Visible Language

54 . 1+2

the journal of visual communication research

Peterson et al.

Gaze-based HCI applications

Eye-tracking technology as a mode of user input is explored through a novel framework, the "tech receptivity interval, which distinguishes infancy versus maturity levels of acceptance of emerging technologies.

implications for designing for emerging technologies

Lonsdale et al.

Enhancing Bowel Cancer Surgery Recovery

Integrating text and visualization using research-based information design principles and user-centered methods significantly improved the accessing, finding, and understanding medical information than a text-heavy presentation of the same information.

implications for designing information, particularly for healthcare

Kuraitytė et al.

Impact of Kinetic Typography on Readers' Attention

Kinds of kinetic (moving or morphing) type are compared using eye-tracking to assess which aspects of kinetic type better attract readers' attention.

implications for design of kinetic type to attract and manage attention

Parhami

Writing in Persian: the Intersection of culture and technology

Presentation of features of Persian script that made it difficult to implementation on modern technologies including discussion of the interplay between centuries-old Persian culture/language/script and modern technology finding that the same features that it difficult to design legible and aesthetically pleasing Persian printouts/displays also lead to challenges in automatic text recognition.

implications for type design and design's role in culture/technology conflicts

ISSN 0022-2224

Published continuously since 1967.

Visible Language

the journal of visual communication research

54 . 1+2

06 . 29

Marshall et al.

Braille in an evolving technological world

Surveys of visually impaired Braille readers show preferences for different technologies in different contexts: audio books for leisure reading; screen readers for on-line resources; Braille for digesting challenging content, while finding information in public spaces and packaging still lacks effective reading supports.

implications for designing for the visually impaired

30 . 65

Nonaka et al.

Linguistic and cultural design features of sign language in Japan

Historic evaluation of the design of Japanese sign language affirms the superiority of user-centered design methods that appreciate language differences, particularly the difference between sign language and signed expressions of spoken language.

important for designers of sign-to-voice/voice-to-sign technologies and for designers of visual symbols versus visual representations of spoken language

66 . 97

Serin et al.

A Latin-script typeface for education of individuals with Autism

Surveys of educators of Autistic children supported by literature review guided the design of a typeface to support teaching Autistic individuals to read.

implications for typeface design and user-centered/inclusive design methods

Before there was reading there was seeing.

People navigate the world and probe life's meaning through visible language. *Visible Language* has been concerned with ideas that help define the unique role and properties of visual communication. A basic premise of the journal has been that visual design is a means of communication that must be defined and explored on its own terms. This journal is devoted to enhancing people's experience through the advancement of research and practice of visual communication.

Published triannually in April, August, and December

website:

http://visiblelanguagejournal.com

send correspondence to:

Mike Zender, *Editor*
College of Design, Architecture, Art, and Planning
University of Cincinnati
PO Box 210016
Cincinnati, OH 45221-0016
mike.zender@uc.edu

Dr. Maria dos Santos Lonsdale, *Associate Editor - Typography*
m.lonsdale@leeds.ac.uk.

Professor Matthew Wizinsky, *Associate Editor*
matthew.wizinsky@uc.edu

Merald Wrolstad, *Founder*

Sharon Poggenpohl, *Editor Emeritus*

University of Cincinnati, School of Design, *Publisher*
© Copyright 2020 by the University of Cincinnati

Visible Language

the journal of visual communication research

. 1 + 2

special issue

Inclusive Design

April + August 2020

Advisory Board

Naomi Baron – *The American University, Washington, D.C.*
Michael Bierut – *Pentagram, New York, NY*
Charles Bigelow – *Type designer*
Matthew Carter – *Carter & Cone Type, Cambridge, MA*
Keith Crutcher – *Cincinnati, OH*
Mary Dyson – *University of Reading, UK*
Jorge Frascara – *University of Alberta, Canada*
Ken Friedman – *Swinburne University of Technology, Melbourne, Australia*
Michael Golec – *School of the Art Institute of Chicago, Chicago, IL*
Judith Gregory – *University of California-Irvine, Irvine, CA*
Kevin Larson – *Microsoft Advanced Reading Technologies*
Aaron Marcus – *Aaron Marcus & Associates, Berkeley, CA*
Tom Ockerse – *Rhode Island School of Design, Providence, RI*
Sharon Poggenpohl – *Estes Park, CO*
Michael Renner – *The Basel School of Design – Visual Communication Institute, Academy of Art and Design, HGK FHNW*
Stan Ruecker – *IIT, Chicago, IL*
Katie Salen – *DePaul University, Chicago, IL*
Peter Storkerson – *Champaign, IL*
Karl van der Waarde – *Swinburne University of Technology, Melbourne*
Mike Zender – *University of Cincinnati, Cincinnati, OH*

Visible Language

Contents

Readers' experiences of Braille in an evolving technological world.
Laura Marshall
Jeanne-Louise Moyse
06 — 29

Linguistic and cultural design features of the manual syllabary in Japan
Angela M. Nonaka
Jean Ann
Keiko Sagara
30 — 65

A Latin-script typeface, based on special education teachers' opinions, to use in literacy education of individuals with autism.
Efecan Serin
Aprigio Luis Moreira Morgado
Ricardo Santos
66 — 97

Anticipating Gaze-Based HCI Applications with the Tech Receptivity Interval: Eye Tracking as Input
Matthew Peterson
Brad Tober
Deborah Littlejohn Mac Hill
98 — 127

Enhancing Bowel Cancer Surgery Recovery through Information Design:
The impact of combining design and cognitive principles with user-centered research methods, on patient understanding of surgery recovery information.
Maria dos Santos Lonsdale
Stephanie Sciberras
Hyejin Ha
Stephen J Chapman
128 — 169

Impact of Kinetic Typography on Readers' Attention
Milda Kuraitytė
Ann Bessemans
Erik Nuyts
170 — 185

Computers and Challenges of Writing in Persian:
Explorations at the Intersection of Culture and Technology
Behrooz Parhami
186 — 223

Computers and Challenges of Writing in Persian:

Explorations at the Intersection of Culture and Technology

Behrooz Parhami

The Persian script has presented some difficulties, ever since printing presses were introduced in Iran in the 1600s. The appearance of typewriters more than a century ago created additional problems and the introduction of digital computers in the late 1960s added to the design challenges. These difficulties persisted, until high-resolution dot-matrix printers and display devices offered greater flexibility to font designers and the expansion of the computer market in the Middle East attracted investments on improving the Persian script for computers. Nevertheless, certain peculiarities of the Persian script have led to legibility and aesthetic quality issues that persist in many cases. In this paper, I enumerate some of the features of the Persian script that made it a poor match to implementation on modern technologies and review the challenges presented by, and some of the solutions proposed for, each new generation of printing and display devices. Before delving into the technical challenges and solutions, however, I discuss the sociocultural significance of the interplay between centuries-old Persian culture/language/script and modern technology, along with associated research problems. Interestingly, the same features that make legible and aesthetically pleasing Persian printing/displaying difficult also lead to challenges in automatic text recognition. I conclude with an overview of current state of the art and areas that still need further work.

K e y w o r d s **–**

Aesthetics;
Arabic script;
Computer display;
Computer printing;
Dot-matrix font;
Font design;
Keyboard; Legibility;
Line-printer;
Movable type;
Persian script;
Printing;
Technology transfer;
Typewriter

1. Introduction

In this paper, I summarize the results of nearly two decades of research and development, while I was working in Iran, and a few years thereafter, on adapting computer technology to the Persian script. The work was a team effort involving multiple colleagues at Sharif (formerly Arya-Mehr) University of Technology, researchers/engineers at other organizations that were developing computer applications in Iran, and several computer vendors targeting the Iranian market. My motivation is to create a formal recorded history of these efforts and, more importantly, present a picture of the state of the art in Persian printing and display technologies that would serve as a baseline for charting a path to solving a myriad of tough problems that remain unsolved or poorly solved after five decades.

In fact, the original title for my UCLA talks (Parhami, 2017), on which this paper is based, was "Fifty Years of Poor Penmanship: How Computers Struggled to Learn the Persian Script." The fifty-year period alluded to in this alternate title began in the late 1960s, when I was an undergraduate student at Tehran University's College of Engineering. In those days, Tehran University had just installed an IBM 1620 (International Business Machines Corp., 1959) at its computer center across the street from the main campus in an effort to mechanize administrative data processing tasks and also offer the students access to computing capabilities. At about the same time, major banks and large governmental organizations were developing customer databases, as well as payroll and personnel information systems (Encyclopedia Iranica, 2011), all of which needed to deal with information, such as employee/customer names and street addresses, in Persian.

A lot has happened in the five decades since then. My start in computing, when I was a junior engineering student in 1967, was learning to program in the Fortran language, preparing programs on punched cards and receiving the output on scrolls of fan-fold paper, a day or so later, iterating until program bugs had been removed. No Persian printing was involved at the time and there was no display unit for us student users. Organizations that did use the Persian script at the time, had to tolerate very low-quality output. The Persian script produced by computer printers and display devices has improved substantially, thanks to the diligence of the early researchers and the rapid development of digital computing in Iran (and Arab countries, using pretty much the same script), which attracted investments and technical know-how to solving the problems.

I hope I succeed in conveying the scope of the problems in the early years and the hard work and ingenuity that went into solving them in many iterations, while at the same time adapting to rapid changes in computer printing and display technologies.

2. Iran Meets Modern Technology

Iran (formerly Persia) has a storied past in science and philosophy. Both domains flourished in Iran and several other Middle Eastern countries during the Islamic Golden Age (Renima, Tiliouine, and Estes, 2016), between the 8th and the 14th centuries, when luminaries such as Omar Khayyam (poet, mystic, and scientist, who wrote an early treatise on algebra), Al-Khwarizmi (a polymath from whose name the word "algorithm" derives), and Avicenna (influential philosopher, regarded as the father of modern medicine) lived. Twenty-first century Iran is experiencing a renaissance in science and technology, in part due to economic sanctions that have forced the country to produce many ideas and products internally (Ashtarian, 2015). However, for some six centuries, science and technology in Iran fell on hard times and scientists and other elites, starving for resources, were also viewed with disdain and suspicion. As a result, Iran moved toward a consumption society, which imported rather than produced science and technology. Scientific and technical textbooks were primarily translated from French and English, and higher-education programs were copied from the West, with little consideration for local needs. While, in principle, this state of affairs should have made the acceptance of computer and other modern technologies simpler, in reality, much resistance had to be overcome.

