Cover Page

Universiteit Leiden

Leiden University
Repository

The handle http://hdl.handle.net/1887/20022 holds various files of this Leiden University dissertation.

**Author**: Chahine, Nadine
**Title**: Reading Arabic : legibility studies for the Arabic script
Date: 2012-10-25

Chapter 5

# Eye Movement and Legibility Studies

The previous chapters have looked at the Arabic script, its calligraphic and typographic developments, and the state of Arabic type design today. All of that was being analyzed and observed with the eyes of a designer. This chapter is a departure from that arc. Here we turn to the study of reading and eye movement. It is a literature review chapter belonging to the domain of psycholinguistics. The aim is to understand eye movement in reading, the effects of the legibility of type on eye movement, and what legibility studies have discovered so far.

The chapter will attempt to give an overview of the findings of eye movement studies. It starts with the pattern of eye movement in reading, its characteristics, what facilitates, and what hinders it. It also investigates the effects of language and different script systems on reading. It explores how the nature of the content affects the reader's eye movements, and how the reader in turn can react differently to the text. It looks at what is in focus, and what is not, and how these two interact and drive the eye forward. That leads to the questions of where and when to move the eyes, and the big debate of eye movement control.

Having studied the patterns of eye movement and how the characteristics of both the reader and the linguistic material being read affect reading measures, the final section shifts to the effects of the visual characteristics of the text on eye movement and reading measures. That section will give an overview of legibility findings, as well as methods of carrying out legibility studies.

Why is this chapter important to this dissertation? There is no Arabic in it, but it is the middle piece of the puzzle. For one to understand how the visual characteristics of a script affect reading, it is important to understand how reading works, and how different scripts and languages affect the reading process, and what makes a legible design. The reading and legibility studies specific to Arabic have been relegated to a separate chapter coming up next, but for now, it is time to get back to reading.

## The Mechanics of Eye Movement

In its simplest definition, eye movement is made up of stops and jumps, called fixations and saccades respectively. Fixations are the instances when the eyes stay relatively still in order to focus on a visual stimulus. Saccades are fast movements during which reduced information intake is obtained due to what is generally referred to as saccadic suppression (Rayner & Pollatsek, 1983). The eyes move so quickly that any information taken in during a saccade would be seriously blurred; the masking effect of clear and sharp information acquired before and after a saccade removes the perception of blurring (Brooks, Impelman, & Lum, 1981).

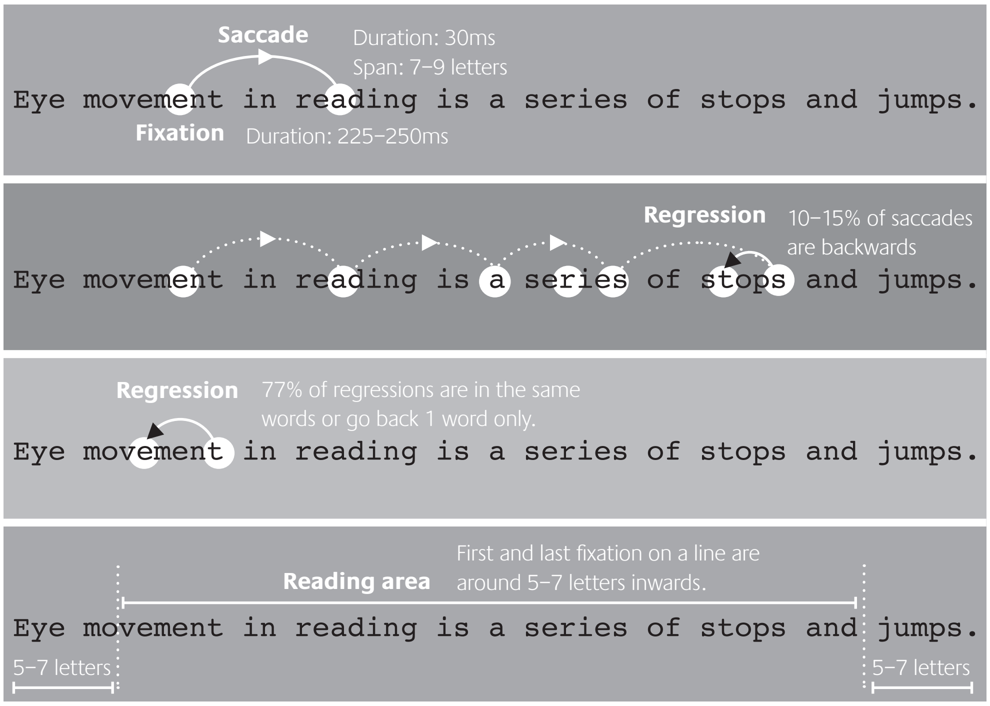


Fig. 5.1 Some of the charactersitics of eye movement in reading.

It is necessary to frequently move our eyes due to limitations in our visual acuity. When looking straight ahead, one has around 2 degrees of a sharp visual (corresponding to objects falling on the fovea), then up to 5 degrees of less clear vision (for objects falling on the parafovea), and then it gets to be of much poorer quality as it extends out (for objects falling in the periphery) (Rayner, 1998). It is then necessary to keep moving the eyes so that the visual that we want to focus on falls onto the fovea, which is the central part of the eye's retina.

The movement of the eyes is tied to how the visual system behaves during reading. However, reading is a much more complex process, and the process of reading involves many different stages of analysis. This will be discussed later on in this chapter. For now, it is good to keep in mind a general overview of what is going on in the cognitive processing of text:

> "First, visual information is obtained, and the orthography (letter identity and word length), phonology (sounds), and morphology (units of meaning, grammatical gender, etc.) of the word are analyzed. Then the lexical representation (the abstract representation of the word form) is accessed. Finally, the semantic (word meaning) and syntactic (grammatical role) representations of the word are accessed and integrated into the meaning of the sentence" (Schotter, Angele, & Rayner, 2012, pp. 7–8).

## Eye Movement in Reading

Eye movements in reading[1] are also comprised of fixations and saccades (Fig 5.1). Fixations in English usually last for around 225–250 ms but can vary greatly from as short as 50 ms to 600 ms or more; the average saccade is 7–9 letters but can also

1 For the rest of this chapter, all discussions of reading refer to silent reading, and when unspecified, usually refer to the reading of English since most of the work done in this field has been done with the English language.

range from as little as 1 letter to up to 20 letters or more (Rayner, 2009). The average saccade corresponds to a 2 degree move and usually takes around 30 ms to execute (Rayner, 1978). These values, though, are averages and can vary across individuals. The difficulty levels of the text material and the reading skills also have a direct effect on reading measures; the more difficult the material is, the longer the fixations, the shorter the saccades, and the more regressions are made (Rayner, 1998). Indeed, the variability seen in fixation durations is mainly due to the level of difficulty of the text being processed (Rayner & Duffy, 1986) and is reflective of on-line processing (Rayner, 1998). Reading measures are also affected by the nature of the script and the typographical qualities of the typeface used as will be discussed later in the chapter.

Saccades are usually forward movements that move the eyes forward through the text to bring in new stimuli to the fovea. However, around 10–15% of saccades bring the eyes backwards in what is usually referred to as regressions (Fig. 5.1). Regressions usually go back one word either because of oculomotor errors where the eyes landed too far from their intended target, or because of the need for further lexical processing (Rayner, Chace, Slattery, & Ashby, 2006). However, the eyes do tend to go back even further into the text when faced with comprehension difficulties (Rayner, 2009). Regressions within the same word also occur, and those are due to difficulties processing the currently fixated word (Rayner, 1998).

Around 77% of regressions within the same line of text are going back a few characters within the same word, or to the word directly preceding it (Vitu, 2005) (Fig. 5.1). Within-word regressions are around 2 letter spaces and inter-word regressions are 3–4 letter spaces (Vitu & McConkie, 2000). Both of these kinds of regressions are not sent randomly back in the text but rather to specific word positions (Vitu, 2005), and they tend to land in the center of the word (Inhoff, Weger, & Radach, 2005). On the other hand, long-range regressions are directed in the general direction of the text to be reread, guided by the reader's knowledge of content and word order, and are then followed by another saccade that brings the eyes to the intended location (Inhoff et al., 2005).

It is also interesting to note that readers can recall the spatial location of information that they have just read, (Inhoff et al., 2005) and this might assist in the planning and execution of regressions. Studies have also shown that a regression is more likely if the directly preceding saccade was long, if the origin word[2] was of low frequency, or if a word had been skipped; this is especially so, if the skipped word was long or of low frequency (Vitu, 2005).

Other than comprehension difficulties and mislocated fixations—and research has shown that around 20% of fixations are mislocated (Engbert, Nuthmann, & Kliegl, 2007)—regressions could be due to ambiguities within the text. Reading could be proceeding smoothly until the reader comes across a word that reveals a different interpretation of the sentence is needed; tests have shown that readers in such cases often made a regression as soon as they came across the disambiguating word (Frazier & Rayner, 1982). However, the pattern of regressions did not show that the readers went back to the beginning of the sentence, nor did it show a systematic backtracking until the source of error was found. Rather, the pattern was indicative of a system of selective re-analysis using whatever available information about the error that there is (Frazier & Rayner, 1982).

Other forms of saccades are return sweeps. These are movements that bring the eye from the end of one line to the beginning of the next one. In left-to-right

**2** If the regression happened after fixating on word n, and there was no previous word skipping, the origin word would be word n-1 (Vitu, 2005).

scripts, this would be a right-to-left movement and vice versa. These are column-wide movements and are therefore often-long ones. The eyes frequently undershoot and end up doing small corrective movements towards the beginning of the line; the first and last fixations on a line are not at the extremes but at around 5–7 letter spaces inwards (Fig. 5.1) and so roughly 20% of the text fall outside the reading area (Rayner, 1998). Words at the line extremes are not the only ones being skipped. Within the text falling between the 2 extremes, other words are also being skipped. In general, only 85% of content words are fixated, while the number drops to 35% of function words; in other words, 65% of function words are skipped (Rayner, 2009).

## Tremors, Drifts, Microsaccades

Even during a fixation, the eyes are never really still (Rayner, 1998) and there are 3 types of movements other than the saccades already discussed: drifts, microsaccades, and nystagmus (also referred to as tremors).

During a fixation, the eyes make small movements that are generally involuntary and rarely more than 1 degree in amplitude (Engbert & Kliegl, 2004). The eyes drift with slow and small movements (drifts) due to lack of perfect control of eye movement by the nervous system (Rayner, 1998). Microsaccades, which are fast and small involuntary movements, bring back the eyes to the original position (Rayner, 1998). Microsaccades are about 25 ms in duration (Martinez-Conde, Macknik, & Hubel, 2004) and help in reducing binocular disparity (Engbert & Kliegl, 2004) which will be discussed in the following section.

Nystagmus are small but constant tremors. Their nature is not clear but is assumed that the eye keeps moving so as to maintain perception and to help keep the nerve cells in the retina active (Rayner, 1998). If the eyes were perfectly still, the visual perception would fade because of neural adaptation (Martinez-Conde et al., 2004). These tremors are the smallest of all eye movements, are independent in the 2 eyes, and are simultaneous with drifts (Martinez-Conde et al., 2004).

## Binocular Fixation Disparity

Contrary to what one might expect, the two eyes do not land on the same letter during reading. Instead, each eye can fixate on a different letter. This phenomenon is referred to as binocular fixation disparity. In some cases, the eyes are even crossed: the left eye lands to the right of where the right eye has landed.

Liversedge, White, Findlay, and Rayner (2006) have shown that the eyes are in fact aligned for only 53% of the time and, in these cases, are within 1 character of one another. This roughly corresponds to 0.29 degrees of the visual angle. For the remaining 47% of cases, the disparity is close to 2 characters. These values refer to disparity measured at the end of the fixation. Binocular disparity is larger at the beginning of a fixation, where vergence movements help to bring down disparity during the fixation. The magnitude of these vergence movements is positively correlated with the fixation duration (Liversedge et al., 2006). The velocity of these vergence movements is positively correlated with the length of saccades, with faster vergence movements after long saccades (Collewijn, Erkelens, & Steinman, 1988) and after reading for comprehension (Hendriks, 1996).

Binocular disparity is more obvious in children than adults, and with a higher number of crossed fixations (Blythe et al., 2006), but this is not affected by either the visual or the linguistic characteristics of the text, or whether the task was reading or not even linguistic in nature (Juhasz, Liversedge, White, & Rayner, 2006). Furthermore, binocular disparity was found to have no influence on fixation times (Juhasz et al., 2006) and was not affected by the frequency of the currently fixated word (Blythe et al., 2006). However, binocular coordination was found to have a more critical role in the foveal than in the parafoveal region (Blythe, Liversedge, & Findlay, 2010). The "effective fusion range," needed in order to form a "single unified perceptual representation" of the text, is when binocular disparity is 1 character apart; more than that results in decreased accuracy, and more fixations and trials (Blythe et al., 2010). This would have a resulting effect on the probability of refixations.

## Effect of Script on Reading Measures

The writing system is one of the major contributing factors to the differences in reading measures across different languages. For example, the Chinese script is ideographic where words are made up of one or two character so the script density is quite high compared to English. The average fixation duration and regression rate for Chinese are similar to that of English; the average saccade length, though, is quite different where the readers move their eyes an average of 2–3 characters, rather than 7–9 (Rayner, 2009). Similarly, saccade length in Japanese is about 2–5 characters (Ikeda & Saida, 1978). Of the different Japanese script systems, katakana requires the longest fixation duration and shortest saccade length (in terms of space rather than character count) and kanji-based text requiring the longest saccade and shortest fixation duration (Osaka, 1989). It is possible that katakana is the most labor intensive since it is used for the transliteration of foreign words, which are by nature less frequent. It is though puzzling why kanji, which is based on Chinese logographic characters and is the most linguistically dense, results in longer saccades[3]. It is the only exception to the Rayner's (2009) observation that the increased density of linguistic information decreases the saccade length.

This effect of linguistic density on reading is also seen in Hebrew. Like Arabic, everyday Hebrew text is unvocalized (undotted, in this case). Because the vowels are dropped out, Hebrew texts are then denser than English ones. As can be expected, saccade lengths in Hebrew are shorter than English ones, with 5.5 letter spaces on average; the speed of acquisition of information (as in words per minute) were similar but the Hebrew fixation duration average is slightly higher than that of English (Pollatsek, Bolozky, Well, & Rayner, 1981).

All these results with regards to linguistic density are in line with the findings of Morrison and Rayner (1981) which showed that the saccade length in fact depends on character spaces and not visual angle. They showed that when type size is maintained constant, the saccade length stays the same even when the distance from the stimulus is doubled. They conclude that the number of characters[4] is an appropriate measure for saccade length (R. Morrison & Rayner, 1981).

---

**3** The relative linguistic density within the 3 Japanese writing systems and their effect on saccade length remains to be studied.

**4** It is necessary to point out here that these tests are usually conducted in monospaced typefaces so all characters have the same advance width. Monospaced typefaces by definition have equal widths for all characters resulting in wide i's and tight m's. The fact that all characters have the same widths means that it is ok to have the character count as a measure of saccades as the character width is always constant across the typeface so words with thin letters like the word "lit" will have the same width as words with wide letters like "mum." In a proportionally spaced typeface, this would not be the case.

