakes) while the least mistakes have been made with the words that had the same word shape, but different letter shape (than-tdan; 8% of mistakes) and the words with a different word and letter shape than the original (than-tman; 10% of mistakes). The results show that the demonstrations of the word shape effect (Haber & Schindler, 1981) and the supraletter effect (Monk & Hulme, 1983; Wheeler, 1970) from the previous studies has to be assigned to the effect that is pointed to because of the features of individual letters.

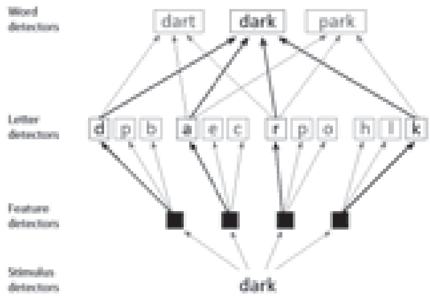
Furthermore, Paap et al. (1984) in this study indicates of the many noise factors and the lack of the internal validity for different studies in which the hypothesis stating that the word shape eases the lexic entry is distorted by using the lower-case and the upper-case letters and thereby challenged. Therefore, Paap et al. through a series of three experiments (one tachistoscopic and two lexic decisions) "which evade the trap related to abnormal transformations" disproves another argument by Haber and Schindler (1981) that the effect of the word shape corresponds to the well know high-frequency words.

Revising the earlier and analyzing the newly obtained research results, Paapet al. assume possibility that "fast, automatic word recognition is mediated by the activation of abstract letter identities. Because the bottom-up activation of abstract letter units would have to be driven by the detection of infraletterfeatures, supraletter features like length and shape may have no influence on automatic word recognition" (Paap, Newsome, & Noel, 1984).

Parallel Letter Recognition

Despite the convictions of many psychologists and typography practicioners, that word and letter patterns are of essential importance for word recognition, in recent decades a new favorizedmodel that points out that along with the causality of the semantic, syntaxic and phonological restrictions of recognition, grapheme recognition occurs through Parallel Letter Recognition within the word itself.

According to the PLR model, the letters within words are recognized at the same time, and the letter information (i.e. the characteristic glyph features) are used for word recognition (McClelland & Rumelhart, 1981; Rayner & Pollatsek, 1989). For example, a participant was displayed the word "WORK". The first step is the letter feature detection (the horizontal, vertical and diagonal strokes, etc.). These features are then sent to the level where every letter from a word is recognized simultaneously. After this comes the word detector level (Figure 1). So the final level involves the word detectors that function in the same way as the letter detectors, identifying features, in this case letters and combining them into words.



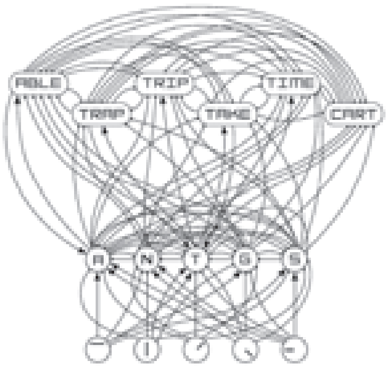
» **Figure 1:** *The letter 'd', for example, triggers activation of all the words that have this letter as their first letter, the letter 'a' of all the letters that have this letter as their second letter, etc. If a certain word has the most activations from the letter detection level (among the displayed words), that word is postulated as predefined, in this example the word 'dark'.*

What happens at the word detector level is not exactly identified or clear. However, it is clear that there is a secondary process occurring parallel with the bottom-up processing at the word detector level that consists of the top-down lexical stimulus processing depending on the context (McClelland & Rumelhart, 1981). The Parallel distributed processing – PDP in the PLR model explains the WSE. While the letters individually need to be exclusively identified by the information gained from the letter detection level, words gain information from both word and letter detector levels, which gives the words a greater advantage to be recognized than and individual letter (McClelland & Rumelhart, 1981; Rayner & Pollatsek, The Psychology of Reading, 1989).

Neural Network Modeling

McClelland & Rumelhart have made the first connectional model of processing individual words under the title of Interactive Activation Model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). The basis of this model is the biological principle. McCulloch and Pitts (1943, 1947; as cited in Larson, 2005) have shown that the neurons gather the data of other neurons. The neurons are connected by the so called synapses. The strength of a synaptic link (determined by the data flow through it) increases during the indication of an excitation event and decreases during the indication of an inhibiting event. When a neuron receives more excitation signals, it becomes active. Figure 2 illustrates the processing method of this model.