Computer technology in the West developed organically from gradual transformation of military electronic systems used in the World Wars, relay-based telephony equipment, programmable textile machines, and mechanical calculators, to cite a few examples (Williams, 1997). Nevertheless, the development of electronic calculating machines came as a major shock to many. Even the inventors and innovators who brought these machines to bear did not at first realize the amazing potential of, and breadth of applications for, the new technology. Iran did not participate in the development of the just-cited prerequisite technologies, hence the greater shock and resistance. This is why Iran of the mid-20th century had to be reintroduced to notions and products of modern science and technology, including computers, almost from scratch.

When I worked at Sharif (formerly Arya-Mehr) University of Technology in Tehran, Iran, during 1974-1988 (the final two years on sabbatical leave in Canada), one track of my research dealt with technology transfer. Specifically, my colleagues and I were interested in adapting the then-new computer technology to Iran's economy, culture, and language (Mavaddat and Parhami, 1977). We preferred to use the term "technology transplantation" to better highlight the challenges of adaptation and acceptance. If not introduced in the right way, a society may reject a new technology, in much the same way that human body could reject a new organ that is ill-suited to its expectations, needs, and immune response. This is where the notion of "appropriate technology," that matches the social and economic conditions of a particular geographic area and promotes self-sufficiency, while also taking environmental factors into account, comes into play (Dunn, 1979). The preceding discussion defines the first research area (RA1) which guided my scholarly endeavors: What are the appropriate forms of computer technology for Iran's economy, workforce, language, and cultural heritage?

RA1 is one of 8 technology transplantation topics identified in Figure 1. This paper deals primarily with RA8 (devising suitable methods for outputting Persian information from a computer), with some attention paid to RA5 (appropriate methods and technologies for feeding Persian information to a computer). So, for the sake of completeness, I briefly discuss the other areas in the rest of this section. Please refer to Figure 1 as you read through the following discussion.

RA2 (adapting computers to the Iranian culture) is a key area that is closely tied to the design of suitable native educational programs (Parhami, 1985, 1986). In the late 1970s and early 1980s, I wrote a number of columns for Iranian newspapers and trade publications on the various aspects of cultural and linguistic adaptation (e.g., Parhami, 1978c). I also contributed a steady stream of articles to *Computer Report* (*Gozaresh-e Computer*) the monthly magazine of the Informatics Society of Iran. I intend to collect these thoughts into a separate article. Much societal good will can be generated toward computers and other technological symbols by showing that they can help in culturally important domains, such as calligraphy, archaeological restoration, literary analysis, or even generating designs for Persian carpets.

Adapting computers to the Persian language (the right-hand side of Figure 1) is where I spent much of my effort while working in Iran. The first item of business in this domain, RA3, was to make sure we could talk about computers, computer technology, and various application domains, using appropriate terminology within the framework of the Persian language (Parhami and Daie, 1981). Coming up with new terms, while ensuring technical and linguistic correctness, as well as acceptability by the public, is a rather difficult process, which is still ongoing in Iran. Many precise and linguistically-sound proposed terms (e.g., "raayaaneh" for "computer") have been pretty much ignored, while others (such as "reez-pardaazeshgar" for "microprocessor") have been widely accepted. Pragmatism and flexibility must be exercised in dealing with certain established terms (such as "electronic") that have found their ways into many languages. Such pragmatism clashes with Persian nationalism, which advocates for the purification of Persian by getting rid of all foreign words, be they European or Arabic (Kia, 1998).

Next, we need to communicate with computers to provide them with guidelines and instructions for doing what we want them to do. In this domain, RA4, language considerations are secondary. While there was some effort in the 1970s to "Persianize" programming languages (particularly the COBOL language, which was verbose and used a large number of English words in its original form), a consensus soon developed that what is difficult in programming isn't the learning of a few English terms, such as knowing what "while $x > 0$" means, but rather the mastering of programming notions and constructs. With a couple of exceptions, this line of research was all but abandoned by the turn of the century (Wikipedia, 2019), although it is now returning in the form of Persian voice commands (discussed under input below).

The rightmost branch of the tree in Figure 1, handling Persian language information, is the most-relevant to the subject matter of this paper (Parhami and Mavaddat, 1977; Parhami, 1978b, 1981). Processing of Persian information requires that information be input to the computer, using keyboards (Parhami, 1984b), natural-language audio, completed forms, or scanned documents fed to optical character recognition systems (Parhami and Taraghi, 1981), collectively constituting RA5. Gradual shift of focus from keyboards and other mechanical input devices to voice communication, as in digital assistants, gesture recognition, or even direct interfacing of computers with human brain, ties this research area with artificial intelligence and machine learning, both very active research and development areas nowadays.

Then, we need information storage and interchange codes, and perhaps data-compression and other storage support mechanisms, to keep the information and to send it between various storage and processing devices as efficiently as possible (RA6). Standardization is always hard, particularly where, as in the case of Persian, strong cultural elements are involved. Typical standards committees are composed of representatives of user groups, vendors, government entities, and consulting engineers, each member with his/her own agenda, biases, and preferences. Compromises on the part of stakeholders are needed to come up with a happy medium that is often anything but happy for most participants! This is why it took many years and several iterations, interrupted by the Islamic Revolution, to converge to an Iranian National Standard Information Code (Iran Plan and Budget Organization, 1980), with adoption, after many adjustments, taking several more years.

Storage of information is closely linked with its processing (RA7), which brings about the need for software and algorithms that may be language-dependent. A good example is

provided by the difficulties in the alphabetical sorting of Persian names, given multiple versions of letters and spelling variations, particularly for composite names. Peculiarities of the Persian script exacerbates the already difficult problems of sorting in general, particularly when special symbols are involved (Lemire, 2019). Activities such as content analysis and style-matching need to be undertaken in collaboration with experts in other fields, such as anthropology, law, literature, and sociology. Finally, translation of texts to/from Persian falls in this domain. Automatic translation, long a pipe dream for researchers, has assumed greater significance in recent years, given the global connectivity and information explosion in many world languages on the Internet (Doherty, 2016).

Last, but not least, is the problem of output production by computers, RA8, as shown in the highlighted box on the bottom-right of Figure 1. When the computer interacts with the outside world by directly controlling devices or systems, linguistic factors do not come into play, but when the output is printed, displayed, or spoken, language becomes a prominent factor. I will present a detailed discussion on producing printed and displayed Persian texts in the following sections of the paper, focusing on pertinent considerations, such as the legibility and aesthetic quality of the resulting script. Spoken output has seen limited research in the case of Persian, but perhaps this situation will change over the next few years.

FIGURE 1

Economic, cultural, and linguistic aspects of computing in Iran, defining 8 research areas.

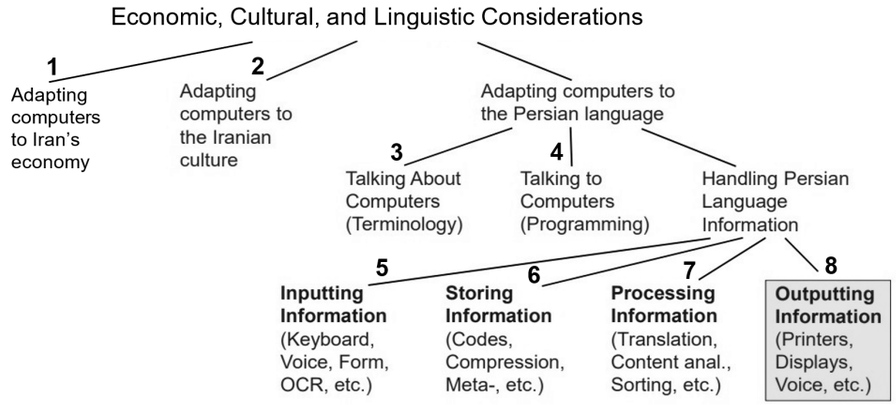


FIGURE 2

Calligraphic art by Farokh Mahjoubi, left, and uncredited handwritten Nastaliq.

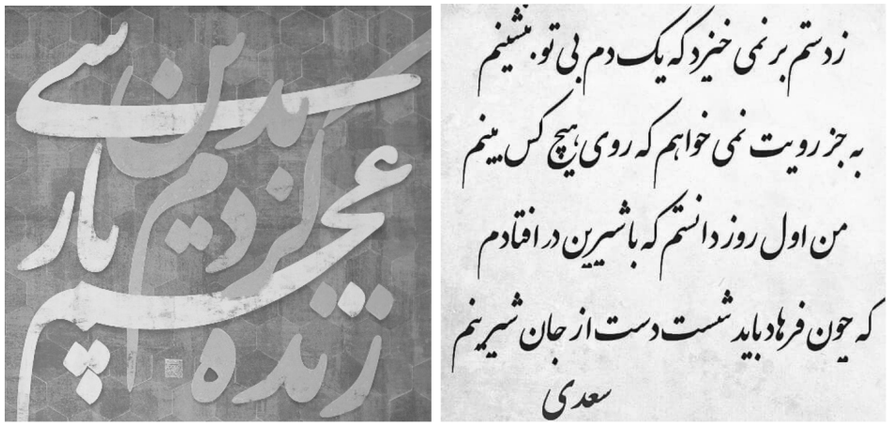
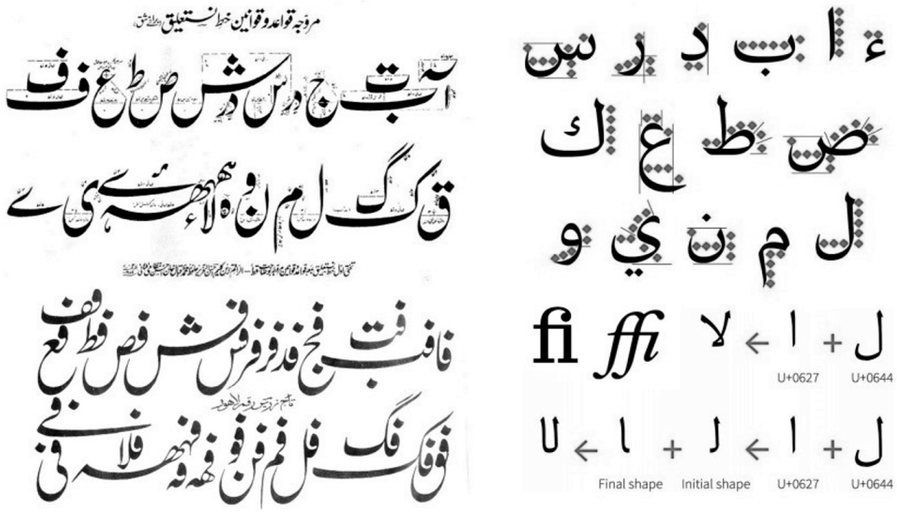


FIGURE 3

Persian script: Rules for beautiful writing of letters and letter combinations



3. A Brief History of the Persian Script

The modern Persian script is believed to be about 1200 years old, when it was adapted from Arabic (Lazard, 2008). The Persian language itself is much older and, along with the script, came in previous versions of Old Persian and Middle Persian before then. For much of this 1200-year history, the script was written manually by scribes, who were hired by poets and others who wanted their writings archived and disseminated. The script used for this purpose was often Nastaliq, written with an eye toward legibility and beauty (Figure 2, right panel). Each book produced in this way was one-of-a-kind, though a kind of primitive printing, consisting of etching the text on stone or wood and pressing paper or parchment against it after inking, was used on occasion to produce multiple copies.