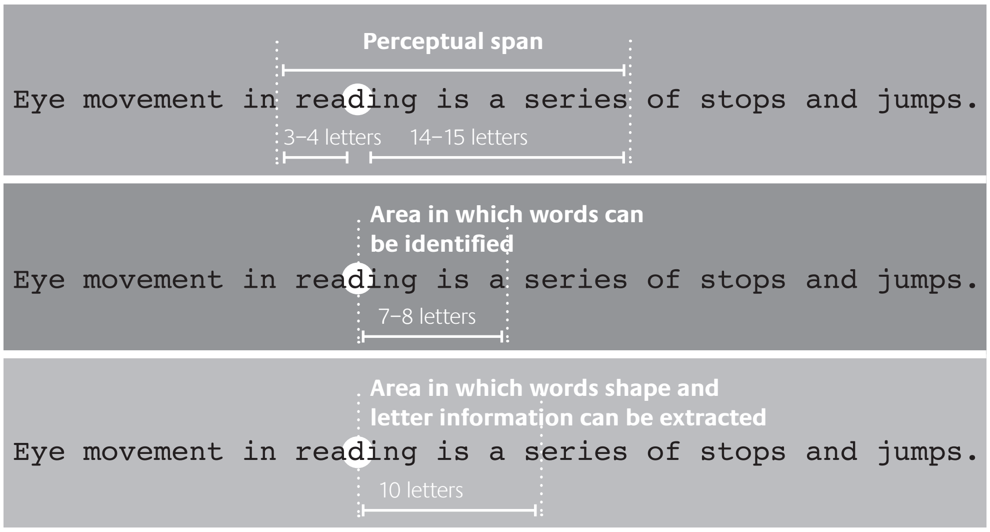


Fig. 5.2 Some of the charactersitics of parafoveal preview in reading.

## Attention, Perception, and the Parafovea

When we read, it appears as if the whole line is in focus. In reality, only a part of the line is within our perceptual span[5] (Fig. 5.2), also called the field of effective vision. In English, this field is asymmetric and extends to 3–4 characters to the left of the fixation (McConkie & Rayner, 1976) and 14–15 characters to the right of it (McConkie & Rayner, 1975). A reduction in the field of effective vision, manipulated via the moving window technique[6], results in longer fixation times, more forward saccades but has no effect on the number of regressions (McConkie & Rayner, 1975). The area in which words can be identified, usually referred to as word identification span (Fig. 5.2), is smaller than the perceptual span and is around 7–8 characters right of the fixation (Rayner, Well, Pollatsek, & Bertera, 1982).

Though the field of effective vision spreads to 15 letter spaces to the right of the fixation, the reader does not obtain word shape or "specific letter information" (Fig. 5.2) until around 10 letter spaces right from the fixation (McConkie & Rayner, 1975). The field of effective vision is bound by the currently fixated line and does not extend to the lines underneath; the only exceptions are in the case of a visual search task, or if the text is arranged to read vertically as in Chinese and some versions of Japanese (Rayner, 1998). However, there is a vertical downward pull of saccadic trajectories that is reflective of a top-to-bottom orientation of attention (Inhoff, Seymour, Schad, & Greenberg, 2010).

---

5 Studies related to the perceptual span usually make use of the moving window technique to calculate and manipulate the size of the area of effective vision.

6 This quote from Rayner gives further details on the technique: "In the moving window technique (McConkie & Rayner. 1975), the text is perturbed except in an experimenter-defined window region around the point of fixation. Wherever the reader looks, the text is visible, while outside of the window area, the text is perturbed in some way. Readers are free to move their eyes whenever and wherever they wish, but the amount of useful information that is available on each fixation is controlled by the experimenter. Each time the eyes move, a new region of text is exposed while the region previously fixated is perturbed. In some cases, the window is defined in terms of letter spaces, whereas, in other cases, the window coincides with word boundaries.

"Sometimes the spaces between words outside of the window are preserved, and other times they are filled in, and sometimes the text is perturbed outside the window only on selected fixations. The assumption with this technique is that when the window is as large as the region from which the reader can obtain information, there is no difference between reading in that situation and when there is no window" (Rayner, 1998).

The perceptual span, then, is asymmetric (McConkie & Rayner, 1976) and biased in the direction that the attention is moving in (Reichle, Rayner, & Pollatsek, 2003) as can be seen in the perceptual spans of Israeli subjects reading Hebrew and English (Pollatsek et al., 1981). Pollatsek et al. (1981) very clearly demonstrated that the perceptual span in Hebrew is also asymmetric but that it is biased to the left rather than to the right, a result which corresponds to the direction of reading and therefore attention. This result is interesting for several reasons: 1. The same readers showed a biased perceptual span to the right when reading English, and so the perceptual span is not a product of reading habit but reading direction; 2. As argued by the authors, this is valid proof that the asymmetry is due to reading direction and not hemispheric specialization and thus quelling the argument that the span is biased to the right so as to bring in more information to the left side of the brain; 3. This is solid evidence to the effect of the script system on the mechanics of reading.

Studies of the perceptual span across various scripts have shown that the size of the field of effective vision is also script dependent. For example, in Japanese, the perceptual span is around 10 characters (Ikeda & Saida, 1978), though later studies have put that number as 7 characters for kanji-hirakana[7] mixed text and 5 characters for hirakana (Osaka, 1992). The same study also found shorter saccade lengths for hirakana. Both these are puzzling since the kanji letters are more visually complex and one would expect that to translate into a tighter the perceptual span, since it usually does have a decreasing affect on saccade length (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006) as will be discussed later. The 7 characters for the kanji-hirakana mix correspond to having a preview of the currently fixated word and the beginning of the next one, which is less than what one usually gets during the reading of English texts (Inhoff & Liu, 1998).

In Chinese, the perceptual span is also asymmetric and extends one character to the left and 2–3 characters to the right (Inhoff & Liu, 1998). Inhoff and Liu (1998) have demonstrated that Chinese readers need a valid preview of up to 3 characters to the right in order to achieve normal saccade lengths, landing locations, and gaze durations, but up to 2 to achieve normal fixation durations. The 3 characters correspond to the currently fixated word and the one to the right of it, thus larger than the Japanese kanji-hirakana perceptual span (Inhoff & Liu, 1998).

Similar to Rayner's (2009) conclusion regarding the effect of script density on saccade length, the perceptual span seems to be affected by the density of linguistic content being presented, and that is script dependent. The Japanese results are not fully in keeping with a smaller perceptual span in terms of number of words previewed, but it is possible that that is due to the mix between kanji and hirakana and effectively reading in 2 different script systems simultaneously. This would certainly benefit from further research.

The perceptual span is not only affected by script and linguistic density but by the strength of the working memory and the reading speed: A study by Osaka and Osaka (2002) showed that readers with a larger perceptual span had better reading measures (such as shorter fixation durations and reading times) than subjects with a shorter perceptual span. The high performance group still had better performance when their preview window was reduced, suggesting a better ability to integrate information coupled with a better working memory (Osaka & Osaka, 2002). In another study, subjects who were fast readers turned out to have a larger perceptual

---

**7** "Hirakana are phonetic symbols for syllables, whereas kanji are essentially nonphonetic symbols, or ideograms, representing lexical morphemes. Kanji characters are more complex patterns involving relatively higher-spatial-frequency components and are more iconic in form" (Osaka, 1992).

span than slow ones whose perceptual span comprised the fixated word and one right after it only (Rayner, Slattery, & Bélanger, 2010).

When it comes to beginning readers, Rayner (1986) found that their perceptual span is smaller, asymmetrical to the right, and goes for about 11 letter spaces to the right of fixation. An interesting finding from the same study was that beginner readers did not make use of the word boundaries in the parafovea in order to program the next fixation. However, perhaps the most interesting findings in this study was that the difficulty of the text reduced the size of the perceptual span and that the span is symmetrical to start with but gets biased after just one year of reading (Rayner, 1986). This is further evidence that the perceptual span is not a product of visual acuity but rather a result of the script and linguistic characteristics of the text being read. This was further supported by the finding that the amount of parafoveal benefit extracted during a fixation and the size of the perceptual scan are both dependent on the specifics and the difficulty level of the reading task at hand (Inhoff, Pollatsek, Posner, & Rayner, 1989). Perhaps the most eloquent description with regards to the perceptual span is given by this quote: "... attention and ongoing processing constraints, and not visual acuity, determine how much information can be obtained on each eye fixation in reading" (Rayner, 2009).

A last note regarding the size of the perceptual span in beginning readers, Rayner (1986) was able to show that it is not its smaller size that results in slower reading in beginning readers. Rather it is that the longer fixation durations, shorter saccades, and the larger number of fixations and regressions are due to the beginner's extra effort needed for word recognition and cognitive processes. He further concludes that the eye movement measures are a reflection of that rather than the cause of it.

As to older readers, their perceptual span is smaller and more symmetrical, indicating a less efficient capability of processing words in the parafovea (Rayner, Castelhano, & Yang, 2009). The question here would be, is this due to lower visual acuity or working memory? The authors maintain that it is due to the loss in visual acuity with age, and the increased time needed for foveal processing, which leads to less capacity to give to parafoveal processing (Rayner et al., 2009). As such, visual acuity is not the defining factor in how big the perceptual span is. However, it is a modulating factor during the lifetime of the reader.

Another argument against the role of visual acuity in defining the perceptual span was given by Miellet, O'Donnell, and Sereno (2009). They systematically increased the size of characters as they were farther away from the fixation location to compensate for the loss in visual acuity. What they found was that the perceptual span remained at 15 characters to the right of fixation and so impervious to the improvement to the visual quality of the stimulus in the parafovea (Miellet et al., 2009). This is consistent with an earlier study that had also found that increasing letter sizes in the parafovea did not take away the viewing position effects and did not improve reading performance (T. Nazir, Jacobs, & O'Regan, 1998).

## Parafoveal Preview Benefit and Effects

Fixation durations can decrease by around 30–50 ms if the reader has a valid preview of the word that comes to the right of the currently fixated word. This is usually referred to as the parafoveal preview benefit.

It involves collecting information regarding the position of letters within words, the abstract coding of letters (doesn't matter if uppercase or lowercase), as well as orthographic and phonological coding. This information is integrated across saccades (Rayner, 2009) and can affect the duration of the fixation on word n+1.

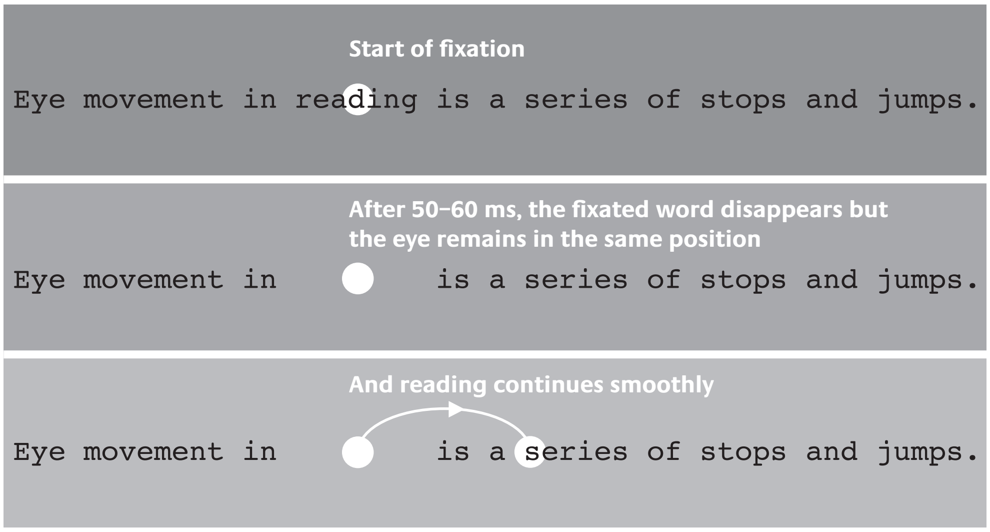


Fig. 5.3 It only takes 50–60 ms for the visual characteristics of a word to be extracted.

However, some characteristics of the text are not. Studies have shown that neither semantic information (Altarriba, Kambe, Pollatsek, & Rayner, 2001) nor low level visual featural information such as the letter case (Rayner, McConkie, & Zola, 1980) are integrated across saccades. Therefore, the extracted visual information from the preview of word n+1 has no effect on the fixation duration on that word (Rayner et al., 1980). This is in line with the findings that when letters are identified via their features, the visual information is very quickly stripped and abstract letter codes remain (Grainger, 2008).

With regards to the integration of morphological information, there is no evidence for that in English (Juhasz, White, Liversedge, & Rayner, 2008) or Finnish (Bertram & Hyönä, 2007), but there is in Hebrew (Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003). Similar to Arabic, Hebrew verbs and most nouns are built around 2 morphological entities, the root and the pattern. The root holds the semantic meaning of the word and was found to have a priming effect when word n+1 had the same root as word n (Deutsch et al., 2003).

The processing of words in the parafovea occurs irrespective of word position in the sentence, and so there is a preview benefit for words at the beginning, in the middle, or at the end of sentences (White, Warren, & Reichle, 2011). The preview benefit is higher when the currently fixated word is easy to process and lower when it is difficult (Henderson & Ferreira, 1990). Moreover, the effect of the difficulty of processing the currently fixated word on the preview benefit is not a spillover that is the result of a fixation that was too short; in fact, this effect was seen for both short and long fixations (White, Rayner, & Liversedge, 2005).

Quite interesting and contrary to what one might expect, **the reader needs only around 50–60 ms of exposure to the currently fixated word in order for reading to continue smoothly (Fig. 5.3).** That is to say, if a word was masked or disappeared after 50 ms of the eyes landing there, reading would continue without any disruption (Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981). This is not to imply that word processing is done within that short of a time frame, but rather than the visual information that needs to be extracted from the word is done very quickly and very early on in the reading process (Reichle et al., 2003). The eyes remain in place even when the target word has disappeared, and the eye will move on within the same pattern of eye movement as if the target word was still there (Rayner, Liversedge, & White, 2006). The ability to continue reading even when the target word disappeared after such a short time was also valid for reading in Chinese, and with similar exposure times (Zhifang, Guoli, & Xuejun, 2011).

Also interesting was the finding by Rayner, Liversedge, and White (2006) that if the word right of the fixation (n+1) were to be masked or disappeared when the eyes land on word n, or even after 60 ms of landing there, reading rates are disrupted. Furthermore, the reader needed to have a valid preview of word n+1 throughout the fixation on word n, or at least for the whole duration after 60 ms of landing there and the masking of word n+1 had a more disruptive effect than its disappearance (Rayner, Liversedge, et al., 2006). This is further indication of the importance of parafoveal preview.

The next question that follows: how much of word n+1 does one need in order to get a parafoveal preview benefit? A study looking at what would constitute useful information in the parafovea found that readers only needed to have the first 3 letters of word n+1 in order to get almost the same preview benefit as they would have if shown the full word (Rayner et al., 1982). This implies that the preview benefit is not necessarily tied to the identification of whole words but rather to the beginning part, most likely as an initiator for lexical access (Rayner, 1998). The parafoveal preview benefit is also responsible for skipping words. These are usually very short (for example, function words such as of, the, and etc.) or highly predictable by the context (Rayner, 2009). In the case of highly predictable words in the parafovea, the parafoveal preview extends to more than just 3 letters; moreover, readers are just as likely to skip a nonword if it is visually similar to a highly predictable word according to the sentence context (Balota, Pollatsek, & Rayner, 1985).

The idea that one does not need a full preview of the word for parafoveal preview benefits to kick in is not surprising given the results that have shown that neither semantic nor morphological information are integrated across saccades. It is also then not surprising that the frequency of word n+1 yields no preview benefit for the fixation duration on word n (Rayner, Fischer, & Pollatsek, 1998). However, recent studies have shown that the frequency of word n+1 does play a role (Reingold, Reichle, Glaholt, & Sheridan, in press). They tested the effect of the frequency of word n+1 on the fixation on word n and found that this effect starts to become evident at 145 ms from the start of the fixation.

While readers of English do not get a preview benefit from word n+2 (Rayner, Juhasz, & Brown, 2007), studies have shown that Chinese readers do obtain a preview benefit from up to 2 words to the right of fixation. The Chinese script is based on box-like characters that, unlike alphabetic scripts, are the morphemic entities of the Chinese language. Words are usually made up of 1 or 2 characters and there are no word spaces. The parafoveal preview effect was seen for both words n+1 and n+2 irrespective if these made up 1 or 2 words (J. Yang, Wang, Xu, & Rayner, 2009). Further research has shown, though, that this is valid when word n+1 is a high frequency word; in the case when word n+1 is a low frequency word, this benefit will disappear (J. Yang, Rayner, Li, & Wang). Furthermore, neither semantic nor orthographic information can be extracted from the preview benefit of word n+2 (Yan, Kliegl, Shu, Pan, & Zhou, 2010).