Unlike the former in which the links were pondered in advance through the insight into the characteristics of the stimulus, the Seidenber and McClelland model, despite having some characteristics of the former one, does not start of from the apriori values provided by the ponders, but rather they increase by increasing the number of iterations. This represents the core of the learning principle and bases on the notion that if the information that arrive from the synapse are important, the connection between the two neurons becomes physically stronger and if the information is less important or completely unimportant, the connection becomes weaker or even severed (Hebb, 1989).



» **Figure 2:** *McClelland & Rumelhart's Interactive Activation model, letter shape nodes that are contained in letter T and other interconnections. In this example only two knots from the left were activated because they contain horizontal and vertical strokes characteristic to the letter T; the links are excitatory and therefore represented by an arrow. Three knots to the right are the inactive part because they do not match the strokes of the letter T. These links are inhibitory and are therefore represented as circles. Every knot on the shape detection level is connected to other knots on a shape detection level. The letters A, T and S in this example receive excitatory activation from the two knots to the left in the previous level because for all three letters of this font the characteristic trait is the horizontal stroke. According to the information gathered by the shape detector, the knot with the letter T is the most active because it receives the greatest amount of excitatory activation. The letter T sends the most excitatory information afterwards with words beginning with the letter T, while the inhibitory activation is being sent through other words. Then the knots with active words send the inhibitory activation through all other words and letters and the excitatory activation only to the knot with the letter T.*

The researchers of this field of studies, McClelland, Seidenberg, Plaut and Patterson, have made a significant contribution to the development of the neural network model of reading by which we explain numerous processes during reading. Later models contributed to a more precise explanation how the graphemic information can be transformed to phonological information, emphasizing the importance of this process as the key element of word recognition.

Empirical findings on differentiation of letters

One of the most constructive/beneficial researchers from the early stages of legibility research, Miles Tinker, in his book Legibility of print (1964, as cited inBeier, 2009) he summarized the findings from the research concerning relative legibility of the letters of the alphabet from 1885 up to 1928. The relative legibility of the letters of the alphabet during this period had been researched by many methods (short exposure, distance, parafoveal vision) whereby the results acquired were expected to be different. However, Tinker grouped the findings that were consistent with seven studies. Among the most legible letters, the following were selected:'d', 'm', 'p', 'q', 'w';letters of mediocre legibility:'j', 'r', 'v', 'x', 'y'; least legible letters: 'c', 'e', 'i', 'n', 'l' (Tinker, 1964; as cited inBeier, 2009). After examining the studies, Tinker concludes that some letter features affect the legibility of the typeface. In his view, ascenders i.e. descenders of letters 'd', 'b', 'p', 'q', i 'k' make these letters the most distinguished. His findings differ from recentfindings which will be discussed later in the paper (cf. Courrieu, Farioli, & Grainger, 2004).

Letterrecognition research, i.e.feature distinctivness, gained popularity with the emergence of the information processing theory. The confusion matrix is a traditional method of research of the distinctive letter features. A typical experiment for generating confusion matrix consists of isolated letters which are presented in the constricted conditions (short exposure and/or energy masking) and that can lead to the problem of identification. Error rate (e.g. reporting letter F when letter E is displayed) indicates that visual similarities are the result of the common features whereanalyzing confusion patterns the group of features that help identify letters is revealed (e g, Bouma, 1971; Geyer, 1977; Gilmore, et al. 1979). Among the 70 studies of this subject published so far, feature lists for Roman letters have been formed, that usually consist of lines of different directions or curvature (Grainger, Rey, & Dufau, 2008). The first accurate description of distinct features of the capital letters of the Roman alphabet was given by Eleanor Gibson at the end of the 1960'.Gibson discusses a list of 12 distinct features through four descriptive factors: the straights, the curves, redundancy, discontinuity (Gibson, 1969, as cited inGeyer &DeWald, 1973). Keren and Baggen (1981) discover 14 distinct features of the capital Roman letters, according to the confusion matrix reported by Gilmore (1979). The different estimated values imply that some of these 14 features are more significant than others. The highest estimated values are for the following features: (3) parallel vertical lines, (6) single straight vertical line, (10) diagonal lines that are not diagonal lines (excluding those that are between parallel straight lines), (14) nonclosed letterstanding on a borad base. However, none of the above mentioned feature lists (Gibson, Geyer &DeWald, Keren&Baggen) can be considered widely applicable because each one of them is only valid for a specific stimulus i.e. a font used in the research, making the findings of these researchers inconsistent in a good measure, which is the subject of discussion in this paper.