The vast importance placed by Iranians on beauty in Persian script led to the appearance and quick spread of calligraphic art (Figure 2, left panel). In calligraphy, the script's legibility is secondary, the prime focus being the proportions and interactions of textual elements and the color scheme. The perfection of this art form led to kings and architects becoming keen on incorporating calligraphic art on mosques, palaces, and other important buildings. In the process, artists developed special scripts that would make it easy to write Persian and Arabic texts, using tiles and other construction elements. In our subsequent discussion of dot-matrix displays and printers, we will return to the use of "pixel fonts," where a black or white tile represents a pixel in forming the desired text with an array of tiles.

With regard to technological adaptation, the Persian script went through three phases of change, as the printing press, typewriters, and eventually computers entered the scene. The changes were cumulative, in the sense of the changes made in response to the printing press persisting through the adaptation to typewriters, and the latter changes affecting computer printing. Sections 4 and 5 outline the first two phases, 300-400 and 100-120 years ago, respectively, with the third phase (which is my main focus) beginning around 50 years ago, covered in Sections 6-9. Much of what I write about printing the Persian script also applies to generating it on display devices, but I will devote the latter part of Section 8 to the specific challenges and successes in displaying the Persian script. Bear in mind that nearly everything I state about Persian script is also applicable to Arabic.

Before the need arose for adapting the Persian script to modern technology, two major scripts Nastaliq and Naskh, were in common use, alongside many other less-popular ones. Rules for writing beautifully in various scripts were passed on from masters to students. For example, sizes and alternative forms for the letters of the alphabet were often specified in terms of the number of dots, horizontally and vertically (Figure 3, right panel). Each master had a unique set of rules that set him apart in terms of style and reputation. Just as in English calligraphy, where certain combination of letters such as "fi" and "ffi" are specified as a unit, much like a single letter, Persian calligraphy has its rules for letter combinations, such as "laam-aleph" (the letter "laam" followed by the letter "aleph," to form the combination "laa," which looks quite different from "laam" and "aleph" letters put together), with such combinations being much more prevalent in the Nastaliq script (Figure 3, left panel).

The Naskh script has become much more common than Nastaliq and has developed into a multitude of forms, because of its better match to the capabilities and limitations of modern display and printing technologies, as discussed in Sections 4 and 5. However, Nastaliq still holds a special place in the hearts of Persian-speaking communities worldwide. For example,

the fanciest books of poetry are commonly produced in Nastaliq, which requires a calligrapher to write the entire book by hand, before it goes to offset printing. Flyers, posters, and brochures, which use Naskh-based print scripts for the most part, tend to include bits and pieces of Nastaliq for emphasis, such as in titles and headings, or for decorative purposes. Some progress has been made in producing the Nastaliq script on computer displays and printers, as discussed near the end of Section 8, but the script quality is not yet at a level to threaten the livelihood of master calligraphers.

It is worth noting that book production in Persian has a much longer history than the printing press or typewriters (subjects of Sections 4 and 5, respectively). In 2014, the US Library of Congress sponsored an exhibition, featuring an extensive lecture series, to celebrate 1000 years of the Persian book (Library of Congress, 2014). An interesting history of publishing in the Persian language exists (Moradi et al., 2013), which despite packing much useful information, unfortunately lacks any photos or diagrams to convey the machines and mechanisms used for printing.

4. The Printing Press Goes to Iran

The printing press arrived in Iran some 400 years ago. In 1618, Shah Abbas I was presented with a set of Arabic and Persian letters and expressed interest in procuring them for his country (Floor, 1990). A printing press and associated types arrived in Isfahan in 1629, but there is no record that the machine was ever used. Printing machines were not used widely for another century, in part because non-trivial adaptations were required. Invention of the Stanhope hand-press in 1800 revolutionized the printing industry, given its relatively small size and ease of use. Iranians traveling in Europe and Russia brought the Stanhope hand-press to Tabriz in 1816 (Green, 2010). Within a couple of years, the first Persian books were printed in Tabriz and, shortly thereafter, in Isfahan and Tehran.

Movable-type printing was based on rectangular metal blocks each of which had the shape of a symbol embossed on one surface. Combining the blocks horizontally formed the text to be imprinted on a line (Figure 4). The typesetter would have in front of him large trays of types, with compartments for different letters and symbols. Multiple trays were needed to house different sets of fonts, such as those with different font sizes, italics, boldface, and so on. Several such lines were completed by the typesetter on a small hand-held tray and, when the small tray was filled, its contents were transferred to a larger tray and the setting of the next chunk of text began. This process continued, until 8 complete pages had been set onto a large tray. These 8 pages would be printed on one side of a large sheet of paper (lower-left panel of Figure 4). Another 8 pages were set for what was to appear on the back side of the large sheet, with page numbers arranged as shown in the top-left panel of Figure 5.

The sheet of paper with 16 pages of content would then be folded and put together with other folded sheets to form a complete book, whose number of pages was a multiple of 16 (sometimes necessitating blank pages at the end). These 16-page blocks were sewn together along the book's spine. The edges of the folded sheets were then cut with special guillotines to form the completed book, which then underwent the process of binding to put a hard or soft cover on it. While some books are still put together by sewing the spine, albeit mechanically, rather then by hand, modern book-making usually entails the use of glue, which is cheaper, faster, and results in flatter spines. Imperfect cutting in some older books led to the

edges of some pages remaining connected to each other, necessitating manual separation by the reader.

FIGURE 4.

Hand-typesetting with movable type, the tray holding the typeset text, and the printing process.

FIGURE 5.

Image transfer to large sheets of paper, folding the sheets, and book-binding.

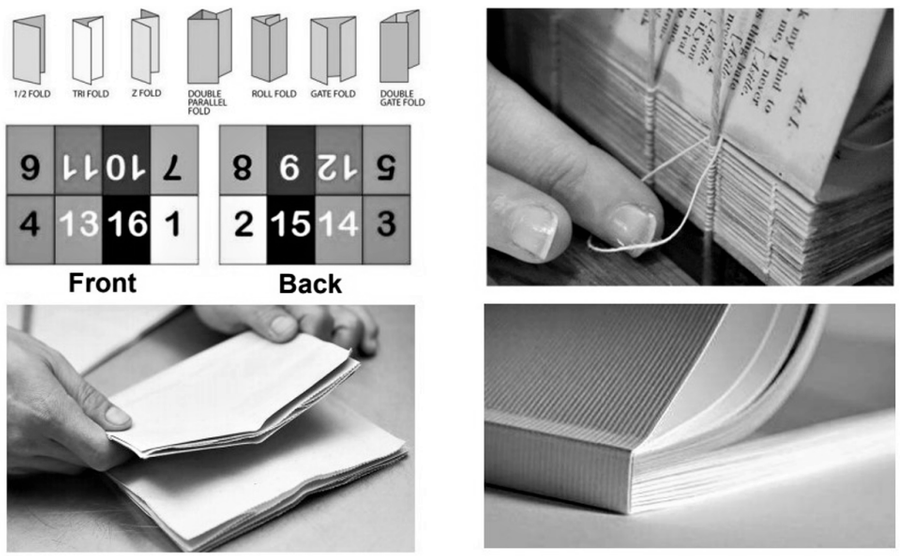


FIGURE 6

Movable Persian type in the compartments of the typesetter's tray.

The first challenge in Persian printing was to make the needed metal blocks holding the letters and other symbols (Figure 6). This wasn't a straightforward process. Whereas the Latin alphabet consists of individual letters that are set next to one another, with a small gap between them, Persian letters needed to be connected to each other. Furthermore, the handwritten script was not readily decomposable into separate, non-overlapping symbols. Some examples are provided in the left panel of Figure 7. The word "mojtame'e," in which the first two letters "meem" and "jeh" are almost completely overlapped horizontally, as are the next two letters "teh" and "meem," must be stretched out, so that letters appear side by side, rather than on top of one another.

So, calligraphers were recruited to design new fonts for Persian in which composability was placed ahead of aesthetic quality as a design consideration. The same word "mojtame'e" now appeared as shown in the center-top part of Figure 7, with arrows pointing to its five constituent letters placed side by side. A similar problem arises for certain wide and/or tall Persian letters, whose handwritten forms might overlap with adjacent letters. In the word "changaal," for example, shown at the bottom-left of Figure 7, the upper part of the letter "gae" extends to the left to overlap with the letter "noon" that precedes it. All such problems were solved by compromises in font design that traded off aesthetic quality, and in some cases readability, for composability.

Not all features of the Persian script cause difficulties in typesetting: there is also one helpful feature. Many Persian letters can be extended or elongated, as needed to fit the available space. This feature comes in handy when typesetting poems, which are traditionally formatted into two columns consisting of the first and second half-verses, with each column left- and right-justified. We see in the example supplied in the right panel of Figure 7 that the letters "feh" and "yeh" in the circled word "aafaridam" have been elongated to ensure a justified left margin for the poem. The elongation mechanism is a dash-like symbol, that, like Latin dashes, comes in different lengths and is aligned with the horizontal line, the connections axis, on which all inter-letter connections take place. So, any letter to be connected to the preceding letter begins on the connection axis, and any letter to be connected to the following letter ends on the connection axis.

Besides connectivity of letters and horizontal overlap, certain other difficulties are inherent in the Persian script. Unlike the Latin script, in which letters have regular geometric shapes of comparable sizes (top half of Figure 8), Persian letters are more curved and exhibit wider variations in width and height. Generally speaking, the Persian script is more compact horizontally, in the sense that the Persian translation of an English sentence tends to take up less space horizontally. This horizontal compactness is mainly due to short vowels not being written out in Persian. However, the Persian script takes up more space vertically, in part due to the need for more line spacing to keep the letters with tall lower parts on one line separated from letters on the following line that have tall upper parts. As a result, the metal type blocks for Persian have greater variations in size, making their manual handling by the typesetter somewhat more difficult.

