Computers in Human Behavior 39 (2014) 177–186

Contents lists available at ScienceDirect

Computers in Human Behavior

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

A calligraphic based scheme to justify Arabic text improving readability and comprehension

Aqil M. Azmi [a,*], Abeer Alsaiari [b]

[a] Department of Computer Science, King Saud University, Riyadh, Saudi Arabia
[b] Department of Computer Science, Taiba University, Madina, Saudi Arabia

ARTICLE INFO

Article history:

Keywords:
Arabic typography
Font
Kashida
Text justification
Readability
Comprehension

ABSTRACT

Studies have shown a correlation between reading comprehension and the visual appearance of the displayed text. One of the factors that affect the visual look of a text is its alignment. The purpose of this paper is to develop and implement a sophisticated algorithm to output a properly justified Arabic text. Most of the tools geared for e-document have not been tailored with Arabic in mind. And so, these either violate several calligraphic rules, or are a far cry from the aesthetics developed by the centuries old tradition of Arabic calligraphy. The scheme we developed is more realistic calligraphically and more pleasing aesthetically. It is a two-step process. Lines are populated with whole words, afterwards we use alternate form of the letters to compress or stretch the line as needed. In the second step we use kashida (elongation of the connecting line between the letters) to fill in the remaining gaps. There are strict rules which dictate which, when, and the minimum/maximum length of the kashida a word can have. We tested our justified Arabic text on university students. The experiment revealed the participants were able to read faster and had a better comprehension when presented with our justified text. The scheme we devised could be extended to other languages which share the basic Arabic script, e.g. Persian, and Urdu.

1. Introduction

Typography (from the Greek words *typos* impression, and *graphia* graphy) is the art and technique of arranging type to make the language visible. This arrangement involves the selection of typefaces, point size, width of the line, the spacing between lines (interline spacing), spacing between groups of letters, the spacing between pairs of letters (kerning), and text justification. The typography of texts is of interest for two main reasons. First, the typography should not interfere with the reader's ability to understand the text; and second, the visual appearance of the text may influence the motivation to read. Just imagine reading this paper in cursive script. There are many studies in the English language that have reported a strong link between reading rate, accuracy, and comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Tan & Nicholson, 1997).

Early studies examined typography readability for printed matters, with later studies tackling the e-texts or computer displays. The majority of the studies were for English typographies, e.g. (Bernard, Lida, Riley, Hackler, & Janzen, 2002). The studies have shown that certain typographical parameters such as typeface, margins and spacing do affect the readability of online text (Dyson, 2004). Campbell, Marchetti, and Mewhort (1981) compared the reading speed of 156 participants for right justified English text using three different techniques: no justification, right justified using fixed character spacing, and right justified using variable character spacing. Of these, the authors reported an improved reading speed for variable-spacing right justified text. However, these guidelines are for English text and they are not applicable to Arabic script due to orthographic differences (Ganayim & Ibrahim, 2013).

The Arabic language is home to over three hundred million people. The Arabic script is more widely used, with over a billion and half Muslims who read the Qur'an in its original Arabic script. And just as most of the European languages uses Roman based alphabet, so is the case with languages such as Persian, Urdu, Ottoman (Old Turkish) ... etc. all of these use Arabic based alphabet. There are 28 basic letters in the Arabic script which includes 3 long vowels. Most of the letters in Arabic assume four different forms depending on the context: initial, middle, final and isolated. In addition there are short vowels (a, i, u), more commonly known as diacritical markings, a total combination of 13. These markings,

* Corresponding author. Tel.: +966 011 467 6574.
E-mail addresses: aqil@ksu.edu.sa (A.M. Azmi), abeer_alsaiari@yahoo.com (A. Alsaiari).

http://dx.doi.org/10.1016/j.chb.2014.07.003
0747-5632/

which are placed either above or below the letter, are used to clarify the sense and meaning of the word. For example, the Arabic word for *flag*, *science*, or *taught* is (علم). There are two ways to disambiguate, use diacritical markings, or through the context. The former is a fool proof scheme, however, it is absent in modern writing. This leaves us with the context, which the natives are fluent at. Though there remain cases where either interpretation is plausible (Azmi & Almajed, 2013). Consider the standalone example, (أكلت الخبز) which could either mean *I ate the bread*, *you ate the bread* (masculine), *you ate the bread* (feminine), or *the bread was eaten*. Clearly this is an important reason why most religious books are full of diacritical markings. The books for those learning Arabic as a second language are also dotted with diacritical markings.