Finally, the answer as to why the parafoveal preview is so important to reading could come from the way language itself is organized. As McDonald and Shillcock (2003) explain, a reader's mastery of a language is dependent on his/her grasp of "word-to-word contingency statistics," the probability of one word occurring after another such as the preposition *on* occurring after the verb *rely*. This statistical knowledge is used by the brain during on-line processes and helps in forming predictions for upcoming words, contributing thus to efficiency in reading (McDonald &

| Younger Readers | Older Readers |
| --- | --- |
| More likely to skip frequent words | Slower in reading<br>More fixations<br>More regressions<br>Longer saccades<br>Less likely to have multiple fixations on frequent words<br>More likely to skip a target word, especialy if it is frequent<br>Need a larger font size |

Shillcock, 2003). This effect was thought to be independent of predictability effects, but later studies were able to show that it is not (Frisson, Rayner, & Pickering, 2005).

### Parafoveal-on-foveal Effects

The extent to which the word right of fixation affects the fixation duration of the currently fixated word is called the parafoveal-on-foveal effects. Results of various studies regarding the effects of the lexical and word frequency characteristics of word n+1 on the fixation duration on word n have so far been controversial and with mixed results (Rayner, 2009).

With regards to the effect of orthographic characteristics, there has been some cautious agreement. Several studies found an effect of the orthography of the word right of fixation on the fixation duration of the fixated word (Rayner, 2009). For example, Drieghe et al. found that a longer word in the parafovea led to a shorter fixation duration on the currently fixated word (Drieghe, Brysbaert, & Desmet, 2005). Another study by Drieghe, Rayner, and Pollatsek (2008) also showed evidence for parafoveal-on-foveal effects with results pointing to longer fixation times on word n when the word right of fixation (n+1) was orthographically illegal, the fixation position was very close to n+1, and the fixation followed a long saccade. This effect was attributed to mislocated (undershot) fixations. The frequency of word n did not appear to have any effect in that study (Drieghe et al., 2008). Other studies such as the one carried out by Rayner et al. did not find evidence for parafoveal-on-foveal effects (Rayner et al., 2007).

Researchers, therefore, are still far from reaching an agreement regarding the effect of word n+1 on the fixation on word n. Irrespective of what is or is not being processed of the parafoveal word, it seems that the reader has certain expectations of what that word needs to be; when the parafoveal word is implausible, i.e. against the expectations set by the text, fixation durations will be increased (Murray, 1998).

## Effects of Age and Reading Skills on Reading

### Age

Reading measures, strategies and habits seem to change through one's lifetime. Studies have shown that older readers are slower in reading and make more fixations and regressions (Kemper, Crow, & Kemtes, 2004) and have longer saccades (Rayner, Reichle, et al., 2006).

When two age groups were tested for the effect of word length, frequency, and predictability on eye movement, a 47 average year in age difference still yielded many similarities though there were also a number of differences. Older readers read more slowly and were more likely to regress. They responded more consistently to word frequency. The effect of frequency on the probability of word inspection also gave different results. It increased skipping in the younger group, and decreased the likelihood of multiple fixations in the older group (Kliegl, Grabner, Rolfs, & Engbert, 2004). Older readers, though, are generally more likely to skip target words than younger readers. This is much more pronounced for high frequency target words than low frequency ones and results in longer saccades for the older readers. Younger readers are equally inclined to skip either (Rayner, Reichle, et al., 2006).

Overall, it seems possible that older readers adopt a probabilistic reading strategy in that they use parafoveal information to skip words in order to get through the text quicker; however, this results in them having to do more regressions to clarify the areas of the text that have not been fully processed (Rayner, Reichle, et al., 2006).

In relation to legibility researchers found no interaction between age and font style, but did show an effect for size: older readers needed larger sizes for minimum acuity levels (Connolly, 1998). In other words, the older the readers are, the less able they are to read type in small sizes.

### Reading Skills

Reading skills also play a role in the process of reading. Ashby, Rayner, and Clifton (2005) compared the average reader with the highly skilled one and their results were quite interesting: 1. The average reader needed more time to recognize low frequency words that were unpredictable in the sentence context. 2. When the context was predictable, they relied on the context to obtain meaning and moved their eyes away from these low frequency words even before they were fully processed. 3. The highly skilled readers took a different strategy by rereading earlier parts of the sentence and working to integrate these words into the overall picture.

## Effects of Language on Reading Measures

### Difficulty

As mentioned early on in this chapter, the content of the text being read has an effect on reading measures. Difficult text results in longer fixations, shorter saccades, and more regressions; indeed, researchers find that fixation duration is a good indicator of the ease or difficulty that the reader is facing in processing the fixated word (Rayner & Duffy, 1986). However, difficulty level is not the only way in which of the linguistic content affects reading. Below are other factors.

### Word Length

Longer words get skipped less often and are more likely to have multiple fixations (Kliegl et al., 2004). The number of fixations on a word are positively related to its length (Rayner & McConkie, 1976). A recent study (Hautala, Hyönä, & Aro, 2011) that tested the effect of word length and character count in both monospaced (Courier) and proportional width (Arial) typefaces found that the number of letters increases the fixation duration but has no effect on the skipping probability. On the other hand, the spatial length of a word played a role in fixation positions and the likelihood of skipping (Hautala et al., 2011). There are many typographic variables that come into play when one tests with a sans serif vs. a monospaced serif typeface so these results need to be looked at with caution.

## Predictability

Subjects were found to be more likely to skip a target word when it was highly constrained by the context; when they did fixate on it, the fixation duration was lower for a predictable word than for an unpredictable one (Ehrlich & Rayner, 1981). Predictable words are also less likely to have multiple fixations (Kliegl et al., 2004). Predictability effects also extended to the parafovea. Parafoveal words that were highly predictable enabled a more detailed "parafoveal visual information" to be obtained (Balota et al., 1985).

Research involving Chinese readers gave similar results. Fixation durations on low predictable words were higher than medium or high predictable ones, and they were more likely to be fixated on (Rayner, Li, Juhasz, & Yan, 2005).

## Skipping

As mentioned in earlier sections, shorter words get skipped more often than longer ones and the skipping probability is around 80% for a one-letter word, 60% for a three-letter word, 45% for a four-letter word, and 30% for a five-letter word (Vitu, O'Regan, Inhoff, & Topolski, 1995). These averages increase if the launch site was 2–4 letter spaces away from the initial letter of the skipped word, and they drop if the launch site was farther than 4 letter spaces (Vitu et al., 1995). High frequency words get skipped more than low frequency ones (Kliegl et al., 2004).

Predictable words are also skipped more frequently than similar looking words that are not as predictable (Drieghe, Rayner, & Pollatsek, 2005). Though both word length and predictability have strong effects on skipping probability, these two variables do not interact i.e. they are independent of one another (Rayner, Slattery, Drieghe, & Liversedge, 2011). As for the effect of syllables, and when controlled for overall word length, a mono-syllabic word is 5% more likely to be skipped than a bi-syllabic word (Fitzsimmons & Drieghe, 2011).

There is, also, another reason why words are skipped. The reason is simply due to errors in programming. Sometimes there are oculomotor errors in programming and executing saccades. This results in overshooting when the target word is near, or undershooting when the target is far (McConkie, Kerr, & Dyre, 1994). In both cases, the word n+1 ends up being skipped by mistake.

Though the preview benefit is tied to word skipping (Rayner, 2009), and the more difficult a word is, the less preview benefit the reader gets (Reichle et al., 2003), it has been shown that the foveal load (the difficulty of processing of word n) has no effect on the probability of skipping the folowing word n+1 (White, 2007). As to what happens after skipping, as reported in Brysbaert, Drieghe, & Vitu (2005), there are two possible scenarios. If the saccade was launched from word n, skipped over n+1 and landed on n+2, empirical evidence points to longer fixation durations for the word n+2. Alternatively, the fixation on word n+2 is rapidly followed by a regression back to word n+1 (Brysbaert et al., 2005).

## More on the linguistic/conceptual difficulty of the text

As mentioned earlier, linguistically difficult text results in higher average fixation durations, a larger total number of fixations, and the overall reading time. Nevertheless, the size of the effect of average fixation duration is quite small. The extra reading time needed for difficult content results from the increase in the number of fixations (Rayner, Chace, et al., 2006). Difficult text also results in shorter saccades and more frequent regressions (Rayner, 1998).

## Frequency

Low frequency words require longer fixation times (when matched for word length) and increase the gaze duration on the word right of fixation (Rayner & Duffy, 1986). High frequency words are also more likely to be skipped (Kliegl et al., 2004). Therefore, not only do high frequency words require shorter fixation duration, they also provide readers with a more effective preview benefit (Inhoff & Rayner, 1986).

In the studies of reading disappearing text as mentioned earlier, frequency effects still had an effect on fixations times. High frequency words required shorter fixation durations, and this effect remained in place, even when the target word n disappeared after 60 ms; the eyes remained in place even when the target was no longer there (Rayner, Liversedge, et al., 2006). Readers are also more likely to refixate low frequency words (Rayner, Sereno, & Raney, 1996).

## Effects of Morphology

Are words recognized as whole-words or do root morphemes play a role? The simple answer is that it depends on the language. For example, Beauvillain (1996) has shown that the root plays a role in French and that the manipulation of the frequency of the root has shown effects in derivational words. In the case for suffixed words, the root was integrated and accessed before the whole word was. Prefixed words did not show this pattern and the root showed frequency effects at a later stage in the word processing and so the whole word was processed before the root. Whole-word frequency effects were also found which gives rise to the conclusion that French words are simultaneously broken down into their root morphemes and are also accessed as a full unit (Beauvillain, 1996).

The dual route is also in effect in the Dutch language. Dutch nouns take on the –en ending in plural, which is an ending shared with verbs as well; studies have shown that highly frequent plural nouns are processed via the whole-word route while the low frequency ones are not (Baayen, Dijkstra, & Schreuder, 1997). The authors of that study suggest that this whole-word approach to frequent plural nouns is more efficient since it reduces the time cost of the –en ambiguous ending.

As for words based on derivational morphology, the decomposition of Dutch words into their morphemes seems to be an efficient process that happens in a left-to-right fashion where suffixes and prefixes yield different results (Bergman, Hudsonxs, & Eling, 1988). The study also showed that the root extraction seemed to be automatic. Thus, the dual access route is evident in both derived and inflected Dutch words.

When it comes to English, the results are more mixed and rather outside the scope of this paper[8]. It is possible that the nature of the language studies has a role to play here. English is half Germanic in origin, and this might set it apart from other Roman languages. Niswander et al. (2000) looked at the role of the root morpheme in derived and inflected English words. The results were the following: The frequency of the root had an effect on the processing of derived words in English, and the root is in some cases processed before the whole word is, though both play a role in lexical access. However, that effect was not always seen in inflected words. It was seen in inflected nouns (with clear effects on fixation duration and gaze duration), but the case for verbs whose root is a noun was more complex to analyze. The effect of the frequency of the root had a spillover effect, and thus can be seen as having a tangible rather than ephemeral effect on word processing. These results suggest that word processing can proceed along both the "direct access" and root based routes (Niswander et al., 2000).

Similar results have been seen with compound words. A compound word is a combination of 2 or more words that come together to form a new word such as school and child forming schoolchild. This is a common phenomenon in languages, such as English, German, and Finnish. They are especially common in the Finnish language and make up around 60% of the Finnish dictionary (Hyönä, Pollatsek, & Bertram, 2005). As reported by Hyönä et al. (2005), compound words are identified via the whole-word route (direct access) and the decomposition route (similar to root based route) operating in parallel. However, they did find a dominance of the decomposition route for long compounds, and the whole-word route for the short compounds as the latter would fall within the foveal vision within just one fixation (Hyönä et al., 2005). In either case, the frequency of the first constituent influenced the first fixation duration and gaze duration on the compound word (Hyönä & Pollatsek, 1998).

As to compound words in English, while there is evidence for a priming effect via the constituents, there was no evidence to show that the meaning of the compound word is being accessed through any of its constituents (Frisson, Niswander-Klement, & Pollatsek, 2008). The compound words are accessed as a whole, as long as there is a lexical index for such a compound word (Frisson et al., 2008).

Morphology certainly plays a role in reading Hebrew (Deutsch et al., 2003), and this will be further discussed in the next chapter because of the linguistic similarity between Hebrew and Arabic. Morphology also plays a role in reading Chinese, with a stronger preview benefit coming from a character that shared the same morpheme as the target word than from one that was orthographically similar to the target word but represented another morpheme (Yen, Tsai, Tzeng, & Hung, 2008).

## Homographs (or Ambiguous Words)

Homographs are words that have more than one meaning. In eye movement literature, they are usually referred to as ambiguous words. These can be either balanced or biased: balanced, if both meanings are equally common, and biased if one is more common than the other.

When reading homographs, all possible meanings are momentarily accessed whether they are relevant to the sentence context or not; this is the case even in the presence of a strong bias in the context that preceded the ambiguous word (Swinney, 1979). This access is retained until 3 syllables after the ambiguous word at

8 For a comprehensive review, refer to the text by Niswander, Pollatsek, and Rayner (2000)

which point only the relevant meanings of the word are able to prime, and this suggests a very fast postaccess decision process (Swinney, 1979).

In the case of balanced homographs, the research by Rayner and Duffy (1986) has shown interesting results: If one meaning was dominant and the context was neutral, no extra fixation time on target word was found. In the case when the context called for the less common meaning, and the disambiguating information was encountered before the homograph, again there was no disruption to reading. However, when the disambiguating information was encountered after the homograph, there was an increase in fixation duration on the disambiguating word and often a regression to an earlier part of the text. If both meanings were equally frequent and the context is neutral, then there is also a significant increase in fixation times. The longer fixation duration is likely due to the fact that ambiguous words have more complex representations because they hold more than one meaning and hence require longer fixations (Rayner & Duffy, 1986).

What if the two meanings of an ambiguous word were a verb and a noun? It seems that the subordinate bias effect, which states that there will be longer fixation times when coming across a less frequent meaning of an ambiguous word, disappears if one meaning is a verb and the other a noun (Clifton, Staub, & Rayner, 2007). In such a case it seems that syntactic context (assigning if a word is a verb or a noun) precedes lexical access (Clifton et al., 2007).

## Eye Movement Control: Where and When To Move the Eyes

### Overview

Manipulations of the moving window paradigm and the delay of text onset have shown that fixation duration and saccade length vary independently of one another leading to the conclusion that the decisions of where and when to move the eyes are independent of one another (Rayner & Pollatsek, 1981). Other studies have also shown that the lengths of consecutive saccades do not correlate with one another; neither do the durations of successive fixation, nor do fixation durations and saccade lengths correlate with each other (Rayner & McConkie, 1976). However, there are cases, such as at the end of a wrap-up clause, where higher-end cognitive processes are controlling both decisions of where and when to move the eyes; other examples include regressions and refixations that are due to comprehension problems, and the skipping of high frequency or highly predictable words (Rayner, Kambe, & Duffy, 2000).

With regards to when to move the eyes, and as shown previously in this chapter, linguistic effects and cognitive processes seem to play a large role, and therefore it is very hard to argue for a reading model that does not account for these effects. The decision of when to move the eyes is in reality a question of the models of eye movement control in reading. Evidence, then, points to the conclusion that cognitive processes influence when to move the eyes (Rayner, 2009).

However, there are three competing theories: “minimal control, visual/oculomotor control, and cognitive control” (Rayner, Liversedge, White, & Vergilino-Perez, 2003). Other researchers will simply recognize two schools of thought: one claiming that low level visual cues, text properties, and oculomotor system guide eye movement in reading, while the other claims that it is the on-going cognitive processes that control eye movement in reading (Reichle et al., 2003). The latter seems a better fit to describe viable models for eye movement in reading given all the data that supports an effect of the linguistic content on reading. The minimal control theory posits that the text context (the content of reading material) has no effect of fixation duration or saccade length and that it is the layout of text that affects saccade length. The author recognizes that "no one would be willing to accept this model as correct" (Suppes, 1990). Given all the studies showing the effect of variables such as word frequency and predictability of fixation duration and skipping probability (and thus saccade lengths), this model is easily refuted by an overwhelming amount of evidence. This leaves only 2 theories to contend with, best represented by the E-Z Reader and SWIFT models which are the most developed models to date (Reichle, Warren, & McConnell, 2009). These will be discussed in the following section.