This wider variation in letter sizes has the side effect of making fixed-width type much less legible than its Latin-script counterpart (more on this, when I introduce typewriter fonts). The bottom half of Figure 8 contains a few verses of a poem, set with fixed-width and variable-width Persian fonts. The difference in legibility, aesthetic quality, and compactness are quite evident. The Persian alphabet has four additional letters compared with the Arabic alphabet, but these added letters have shapes that are very similar to existing Arabic letters, so font design and considerations of legibility and aesthetic quality aren't any different between the two languages, and we can pretty much ignore in our discussion specific mention of Arabic. One way to impart the difficulties discussed so far to non-Persian-speakers it to liken them to those encountered in printing Latin calligraphic script, which has both the problem of connectivity between adjacent letters and wide variations in letter sizes, particularly between initial-capital and regular letters.

The problems of Persian typesetting were transformed and somewhat eased when Linotype machines arrived on the scene in the 1950s, decades after their use for Arabic typesetting (Ross, 2017). Linotype, an invention of Ottmar Mergenthaler, a German immigrant to the United States, was a hot-metal typesetting system that replaced the work of hand typesetting with automatic casting of an entire line of text, hence its name (Kahan, 2000). Text was entered through a keyboard and the machine assembled matrices that were essentially molds for entire lines of text, thus enabling a significant reduction in typesetting personnel. Iran's *Ettela'at* newspaper commissioned the manufacture of Persian font matrices from Linotype in 1957 and successfully integrated the resulting text blocks into mostly hand-typeset text in the paper (Nemeth, 2017).

Linotype was influential in the printing business, particularly newspaper publishing, for a couple of decades, eventually giving way, in the 1970s and 1980s, to photo-typesetting. While a labor-saving invention, Linotype did not have a significant effect on the design of Persian print fonts and challenges therein, given that it also relied on juxtaposing pre-designed symbols with fixed shapes to form lines of text. In fact, the success of Linotype was judged based on how close it came to replicating the appearance of hand-set type (Nemeth, 2017).

Linotype affected the print quality indirectly, by reducing the error-prone labor of hand-type-setting and proofreading and allowing newspapers to carry more pages, while still meeting strict publication deadlines. Many font designers contributed to the Linotype library of fonts and later to more diverse families of fonts (Devroye, 2019a).

Photo-typesetting made the printing process less bulky and more economical. It formed the desired page content on film, which was then used to transfer ink to paper via the use of light-sensitive paper or offset printing (Twyman, 1999). For many years during my academic tenure in Iran, I had first-hand experience with offset printing, which entailed assembling the contents of my class notes or a scientific journal I edited onto sheets of paper, via typewriting and cut-and-paste of text and images. The sheets were then delivered to the print shop, which photographed them, transferred the images onto aluminum sheets, which were then used to print the galleys onto paper. Offset printing constituted the final step in the transition from movable-type fonts to computer-based digital printing.

The samples of newspaper front pages depicted in Figure 9 provides a sense of modern Persian print scripts of various designs and sizes, collected in July 2017, upon the passing of Iranian-born mathematician and Fields-Medal winner Maryam Mirzakhani. Even though headline fonts are significantly larger than normal-text fonts in print media, legibility issues are pretty much the same, given that headlines must be readable from a much larger distance.

FIGURE 7

Adaptation of the Persian script to movable-type print technology.

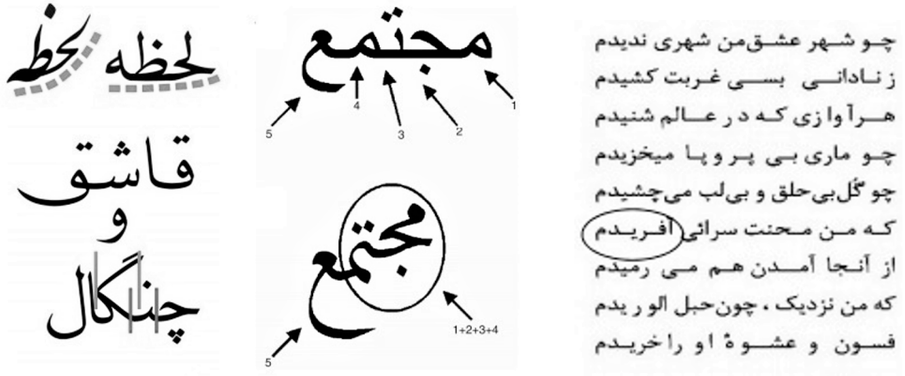


FIGURE 8

Comparing the Latin script and fonts with Persian/Arabic script.

FIGURE 9

Modern Persian print fonts used in today's newspapers.

5. Persian Typewriters

Typewriters entered Iran around 120 years ago, but again it took a while for them to be used widely. Another set of adaptations were needed to allow Persian typewriting. In print type, each Persian letter has up to four shapes, corresponding to it appearing alone, at the beginning of a word, in the middle of a word, or at the end. The solo, initial, middle, and end forms of some letters, such as "ein," shown at the top-left of Figure 10 are quite different, requiring 4 typewriter keys (or two keys, without or with shift) to cover them. With 32 letters and 4 shapes, even ignoring digits and special symbols, 64 keys would be needed, along with the shift key. Fortunately, however, for most Persian letters, the initial and middle forms, and the solo and end forms, are sufficiently similar to allow their combining, with no great harm to the resulting script's readability or aesthetic quality.

For example, the letter "meem" will not be damaged too much if it is given two forms rather than four. Ditto for the letters "beh" (bottom of Figure 10) and the letter "heh" (left edge of Figure 10), with slightly more compromise. This combining process allowed Persian typewriter keyboards to have only two forms for each letter, the only exceptions being the letters "ein" and "ghein."

A breakthrough in typewriter technology occurred when smart electronic word-processors and associated software were introduced. Such smart input devices needed just one form of each letter on the keyboard, with the actual shape of the printed letter determined based on context (the previous and next letter). For example, if you want to type the word "kamtar," composed of the four letters "keh," "meem," "teh," and "reh" (recall that short vowels such as "a" are not printed, except perhaps as a diacritical mark, as shown near the top-right of Figure 10), you begin by entering "keh." At this point, the typewriter uses the solo shape, given that it has no information on what will come next. A display unit actually displays the solo "keh" and changes it later, if needed. A typewriter must delay the printing of each symbol by one position, until it can determine the correct shape. Once you type the second letter "meem," which connects to "keh," the initial shape of "keh" is established, but the shape of "meem" depends on what comes next. This process continues until the end of a word (hitting of the space key) or line (return key) is encountered.

Mechanical Persian typewriters had keyboards such as the one shown in the left panel of Figure 11. It was recognized early on that the arrangement of keys on the keyboard should be standardized, so that typing skills, once acquired, do not have to be re-learned. So, a standard layout was agreed upon and published (Institute of Standards and Industrial Research of Iran, 1976). Over the years, the print mechanism, which initially consisted of hammers hitting the ribbon and paper upon depression of a key, underwent improvements.

One of the highest-quality fonts for typewriters was offered by IBM in its Selectric line of electric typewriters (Reisinger, 2011), which featured an ingenious golf-ball print mechanism (right panel of Figure 11). When a key was pressed, the golf-ball mechanism would tilt and rotate to align the selected symbol with the print position before striking the golf-ball against the ribbon and paper. The golf-ball was easily removable for replacement with another golf-ball bearing a different font (italic, boldface, symbol, etc.), making is easy to typewrite technical manuscripts involving multiple typefaces and equations. Even multiple languages could be easily incorporated in the same document. In fact, I used such a typewriter to produce my first textbook, *Computer Appreciation* (Parhami, 1984a), sample pages of which appear in Figure 12.

Given that offering a new typeface for the IBM Selectric entailed only the design and production of a new golf-ball print mechanism, a large number of typefaces were developed over time, making the Selectric highly popular. These fonts are catalogued in a variety of online sources (e.g., Devroye, 2019b; Monk, 2013), which unfortunately do not offer any information on the Persian/Farsi typefaces. I recall having access to 6 or 7 different golf-balls for Persian, including italics, boldface, and a couple of "fancy" fonts. IBM Selectric fonts came in 10-per-inch and 12-per-inch pitches, and the typewriter carriage could be set to move from left to right or right to left.

In addition to its advanced and high-quality print mechanism, later models of IBM Selectric had a storage buffer for storing the content of an entire typed line and an erasing capability. When the backspace key was pressed, the previous printed letter was read out from the buffer and was printed again, but with a white erasing ribbon instead of the black printing one. This caused the letter to be erased, providing a simple mechanism for correcting errors, without the mess of white-out liquid and its smudging problem. IBM Selectric can be viewed as an important stepping stone in the transition from simple or "dumb" Persian typewriters to full-blown Persian word-processors of later years, which eventually evolved into Persian desktop publishing systems (Sanati, 2018).

There was a push in the mid-1970s within Iran's academic circles and the government's Plan and Budget Organization to devise a unified standard layout for both typewriter and computer keyboards. An attempt in this direction, which was never approved as a standard or implemented, is depicted in the left panel of Figure 13 (Parhami, 1984b). Taking a cue from the "QWERTY" name for standard Latin keyboards, the group working on the standard under the auspices of Iran Plan and Budget Organization called this keyboard layout "Zood-Gozar," based on the letters appearing in the middle section of the bottom row of keys. There was also a proposed standard information interchange code (Iran Plan and Budget Organization, 1980; Parhami, 1984b), an extension of ISO standard code to the Persian language, which was also not adopted, as we hit the strikes and disruptions of the last couple of years before Iran's Islamic Revolution, when government activities came to a stand-still.

FIGURE 10

Persian letters have four versions in print, two in typewriters, and one with smart text editors.

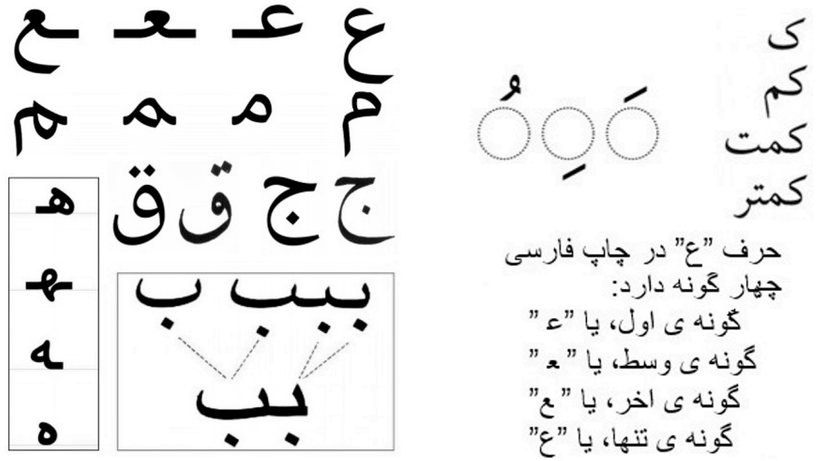


FIGURE 11

An early Persian typewriter keyboard and IBM Selectric golf-ball print mechanism for Persian.