In the legibility study of the Cyrillic and Roman letters, Rot and Kostic (1987) have conducted four experiments in which the legibility of individual capital and lowercase letters was explored. In their work Rot and Kostic emphasize that during the consideration and definition of the recognizablefactors of letterforms, they start with the assumption that the factors need to be mutual for both alphabets, in other words they have focused on discovering the global parameters. A study was conducted in the quest to answer the two basic questions: a) Are there letters in both Cyrillic and Roman alphabets that are relatively consistently legible, i.e. less legible than others, b) which grapheme characteristics influence their legibility?

Measuring the response times for a given stimuli, the authors have concluded that the most legible Cyrillic lowercase letters can be categorized by the following order: 'љ', 'ж', 'ш', 'њ', 'т', 'к', 'и', 'ц', 'о'; while the least legible would be 'б', 'в', 'р', 'ч', 'с', 'з'. The most legible Roman lowercase letters are: 'j', 'k', 'č', 'i', 'o', 'ž', 'r', 'š', 'f', 'e', 't'; and the least legible: 'b', 'g', 'đ', 'u', 'dž', 'h', 'v', 'l', 'a'.

The results of the legibility studies of the capitals show that the most legible capital letters of the Roman alphabet range in the following order: 'Š', 'O', 'A', 'K', 'E', 'L', 'S', 'M', 'R', 'T'; and the least legible are 'H', 'C', 'B', 'P', 'DŽ', 'LJ', 'F', 'Č', 'V', 'J'. The most legible Cyrillic capitals are 'Ш', 'Л', 'Њ', 'А', 'Е', 'М', 'О'; and the least legible are 'Х', 'Ц', 'У', 'Џ', 'Ф', 'Н', 'З', 'В', 'Ђ', 'И', 'Р', 'Б', 'Ћ'.

Making a qualitative analysis of the results of this study, Rot and Kostic (1987), draw a conclusion that the legibility of the letters greatly depends on their graphemiccharacteristics, and the most important difference determinant of the legibility is defined by the dimension of the straight and curved lines. "The straight lines, especially the vertical ones and with them the straight and sharp corners, are the most important factor of greater legibility" (Rot & Kostić, 1987). Alongside this, the legibility is influenced by the special letter parts perceived as additions, such as the diacritic symbols found in the letters 'č', 'š', 'ž', and the letters with a point above them like 'j', 'i', and the lateral additions of the letters 'њ', 'љ', 'ц'. Therefore, the discontinuity of the line as a factor and the factor of the asymmetric form, as well as the factor of the unusual form compared to the other letters (for example 'ж', 'ш', 'Ш', 'S'). The given results are somewhat in accordance with the Keren&Baggen results (1981). These authors also emphasize the presence of the straight vertical lines as the necessary characteristics for discerning of the graphemes. Rot and Kostic state that the factors that decrease the letter identification are the curved strokes andgraphemic similarity (found in letters such as 'b' and 'd'), that is the decreased legibility is pronounced by graphemic complexity (the digraphs characteristic of the Serbian and Croatian Latinalphabet: 'lj', 'nj', 'dž') a weaker visibility of the discernable parts (for example: 'l', 'v', 'ć'). which is to some extent in accordance with the findings of recent studies. (cf. Courrieu, Farioli, & Grainger, 2004).

Courrieu et al. (2004) exemplify a simple, precise and mathematically based method of measurement of the similarities of the letter pairs based on the response time of the participant to these stimuli. The participants had a task to respond by pressing a button each time they were presented with a different stimulus, and not to respond each time they were presented a same or similar stimulus (go/no-go task). All of the lowercase Roman letters were shown in Arial size 12pt and with them a 27x27 matrix was formed so that all the combinations of letters were made. The authors of this work using Principal component analysis (PCA) come to a categorization of the letter characters into classes of similarity of letter properties, contrast classes as well as the most characteristic features of the class.

The authors define 25 dimensions, for each of the dimensions a single main class of feature similarity and a contrast class with the prominent similar feature for each main class. A contrast class encompasses a character set that do not share a similar feature as the main class.

Courrieu et al. (2004) present the following classes:

1. Main class of feature similarity—a circle with an ascender or descender('b', 'd', 'g', 'p', 'q'); main contrasting class ('r', 's', 'x')
2. Small curvilinear shapes ('a', 'c', 'e', 'o', 's'); contrasting class ('h', 'k', 'y')
3. Repeating vertical strokes class ('m', 'n', 'u', 'w'); — ('f', 'i', 'j', 'l', 'p', 'q', 't')
4. Asubclass of the first class with ascenders ('b', 'd'); — ('g', 'u')
5. V-shaped class ('v', 'y'); — ('h', 'u')
6. An ascenderwitha horizontal bar('f', 't'); — ()
7. The n-shape class ('h', 'n'); — ('v', 'z')
8. Four corners and square diagonal class ('x', 'z');— ()
9. The i-shape class ('i','j');— ('w', 'y')
10. Reversible two storey 'a' class ('a', 'g'); — ('q')
11. The upper left arc class ('c', 'r'); —
12. The character with the same but interrupted stroke class ('i', 'l'); — ('x')
13. Sigmoid shape class ('s', 'z'); — ('u')
14. The subclass of the first class with the left rounded segment ('b', 'p'); — ('c', 'd', 'f')