As to the decision of where to move the eyes, it depends on 2 factors: the length of the word currently fixated and the word spaces between the words after it, i.e. the lengths of word(s) that come right after (Reichle et al., 2003). The calculation of word length is processed in the parafovea, and the extraction of this information is an independent process from lexical processing used for the selection of upcoming target words and the programming of saccades (Inhoff, Eiter, Radach, & Juhasz, 2003). For example, very long words will result in a longer saccade. Short words will also result in longer saccades as they will be skipped over (Juhasz et al., 2008). Moreover, studies have shown that the parafoveal preview of upcoming word lengths works to send the eyes to the center of the target word (Juhasz et al., 2008). A study found that if the launch site from word n is towards the end of the word, and the word n+1 is of high frequency and predictability, the landing position will shift towards the end of word n+1 (Lavigne, Vitu, & d'Ydewalle, 2000). However, subsequent testing found these results to be invalid, and rather confirm the conclusion that the decision of where to move the eyes depends on low level information and not on cognitive processing (Rayner, Binder, Ashby, & Pollatsek, 2001).

Several studies added spaces in between words in script systems that did not usually allow word spaces, and the results showed improved reading measures in these trials, which goes to show that word spaces are beneficial even when they are orthographically illegal (Rayner, 2009). This was the case for Thai and long German words but not for Chinese, possibly because Chinese words are shorter and therefore fall almost entirely in the fovea in which case visual acuity is much better (Schotter et al., 2012). Similarly, removing word spaces from English results in a 40–50% reduction in the reading rate (Rayner et al., 1998). It also results in a disruption of the landing positions; readers are better able to send their eyes to the optimal viewing position when the text includes word spaces (Rayner et al., 1998). This could partially explain the increase in fixation durations. In contrast, using double rather than single word spaces actually improves reading performance (Drieghe, Brysbaert, et al., 2005).

The optimal landing position in a word corresponds to the fixation position that yields the shortest gaze duration or the least likelihood for a refixation, and this position is usually found in the middle of the word (Vitu, O'Regan, & Mittau, 1990). However, its effect are seen more strongly in the reading of isolated words and less so in the reading of continuous texts (Vitu et al., 1990). However, the eyes do not always land in the optimal position. In fact, the preferred viewing location, the area where fixations tend to fall, is halfway between the beginning of the word and its center (Rayner, 1979). When the eyes land at the beginning or end of a word, there is a higher likelihood of a refixation than if the eyes landed in the middle of the word (Rayner et al., 1996).

Lastly, the interesting aspect about where to send the eyes is the learned strategy of directing the eyes towards the center or left of center of the target word. This is relevant since the preferred viewing locations are landing in a position between the

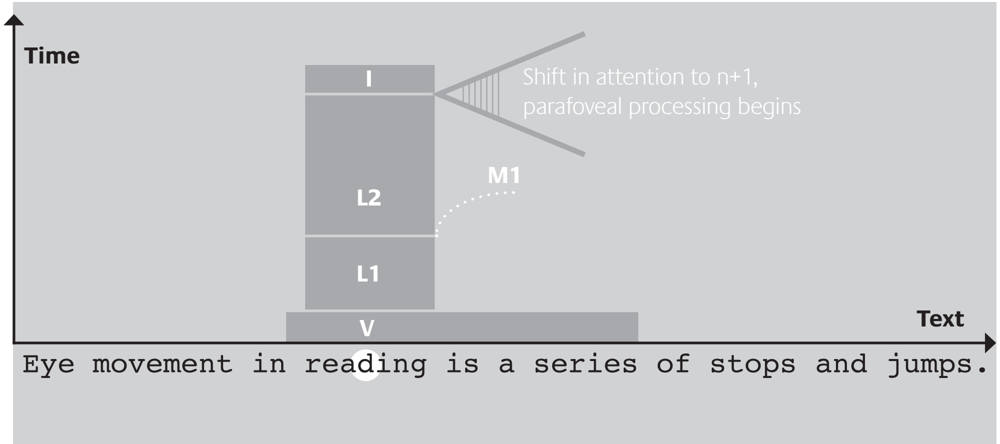


Fig. 5.4 The different stages of reading according to the E-Z model:

V: Visual analysis
L1: Familiarity check
L2: Lexical access
I: Integration of meaning
M1: Labile stage of saccade programming
M2: (not shown) non-Labile stage of saccade programming.

beginning and the middle of the word, rather than an absolute value that is independent of word length. The reason behind leftward bias is still an open question. However, there is evidence to indicate that certain parts of the retina get trained during reading, and this training kicks into effect when the eye fixates on a word, but not on a non-word (T. A. Nazir, Ben-Boutayab, Decoppet, Deutsch, & Frost, 2004).

## Reading Models

One of the key questions that reading models deal with is the question of word processing. Specifically, are words being processed sequentially one after the other, or can several words be processed at the same time? This would be what the literature refers to as serial vs. parallel processing, the most well known examples being the E-Z Reader and the SWIFT reading models. The other key question is: what drives eye movement? As briefly mentioned earlier, there are several competing theories for how to account for the empirical data that has been recorded in the various studies. These models differ on these issues: the eyes being driven by lexical processing (whether serial or parallel) or via oculomotor constraints (Rayner, 2009). In other words, the different reading models are trying to answer the question: What is driving eye movement (cognitive processes or the oculomotor system), and what is the nature of this drive (serial or parallel)?

One might wonder at how it is possible that the oculomotor system would be the one to drive eye movements instead of the content of what is actually being read. Such models refer to studies that show that the pattern of eye movements remains similar when reading actual text as when reading letter strings or when searching for specific letters (Vitu et al., 1995). They argue then that eye movement is carried out via predetermined strategies (Vitu et al., 1995). However, subsequent testing has shown that in reality, a visual search or reading a string of letters resulted in longer fixations, shorter saccades, and more likelihood of word skipping (Rayner & Fischer, 1996). Rayner and Fischer (1996) argue that the previous study was able to account for some aspects of eye movement but that these are a small part of a much larger set of cases to be accounted for.

Of the oculomotor control reading models, the Competition-Interaction model is based on the following assumptions as listed by Yang and McConkie (2001): There is a race between the programming and cancelling of saccades based on the signals that are being fed in. When the eye is fixated upon a word, it is being inhibited from movement by a “central-vision-related fixate center” which is in turn being affected by higher-level processing. This center keeps the eye in place, thus influencing the timing of saccades. When that competition or race is lost, with the activation level of a word dropping below a certain level, the order for a saccade is given. The movement-inhibiting activation increases during the saccade so that by the time it lands on the target word, that inhibition is fully in place. The reader has learned a strategy for effective reading, and this includes the activation of a move command after a certain time has lapsed in a fixation. The word being fixated needs to have a certain set of visual characteristics in order for the inhibition to recede. If it does not, then the inhibition will instead increase with time. Similarly, in case the higher level processes are facing difficulties, inhibition signals stop the programming of another saccade. However, the fixation durations vary most of the times due to "physiological processes (time to resolve competition, random waiting times) that have little relation to the current cognitive activity" (S. N. Yang & McConkie, 2001, p. 3584).

The E-Z Reader model for eye movement in reading is a cognitive-control SAS model, short for *sequential attention shifts*. It is able to account for fixation durations (how long or how short of time needed for specific kinds of words), the probabilities for skipping and refixations, global eye movement measures, as well as local processing (Rayner, 2009). Its main assumptions are (a) that the lexical processing of words is strictly serial i.e. one word after the other, and (b) that this is the main driver behind eye movement though the model does allow that low level "pre-attentive" processing in the visual field happens in parallel (Rayner, Reichle, et al., 2006). This would benefit from low-level visual cues such as word spaces and word lengths, and this extends to an area that is larger than the one where lexical information can be extracted; cited evidence for that includes the ability to program roughly accurate return sweeps (Pollatsek, Reichle, & Rayner, 2006).

According to the E-Z reader (Fig. 5.4), word recognition starts with a visual analysis stage (V) which analyzes the visual characteristics of all the words in the perceptual span (Schotter et al., 2012). This is then followed by lexical processing which is completed in 2 stages[9]. The first stage (L1) is a rapid indexing of the familiarity of a word, while the second stage (L2) is the extraction of the relevant orthographic, phonological, and/or semantic code for linguistic processing. The completion of the first stage triggers the programming of a saccade. The completion of stage 2 triggers a shift in attention on to the next word (Rayner, Reichle, et al., 2006). Attention, in this case, does not refer to "spatial orientation... [But] to the process of integrating features that allows individual words to be identified" (Reichle et al., 2003, p. 453) which in effect decouples attention from eye movement. This shift in attention is when the parafoveal preview begins (Reichle et al., 2003) and when the familiarity check for word n+1 begins (Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008). The latest version of E-Z Reader (v10) specifies this shift in attention as a distinct phase that is assumed to take around 20 ms to complete (Pollatsek et al., 2008).

As explained by Pollatsek el al. (2008), the E-Z Reader 10 also accounts for postlexical word integration into the sentence construct by adding a specific stage (I) for that, right after L2 is completed. This integration of a word's meaning into that of the sentence roughly calculated to take around 50 ms. This stage might be completed with the eyes still on word n, or when they have already moved on to word n+1. The authors argue though that such a stage is included in this version of the model as a "placeholder" with which to test the effect of postlexical integration. They see this as a starting point for addressing the effect of comprehension difficulties on eye movement. Reichle, Warren, and McConnell (2009) again explain that

**9** It is important to note that the authors do not see the mentioned processes as 2 discrete stages and acknowledge that this separation is better for the modeling (Rayner, Reichle, et al., 2006).

this stage is a "placeholder for a deeper theory of postlexical language processing during reading" and so is a "tentative" step that goes to account how postlexical processing affects eye movement in reading (Reichle et al., 2009, p.6).

It is important here to note that saccade latency, which is the time needed to program a saccade, takes on average a minimum of 175–200 ms (Rayner, Slowiaczek, Clifton, & Bertera, 1983), which means that the trigger for saccade programming happens very early on in a fixation (Reichle et al., 2003). The division of lexical processing into an early and a late stage seems to accommodate that observation. Note, though, that the word identification process is not done in 2 stages but 3: early stage of visual processing, then the 2 stages of lexical processing (Pollatsek et al., 2006).

An important aspect of the E-Z Reader is that it does not offer a deep analysis of higher level linguistic processing (Reichle et al., 2003), though the addition of stage I in version 10 seems to go in the direction of accommodating that factor (Pollatsek et al., 2008). As Reichle et al. (2003) explain, the influence of that level of processing on eye movement is seen in the cases when things go wrong in text comprehension and the readers then stop or regress to earlier sections of the text. The model, then, is more concerned in the instances when comprehension is proceeding smoothly and is at "default" setting. A resulting shortcoming is that it cannot account for interword regressions (Reichle et al., 2003), nor can it fully explain what happens with long distance regressions (Reichle et al., 2009).

Another important aspect of the E-Z Reader is how it accounts for the programming of saccades. These are done in 2 stages: a labile stage (M1) where saccades can still be cancelled, and a non-labile stage (M2) where saccades cannot be cancelled. Saccades are assumed to take 25 ms to execute and are targeted towards the middle of the word. The actual fixation positions are subject to a systematic motor error with the specific pattern of overshooting the target on short saccades and undershooting on long ones (Pollatsek et al., 2008). As to where to fixate, the E-Z Reader projects that this depends on "linguistic, visual, and oculomotor factors" (Reichle et al., 2003, p. 459).

The SWIFT reading model is a mathematical model that aims to give a feasible "psychological and neurophysiological" account for the known pattern of eye movement (Engbert, Nuthmann, Richter, & Kliegl, 2005). Engbert, Nuthmann, Richter, and Kliegl (2005) outline the core principles of SWIFT-II in the following points:

- Several words are being processed in parallel, and in competition based on their different levels of activation.
- The decision of when and where to move the eyes are "decoupled."
- The programming of saccades is controlled by an independent timer that is "modulated by a foveal inhibition process" which allows for more time to inspect difficult words. This inhibition includes a time delay (this point is a new addition in version II).
- This program takes place in 2 stages: first is the labile stage during which a saccade is being programmed and can still be cancelled. The second is the nonlabile stage at the point after which saccades cannot be canceled. This is similar to what E-Z Reader also proposes.
- The oculomotor system intrinsically produces both systematic and random errors.

- Saccades are often mislocated, and in these cases a new saccade will be immediately programmed.

- The time needed to program a saccade (saccade latency) depends on the intended length of the saccade to be programmed. Both of these take place at the end of the labile stage, and so, only the duration of the nonlabile stage is affected by the length of the upcoming saccade.

Overall, the authors sum up the process as such: Words are processed in parallel with different level of lexical activation. This helps to keep track of the state of word processing, and to select the next target and the decision of when to move the eyes. These 2 decisions are made via "separate pathways." Foveal inhibition is able to delay saccade timing in case more time is needed for processing. The programming of saccades happens in 2 stages, and the target is selected at the end of the labile stage (Engbert et al., 2005). The selection of the target is random and based on the activation level the multiple words being processed (Laubrock, Kliegl, & Engbert, 2006). As to the activation level, its maximum level is governed by word frequency only, and word predictability is assumed to function only within the lexical preprocessing step (Laubrock et al., 2006).

The SWIFT reading model is proposed to be able to "reproduce a number of well-established measures of eye-movement control during reading, average fixation durations and fixation probabilities, distributions of within-word landing positions, and interword regressions. Finally, the SWIFT model can explain the IOVP effect of fixation durations based on error correction of mislocated fixations." (Engbert et al., 2005, p. 805). However, SWIFT is still unable to account for the number of refixations as seen in the body of evidence collected in eye tracking (Richter, Engbert, & Kliegl, 2006). The latest version of this model, zSWIFT, is now able to account for the pattern of regressions and refixations, though not on a quantitative measure (Nuthmann & Engbert, 2009).

It is quite interesting to compare the serial and parallel accounts of reading and eye movement, and it is quite possible that the processing of words goes in parallel for tasks that do not require high level cognitive processing (such is in a visual search). However, researchers argue that word processing is serial in the cases where one is reading for meaning (Reichle, Vanyukov, Laurent, & Warren, 2008).

It is not the point of this review to go very deep into the debate surrounding the various reading models. However, given all the evidence reviewed in the previous sections regarding the effects of language on reading metrics, it is difficult to accept a reading model that does not account for these findings. When reading for comprehension, one would expect that the "comprehending" of the text is what is driving the eyes forward throught the text.

A serial account of word processing seems more logical and reasonable, and in tune with how one understands the spoken language, or how one formulates thoughts. It is interesting that there might be a good chance that Chinese words are processed in parallel rather than serially (Inhoff & Wu, 2005), but that could very well be due to the absence of word spacing and so the parallel processing is that of different word components and not words themselves. In such a case, this is not too far from the attentional beam that is discussed in the E-Z Reader. In either case, the effects of language and script are very much seen and felt in eye movement measures and mechanics. It is then interesting for such research to be developed for a more varied and international approach to reading studies.

# Effects of the Stimulus

Reading is an activity involving two parties: the reader and that which is being read. The previous sections have looked at the various effects of the reader's age and reading skills level as well as the linguistic characteristics of the text being read. The next sections now turns to address the effects of the visual characteristics of the reading material, including aspects such as its size, the medium on which it is displayed, the patterns and features involved, and the distance from the reader.