FIGURE 12

IBM Selectric's Persian script, as used to typeset a book by the author.

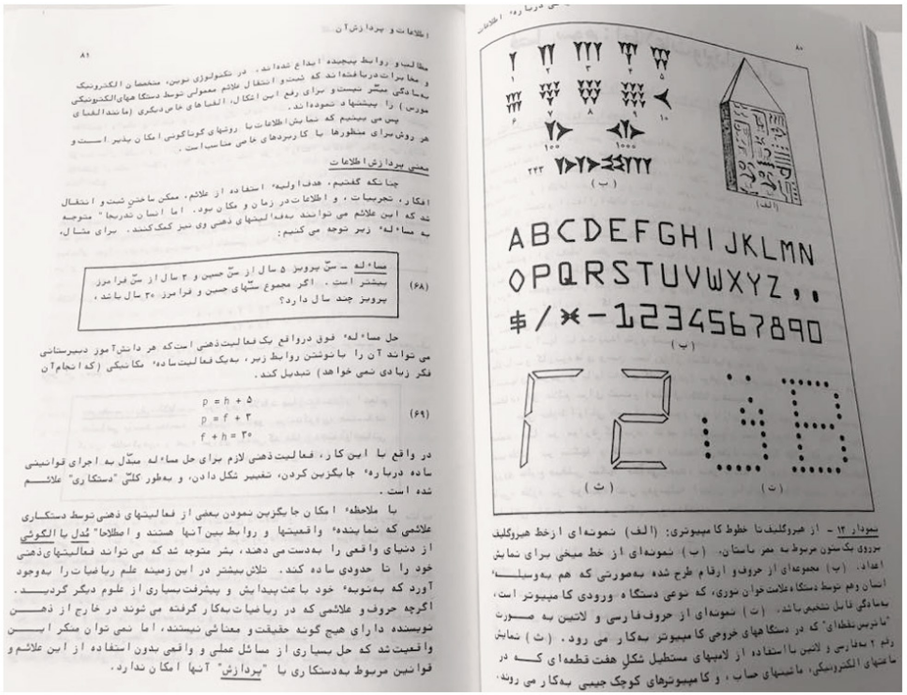
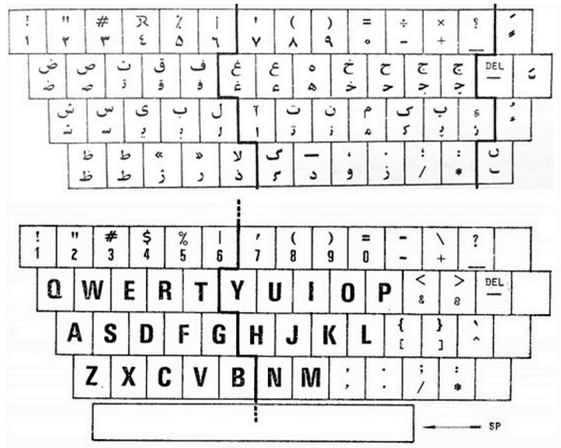


FIGURE 13

Proposed standard Persian keyboard and information-interchange code.



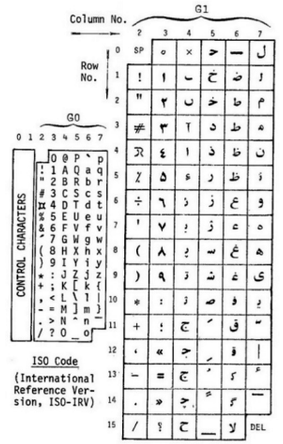


FIGURE 14

Drum printer mechanism, sample of the resulting Persian script, and an early line-printer.

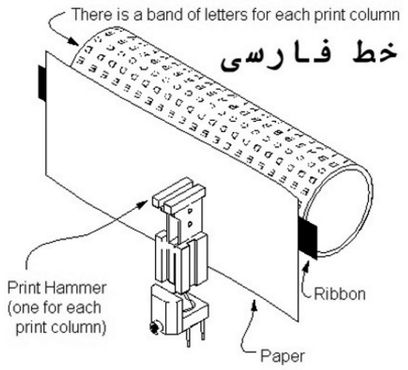
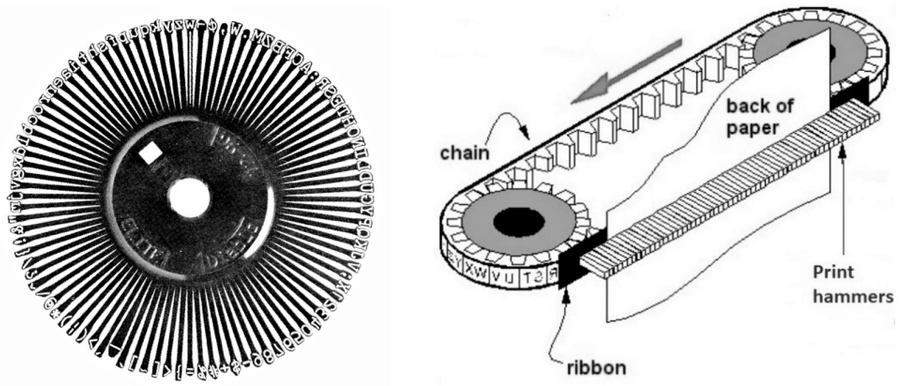


FIGURE 15

Chain (right) and daisy-wheel print mechanisms led to more compact printers.



6. Early Computer Printers and Displays

The earliest computer printers were known as line-printers, because they printed an entire line of text (typically on a wide paper, holding 132 columns) at once, advancing to the next line via moving of the continuous-feed fan-fold paper (right panel of Figure 14). The print mechanism in the earliest line-printers was a drum (left panel of Figure 14), with one ring of symbols for each of the print positions and each ring containing the entire printable character set. As the drum rotated at high speed, every letter eventually got aligned with the position of the line being printed. At that instant, a hammer would hit the paper and ribbon against the drum, creating the image of the desired symbol. In one drum rotation, taking a few milliseconds, one line would be printed. A drawback of drum printers for the Persian script was the gaps between adjacent symbols, necessitated by the placing of letters on the drum and also the spacing needed between hammers. The resulting fixed-width script appears near the center-top of Figure 14. Slight errors in the timing of the hammer strikes would lead to the letters on the line to be mis-aligned, further deteriorating the script's legibility and aesthetic quality.

An innovation, that improved the print legibility and quality but was short-lived because of other problems it created, was switching the direction of lines and columns on a drum printer. With this variation, print lines would go along the perimeter of the drum rather than across the length of it. Consecutive Persian letters were now printed by the same hammer that could strike at any time, thus, at least in theory, removing the gap between symbols completely. However, errors in the timing of hammer strikes was still problematic, as was the need to print many times, even for pages that contained only 1 or 2 lines of text (the normal line-printer would skip the rest of the page once the few needed lines had been printed).

I conclude my discussion of early computer printers by introducing two other common print mechanisms as examples. Chain and daisy-wheel printers, whose print mechanisms are depicted in Figure 15, were cheaper and less bulky than drum printers, but they suffered from similar script-quality problems. In both cases, hammers would hit the embossed letters, on chain links or on the petal-like arms of the daisy-wheel, to print the corresponding letter on paper.

Before discussing how computer printing technology advanced beyond the early drum, chain, and daisy-wheel print mechanisms, let me discuss the technologies for outputting information on display units. One of the earliest and simplest display units for numerical data was the 7-segment display, shown at the top of Figure 16. Seven small lights, later made of light-emitting diodes (LEDs), would be turned off and on to illuminate the shapes of decimal digits, which we need on calculators, clocks, thermometers, and the like. This kind of line-segment display units were later extended to letters and other symbols (bottom of Figure 16), using, for example, 9 segments, 14 segments, or other designs, with the results being acceptable for upper-case Latin letters, but rather deficient when used to display lower-case letters and certain special symbols.

We, at Sharif University of Technology, tried our hands at creating numeric and alphanumeric line-segment displays for Persian, with the proposals shown in Figure 17 (Parhami, 1978b). We found out that a reasonably readable set of numerals can be created using 7 curved segments (top of Figure 17) and that no fewer than 18 segments were needed to approach a barely readable set of letters (bottom of Figure 17). Both proposals remained at the research and prototype stage, as the approach was overtaken by the more flexible dot-matrix display technology. A 7-by-5 matrix of lights (7 rows, 5 columns), for example, was adequate for producing a readable Latin script consisting of numerals and uppercase letters for use on airport information boards, stadium scoreboards, and other environments where alphanumeric information was to be displayed in public places. As the capabilities for making smaller dots/lights and to place them closer to each other developed, the quality of the resulting script improved through the use of 9-by-7 and even larger dot matrices.

Early CRT display units either generated desired textual and graphic elements by drawing lines on the screen (the way we write by drawing straight and curved strokes) or were equipped with special character-generator hardware that would direct the electron beam hitting the phosphorus screen to form the shape of a desired character. Eventually, CRT designs settled on the vector scheme, where a buffer would hold the dark/light or 0/1 pattern to be displayed on a dot matrix and the data in the display buffer was used to turn off or on the electron beam, as it scanned the surface of the display, according to the 0s and 1s in the display buffer. Color displays essentially held three copies of the said mechanism, one for each of the primary colors. Eventually, CRT displays would give way to flat-screen displays of various kinds, but the dot-matrix display method is here to stay.

FIGURE 16

Line-segment display devices for numerals, Latin alphabet, and special symbols.

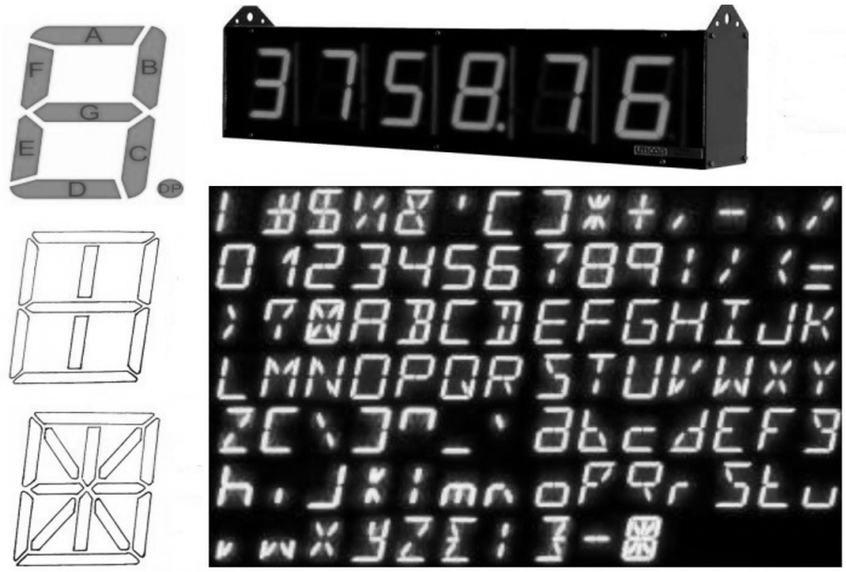


FIGURE 17

Proposed line-segment display devices for Persian numerals and letters.