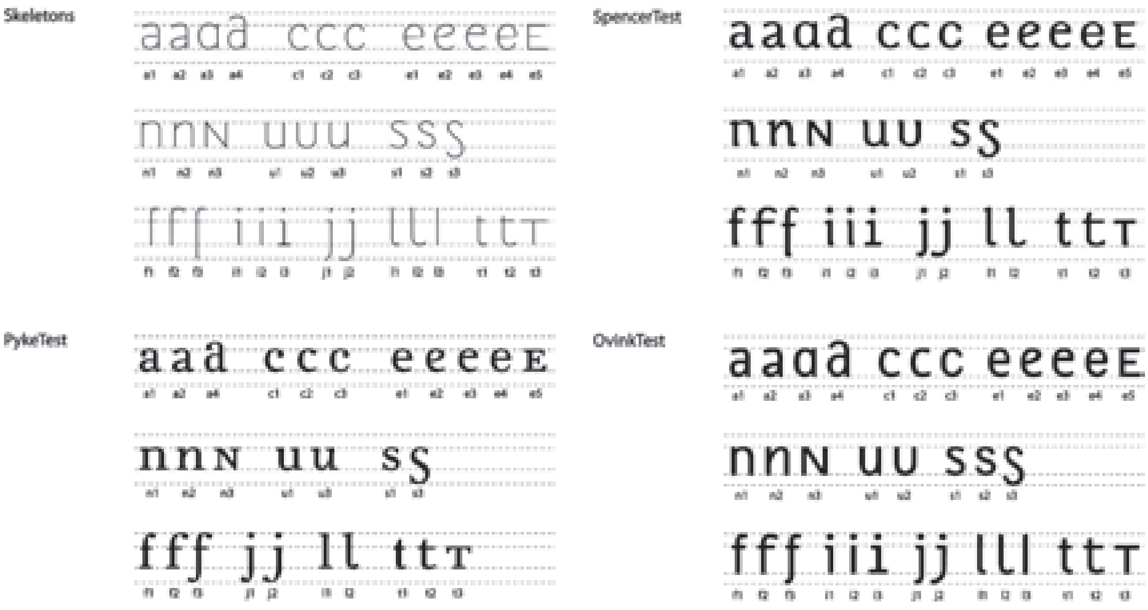
The remaining 11 classes refer to the other distinct features of the individual characters (i.e. letters).

Using the Bubbles Classification Image Technique Fiset, et al. (2008) discover which parts of the individual glyphs are most important for their recognition. The studied typeface was Arial. Their study provides evidence that the terminals of the glyphs are the features most important for letter recognition. The authors also emphasize that it is exactly these terminals that assist the reader to differentiate between the visually similar characters. The logics behind the Classification Image Technique are simple. If we deprave the observer of the visual information necessary to perform a task by masking them, this will greatly diminish his performance. In contrast, if we deprave him of information non-essential for performing a task, his performance will be undisturbed. For example: to make a difference between the letters ‘F’ and ‘E’, the information necessary is in contained within the lower part of the letter. If we conceal the lower part of the letter ‘E’, there will be no difference between these two letters. However, if we conceal the upper part of the letter ‘E’ the difference will be visible, and the observer will clearly recognize the difference and which letter is in question. The authors have made 213 masks containing the following 10 features for each of the letters of the alphabet: (1) verticals, (2) horizontals, (3) Slants Tilted Left, (4) Slants Tilted Right, (5) Curves Opened Top, (6) /Curves Opened Bottom, (7) Curves Opened Left, (8) Curves Opened Right, (9) Intersections i (10) terminations.

SofieBeier has sumarized the findings of five studies in her dissertation Typeface Legibility:Towards defining familiarity (2009) with the focus on the misrecognition of the lowercase letters. Comparing the findings of Sanford, Bouma, Tinker, Geyer and Dockeray, she states that the different typographic alphabets were used as stimuli, therefore, the misrecognition of the letter pairs were inconsistent between the compared findings. Other than that, Beier points out that with the review of the data an error repetition pattern is noticeable with the two main groups of the problematic letters. One group is the letters whose x-height and standard width are the same, different only by their straight and curved strokes (‘e’, ‘c’, ‘a’, ‘s’, ‘n’, ‘u’, ‘o’); the other group consists of the narrow lettes with one vertical stroke and small width (‘i’, ‘j’, ‘l’, ‘t’, ‘f’) (Beier, 2009). These differentiated groups of letters will be the subject of her further research.