## Manipulation of the Visual Characteristics of Text

In order to get a better understanding of the reading process and what sort of information is being processed at the various stages of a fixation, researchers often manipulate specific characteristics of the text. These studies are quite instrumental in drawing a full picture of the process of reading as will be later discussed in the chapter. Nevertheless, the results are also interesting in and by themselves. For example, Reingold and Rayner (2006) made a series of tests that showed that reduced contrast between the target word and its background significantly increases the fixation times on the target word but has only a negligible effect on the fixation durations of word n+1. Alternating cases, using a mix of uppercase and lowercase letters of the target word n, also increases fixation times on word n though to a lesser extent. It also increases the fixation times on word n+1, and to a much stronger effect than the decreased contrast. Increased typeface weight had weak influences on the processing of target word n but did exhibit stronger ones on word n+1 (Reingold & Rayner, 2006; Wang & Inhoff, 2010). This last finding is interesting in the context of typesetting. Every time one sets a word in bold, it increases the time needed to read the word that comes right after. Another interesting finding of the Reingold and Rayner study is that the effect of the quality of the visual stimulus has an impact on the first stage of lexical processing but not on the second one. The E-Z Reader model allows for 2 stages of lexical processing (Eyal M. Reingold & Rayner, 2006), and this study has pinpointed the time stamp of when the visual characteristics play a role in lexical processing.

Studies have also found that any degradation in the visibility, as in blurring for example, of a few letters in a word will reduce the reading rate, which is the number of words read per a unit of time. The disruption is worst when the letters at word extremes are degraded (Jordan, Thomas, Patching, & Scott-Brown, 2003).

## Reading on Screen

Early testing of reading on screen showed that it is slower than reading from paper (Gould, Alfaro, Barnes, et al., 1987). Certainly reading on a TV screen was possible but it was 28.5% slower than reading from paper (Muter, Latrémouille, Treurniet, & Beam, 1982). The intermittent light shining into the eyes was one possible reason for slower reading since the frequency of the light was disturbing the ocular motor control (Wilkins, 1986).

However, later research found that reading on screen could be as fast as reading on paper, and yielded similar comprehension levels if several factors were met (Gould, Alfaro, Finn, Haupt, & Minuto, 1987):

- The quality of the display needs to be good with high enough resolution.

- The typefaces used need to be similar to those used in print rather than dot matrix fonts used in CRT displays.

- There needs to be ample contrast of dark text on a light background, and the font should be anti-aliased (adding grey pixels to achieve a smooth effect rather than a jagged edge at the curves) (Gould, Alfaro, Finn, et al., 1987).

The smoothness of the outlines was further supported by the testing of the effects of ClearType hinting on LCD screens. ClearType employs a sub-pixel rendering technology that smoothens the ragged edges on curves on PC computer screens. When faced with a choice of ClearType vs. grey scale rendering (as visible on Windows XP), readers had a preference for ClearType (J. Sheedy, Tai, Subbaram, Gowrisankaran, & Hayes, 2008). Studies have shown that there is an advantage to reading in ClearType, showing higher efficiency in word processing with lower reading times, fewer fixations and shorter fixation durations; however, these effects were more relevant for more "intricate" typefaces such as ones with curves or in italics (Slattery & Rayner, 2009). Another interesting finding was that this improved reading performance did not have any tradeoffs in reading comprehension (Slattery & Rayner, 2009).

Increasing pixel density (having more pixels to render characters on screen) also results in better performance in letter identification; this is most profoundly felt in low resolutions (J. E. Sheedy, Subbaram, Zimmerman, & Hayes, 2005). In fact, this specific study found that if conditions were equalized between LCD and CRT screens and paper, no main effect for display was found (J. E. Sheedy et al., 2005). The authors claim that letter identification could be equal among these conditions, but that cannot really be asserted with full confidence, as the failure to find a main effect for a display does not necessarily mean that there is no effect at all. Simply, it can only mean that no effect was found. The study found an improved performance with ClearType turned on (J. E. Sheedy et al., 2005), as has been demonstrated previously.

With regards to the typographic presentation of text on screen, the effect of line length and column width have been shown to affect reading speed. Short line lengths of 25 characters per line negatively influence the reading rates compared to the medium and long lines; long lines (100 characters per line) only held an advantage over the medium length (55 characters per line) when the subjects are asked to skim through the text (Dyson & Haselgrove, 2001). The study also found that the medium line length resulted in the highest comprehension rates of the three conditions (Dyson & Haselgrove, 2001). Comprehension was also higher when text was presented on screen in three columns in a paged design (requiring no scrolling) rather than one column that did require the reader to scroll through (Dyson & Kipping, 1997). When the single column did not require scrolling, its reading rate was faster than that of the 3 columns design (Dyson & Kipping, 1997). This is consistent with an earlier study that also found that readers were more efficient in reading static pages than ones that required scrolling (Kolers, Duchnicky, & Ferguson, 1981) and a later study that showed that longer lines lead to better scanning but that shorter lines are more in favor (Ling & Schaik, 2006).

Though reading on screen might proceed as fast as that on paper, recent studies have shown that an increasing number of people are suffering from what is being referred to as computer vision syndrome. This is a collection of symptoms that include "eyestrain, tired eyes, irritation, redness, blurred vision, and double vision" that are mostly due to dry eyes (Blehm, Vishnu, Khattak, Mitra, & Yee, 2005). One of the side effects of screen flicker is an increase in saccades that fall short of their target and premature triggering of saccades that results in inaccurate landing positions (Kennedy & Murray, 1991).

## Patterns and Features

Text is a series of letters strung together, repeated on and on, making a pattern out of these components. These letters make words, but can we recognize words if we were unable to recognize the component letters? A study has shown that a word cannot be read if its individual letters are not individually identifiable. Even the 5 most common 3-letter English words are recognized as a pattern of features rather as one feature. A word, therefore, is never one feature. Everything that one sees, is a pattern of features and this limits our efficiency in reading where efficiency is inversely proportional to word length (Pelli, Farell, & Moore, 2003).

In the Latin script, terminations, or the areas where strokes begin or end, have been shown to be the most important features in letter identification and differentiation from other letters; other features include curves, the direction these curves open to, lines, intersections etc. (Fiset et al., 2008). A feature can be defined as "an image, or image component, and suppose that there are several possible features, so that the signal to be identified can be described as a sum of independently detectable features" (Pelli, Burns, Farell, & Moore-Page, 2006, p.4647).

Efficiency is the unit of measurement when looking at letter identification and is calculated as the "ratio of thresholds of ideal[10] and human observers… [which is] a pure measure of human ability" (Pelli et al., 2006, p.4649). Pelli et al. looked at various script systems and tested the efficiency of letter recognition. The scripts in order of most to least efficient in letter identification were listed as: Latin, Hebrew, Devanagari, Armenian, Arabic, and Chinese. Arabic and Chinese are the least efficient in letter identification and that is most likely because of the complexity of the letters themselves (Pelli et al., 2006). Note here that the test was carried out with individual (and isolated) letterforms so the role of Arabic orthography has not yet come into play. This study showed that complexity is inversely proportional to the efficiency of letter identification, and that it is an excellent predictor at that (Pelli et al., 2006).

Moreover, the study has shown that any modification that added complexity to letterforms also had an inverse effect on the efficiency of their recognition (Pelli et al., 2006). In fact, the authors concluded that "**simple forms are seen efficiently, complex forms inefficiently, as though they could only be seen by means of independent detection of multiple simple features**"(Pelli et al., 2006, p. 4665). It is, however, perplexing how the study found Devenagari to be less visually complex than the isolated Arabic letters as the forms of the latter are much more intricate and dense.

An important factor to keep in mind regarding shape identification is the degree to which this particular shape is similar to other shapes with which it might be confused (Attneave & Arnoult, 1956). **This is possibly the reason why Arabic had such low efficiency marks since so many of its letters share the same basic shape and it is only the number and position of dots that differentiate them.**

Within the same study by Pelli et al. (2006), one letter of a script typeface (Kuenstler) was calculated to be as complex as a common 5 lettered English word set in a typical Serif typeface (ITC Bookman). Subjects were shown individual letters of various scripts, usually for 200ms and were then shown a line up of 26 letters including the one already shown. The task was to identify which letter was shown at first. Evidence showed that it took only hours to learn to master such a task and that years of reading in this script, up to reading a billion letters over the span of 40 years, did not add to the efficiency of letter identification. It is important though to

**10** The ideal observer is in this case programmed into the software used in these tests and serves as an absolute scale to test human performance against (Geisler, 1989).

add that the subjects did not need to remember which letter it was, but simply to pick it out of 26 possible options (Pelli et al., 2006).

Another interesting finding of the same paper was that letters and words are identified in the same way we identify everyday objects whether they are inert or living: by detecting independent features, usually around 7 features per letter. This is usually referred to as the probability summation, which posits that the probability of not identifying a stimulus made of several features is equal to the product of probability of not identifying its individual features (Pelli et al., 2006). It supposes that a stimulus is identified via "any one of its features, and that the features are detected independently" (Pelli et al., 2006, p. 4669). Indeed, this theory has become the "null hypothesis of visual perception, an extremely simple model that accounts for much that we know" (Pelli et al., 2006, p. 4647).

### The Question of Word Shape

The role that the word shape plays in reading has been tested via the preview benefit as in a study by Rayner, McConkie, and Zola (1980). Words were presented parafoveally to the subjects, and when the eyes moved towards that target, the words were changed. The results were quite interesting when it came to the question of word shape and the integration of visual information across saccades as briefly mentioned earlier in this chapter. Of interest, here, is the finding that the words "ROUGH" and "rough" had equal facilitation for the target word "ROUGH," even though the word shape and visual characteristics are different. Moreover, readers did not benefit from parafoveal previews of a word that had the same word shape but different letters (Rayner et al., 1980).

Actually, using the display change to test if the subjects were even aware of this switch, researchers found that subjects were more likely to notice a difference when the letter codes were changed, rather than when the lettercase was (Slattery, Angele, & Rayner, 2011). The one situation in which the word shape does play a role is in acronyms. Acronyms set in uppercase in a normal sentence case context were read as acronyms (i.e. read out by the initials), but these were mistaken for normal words (i.e. read out as whole words) when set in an all caps sentence (Slattery, Schotter, Berry, & Rayner, 2011).

Pelli and Tillman (2007) also looked at the reading process and questioned the methods used in word identification: 1. Letter by letter, 2. as a whole word or 3. based on the sentence context. They found that letter by letter identification accounted for 62% of the reading rate, identification of whole words accounted for 16%, and the context accounted for 22% but that rate was variable across readers. They found that these 3 processes are dissociated from one another but that their effect is additive. The conclusion they get to is that words, like objects, are recognized by parts, as wholes, and depending on the context (Pelli & Tillman, 2007).

## The Legibility of Type

### Defining Legibility

A definition of legibility that deals only with the visual characteristics of text is inherently flawed, for legibility is a measure related to reading and recognition. Legibility operates in the process starting with the viewing of a linguistic symbol,

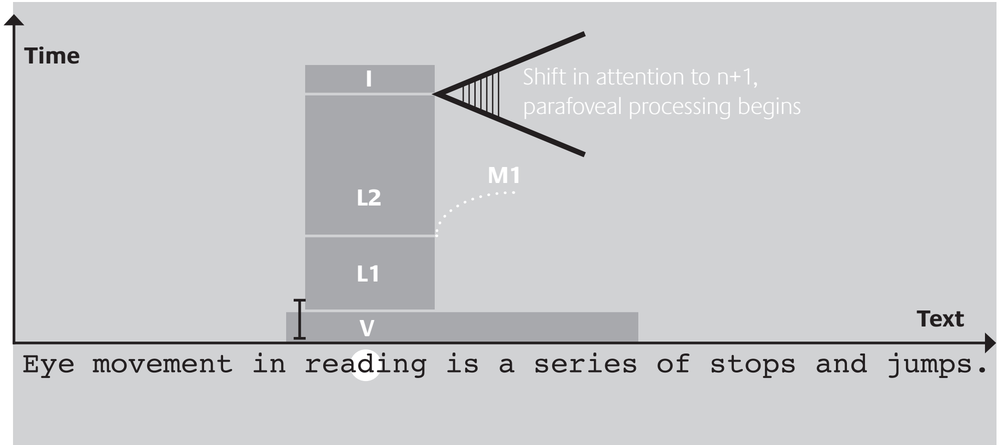


Fig. 5.5 Legibility effects are seen in the first 50–60 ms of a fixation and during the parafoveal processing of word n+1.

and obtaining meaning. It cannot, therefore, be contained in the pure visual characteristic of that symbol. As such, it is relevant and necessary to define legibility within the context of the reading process.

As mentioned earlier, it is in the first 50–60 ms of a fixation that the reader needs the visual stimulus to exist in order to extract relevant visual information, and that if the fixated word were to disappear after that period, reading would proceed as normal (Rayner et al., 1981). It is safe then to assume that the visual characteristics of a fixated word show their effect in that early time frame of the fixation. Legibility effects, then, are felt early on in a fixation. This, however, is only one part of the overall legibility effects.

Starting with the E-Z Reader model, there are different stages in word identification. An important stage is the shift in pre-attentive processing onto word n+1. This low level processing extracts abstract letter codes that contribute towards the preview benefit. The visual characteristics of a typeface are then very likely able to facilitate the preview benefit, in turn aiding in reducing fixation times and increasing saccade lengths. **Legibility effects, then, operate during two distinct processing stages: The first 50–60 ms of a fixation, and during the "pre-attentive" processing of parafoveal words.**

This localization of legibility effects has already been demonstrated. One study has clearly shown that typographical manipulations have an effect during the stage L1 of lexical processing but not during L2 (Reingold & Rayner, 2006). Similarly, the use of Bold weight on word n+1 gave higher fixation durations for n+1 (Reingold & Rayner, 2006), most likely because of lower performance in the parafoveal processing of word n+1.

As to the definition of legibility, the most convincing to date is that by Slattery and Rayner: "how easy the letters in a word are to encode" (Slattery & Rayner, 2009). In other words, **legibility is how easy it is to extract the visual information of the fixated word in order for lexical processing to begin.**

The ease of encoding the visual characteristics of words via their letter codes manifests itself in faster fixation times. The facilitation of the preview benefit results in shorter fixation times (due to the facilitation of extracting letter codes from word n+1), a wider perceptual span, and wider saccades.

The wording here is interesting to dig deep into. Is it decoding or encoding? A dictionary definition gives: "Decode: To extract the underlying meaning from" and "Encode: to convert (a nerve signal) into a form that can be received by the brain" (Farlex, 2012). Another definition that is more relevant to reading as a process:

> "For the simple view, skilled decoding is simply efficient word recognition: the ability to rapidly derive a representation from printed input that allows access to the appropriate entry in the mental lexicon, and thus, the retrieval of semantic information at the word level" (Hoover & Gough, 1990).

One might be tempted to use the word "decode" in such a case, since the purpose of reading is to extract meaning. However, a more accurate term is encoding for legibility has more to do with the recognition of letter code (and making this information available for higher level cognitive processes) rather than the ability to put these together into meaning. In other words, legibility is not about word recognition as in the Simple View of Reading by Hoover and Gough. It is about the ability to recognize the abstract letter codes so that lexical processing can begin.

In the Simple View, reading is the product of word decoding and linguistic comprehension:

> "In the simple view of reading, linguistic comprehension is the ability to take lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations. Reading comprehension involves the same ability, but one that relies on graphic based information arriving through the eye" (Hoover & Gough, 1990).

This again drives home the point that decoding involves lexical access at the word level, which is more than the pure level of what legibility is. As a counter example, imagine that a subject is presented with a word in Spanish and asked to read it aloud. If the subject does not understand Spanish, then the stimulus is not decoded. However, the subject will still be able to read it aloud to a certain degree as he/she will recognize the letterforms. The word would be legible, but not comprehended.

Similarly, if legibility is the ease of decoding, i.e. word recognition, then testing for legibility with non-words is not valid. However, this is again not the case. In addition, the different stages of reading in the EZ model split it into stages where orthography, phonetics, etc. are being processed, and higher level processing (comprehension) comes later. As discussed in the paper, legibility effects come into play early (first 50–60 ms) and not late in the process. Hence, legibility will enable comprehension, but the ease of comprehension is not a measure of legibility.