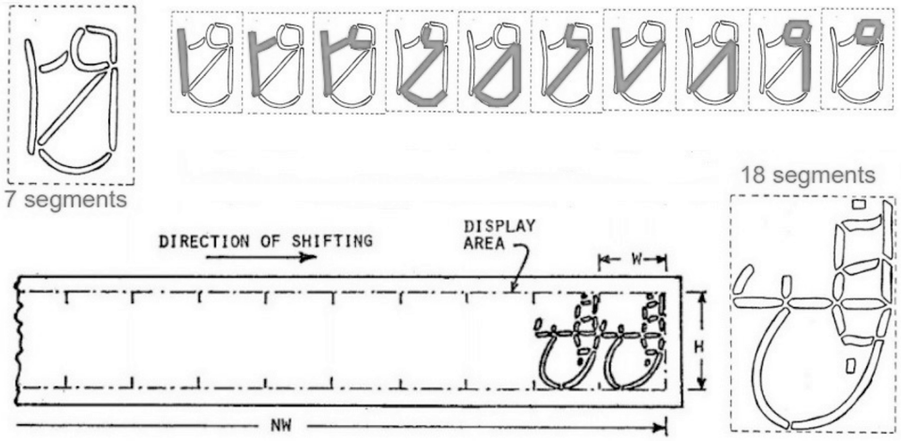
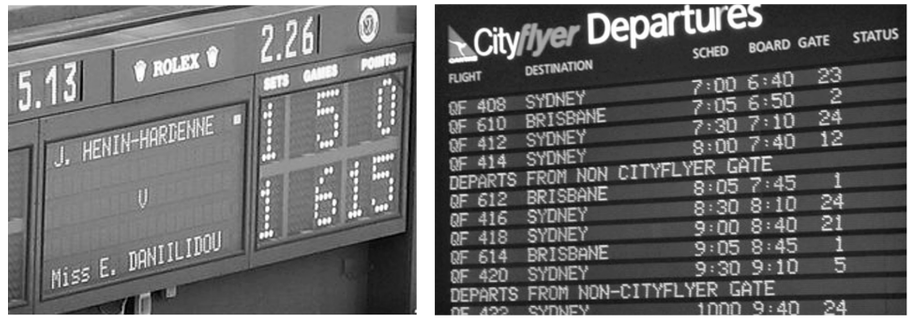


FIGURE 18

Early dot-matrix displays were used at airports and on stadium scoreboards.



7. Progress in Computer Printing Technology

Several other printing technologies came and went, until improvements in dot-matrix printing made all other methods obsolete. The dot-matrix printing method has some elements in common with the square Kufic script, developed by artists to facilitate writing with tiles on the façade of mosques, palaces, and other buildings (left panel in Figure 19). Just as a ceramic artist might have set all the tiles in one column before moving on to the next column, early dot-matrix printers had a column of 7 print needles that would be set to high and low, before being pushed forward to produce contact with a ribbon and paper, proceeding to the next column, either by moving the print mechanism or shifting the paper (right panel of Figure 19). Eventually, more than 7 needles and more than one column of needles, aligned or skewed, would improve the print quality. The needles later gave way to miniature ink-jets that shot a tiny drop of ink toward the paper to form a dot.

Dot-matrix printers offered extreme flexibility. Not only could we design fonts in a variety of ways, using our own preferences and trade-offs, we could also produce maps and geometric shapes, such as lines and circles (Figure 20). As the dots became smaller and more of them could be placed in one square inch of space, print and image resolution improved, so much so that, today, we barely notice the tiny dots that form the letters and images on paper or on screen. Problems arising from the connectivity of Persian letters and their disparate sizes also disappeared. What was a hardware problem (designing print mechanisms and fitting them to the Persian script) is now a software exercise: specifying the light and dark dots that form each Persian character.

My colleagues and I, therefore, got busy producing dot-matrix character sets for the Persian alphabet, as had been done previously for the Latin alphabet (Figure 21). Once a dot-matrix alphabet had been created, we would copy the letters on small pieces of paper and would juxtapose them to write words and sentences (middle part of Figure 21). If the quality wasn't to our satisfaction, we went back to the drawing board and made changes to letter designs. This process was eventually mechanized and we created the sample texts on a screen rather than by arranging pieces of paper. As dot-matrix size and resolution grew, more legible and aesthetically pleasing symbols could be created.

The impact of dot-matrix size on font quality is seen on the left half of Figure 22, where the letter "R" is generated in two matrices: a 7-by-5 matrix and a 16-by-16 matrix with overlapped dots that leads to smoother curved edges. Interestingly, the legibility of a script is connected with the ease of recognizing it automatically by a computer program. In an early Persian character-recognition study using newspaper headlines (because, at the time, we did not have high-resolution scanners to use normal printed text), we solved the problem of how to decompose chunks of printed text into their constituent letters (Parhami and Taraghi, 1981). The result for the first text chunk in the word "Tehran" is depicted in the right panel of Figure 22, where the first three letters have been separated before recognition. The results of this study not only showed the feasibility of automatic decomposition and recognition, but also helped us in the design of dot-matrix fonts with better legibility. Since our early work of the late 1970s, the field has advanced by leaps and bounds, so much so that a Google search for "Persian character recognition" produces more than 6 million hits.

Dot-matrix display and printing finally realized the goal of high-quality Persian output, while also allowing certain exotic fonts (Parhami, 1978a). As seen in Figures 23 and 24, the change in the Persian script quality was already significant from early line-printers and rudimentary dot-matrix printers of the 1970s to second-generation dot-matrix printers of the 1980s. This improvement trend has continued unabated in the three decades since, to the extent that today, we barely notice the collection of dots that forms each displayed or printed symbol. I will present the latest in Persian output-script production in Section 8, but the reader can preview the modern script quality in Figures 29 and 30.

The production of high-quality print and display output has also quelled sporadic attempts at reforming the Persian script or changing the alphabet (Borjian and Borjian, 2011), a la what was done in Turkey, to save the society from "backwardness." I consider this outcome quite fortunate, as the right approach to technology transplantation is adapting technology to cultural factors, not the other way around. The extreme of insisting on language purity, which leads to resistance to new concepts and inflexibility in making trade-offs for the sake of producing legible output, is just as misguided as wholesale abandonment of centuries of traditions for short-term technological gain.

FIGURE 19

The square Kufic script for writing with tiles, and early dot-matrix printing technology.

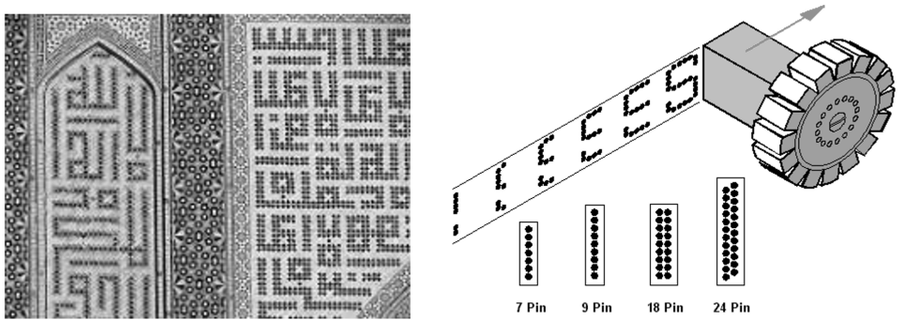


FIGURE 20

Dot-matrix style of printing and display is quite flexible, but it isn't a cure-all.

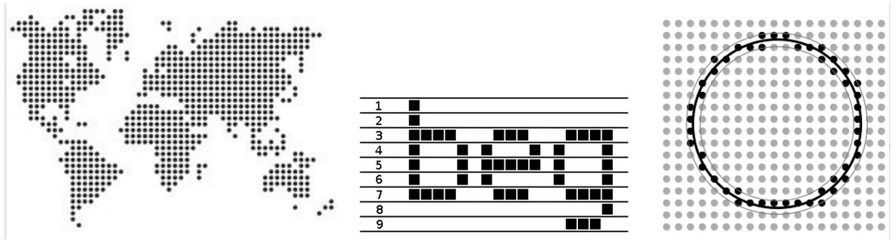


FIGURE 21

Dot-matrix designs for some Persian letters, compared with those for the Latin alphabet.

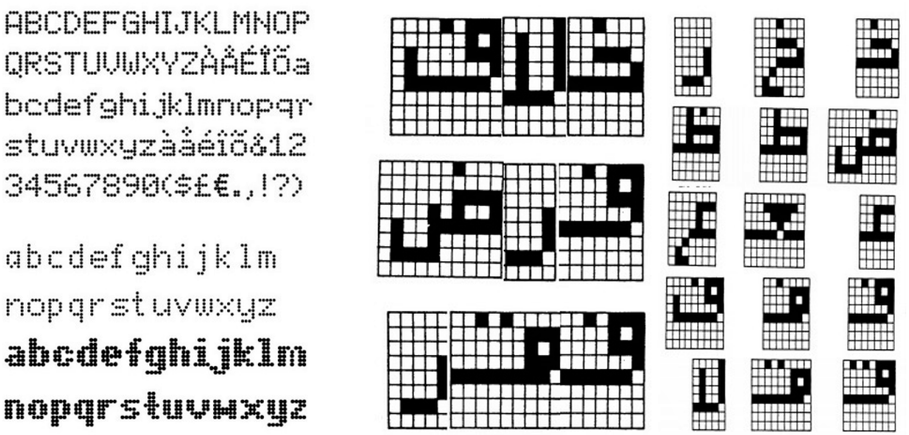


FIGURE 22

Small 7-by-5 vs. large 16-by-16 (overlapped) dot matrices and example of connected letters.

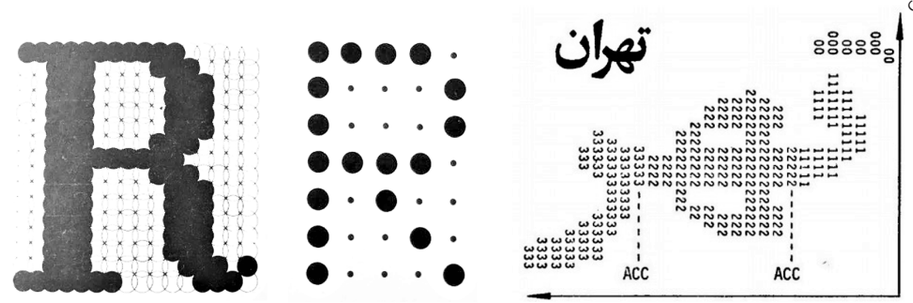


FIGURE 23

Samples of early computer printouts (1970s).