Inspired by the differentiation theory given by Legros& Grant (Legros & Grant, 1916), and significantly later common skeleton model proposed by Adrian Frutiger (1998, pp. 202,203), Beier considers the shapes of the skeleton variations as the subject of her study (Beier & Larson, 2010; Beier, 2009). For research purposes, she designs three typefaces: a serif, a sans serif and a pseudo-serif (Pyke, Ovink, Spencer) with the same skeleton. Beier and Larson (2010) have tested numerous hypotheses with these two carefully prepared experiments: short exposure and distance viewing. With the short exposure experiment in the conditions when the stimuli were placed within the foveal area of vision, no recognition errors were detected, which made the subject of their study the conditional identification errors at the parafoveal area. In the effort to discover the most optimal differentiation of letters with a similar skeleton, they tested the following hypotheses with designed stimuli:

- Horizontal bar on sans serif letter ‘i’ used to emphasize the division of the stem from the dot, which made them expect a better legibility in comparison to the sans serif letters without horizontal bar. The hypothesis was confirmed for distance viewing.



» **Figure 3:** The skeleton and the variations of the skeleton viewed in three different fonts designed by SofiBeier (reproduced, from Beier& Larson, 2010, with the permission of the author).

- Greater differentiation of the letters ‘u’ and ‘n’ should provide better legibility. To test this hypothesis Beier designs variants for these letters: tailless ‘u’ and variations for both letters with a pronounced difference in the height of the junction stroke. This hypothesis was not confirmed.
- According to the Gestalt psychology, our perceptive system tends to close the incomplete shapes by filling the gaps (The Law of Closure). Guided by this theory it might be expected that the smaller aperture of the letters ‘c’ and ‘e’ increases the risk of an error, by closing the gaps and recognizing them as the letter ‘o’. The closed apertures hypothesis was tested with many other versions of the letters ‘e’and‘s’ and the two-storey‘a’. The closed apertures hypothesis was only confirmed in the case of the letter ‘e’, where the stimulus ‘e4’ recorded far greater identification errors than the other letters. The authors think that the hypothesis isn’t confirmed in case of the letter ‘s’, they assign considerable differences between the variants of the letters to the difference in the spine of the letter. The varaint of the letter ‘s1’ has a diagonal spine and a big closed aperture while the variant ‘s2’ has a rounded spine and a closed aperture. The authors state that this finding is inconsistent with the earlier findings of the typography researcher G.W. Ovink (cf.1938, as cited in Beier, 2009), whch increases the doubt in the precision of the interpretation of this result.
- For the one-storey‘a’ a lower legibility was expected compared to a two-storey‘a’ and a misrecognition of the letter ‘o’. The hypothesis was confirmed and frequent identification errors were recorded, i.e. the confusion between the one-story letter ‘a’ with the letter ‘o’ and the letter ‘q’.
- The narrow letters ‘l’, ‘f’, ‘t’,‘j’ i ‘i’ take up little width. If their image was to expand, greater legibility would be expected. For these letters there is a wide and a narrow variation. This hypothesis was confirmed for the letters ‘j’, ‘l’ i‘t’, in the case of the pseudo-serif font (Spencer) it was confirmed for the letter ‘f’ while the letter ‘i’ has no clear value in a widened form.

Discussion

According to the authors, the first epistemological study of the construction of typographical knowledge, Ole Lund (1999) has discovered that within 26 of the examined readability studies (out of 72 published in total until then) each one of them lacks internal validity, besides the various methodological errors. With the exception of a few researchers, the most commonly noticed deficiencies were inadequate usage of the typographic stimuli which according to the author largely stems from the insufficient knowledge of the domain and the matter of the typographic design. Lund concludes that the effort invested in the 72 research studies of the type face legibility studies based on the various operational methods, has not come to fruition, hence we cannot consider that the typographic knowledge has been expanded nor that certain theories have been clarified or set according to the empirical results gained. However, the contribution of the revised studies vary according to quality, states Lund. He emphasizes that the studies of English, Zachrisson, Harris, Vanderplass and de Lange are interesting and rational but not without flaw (English 1944, Zachrisson 1965, Harris 1973, Vanderplas and Vanderplas 1980, and de Lange, as cited in Lund 1999). Besides these, Lund points to the fruitful researchers such as Tinker, Peterson as worthy of mention and with an impressive and imposing research program.