## Methods for Measuring Legibility

Researchers have used various methods to study the legibility of typefaces. The following list is a very brief overview of the types of testing used:

- Speed and comprehension: Reading speed and a comprehension check are quite common methods, as favored by Tinker (1963). The logic in such tests is that more legible type will facilitate the reading process and will result in faster reading. Speed can be measured as in total reading time or the number of words read per minute. This is usually a test of continuous reading.

- Naming: This usually involves the presentation of a word or nonword either in the fovea or parafovea. The faster or more accurate the words are named, the more legible the typeface is. This is a very pure method of measuring legibility as higher-level processes are not involved.

- Threshold visibility: This is a form of a naming test. The threshold is either that of exposure time (the stimulus is presented for a very short period of time, and then masked), or that of distance (the stimulus is presented at a distance from the reader). The shorter the exposure rate, or the farther the distance, the more legible a typeface is supposed to be (Lund, 1999). Another less frequent threshold is that of luminance, with higher legibility for the lower luminance at which characters are still identifiable (Yager, Aquilante, & Plass, 1998).

- Eye tracking: The use of eye tracking to test legibility gives very accurate speed tests. When looking at global reading measures, a more legible design is characterized by shorter fixation times, longer saccades, and a smaller total number of fixations (Slattery & Rayner, 2009).

- Blink rate: How often the reader blinks can be indicative of eye fatigue. Less legible text would hypothetically cause more fatigue, which results in increased blinking (Luckiesh, 1947).

- Search tasks: The reader is given a search task to complete and the faster that is the more legible the design (Lund, 1999).

These methods measure quite different things. Speed of reading and the various eye tracking data are concerned with the reading of continuous text. As long as the different typefaces are being tested with paragraphs of equal difficulty and familiarity levels, then the linguistic effects should not interfere, and the global reading measures are good indicators of legibility.

Naming and threshold visibility are used in the identification of letters or words and as such are in many ways different from reading as a process. On the one hand, they do not involve higher level cognitive processes and are therefore purer measures of legibility, especially if nonwords are used as that neutralizes the effects of word frequency. However, the distance threshold test is a test of visibility and how a typeface degrades over distance. It is then very different from the typical reading process in the fact that both type size and distance are quite different in these settings. This is not to say that either of them is not valid as a test, but rather that they are measuring different aspects of letter identification, and this is to be expanded upon in the following section.

## Letter Identification and Different Script Systems

Letters have a raison d'être, and that is to be identified. They are a visual code that is the building block of written language, and that is true of any kind of writing in any kind of language. It is, therefore, quite relevant to note that letterforms of different scripts have a few things in common. A very interesting paper by Changizi and Shimojo (2005) studied the commonality amongst all writing systems with the exception of ideographic ones like Chinese[11]. The variables looked at the number of strokes needed per letter and the redundancy of the letterforms i.e. how many strokes are actually needed to differentiate one letter from another. Averages, for example, were: Arabic 2, Cyrillic 3.69, Devanagari 3.27, Greek 1.71, and Latin 2.08. The overall average for all 115 scripts combined was roughly 3 strokes per letter (2

**11** The reason for this exclusion had to do with the fact that Chinese letter can represent actual words and so the boundaries between visual codes and actual words is blurred (Changizi & Shimojo, 2005).

for numerical systems) and with around 50% redundancy. This average of 3 strokes was irrespective of the size of the character set of the respective writing systems. The authors offer a very interesting hypothesis: this commonality across various writing systems is actually due to the need of letters to be seen. Therefore, the physiology of the human visual system is influencing the development of writing systems. It is their view that the existing visual system for object identification could be behind the selection and developmental trend of such forms (Changizi & Shimojo, 2005).

This view is certainly logical when one looks at the commonalities of human inventions that have sprung up across different continents. Like the wheels are round and walls are more or less straight, there is a certain level of efficiency to the way letters are drawn: a reasonable number of strokes per character and enough differentiation to keep letter identification more or less simple. Even in the most complex of writing systems, strokes are generally either straight lines or circular in form. Very rarely does one come across squiggly lines or zigzagged strokes. In such case, it is then logical to assume that there is a universal human drive towards efficiency in form, but only up to a certain point. The authors point out that 50% redundancy is the cost of improved differentiation (Changizi & Shimojo, 2005), and that is certainly in keeping with a selective drive for efficiency in visual communication. As shown by Pelli et al. (2006), some script systems are more efficient than others, but it is as interesting to note the differences between scripts as it is to note that which is common.

Indeed, these common elements of the different script systems stem from several important factors, as Changizi, Zhang, Ye, and Shimojo (2006) expand in a follow up paper. The basic principle that they build upon is the various configurations in which a number of strokes can combine to form different shapes. For example, 2 strokes can intersect in a T shape, an L shape, or an X shape. These configurations are topologically identified in such a way that the lengths of the strokes, or the orientation of the shape can vary, but the basic topology remains the same. Having noted that the 115 scripts that they analyzed had an average of 3 strokes per character, the authors developed a topological system of the possible configurations of a 2- and a 3-stroked letter. These configurations were seen to be similar across human visual symbols, in a non-random fashion. The authors also compared these configurations with shorthand symbols (signs developed to support the motor skill). They assessed the efficiency of the consecutive strokes and pen lifts and found no correlation between motor complexity and the distribution of topological configurations that letterforms take which lead to the conclusion that these shapes were "not strongly selected for the motor system" (Changizi et al., 2006, p. 123).

The authors also looked at the relationship of these configurations and the perception of visual complexity. They compared them to trademark symbols (these developed for the visual aspect rather than the motor one). Their results found a strong correlation between visual complexity and the visual configurations and concluded that these signs used in script systems are selected for the visual system and not for the motor one (Changizi et al., 2006).

Next they compared these topological shapes found in script systems to those that are found in nature and found a relationship there, leading to the conclusion that **the selection and evolution of these shapes have developed in accordance to what the human eye is best trained to recognize** (Changizi et al., 2006).

The remaining point to be argued, is that the written symbols need to be simple enough for any given person to be able to learn how to write them. Of this Unger (2012) writes:

> "… Legibility research is really a study of ergonomics… While learning to write and read, such basic elements [such as elementary graphic entities, that

later are used as parts of letters] are both externalized and internalized. Script systems have not only evolved to fit what the human brain and eyes can see, but also what the brain, the eyes and the hand can form. While you write, the eyes and the brain both guide and follow the hand and the writing instrument, making the processes of writing and reading truly reciprocal" (Unger, 2012)

## Letter identification and frequency channels

The recognition of letterforms is carried out via specific frequency channels that are dependent on the scale of the letter, and are influenced by the stroke frequency of the letter itself (Majaj, Pelli, Kurshan, & Palomares, 2002). The authors defined stroke frequency as the average number of strokes that intersect with a horizontal cross section of the letter; the division of that average by the average letter width yields the spatial frequency of the letter (Majaj et al., 2002). The authors found that the selection of the channel for identifying letters did not vary across different sightings of the same character. In fact, the observer was using the same channel to identify letters, rather than choose different channels that might help to reduce noise. They conclude then that the choice of which channel to use is dependent on the letter itself rather than the observer, making letter identification a bottom-up process (Majaj et al., 2002). The spatial frequency channel also does not change if attention was overt, as in a direct fixation, or covert, as in the pre-attentive phase of letter identification (Talgar, Pelli, & Carrasco, 2004). It is also the same frequency (1.6 octave wide) whether the reader is looking at a letter or a word (Majaj, Liang, Martelli, Berger, & Pelli, 2003).

Another interesting finding from the same paper reveals that different aspects of letterforms play different roles according to the size in which the type is set: "Large letters are identified by their edges, small letters by their gross strokes" (Majaj et al., 2002, p. 1181). In other words, when type is set in large sizes, the reader will react to the details of outline treatment. On the other hand, type set in small sizes is identified via the strokes themselves and how they relate to one another rather than how their outline is treated.

## Letters Identification and Brain Functions

Dehaene (2009) reports on several interesting findings related to specific brain areas related to letter identification. He gives the example of a patient who, after suffering a stroke that affected his right visual field, was no longer able to recognize letterforms or words. The patient's linguistic abilities and writing skills were not affected and he was still able to recognize numbers which led to the conclusion that "reading digits relies on anatomical pathways partially distinct from those used for reading letters and words" (Dehaene, 2009, p. 56).

This loss of the ability to read is referred to as pure alexia and the patient mentioned was the first recorded example by neurologist Joseph-Jules Déjerine in 1892. The point to make here is that there are specific areas of the brain that are essential for reading and letter identification, and that an injury there could very much obliterate the possibility of the patient to ever read again.

Indeed, the region of the brain critical for reading is located in the lower back area of the left hemisphere, or more accurately "a few centimeters to the front of the occipital pole, on the bottom side of the left hemisphere" (Dehaene, 2009, p. 62). Research has also shown that this region is strategically placed to act as a channel through which all visual word information flows which is then dispersed to many other areas of the left hemisphere (Dehaene, 2009). This "letterbox" area is also immune to changes in uppercase and lowercase and will show similar activation when one or the other is used as a prime. So the abilities of this region are not purely visual in nature but are able to function on a more abstract level. Moreover, this region remains the same for readers of right to left scripts as well and the location is almost identical in readers of Chinese.

With the help of brain imaging research, it is also possible to detect the time course of activation in the two hemispheres in response to specific reading tasks. Stimuli presented in the right visual field are transmitted to the left hemisphere and vice versa. However, the letterbox region in the brain that specializes in recognizing words is in the left hemisphere. It turns out that for words that are presented in the left visual field, their initial transmission goes to the right hemisphere but after 40 ms, this signal is transmitted onwards to the left hemisphere to this same letterbox. Within 200 ms, these words are processed just like the ones presented in the right visual field (Dehaene, 2009).

Going back to the issue of the commonalities among world scripts, neuroscience offers further evidence to the concept of universality in writing systems. Dehaene (2009) offers these observations:

- Writing systems involve concentrated black on white marks that are probably the most efficient way to transmit visual information to the retina.

- All scripts use a small number of basic shapes which are hierarchically combined to create words, word-parts or letters.

- All the script systems make the assumption that size and location can vary, but rotation cannot. This is most likely due to limitations in the ability of the visual neurons to tolerate only 40 degrees of rotation.

This is further support for Changizi's conclusions that script systems evolved to fit what the human eye, and by extension the human brain, is able to see. It is also what the eye and the brain are able to process efficiently.

## Letter Identification as a Measure of Legibility

Several studies look at the question of legibility via the task of letter identification. In these studies, the experiments tests the relative legibility of typefaces via how easily their individual characters can be identified, as well as the internal legibility of the typeface itself where tests check for the accuracy of character identification and probabilities for misidentification. One study of this sort was carried out by Chaparro, Shaikh, and Chaparro (2006) which looked at the relative legibility of Cambria, Constantia, and Times New Roman. Commonly confused pairs of letters were (!, l), (2, z), and (0, O). Of these 3 typefaces, Cambria had the highest number of correctly identified letters. A typeface in which letters are not confused with one another is most likely a more legible typeface. The test for identifying letters is a good measure of the ease with which one can extract the visual characteristics of what is being read. A word cannot be identified if its individual letters are not. Since the probability of not recognizing a word is the product of the probability of its letters not being recognized (Pelli et al., 2006), this sort of test can give clues as to what constitutes a more legible design.

Another comparative study looked at the relative performance of the letter e and number 0 in 20 typefaces (Fox, Chaparro, & Merkle, 2007). Character identification tasks were run and the typefaces were listed in the order in which these characters were confused with others. Though such a listing is very common in legibility studies (which typeface performed better than another), it is the lesser interesting aspect of this paper because the typefaces were tested at 10 pts. each and so the optical size was varying. This is a possible confounding factor as some typefaces will appear larger and size has an effect on legibility performance (Webster & Tinker, 1935). The really interesting aspect of this study is that the authors calculated different factors and analyzed the letterforms to arrive at specific characteristics that improve the legibility of these characters. That is definitely more informative and relevant than the relative performance of a group of typefaces. This sort of information can help designers to improve the legibility of their design by introducing characteristics that aide in correct character identification. For example, the study found that the lower the crossbar of the lowercase e is, the better it is at being recognized (Fox et al., 2007).

The use of legibility studies to inform type design choices seems to be picking up. A recent paper examined the variations of letter shapes and their effect on letter identification both at a distance and in parafoveal processing (Beier & Larson, 2010). The results, such as better performance if narrow letters were widened or x-height letters took up more of the ascender or descender space, are quite interesting for several reasons: 1. Modifications were made on the same typeface being tested and so the variables were well controlled. 2. Modifications were done systematically and were targeted towards testing specific features. 3. Therefore, it is possible to generalize the findings thus contributing to the external validity of the research.

As to which features are most important in letter identification, and as mentioned earlier in this chapter, letter identification is dependent on specific features of letters. The study by Fiset et al. (2008) showed that stroke terminals were the most important feature in the identification of Latin characters and the ones that carried the most valuable visual information. In terms of lower case, the three most important features were: terminations, horizontals, and slants titled right; for the uppercase, they were: terminations, horizontals, and intersections (Fiset et al., 2008). These are then the most critical features that operate in letter identification.

## Structure and Word Identification

Readers have certain expectations in terms of the regularity of a typeface design that produces efficiency in letter identification; irregularity in typeface characteristics have the opposite effects since the readers are tuning in to these characteristics and the details that they expect to see (Sanocki, 1987). Indeed, mixing fonts results in lower reading performance which suggests that readers develop a schema of characteristics that help them to identify letters more efficiently (Sanocki, 1988). Chinese readers showed the same font tuning process when reading; non-Chinese readers, however, did not font tune when reading Chinese leading to the conclusion that font tuning is specific to the task of reading and that the processing of letters as shapes is different from letter identification processes (Gauthier, Wong, Hayward, & Cheung, 2006).

Recent evidence points to the active role that structure plays in word identification: readers use visual cues based on a structural description of letterforms during word processing (Walker, 2008). The structure of a letterform is the spinal cord around which the letter is drawn, and can be arrived at by drawing the median line in between the 2 edges of the letter outline. The structure of one letter is, by conventions of type design, related to other letters via various relations of size and proportion. The end result is a system of structural relationships around which the whole typeface is built. For example, if the letter m is titled forward (therefore we have an italic or oblique structure), one would expect that the whole typeface is titled forward with the same angle. This knowledge comes from the exposure to typographic practices, and is referred to in this paper as the "translation rules."

The reader, Walker explains, carries forward these translation rules for the typeface that text is set in, and these depend as explained on the underlying structure common across letterforms. These translation rules are time dependent: if there is enough of a time lapse between the presentation of stimuli then the font tuning effect disappears (Walker, 2008). The reader then carries forward expectations of what the structure of the typeface is supposed to be, but if enough time passes, these expectations are no longer present. The Reader can retain the font translation rules in working memory for up to 750 ms, though this seems to be an unconscious decision rather than a conscious strategy (Walker, 2008).

Therefore, it seems that **the features needed for letter identification are not set loosely but rather fixed onto an underlying structure that is modulated by the design of the typeface that the text is set in.** In other words, the features that are critical to letter identification such as terminal endings are not ones that operate in a free grid. These features (or rather body parts when one is referring to letter anatomy) need to be set in the right position that is dictated by the internal structure of the typeface itself.

To give a human related example, if the face is the most important feature for recognizing people, it needs to be placed in its right position: the front of the head, rather than the back of one's feet. Moreover, these features need to conform to the Gestalt law of good continuation and the law of grouping, i. e. that the features are placed close enough in order to be recognized as one whole object (Pelli et al., 2009).

## Crowding

Researchers observed an interesting phenomenon: A letter is less likely to be identified in peripheral vision if it is surrounded by other letters; the farther it is, the higher the "eccentricity" and the less likely it will be identified (Pelli et al., 2007). Crowding, then, is this "excessive feature integration, inappropriately including extra features that spoil recognition of the target object" (Pelli et al., 2007, p. 2).

An interesting aspect of crowding is that it is able to predict the size of the visual span: that the visual span is in effect the span of text that is not crowded (Tillman et al., 2007). Tillman et al. (2007) found that the minimum letter spacing needed for letter identification (so as to avoid crowding where letter identification is no longer possible) predicted the minimum letter spacing as well as the visual span. As to what the visual span is, Pelli (2009) describes it as the "number of letters, in a line of text, that one can identify without moving one's eyes."