Few studies have examined the typography of Arabic texts (Abubaker & Lu, 2012; Alsumait & Al-Osaimi, 2009; Ganayim & Ibrahim, 2013; Hemayssi, Sanchez, Moll, & Field, 2006; Ramadan, 2011). Ramadan (2011) presented a study evaluating college students' readability and comprehension of Arabic e-books. In a series of three experiments involving 49, 31, and 31 participants, where in the first experiment the participants were asked to rate some good Arabic fonts from those found in MS Windows system for reading e-passages. In the second experiment, the participants determined the best font style and font size out of those rated in the first experiment, while in the final experiment they determined the best combination of page layout and background/text color arrangement. The result revealed that Simplified Arabic font, with a point size of 14pt were the best combination between different Arabic font styles and sizes. Also, a single column text with black/white background color was the best combination for the third experiment. Interestingly, this elaborate study overlooked two important factors: text justification, and the diacritical markings; which are likely to influence the readability and the comprehension of the Arabic text. Abubaker and Lu (2012) experimented for the optimum font size and type to read from screen for students aged 10–12. The study involved two fonts (Traditional Arabic, and Simplified Arabic) at four different sizes, 10, 14, 16 and 18pts. A total of 30 students (10 per age group) participated in the study. The authors concluded that font size 16pt and below are inappropriate for children age 10, however, font sizes 14 and 16 are readable for those aged 12 and over. They also recommended avoiding Traditional Arabic for this age group. The study in Ganayim and Ibrahim (2013) involved 210 native Arab students. The authors found that multicolumn text affected the comprehension achievement but not the reading speed. The students reported a better comprehension for single-column text than for the multicolumn. For designing Arabic text, a reading rate of 127 words/min is to be considered. In the same study it was reported that the interline spacing had no relevance on either factor. As we mentioned, there are many facets to typography. Elements such as selection of typeface, point size, and various spacing are fine-tuned by individual typesetters depending on the aesthetics or subject to some cultural conventions. For example, in Arabic and French it is customary to leave a space before a colon or semicolon in a sentence, while in English it is not. Text justification is one element of typography which can be automated; the other is hyphenation, something which is inappropriate for Arabic script. To the best of the authors' knowledge, there is no study on the readability and comprehension of justified Arabic text. We therefore assume the study in Campbell et al. (1981) for the English justified text holding true for Arabic. Having said so, we in this paper develop a sophisticated system to justify Arabic text. A system which takes into account all the peculiarities of the Arabic script and at the same time conforms to the traditional Arabic calligraphy. With the help of 44 participants, all university students, we evaluate our Arabic text justification system using the font we developed for this purpose against 14 point Simplified Arabic font, the best combination according to the study in Ramadan (2011). Another study which backed this font, albeit for young children, is (Abubaker & Lu, 2012). For our experiment we used plain text as well as text with full diacritical markings. The assessment revealed that our font using a sophisticated text justification algorithm beats Simplified Arabic font in both readability and comprehension tests.

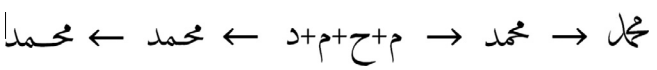

Fig. 1. The four letter name 'Muhammad' at different degrees of stretching (left) or compressing (right).

The process of typesetting languages using the Arabic script is more challenging than typesetting using the Roman script because of the requirement for special needs and strict rules. Standard software on different platforms provide appropriate tools to handle Arabic and other right-to-left languages, but leaves much to be desired in terms of quality. The flow of traditional calligraphy is said to be lost in moveable-type printing and most likely in plain computer typesetting (Andrew, 2008). According to Mulder (2007), the Ottomans did not use printing with moveable type until long after it was in use elsewhere, mostly because of reluctance to compromise the elegance of the text. During the last century there were calls to 'simplify' the Arabic writing system, e.g. one form per letter and even allowing for left-to-right ordering (Széll, 2012). These calls faded as new generation of smart fonts came into existence which makes it possible to mimic some of the features that one desires from a calligraphic point of view (Azmi & Alsaiari, 2009; Széll, 2012). Though Arabic typography got much better in the last two decades, yet, there are still some aspects that need to be addressed so that the quality of Arabic typography can match as closely to the quality of its calligraphy. When justifying a Latin based text, the software rely on: (1) hyphenation; and (2) insertion of extra spaces between words. On the other hand, the Arabic based typesetting software stretches the words horizontally using kashida (كشيدة). Such a solution violates a basic principle of Arabic calligraphy where certain letters can be stretched while others can be compressed (Benatia, Elyaakoubi, & Lazrek, 2006), see Fig. 1. These set of rules as defined by the calligraphers have been maintained ever since and we would like to keep them intact into the digital age. So in developing our Arabic text justification system we aim to develop one that conforms to the rules as put forth by the calligraphers.

The rest of the paper is structured as follows. In the next section we look into the characteristics of Arabic writing. We cover related works in Section 3. In Section 4 we go over our proposed scheme. And in Section 5 we look into our implementation and how it compares to output by other software. Section 6 covers the evaluation of our text justification system. Finally we conclude and suggest some future works in Section 7.

2. Characteristics of Arabic writing

As a language, Arabic predates Islam (Al-Azami, 2011, pp. 123–129). However, it was Islam that boasted it and gave it eminence because Qur'an was revealed in it. During the early period of Islam there were no mandatory rules for Arabic script. The drawing of letters did not follow any particular style or specific rules. Since Islam prohibits the depiction of human form (aniconism), this lead to Islamic art being dominated by decorative geometric patterns, and calligraphy. The latter was specially revered in Islamic arts as it was the primary mean for the preservation of the Qur'an. While the art of calligraphy has often been dominated by men, there were famous women calligraphers who are known to scribe the whole of Qur'an (Al-Munajjid, 1995; Kazan, 2010;