The choice of different types of stimulus (typefaces), and yet again their inadequate manipulation in the studies of feature detection with confusion matrix method, have led to inconsistencies among the results between numerous studies. Consequently, this has slowed down the progress of finding the ultimate list of distinctive features of the letters – the key to letter recognition. Even though many scholars were aware of this, some still compared their results to the results of others and published their findings with inconsistencies neglecting the fact that their stimulus was different (cf. Geyer, 1977; Bouma, 1971). Nonetheless, oversights of this kind have caused scholarly discussions and qolitative analysis of the confusion matrix findings. In his study, Geyer (1977), argues that previous results of the confusion matrix even for the lowercase of the alphabet in the Boume study (1971) unconvincing. He suggests that we must consider comparison of the results of the foveal and eccentric field of vision, based upon the performance of the identification, as an empirical question. In the attempt to present the answer to this question, Geyer compared confusion matrix of the eccentric field of vision published by Bouma (1971) to the confusion matrix of the foveal vision from his findings (Geyer 1977). Apart from the differencies in the data analysis, Milloy (1978) provided a detailed review based on a drastic difference in the choice of stimulus. Milloy points out obvious difference between typefaces Courier 10, used by Bouma in his study, and Tactype Futura demi 5424, used by Geyer. The differences such as: the amplified (slabbed) serif typical for the Courier typeface, and the lack of the serif in the Futura typeface; the two-storey ‘a’ in contrast to one-storey ‘a’; noticable similarities of the one-storey letter ‘a’with the letter ‘o’ ofthe Futura typeface; and many other not listed by Milloy but mentioned by Beier (2009) are sufficient enough not to consider the findings related to this matter and published by Geyer as fully grounded.

Analyzing the findings of five studies with the focus on the misrecognition of lowercase, SofieBeier manages to single out the pattern of the repeated error for the two main groups of the problematic letters. The deviations in confusion matrix findings of Baume (1971) and Geyer (1977), as well as the deviations of these two authors in comparison to the findings of Sanford, Tinker and Dockeray, are very appealing for detailed analysis. Namely, Beier argues that the characteristics such as amplified serifs and monospaced characters of the tested typeface (Courier) in the Boume study (1971) had been crucial for errors in some letter pairs for which other studies do not recogniseany deviations. This observation logically explains that the uniform width of the character may have been the cause in the misidentification of the letter 'n' with the letter 'm', as well as the letter 'v' with the letter 'w'. Contrary to this finding, we recognize the serif feature as oversight which caused misidentification of the letter 'g' with the letter 'q', considering these two characters have very similar skeleton of the letter i.e. in this typeface the letter 'g' is one-storey (Figure 4).

» *Figure 4. Lowercase 'g' and 'q' of the typeface Courier*

That does not exclude the effect of the amplified slab-serif, yet certainly this is not considered as the feature that had caused the error in the first place. Additionally, findingsof Courrieu et al. (2004) supports this assumption. Similarities that caused the error should be perceive as essential for further scholarly research, but also for the research of type design practitioners. In their study Courrieu, Farioli, & Grainger (2004) applying the factor analysis (PCA) have arranged the letter 'g' into the main similarity class along with the letters 'b','d', 'q', 'p', where the common feature is specified as 'a circle with an ascender or descender'. Thereby, Courrie, et al. in the tested font (Arial) identify that „the graphic realisation of the letter 'g' is the one that looks like 'q' and the digit '9' (in other words one-storey 'g').The findings of Courrieu et al. along with most common misrecognition (31% confusion) of the letter 'g' and 'q' in the study of Boume (1971) imply an inquiry to why Beier and Larson (2010) didn't find interest in testing skeleton variability of the two- storey and one-storey letter 'g' allograph. We believe that this issue must be considered as empirical.