## Letter Superiority vs. Word Superiority: Redefining Legibility

At first glance, there seems to be a conflict between two observations related to letter and word identification:

- Word superiority effect, which says that a letter can be better identified in the context of a word than when it is on its own or in a non-word (Reicher, 1969).

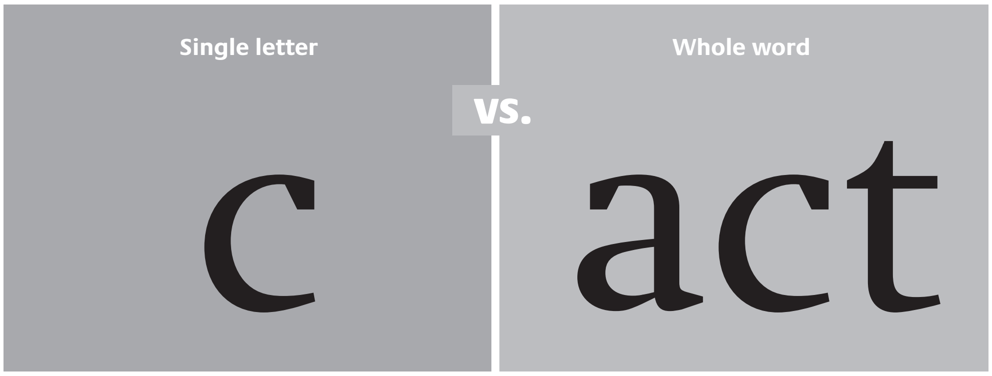


Fig. 5.6 The word superiority effect: A letter can be better identified in the context of a word than on its own.

. Letter superiority effect, which says that in threshold testing, an isolated letter has better visibility than a word (J. E. Sheedy et al., 2005).

This conflict is similar to the apparently opposing results regarding the effect of weight on typeface legibility. As will be discussed later in the chapter, a bold typeface resulted in longer reading times (Slattery & Rayner, 2009), and yet Sheedy et al. (2005) found that weight improved the performance of a typeface in threshold testing. However, the results need not be contradictory. For these to be reconciled, we need to look at the definition of legibility again.

Legibility is the ease with which the visual features can be extracted. So, what is a good measure of "ease"? To answer that, we first need to look at the process of extracting visual information. As mentioned earlier, the viewer uses different frequency channels to identify letterforms based on their size, this channel is dictated by the signal itself and not by the viewer, and that size can be manipulated by changing the viewing distance (Majaj et al., 2002). Thus, it is possible to conclude that text seen at a distance is inducing a different frequency channel than the text at a close range one, as it is generally larger than typical sizes used in body text. So back to the original question, how can one measure "ease"?

It stands to reason, that in the context of typical reading settings as reading from a printed publication or from screen, the ease of letter encoding would translate into shorter fixation durations, lower reading times, and less fixations in total. However, in the case of reading from a distance, which is reading under duress, it is a different scenario. It is not about the speed with which letters can be encoded, but rather the question if they can be encoded at all. The threshold test is then a valid test for reading under duress just as reading measures are good indicators of the legibility of type in continuous reading. The apparent contradiction in the results of word vs. letter superiority, as well as the different indicators for the use of bold, can be reconciled by the observation that the different frequency channels are being used to analyze letterforms, and by Majaj et al. finding that different aspects of letterforms (outline edges vs. gross strokes) are informing letter identification.

If legibility is about the ease of extracting visual information, then in the case of normal reading, speed is a good measure. In the case of constraints on visual acuity (as in reading from a distance), then the threshold measure is good. Both are compatible because the relative size of the text affects the frequency channel used to see it in. In such case, then legibility is relative to the distance from the stimulus, as well as the task required.

It follows then that **legibility is relative rather than absolute**, and depends on relationship of the viewer to the stimulus in terms of distance, task required, and familiarity with the content in both visual and linguistics aspects. Specific aspects of text styles such as italics are less legible whether viewed up close (Slattery & Rayner, 2009) or at a distance (J. E. Sheedy et al., 2005). Other aspects change roles, so text set in bold and viewed at close range will read slower that text set in regular (Slattery & Rayner, 2009), but when viewed from a distance, a bold letter is better distinguished from a regular one (J. E. Sheedy et al., 2005).

J. E. Sheedy et al. (2005) do address the apparent conflict between letter and word superiority and also find that these results are reconcilable, though they offer a somewhat different explanation:

> "The letter superiority effect has been identified at threshold sizes and seems attributable to early visual-processing factors. The letter superiority effect may or may not apply to text presented at suprathreshold size, subject to the same discussion presented earlier about the amount by which text is above threshold. Letters at typical reading size are more above threshold than are words of the same size; hence it can be argued that letters should also be more legible than words at typical reading sizes. However, this requires further study. The word superiority effect, on the other hand, refers to several related findings that individual letters are more accurately read when they appear in a word than when they appear embedded in nonsense strings. A familiar context (a word) makes individual letter recognition easier than does an unfamiliar context (nonsense letter string). This effect is more likely attributable to higher level cognitive factors than those required for simple letter identification. The present results show that individual letters are more legible than words at the simpler and earlier processing stage of letter and word identification." (J. E. Sheedy et al., 2005, p. 812)

There is a further argument to support their claim, and that is the issue of efficiency of feature detection. As mentioned earlier, the efficiency of feature analysis needed for word identification is inversely proportional to the number of letters in a word (Pelli et al., 2003). Therefore, it follows that identifying a single letter is more efficient than identifying a whole word.

Moreover, the word superiority effect is not claiming that the word is superior in being recognized as the name implies, but rather that a word provides a context which aides in faster recognition of letterforms. There is then no conflict between these findings, but rather that they paint a picture that is more complex and nuanced than what one might initially expect.

As to the relative nature of legibility, the conclusion of Beier and Larson (2010) supports this changing aspect of letter identification:

> "The study confirmed the notion that the performance of letter shapes varies according to the situation in which it is presented, and that some features are most important in distance viewing and others are most important in the parafoveal view" (Beier & Larson, 2010).

The changing dynamics of text is nothing new and was expressly addressed almost one century ago:

> "The factors which make it possible to read one style of print at a greater distance than another may not be the same as those which lead to the reading of one type face faster than another under ordinary reading conditions." (Webster & Tinker, 1935)

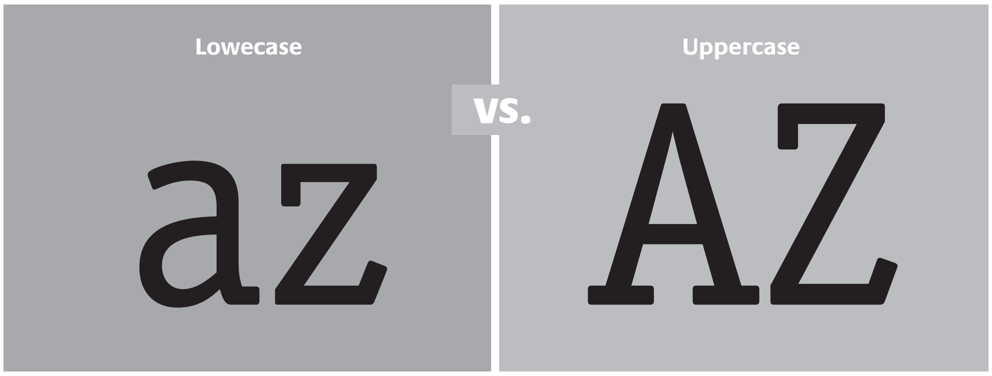


Fig. 5.7 Typographers are interested in the relative performance of uppercase and lowercase text.

Webster and Tinker (1935) think that it is the context clues that are facilitating reading text in paragraphs and that might be the reason behind the different results. Though that might help in the reading process itself, it does not explain why one typeface is more legible than another. The reasons for different performances of typefaces depend on the design of the typeface, the distance from the viewer, and the setting of type itself.

## Uppercase vs. Lowercase

A typeface is not an end product but rather an ingredient in the design process. In terms of legibility, what one does with the typeface (how it is set, which case is used etc.) can be as important as the design of the typeface itself. A few studies looked at the role of case in Latin typesetting. Arditi and Cho found that in threshold sizes, uppercase text performed better than lowercase in both threshold testing and in the speed of reading tests (Arditi & Cho, 2007). The study was done with both uppercase and lowercase set at the same point size so the overall area and the size of the counters of uppercase letters are bigger than those of lowercase ones. The experiment is then relevant to question of optical size. It confirms the consensus amongst typographers that uppercase letters appear larger and more visible than lowercase ones that are set at the same point size and that this is advantageous in the case of very small type (Fig. 5.7). The general convention regarding the style of Latin text is that text set in mixed (sentence) case is easier to read than text set in uppercase:

> ‘‘Lower-case letters have more ‘character’ in terms of variation in shape and the contrasting of ascenders and descenders with short letters. This leads to characteristic word forms that are much easier to read than words in all capitals’’ (Tinker, 1963, p. 34).

Webster and Tinker have also tackled the same issue of different results between visibility from a distance and continuous reading and came away with this conclusion:

> “It was found that material in lower-case letters had more definite word form and was read faster in connected discourse than material in all capitals (uppercase). The latter, however, was perceived at a greater distance from the eye than the lower-case print. The larger outlines of the upper-case letters undoubtedly caused the difference” (Webster & Tinker, 1935)

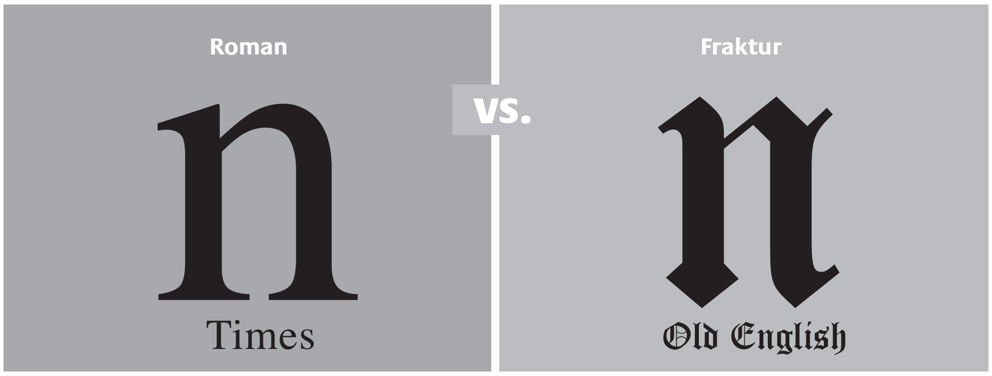


Fig. 5.8 Different styles of typefaces often result in diferent reading speeds.

Later studies have shown that word shape is not facilitator in parafoveal processing but rather it is character code information (Rayner et al., 1980). However, even when the word shape yields no preview benefit, the greater distinction between lower case shapes and features should facilitate the extraction of visual information.

## The Effect of Style

The style of the typeface used has an effect on reading performance. It is though less pronounced than the effect of the difficulty level. A study looking at the different reading measures when reading in Old English vs. Times New Roman found that reading in Old English resulted in a larger number of fixations, higher fixation times and shorter saccades. The font effect was also larger for the older group (almost 54 years older on average) who fared worse with Old English (Fig. 5.8). There was a significant interaction between age and font style for sentence reading time and fixation duration and marginally significant for number of saccades and regressions. Readers were more likely to skip words when reading in Times. (Rayner, Reichle, et al., 2006) which is again indicative that it easier for them to read in Times.

When reviewing legibility studies, the question of external validity comes up. An often cited paper is by Mansfield, Legge, and Bane (1996) which looks at the relative legibility of proportional vs. fixed width typefaces for readers with normal or low vision. The testing is done with Times Roman and Courier Bold. As a comparative analysis of two unrelated typefaces, the findings are relevant. However, it is at the point where one tries to extrapolate these results that drawing conclusions becomes that much harder.

The authors admit that there are many differences between the typefaces and that these could be the cause of any results found (Mansfield et al., 1996, p. 1493). Of these differences, the most obvious is that one is in regular weight while the other is in bold. Another noticeable difference is the very large difference in word space width. Both of these factors have an effect on reading speed as shown earlier in the chapter (Bold weights increase fixation durations, and word spaces play a very important role in the ability to program saccades). More subtle differences include the contrast and modulation of forms in Times, the heavy slab serifs in Courier Bold, and the large difference in the character widths of the two typefaces. As such, the results of this study cannot be generalized, and can be only seen as a comparative analysis of the relative legibility of two typefaces.

Another study looking at the relative threshold performance of typefaces was the already cited study by Sheedy et al. (2005) which tested 6 typefaces (Georgia, Times

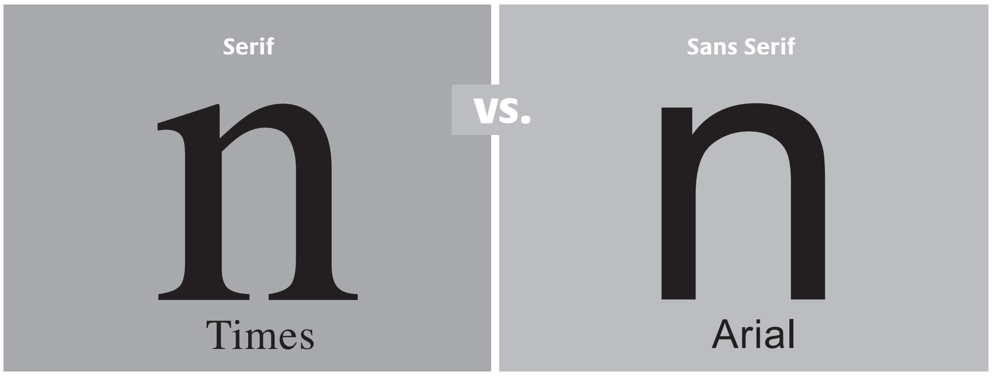


Fig. 5.9 Arial and Times have often been used to test the relative legibility of Serifs and Sans Serifs.

New Roman, Plantin, Verdana, Arial, and Franklin) with Verdana and Arial coming out on top and Franklin at the bottom. In a way, this is again a comparative analysis. However, the study stands out as it is using a small sampling of typefaces to test other factors including rendering, resolution, display type, and font styles such as Bold and Italic. As such, the comparative analysis of the different typefaces is not the end in itself, but rather the means to test other factors, and to see how they interact with font style.

## The Question of Serifs

The question of serifs has taunted legibility researches for decades now. If anything, the results have been inconclusive (Lund, 1999) though that has probably more to do with the test's setup than the research question itself. The problem with such a question is that it takes two of the general categories of Latin typefaces (serif and sans serif) as the starting point of the question. Such categories are convenient as they are based on easily identifiable elements: the presence or absence of terminal in- and out-strokes. What is missed in such broad categories is the fact that many design variables change across these styles (Fig. 5.9). These include contrast (the variation between thin and thick parts of a stroke), axis (derived from the angle of the pen), and the proportion of x-height to ascenders and descenders. So, is the question of legibility of serif typefaces really only about how the terminals are treated, or is it related to the overall package of modulated strokes, tilted axis, and modest proportion of x-height to ascenders?

If terminations really are the most important feature in letter identification (Fiset et al., 2008), then this might explain the perceived[12] superiority of the legibility of serifs over sans serifs in the domain of printed long texts such as newspapers and books. Serifs are terminal treatments that emphasize the beginning and ending of strokes, and in effect are visual accessories that seem to serve the most important feature in character recognition.

It is also possible that the stroke modulation present in the majority of serif typefaces helps to highlight features that aid in differentiating characters. The tilted axis can help break the rigid vertical rhythm. There are many possible reasons why designers often find serif typefaces to be more legible than sans serifs. Familiarity and convention are often cited. However, this is all exasperated by a reversal of fortunes when it comes to reading on screen where the general consensus amongst typographers and designers is that sans serifs are more suitable for low resolutions.