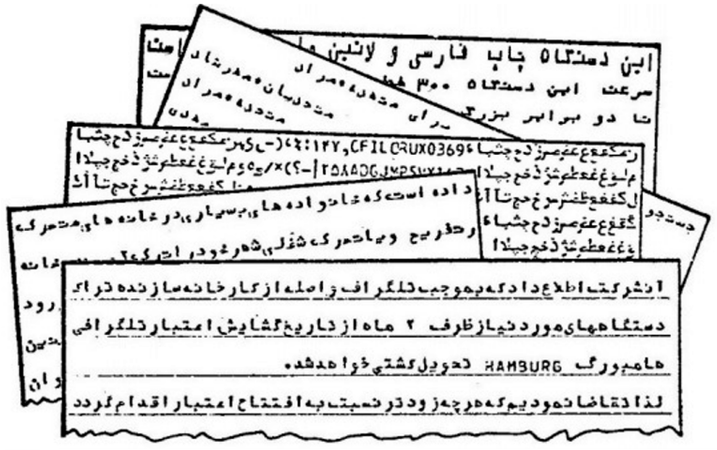
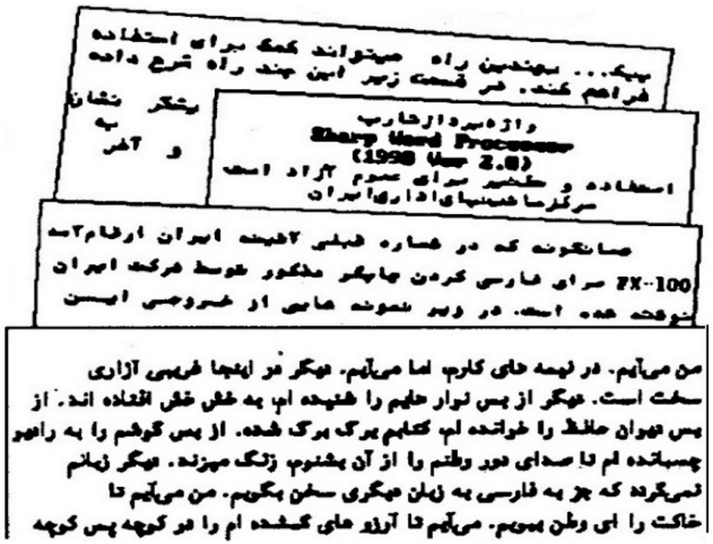


FIGURE 24

Samples of later computer printouts (1980s).



8. Persian/Arabic Scripts on Today's Computers

Much progress has been made in improving the script quality in Persian printers and displays, all of which now use the dot-matrix scheme to form the symbols. To cite an example, the Persian font quality on Facebook shows continued improvement. The improvement does not always come on all platforms at once. In 2017, I made several suggestions for font quality improvement to Facebook and, in late 2017, a new font was rolled out on the mobile Facebook platform, but unfortunately, the desktop platform font (left panel in Figure 25) is still illegible,

unless the image is enlarged significantly. The right panel in Figure 25 shows the font used on an on-line news site, which is significantly better. The lesson here is that users of Persian script should complain loudly at every opportunity in order to get better print and display quality. Now that we have huge data sets of Persian texts, such as posts on social media, there is no excuse not to use the data set to find out about user tastes and preferences, perhaps in a crowd-sourcing format.

As mentioned earlier, legibility problems of Persian fonts lead directly to difficulties in automatic script recognition. The newspaper headline in Figure 26, from a four-decades-old study (Parhami and Taraghi, 1981), depicts some of the problems in both domains: discerning connection points (a), letters that are very similar in shape (b_1, b_2), wide variations in letter widths (c_1, c_2), horizontal overlap between letters on a line (d), and vertical overlap between lines (e), all of which make segmentation of text into its constituent letters difficult. A byproduct of the study was the conclusion that for accurate printed-text recognition, a "pen-width" of 4 dots (left panel in Figure 27) seems to be necessary. This is also a good guideline for highly readable printed and displayed Persian script. With a pen-width of 4 or more, letter features, such as curvatures and holes, become easily noticeable (Figure 27).

Now that we have an idea about the desirable dot-matrix size to accommodate a pen-width of 4 for high-quality fonts, it is also instructive to probe the lower boundary on matrix size to generate a script of adequate quality for certain mobile and embedded applications, where cost factors may inhibit the use of large dot-matrices (Parhami, 1995). The simulation results for fonts in 7-by-5, 7-by-9/2, and 9-by-9/2 are shown in Figure 28 (middle panel, top to bottom). A matrix dimension *m*/2 implies the presence of *m* rows/columns of dots in skewed format, so that the physical dimension of the matrix is roughly *m*/2, despite the fact that there are *m* elements. This kind of skewed arrangement of dots helps with generating fonts of higher quality, when the letters have curved or slanted strokes.

FIGURE 25

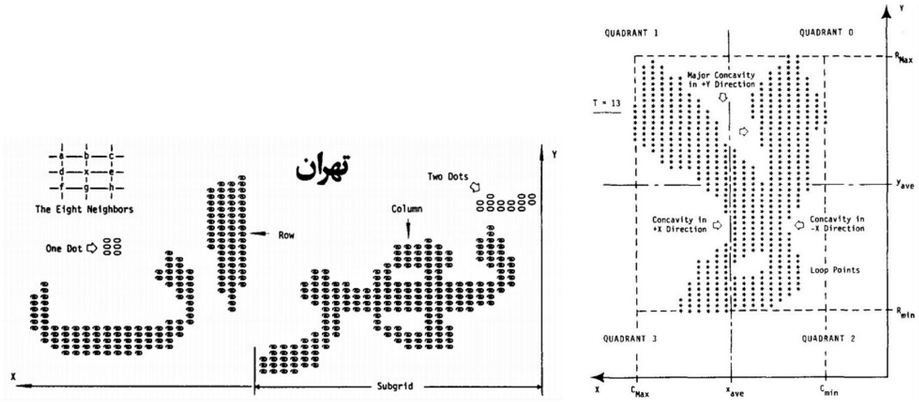
Sample dot-matrix display typefaces from Facebook (left) and an on-line site.

FIGURE 26

Some of the challenges in the automatic recognition of printed Persian texts.

FIGURE 27

Techniques and results from the automatic recognition of printed Persian texts.



Given that certain tall Persian letters are particularly challenging for displaying or printing with small matrices, a separate simulation was conducted with sample texts featuring many tall letters. Partial results are shown in the top-right of Figure 28 for the same three dot-matrix dimensions, along with the sample text represented with variable-width dot-matrices. The study's conclusion was that a 9-by-9/2 dot-matrix is the absolute smallest size for somewhat legible Persian script.

In modern computer applications, scripts need to be displayed on devices of varying sizes, from large desktop screens to tiny displays on mobile devices. So, each font, is designed in various point-sizes (top of Figure 29). Legibility considerations dictate that a 20-point font,

say, not be a linear reduction by a factor of 2 of a 40-point font. In some cases, such as viewing a map on a small screen, the fonts and other image elements are actually shrunk linearly to fit the available space. It is thus important to conduct experiments on font legibility and aesthetic quality when such shrinkage takes place. The bottom half of Figure 29 shows the impact of font-size adjustment/reduction as well as image resizing/shrinkage on the resulting script.

I conclude this section with actual modern Persian fonts, as implemented by Microsoft. The font menu of Microsoft Word, from which the samples in Figure 30 have been drawn, offers multiple font options for the Persian script. The first four options (Arial, Cambria, Helvetica, Times New Roman) generate the same script on my version of Microsoft Word. Perhaps they will be replaced with distinct fonts in future. The other options (Calibri, Courier, Dubai, Dubai Light, Tahoma, Traditional Arabic) exhibit different levels of readability and aesthetic quality. The Calibri and Dubai fonts perhaps strike the best balance between readability and aesthetic quality, as evident from the enlarged samples at the left edge of Figure 30. The Courier font suffers from typical problems of fixed-width fonts for Persian, but if having a constant symbol width is a requirement, it is more than bearable.

Beyond run-of-the-mill printing and display, decorative and artistic writing, using a variety of handwriting styles can be and has been mechanized. Some examples are provided by the on-line site NastaliqOnline.ir, which features an input box for entering a text and a menu to choose the style of writing. Then, upon issuing the "write" command, the site's software algorithm creates the resulting script and presents it in jpg format for copying and pasting into other documents. There are still problems with the script quality, but the site's designers are making improvements on a regular basis. Examples of the resulting scripts are provided in Figures 31 and 32.

FIGURE 28

Some results of a study to determine the minimum dot-matrix size for Persian script..

FIGURE 29

Persian fonts of various point-sizes and the effect of scaling on script readability and quality.

FIGURE 30

Examples of Microsoft Word Persian fonts and their resulting scripts.

FIGURE 31

Nastaliq script generated by the on-line service at NastaliqOnline.ir.

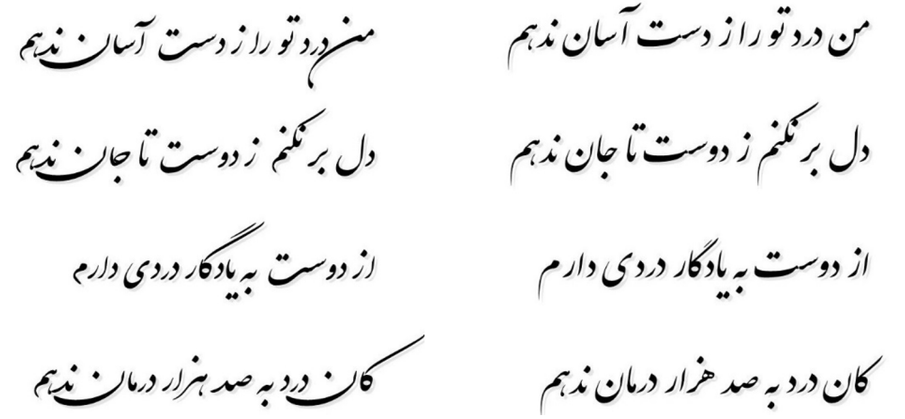
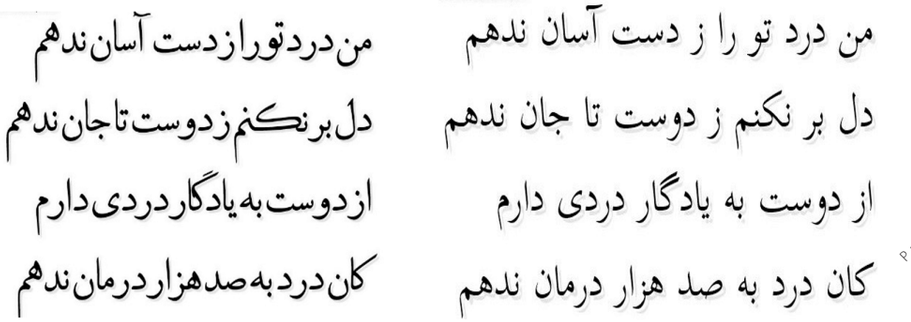


FIGURE 32

Two other machine-generated Persian scripts: Neirizi (left) and Osman-teh.