Faced with the inconsistencies among the findings of other scholars, Beier considers certain findigs of Geyer (1977). Beier implies that the straight stem (without typical curving at the bottom) had probably increased chances why subjects misrecognized the letter 't' during short exposure with the letters 'i' and 'l'. As stated before, the unusual choice of the tested typeface (Tactype Futura demi 5424) in this study had caused different errors in letter recognition. The unusual character of the letter 't' of this typeface had been highly reduced presenting kind of an idiom or "idiosyncracy". Geyer was aware of the fact that there is variability in the configuration of lowercase letters among different typefaces, and that is actually a function of certain typographic style. He believed that the cause of this variability is the level of "ornateness''. Therefore, he tried to find appropriate stimulus for his experiment which in his opinion would "...minimised the influence of any particular type style idiosyncracy"and chose Tactype Futura demi 5424. Futura typeface (designed by Poul Renner) is the typical example of the elementary-constructivistic early German modernism which would explain why this typeface, in its idiosyncracy, among other things, is aestheticaly different. In his book "New Typography", Tschichold (Tschichold, 1928/1995)pointed out: "All the attempts up to now to produce a type for our time are merely 'improvements' on the previous sans serifs: they are still too artistic, too artificial, in the old sense, to fulfil what we need today". This was generalised comment on typeface designed in this manner. Like many experimental psychologists before him, Geyer was not educated proparly in the field of typography in order to deal with this problem. Still, his findings are considered interested for examining this group of sans serif typeface that have a new shape of the skeleton as their common feature, besides upfore described features – one-storey letter 'a' i.e. allograph of this character, and one-storey letter 'g' (previously mentioned allograph). Sanford, Dockeray and Tinker in their research used antique typefaces as stimulus which made their findings more consistent according to which Beier (2009) was able to define a pattern of continuous letter misrecognition.

Conclusions

After a breakthrough in information processing theory, researchers were able to present significant findings on letter recognition i.e. letter distinctive features within a typeface – alphabet. Numerous research consulting confusion matrix (Bouma, 1971; Geyer, 1977; Gilmore, Hersh, Caramazza, & Griffin, 1979; Keren & Baggen, 1981; Geyer & DeWald, 1973) provided qualitative and significant findings which were helpful in many ways – with the development of eye tracker device and neural network modeling (McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989; Rayner, 1998) these findings were beneficial in presenting precise answer to the question: How do we recognize words?

Further research provided groundwork for the pattern recognition theory. However, foundation for the theory of readability and legibility of typefaces fundamental for typographers and type designers was made possible with the findings of Rot and Kostic (1987), supported by the study of Courrieu et al. (2004); epistemological study of typographic design knowledge by Ole Lund (1999); the findings most prominant for type designers were in the studies of Beier and Larson (2010; 2009) which propose new aproach to legibility research of the type letters, and they also, in great extent, certify some feature styles as more legible than others. Not in the least less important within this field were the findings of familiarity, commonality, and font tuning (Sanocki, 1987; Sanocki, 1992; Walker, 2008) which by its nature of its scope could not be reviewed in this research (see Sanocki & Dyson, 2012).

Acknowledgments

This work was supported by the Serbian Ministry of Science and Technological Development, Grant No.:35027 “The development of software model for improvement of knowledge and production in graphic arts industry”

References

1. Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual Word Recognition : The Journey from Features to Meaning (A Travel Update). In M. Traxler, & M. A. Gernsbacher (Eds.), Handbook of Psycholinguistics (2 ed., pp. 285–375). Academic Press.
2. Beier, S. (2009, May). Typeface Legibility-:Towards defining familiarity. PhD thesis. London: The Royal College of Art.
3. Beier, S., & Larson, K. (2010). Design Improvements for Frequently Misrecognized Letters. Information Design Journal, 18(2), 118-137.
4. Bouma, H. (1971). Visual recognition of isolated lower-case letters. Vision Research, 11(5), 459–474.
5. Cattell, J. M. (1886). The Time Taken up by Cerebral Operations. Mind, 11(44), 524-538.
6. Cheetham, D., Poulton, C., & Grimbly, B. (1965). The Case for Research. Design. Design(195), 45–51.
7. Courrieu, P., Farioli, F., & Grainger, J. (2004). Inverse discrimination time as a perceptual distance for alphabetic characters. Visual Cognition, 7(11), 901–919.
8. De Lange, R. W., Esterhuizen, H. L., & Beatty, D. (1993). Performance differences between times and helvetica in a reading task. ELECTRONIC Publishing, 6(3), 241-248.
9. Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008, November). Features for Identification of Uppercase and Lowercase Letters. Psychological Science, 19(11), 1161-1168.
10. Frutiger, A. (1998). Signs and Symbols: Their Design and Meaning. New York: Watson-Guptill Publications.
11. Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. Perception & Psychophysics, 22(5), 457-490.
12. Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. Perception & Psychophysics, 14(3), 471-482.
13. Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. Perception & psychophysics, 25(5), 425–431.
14. Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium. Trends in Cognitive Sciences, 12, 381-387.
15. Haber, R. N., & Schindler, R. M. (1981). Error in proofreading: Evidence of syntactic control of letter processing? Journal of Experimental Psychology: Human Perception and Performance, 7(3), 573-579.
16. Hebb, D. O. (1989). The first stage of perception: Growth of the assembly. In J. Anderson, & E. Rosenfeld (Eds.), Neurocomputing: Foundations of Research (pp. 54-56). The Mit Press.
17. Hebb, D. O. (1989). The first stage of perception: Growth of the assembly. In J. Anderson, & E. Rosenfeld (Eds.), Neurocomputing: Foundations of Research (pp. 54-56). The Mit Press.
18. Huey, E. (1908). The Psychology and Pedagogy of Reading. New York: The Macmillan Company.
19. Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. Perception & psychophysics, 29(3), 234-46.
20. Kostić, A. (2010). Kognitivna psihologija (Drugo ed.). Beograd: Zavod za udžbenike.
21. Larsen, A., & Bundesen, C. (1996). A template-matching pandemonium recognizes unconstrained written characters with high accuracy. Memory & Cognition, 24(2), 136–143.
22. Larson, K. (2005). The science of word recognition. Typo, 2–11.
23. Legros, L., & Grant, J. (1916). Typographical printing surfaces: the technology and mechanism of their production. London: Longmans, Green, and Co.
24. Lund, O. (1999). Knowledge construction in typography: the case of legibility research and the legibility of sans serif typefaces. PhD Thesis. The University of Reading, Department of Typography & Graphic Communication.
25. McClelland, J. L., & Johnston, J. (1977). The role of familiar units in perception of words and nonwords. Attention, Perception, & Psychophysics, 22(3), 249-261.
26. McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375–407.
27. Milloy, D. G. (1978). Comment on Recognition and confusion of the lowercase alphabet. Perception & Psychophysics, 22(2), 190–191.
28. Monk, A. F., & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. Memory& Cognition, 11(1), 16-23.
29. Neisser, U. (1964). Visual search. Scientific American, 210(6), 94-102.

30. Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. Journal of Experimental Psychology: Human Perception and Performance, 10(3), 413-428.
31. Paterson, D. G., & Tinker, M. A. (1932). Studies of typographical factors influencing speed of reading. X. Style of type face. Journal of Applied Psychology, 16(6), 605-613.
32. Paterson, D. G., & Tinker, M. A. (1944). Eye movements in reading optimal and non-optimal typography. Journal of Experimental Psychology, 34(1), 80-83.
33. Poulton, C. E. (1965). Letter differentiation and rate of comprehension in reading. The Journal of applied psychology, 49(5), 358-62.
34. Poulton, C. E. (1972). Size, style, and vertical spacing in the legibility of small typefaces. The Journal of applied psychology, 56(2), 156-161.
35. Rayner, K. (1998). Eye Movements in Reading and Information Processing : 20 Years of Research. Psychological Bulletin, 124(3), 372-422.
36. Rayner, K., & Pollatsek, A. (1989). The Psychology of Reading. New Yersey: Lawrence Erlbaum Associats.
37. Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81(2), 275-280.
38. Rot, N., & Kostić, A. (1986). Čitljivost ćiriličnog i latiničnog alfabeta. Psihologija, 19(1–2), 157–171.
39. Rot, N., & Kostić, A. (1987). Čitljivost slova latinice i ćirilice i njihove grafemske karakteristike. Psihologija, 20(1–2), 3–19.
40. Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 89(1), 60-94.
41. Sanocki, T. (1987). Visual Knowledge Underlying Letter Perception: Font-Specific, Schematic Tuning. Journal of Experimental Psychology Human Perception and Performance, 13(2), 267-278.
42. Sanocki, T. (1992). Effects of font- and letter-specific experience on the perceptual of letters processing. American Journal of Psychology, 105(3), 435-458.
43. Sanocki, T., & Dyson, M. C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. Attention, Perception & Psychophysics, 74(1), 132-145.
44. Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. Psychological review, 96(4), 523-568.
45. Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. Proceedings of the Symposium on Mechanisation of Thought Processes (pp. 511–529). London: Her Majesty's Stationery Office.
46. Sternberg, R. J., & Mio, J. (2008). Cognitive Psychology (5 ed.). Wadsworth Publishing Company.
47. Tinker, M. A., & Paterson, D. G. (1946). Readability of mixed type forms. The Journal of applied psychology, 30(6), 631-637.
48. Tschichold, J. (1928/1995). The New Typography/ The First English Translation of The Revolutionary 1928 Document. London: University of California Press.
49. Walker, P. (2008). Font tuning: A review and new experimental evidence. Visual Cognition, 16(8), 1022-1058.
50. Wheeler, D. (1970). Processes in Word Recognition. Cognitive Psychology, 1, 59-85.