12 This perception is one held amongst typographers for many decades now, though it is not one that is based on empirical research but rather on conventions of design.

One of the shortcomings of the studies regarding the role of serifs in reading is that they were more of comparative performance analysis rather than an isolation of the factor to be studied. So if one was to compare Arial and Times, whatever results generated cannot be really nailed down as due to the presence of serifs because there are several other variables that have not been controlled in the experiment testing. While such research is informative for graphic designers in the process of selecting one typeface or another, it is less informative for type designers in the process of designing a new typeface.

The legibility effect of specific features is an area that easily invites further research. A study that finds Times to be more legible than Arial cannot be generalized to mean that all serif typefaces are more legible than sans serif ones. At the end of the day, it is an issue of external validity and the ability to build theories that can be verified by empirical data. What it does is to actually offer support rather than final proof. In such an avenue of research, this might be the best-case scenario.

On the other hand, there is the other approach, and that is to design a system of typefaces where possible variables are controlled, and serifs are the only factor to change in the design. One such study designed a typeface family specifically set to control parameters that might vary between a typical serif and sans serif typefaces, and found that sans serif typefaces were 20% faster to read in very small sizes (Morris, Aquilante, Yager, & Bigelow, 2002). They also found that this advantage disappeared in large sizes.

The question of serifs poses a question for its place in this paper. The narrative being built here tries to connect findings that are less script specific and rather universal in nature. The question of serifs is one very specific to the Latin script, and Greek and Cyrillic by extension, and is very far from the research purpose of this paper. As such, given the questionability of external validity, supported by Lund's findings (1999), and the wealth of confounding factors in the majority of studies, also reviewed thoroughly by Lund (1999), it was best to learn from the errors committed in these studies while building the experiment, but not to go deep into this topic in this literature review. A few studies stood out, and are briefly covered in the paragraphs below.

One of these studies is a paper by Robinson, Abbamonte, and Evans (1971) in which they outline a proposal as to why serifs are important, and then conduct computer simulations of letter identification that echo the visual detectors in the retina. Of the reasons they initially propose, one finds the typical reasons that one hears in typography and type design education: The serifs lead the eye across the horizontal. Serifs preserve the shape of the letters by emphasizing the terminals, etc. These they refute with the arguments that the eye makes only a few fixations per line so horizontal continuity cannot be such a big factor, and that instead of preserving the shape, serifs can very well be added noise. However, it is their conclusion that really stands out:

> "Serifs are only important in letters which are small enough to be perceived by line detectors: most ordinary print in the texts of books, periodicals, and typewritten material. Larger and/or thicker letters are probably perceived by a different part of the system—the edge detectors. However, the image of large letters are of line-form when viewed from a distance. Serifs are useful not only when the letterforms are physically small, but may also be functional when large letters form a small image on the retina—as, for example, when a billboard is seen from a distance" (Robinson et al., 1971, p. 358).

Though one may question the validity of using a computer simulation to verify the way the human eye detects visual stimuli, this paper stands out especially when viewed with the information that we know today. As discussed earlier, the reader uses different frequency channels when looking at different sizes of text, and letterforms are identified by their gross strokes in small sizes and by their edge in large sizes (Majaj et al., 2002). So, if one were to review that paper from the point of view of research available in the 70s, then yes, there are many questions still open, but it seems that the authors were on to something with their theory. The interesting thing is that their theory provides a neat explanation to a very common typographic practice that students of typography learn very early one: use serif typefaces for text, and sans serifs for headlines.

Legibility studies, though, usually revolve around a human reader. Another exception was a study that looked at the legibility of various typefaces by a machine rather than human reader (Zhang, 2006). It used different OCR readers to analyze texts set in 18 different typefaces. The study shows better results for Sans Serifs and identified commonly misidentified pairs of letters. Though the approach is interesting, one wonders if OCR text analysis is analogous with the reading process.

Connolly (1998) found that the presence or absence of serifs did not play a role in legibility as the top performing typefaces of several tested were 2 serif and 2 sans serif designs; the study did though find better performance and subjective rating for typefaces with open design such as large counters and wide spacing (Connolly, 1998). This is another hint that the relative legibility of serif to sans serifs might be dependent on factors unrelated to the actual treatment of stroke terminals.

In either case, the role of serifs is a valid research construct and definitely an endeavor that will support typographic design using the Latin script. However, and as mentioned earlier, this is a topic that is tangential to the core of this paper and so we leave it here.

## The Effect of the Complexity of Style

Though the effect of the complexity of typeface style on legibility is under represented as a research topic, there are still some very interesting studies that have gone to show that the more complex the style, the less legible it is. For example, Pelli et al. (2006) looked at the effect of complexity (as evident in script typefaces) and defined "perimetric complexity" as the "inside-and-outside perimeter, squared, divided by ink area" (Pelli et al., 2006). Their results showed lower performance for those typefaces. Another study also found lower performance for a script typeface vs. a Roman (Rayner, Reichle, et al., 2006), and in this case the use of a script typeface was specifically done to test font difficulty. A later study by Slattery and Rayner (2009) also looked at the relative legibility of a serif (both in roman and italic) and a script typeface, and found lower performance for script. Comparing Times New Roman to Harrington, and Script MT: Times New Roman had a smaller number of fixations, and with shorter average fixation duration; testing Roman against Italics and Bolds also showed longer reading times for Italics (Slattery & Rayner, 2009).

It is possible then that these effects, worse reading times for script and Italic (Fig. 4.10), are due to complexity of form rather than the familiarity of typefaces as the authors mention. It is especially so given Pelli's findings, as well as the fact that the same paper found longer reading times for Italics and those are arguably almost as familiar as the Roman.

As mentioned earlier on, all visual information needed for reading to continue smoothly is extracted in the first 50-60 ms of a fixation (Rayner et al., 1981). This implies that the role of typography and its effect on foveal processing are limited to that time frame. It is possible then that a complex visual will increase that

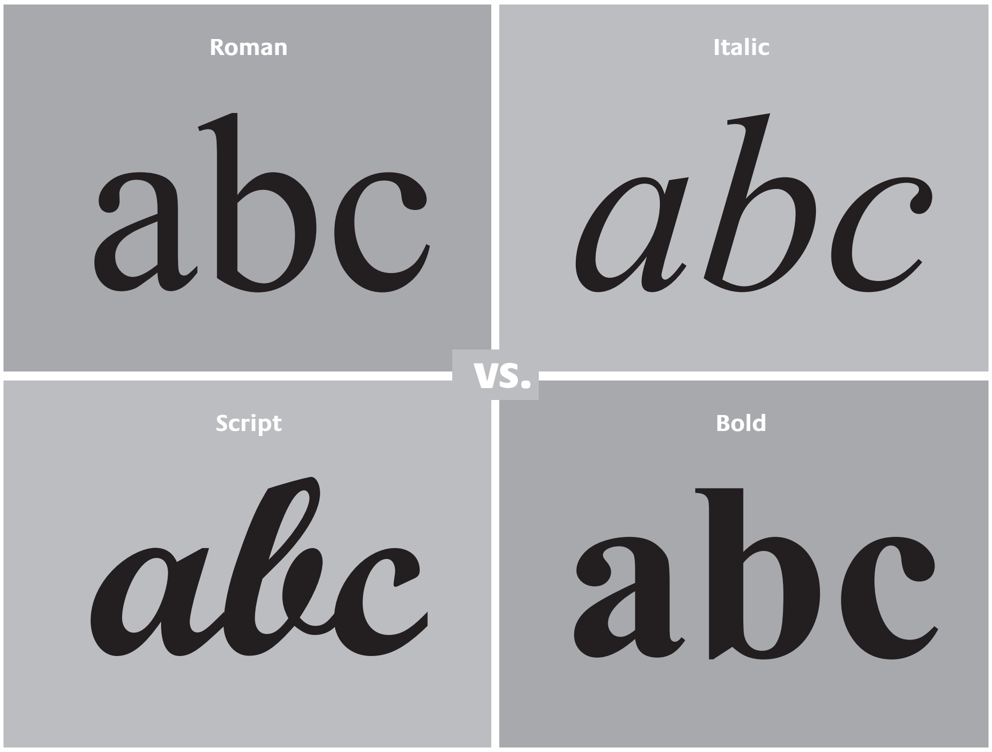


Fig. 5.10 Different levels of complexity result in different reading speeds.

duration, similar to the effect of a low-frequency word (Inhoff & Rayner, 1986), and will also reduce the preview benefit resulting in shorter saccades and longer fixation durations. An interesting study looked at the effects of age on typeface legibility as well as the comparative performance of various typefaces. The results were independent of the age of the reader but specific to small sizes in print. They too showed consistent low reading performance for typefaces that were condensed and complex in form and design (Connolly, 1998).

Perhaps the best summation of the effect of the complexity of the style was put by Rayner et al. (2006). They wrote that as long as text looks normal, typography effects are very small, but "it stands to reason that a font that is more difficult to encode should cause readers to look at words longer than fonts that are easier to encode" (Rayner, Reichle, et al., 2006). This confirms what Webster and Tinker stated in 1935 that the simplicity of outline improves legibility (Webster & Tinker, 1935).

An interesting aspect of complexity of form is the increased number of strokes per character width, leading to higher stroke frequency as mentioned earlier. The reason why this higher frequency negatively impacts reading can very likely be due to crowding. Pelli (2008) claims that for objects to be recognized and differentiated they need to be separated in the visual cortex at a specific minimum distance (6mm in the radial direction, or 1 mm in the circumferential direction); if the objects are closer than that, they will be perceived as a jumble (Pelli, 2008). It follows then, that if the strokes are too close to one another they stand a higher risk of being indistinguishable from one another. This is very likely one of the main reasons why complex shapes require more time to encode.

## The effect of Typographic Variables

A common sentiment echoed by designers is that legibility is not just about the typeface but what one does with the typeface. Typographic variables (Fig. 5.11), such

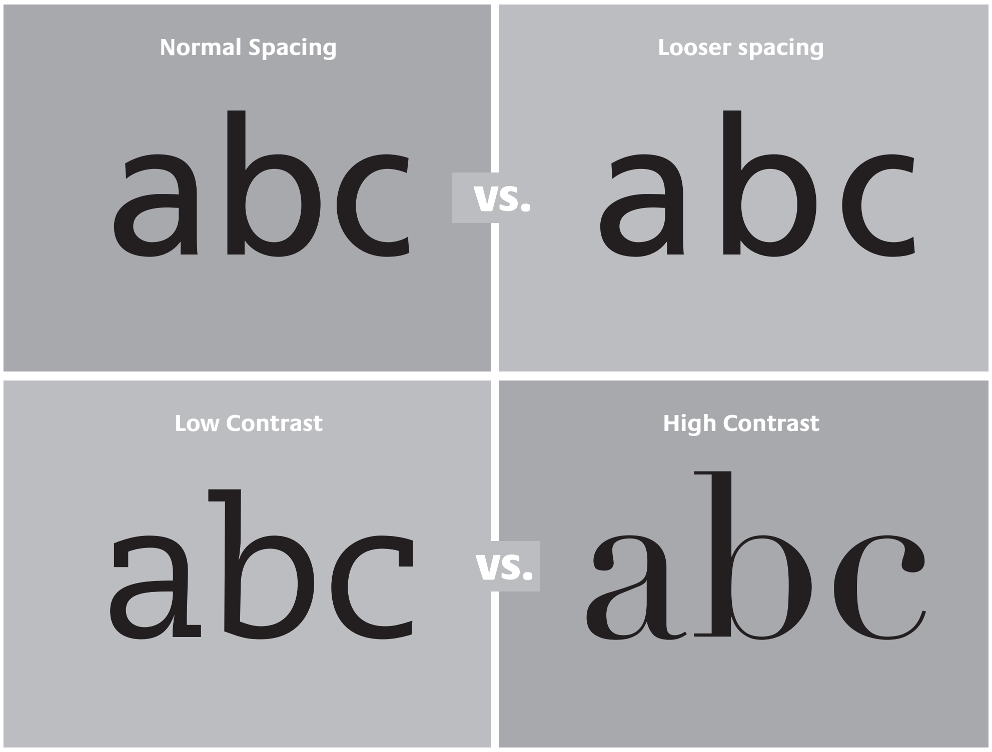


Fig. 5.11 Different typographic variables have been shown to affect legibility.

as type size, line length, word spacing, letter spacing, leading, color and contrast, are all variables that can interact or individually influence the legibility of a typeface. For example, type size can affect legibility when text becomes very close to minimum acuity levels (Mansfield et al., 1996), and research has shown that common text sizes used in print today and historically have been influenced by the properties of our visual system (Legge & Bigelow, 2011). This becomes even more critical as we age, and older readers prefer larger text sizes (Bernard, Chaparro, Mills, & Halcomb, 2003), most likely due to the drop in their visual acuity.

Of the body of work related to legibility research, the work of Miles Tinker stands out in terms of both quality and quantity. Tinker used speed of reading as a measure of legibility, and used comprehension checks to ensure that text was being read for meaning, all the while testing the effects of various typographic variables. His view was that more legible text has a wider perceptual span, longer saccades, and less fixations. He defined perceptual span as the number of words in a line divided by the number of fixations on than line. This has since been proven to be wrong (R. E. Morrison & Inhoff, 1981). Morrison and Inhoff reviewed Tinker's findings and showed that Tinker's perceptual span are directly correlated with saccade length, and therefore mapped his findings of the perceptual span to saccade length (R. E. Morrison & Inhoff, 1981). This enables an analysis of the data with the benefits of more recent findings. Therefore, the following points follow their interpretation of Tinker's data.

Tinker proposed that threshold tasks were not necessarily indicative of higher legibility in reading tasks. For example, uppercase letters were better discriminated at a distance, but functioned poorly in text settings. Text set in all caps resulted in longer reading times, and 12.4 more fixations. The saccades were longer in physical span, but in reality covered a smaller number of letters (7 vs. 8.1) (Tinker, 1963).

Increasing text size from 10 to 14 pts. resulted in reduced reading speed and higher number of fixations. The average fixation duration was reduced, but that was offset by the increase in fixations and reduction in the number of characters covered in a saccade (Tinker, 1963). Decreasing text size to 6 pts also had negative effects with longer fixations and shorter saccades (Tinker, 1963).

In terms of line length, very short line lengths resulted in longer fixations and a higher number of fixations. Very long lines resulted in decreased accuracy in executing return sweeps (Paterson & Tinker, 1942). Extra leading improved the reading of long lines (Tinker, 1963). When type is set with the optimal line length and leading, there was no difference between 9, 10, 11, or 12 pts. (Tinker, 1963). Reduced contrast between text and background also has a negative effect on legibility; it increases fixation times, number of fixations, and regressions (Tinker, 1963).Typefaces, then, have improved legibility "when the size of the letter is increased, when the lines in the letter are widened, when the area of white space around or within the outline of the letter is increased, when the contrast of shading and hair lines is lessened, and when the outline of the letter is made simpler" (Webster & Tinker, 1935).

## Legibility, Complexity, and the Mark of a More Legible Design

This chapter is too long and with too much information for one to be able to adequately summarize all that it has covered. There are, on the other hand, several conclusions to be drawn and are essential to carry forward into the next chapters. These are the corner stones on which the next two chapters are built:

- Reading is affected by the linguistic qualities of the text, as well as the reading abilities of the reader, and the visual characteristics of the text.
- Typefaces have been shown to have a statistically significant effect on various reading measures.
- Legibility is relative, rather than absolute and it is affected by the visual characteristics of text, as well as the distance of the viewer from the stimulus and the task at hand.
- Legibility is the ease with which the reader can extract the visual information of text in order for lexical processing to begin.
- The marks of a more legible design are shorter fixation times, longer saccades, and a smaller total number of fixations.
- Legibility is negatively influenced by the complexity of the writing system and typeface style.

And so with this chapter, we are half way in drawing the arc of our narrative. We have seen how the Arabic script has developed from manuscript to typographic forms, and where it stands today. We have looked at the reading process and what makes a more legible design. Next, we look at how these concepts map out to the Arabic language and script. For that we turn the page.