9. Some Remaining Challenges

After facing five decades of challenges and responding with solutions and improvements, we are still nowhere near the end of the line. Display quality and resolution is continually improving, which is a big help. High-definition (HD) and ultra-high-definition (UHD) displays offer 1920-by-1080 and 3840-by-2160 matrices, respectively, with the latter doubling the resolution in each dimension and, thus, quadrupling the areal resolution. In printers too, even the cheapest models now offer a resolution of 600 dots per inch (DPI) at a minimum. So, hardware capabilities are no longer our main problems. The problems lie in font design and text-processing algorithms for information display and printing.

Put another way, whereas until a couple of decades ago, computer typefaces had to be designed with an eye toward the capabilities and limitations of printing and display devices, we can now return to typeface design with only aesthetics and readability in mind. Thus, much work from the pre-computer era becomes relevant to our endeavor, including the rich legacy of typography over many centuries (Heller, 2001; Keyghobadi, 2012). Any typeface can now

be mapped to suitably large dot-matrices to produce high-quality and easily-readable Persian script. There is much accumulated know-how about fonts and algorithms, which must be put to use in a systematic way, with empirical evaluation and extensive feedback.

The problem of codes for internal representation and transmission of Persian characters, and the resulting incompatibilities between different hardware platforms and software applications, is now pretty much solved with the widespread adoption of the Unicode standard (Unicode.org, 2019). It is still the case, however, that the Persian letter "seen" (س"), which is now consistently represented using Unicode, will have many different language- and culture-dependent embodiments in print and display. So, a text carefully composed and formatted in Persian, may look much less pleasing when printed or displayed on an app or Web site that is optimized for Arabic or Urdu texts, say. In other words, the problem of compatibility and machine-independence has moved from the lower level of bits and bytes to the higher level of formatting and rendering of texts.

Computer word-processing and typesetting systems emerged on the heels of ubiquitous personal computing and, most recently, in connection with smartphones and tablet devices. By providing simpler and friendlier user interfaces, such systems facilitated the production and dissemination of documents containing Persian text. Such products relied on existing Persian fonts, so they did not directly improve script quality, except through comments and feedback resulting from the now much larger user community. An early word processor, Zarnegar (Sanati, 2018), was marketed in Iran and was soon followed by a host of other offerings. Modern word processors and text input/output systems are not tied to particular languages but can handle a variety of scripts by providing a generic bidirectional text-entry mechanism in connection with Unicode. The influential FarsiTeX (Esfahbod and Pournader, 2002) was designed at Sharif University of Technology, Iran, as the bilingual incarnation of Donald Knuth's well-known and widely-used mathematical text processor TeX (Knuth, 1984). Because FarsiTeX had to solve the problem of bidirectionality for dealing with mathematical formulas, it constituted a good foundation for bilingual text processing in general. ArabTeX (Lagally, 1992) is the Arabic counterpart of FarsiTeX.

One aspect of the ease of Persian-language processing is the support provided within popular programming languages and systems for dealing with Persian texts. The level and type of support varies from one language to another and must thus be mastered along with other language details on a case-by-case basis. A Google search with the language name and "Persian text" should reveal the pertinent information. For example, there is a resource page with "Links to Python information in Persian/Iranian/Farsi" (Python.org, 2019). Toolkits for converting Persian texts from incompatible formats into a standard format for ease of processing are also required (e.g., Mohtaj, Roshanfekr, Zafarian, and Asghari, 2018).

Another remaining problem is developing a better understanding of the trade-offs between legibility and aesthetic quality and in-depth studies of formatting issues, especially when bilingual and multilingual output is involved. The first problem can be investigated through the use of crowd-sourcing. The vast data sets of available Persian script (e.g., from Facebook and Twitter posts) can be used for occasional polls, putting two different scripts side-by-side and asking users to rate them. In the area of formatting, the opposite directions of writing in Persian (right to left) and Latin-based (left to right) scripts create problems. If I were to insert a Persian word here, say, فارسی,the formatting might be messed up if the Persian word appears close to a line-break so that it has to be moved to the next line. Punctuation marks, particularly parentheses and brackets, also create similar problems, sometimes moving from their proper places to different positions in the text.

The expanding user base demanding the capability to print and display the Persian script is helping with improvements. However, it is important that users stay active and engaged, providing feedback to the fullest possible extent in order to motivate investment in improving the hardware technologies and software algorithms. An integrated, easy-to-use bilingual keyboard and improved optical character recognition would be important first steps in solving the remaining input problems. Alongside improving the existing facilities for Persian input and output, we must keep an eye on what's coming in the near future: voice input and output. Much work has been done in English voice input, but we are at the beginning of the process of applying such techniques to Persian (Sameti, Veisi, Bahrani, BabaAli, and Hosseinzadeh, 2008). Ditto for Persian voice output, from a textual document (Bokaei et al., 2010).

Display and printing technologies aside, studies are needed to gauge the Persian script itself, in order to develop a better understanding of legibility and aesthetic quality considerations. I will cite just two examples of corresponding studies for English hand-printed and printed scripts, as examples of how the concepts and results can be extended to Persian.

Appendix A in a book (Bailey, 1983) that I have used in a graduate-level course on fault-tolerant computing, where the impact of errors of various kinds on the correct functioning of computer systems is studied, contains an interesting table which provides empirical error rates, when manually-written texts and filled-out forms were keyed-in by operators of early computer input devices. According to that table, the least error-prone hand-printed English alphanumeric symbols, with error rates of less than 1%, are these 8 symbols: W, M, 3, 7, A, 9, E, C. Near the high end of the error spectrum, with error rates of around 5%, we find these 6 symbols: N, 0, 5, J, V, G. Particularly error-prone are the letters Z (13%) and I (25%). Error-proneness is intimately related to legibility, but I am unaware of any study of error-proneness for handwritten or printed Persian scripts.

In studies of reader comprehension of texts, the notion of readability (Zamanian and Heydari, 2012) has been widely studied. Readability, which is different from legibility, is affected by word usage (short vs. long, common vs. rare, etc.), sentence structures (short/simple vs. long/complex), and the physical parameters of the text in terms of font, size, color, and the like. Legibility is a more focused notion that pertains only to the letter/symbol shapes, their spacing, and combined appearance in words, the so-called word "shapes." Again, I don't know of any studies in this area specific to Persian, but efforts to characterize the legibility of Arabic script have begun (Chahine, 2012).

10. Conclusion

I have provided a survey of challenges in producing legible and aesthetically pleasing Persian script by computer output devices and the progression of solution strategies, as printing and display technologies improved over time. Like all technical endeavors, the evolution of Persian script alongside changes in printing and display technologies involved some trial and error, a few outright successes, and instances of backtracking from dead ends. Making mistakes and learning from them are essential to successful engineering.

I am guided in my pursuit and evaluation of solution strategies by two books of Henry Petroski, who has written widely on the nature of engineering and the role of failures in successful design. In one book (Petroski, 1992), the engineer extraordinaire informs us: "No one wants to learn by mistakes, but we cannot learn enough from successes to go beyond the state of the art." In another, equally fascinating book (Petroski, 1994), he traces the evolution of simple gadgets and processes, as stand-ins for more complex systems such as computers that are much harder to understand or analyze, concluding that, designers/engineers are essentially spotters of weaknesses and failures in previous designs and are also equipped to do something about it. I will go one step further and advise the readers that we all should act as failure spotters in designs we encounter and use, whether or not we are equipped to correct the problems. A computer user need not be an expert designer to spot problems, any more than a film critic is required to be a capable director.

I am no longer directly involved in improving the quality of Persian output as a designer, given that my research interests have moved toward computer hardware architecture. But, as a user, I do encounter on occasion problems of poor design and bring those problems to the attention of those who are authorized or equipped to do something about them. In today's connected world, we can all use our blogs, Twitter feeds, and other communication channels to participate in further improvement of the quality of printed and displayed Persian script. I hope that this paper has provided the fundamental basis needed to understand the issues and to be informed users of typographic products and systems in connection with the Persian script.

Acknowledgements

My sincere thanks go to the coordinator of UCLA Bilingual Lecture Series on Iran, Dr. Nayereh Tohidi, and the program's staff and sponsors, for inviting me to deliver a pair of lectures (Parhami, 2017), which motivated me to write this paper. While I played a central role in formulating and categorizing language-related challenges and keeping track of advances and developments for reporting purposes, solving the technical problems discussed here was most certainly not a solo effort. Dr. Farhad Mavaddat and Eng. Armen Nahapetian were two of my colleagues who played key roles during the early years of studying and solving problems related to Persian computing. Our graduate students, as well as researchers/engineers at other universities, government organizations, and private companies also contributed to the solutions. Thoughtful reviewer comments helped make the paper clearer and more complete. As a final side note, my use of the word "Persian" covers a set of related languages that Iranians call "Farsi" and other speakers refer to as "Dari," "Pashto," or "Urdu," depending on their geographic regions and ethnic backgrounds.

References

Ashtarian, K. (2015). Iran. In *UNESCO science report: Towards 2030*. Chapter 15, pp. 388-407. https://unesdoc.unesco.org/ark:/48223/pf0000235406

Bailey, R. W. (1983). *Human Errors in Computer Systems*. Upper Saddle River, NJ: Prentice-Hall.

Bokaei, M. H., Sameti, H., Eghbal-zadeh, H., BabaAli, B., Hosseinzadeh, K. H., Bahrani, M., Veisi, H., and Sanian, A. (2010). Niusha, the first Persian speech-enabled IVR platform, *Proc. 5th IEEE Int'l Symp. Telecommunications*, pp. 591-595.

Borjian, M. and Borjian, H. (2011). Plights of Persian in the modernization era. *Handbook of language and ethnic identity: The success-failure continuum in language and ethnic identity efforts*, Vol. 2, pp. 254-267.

Chahine, N. (2012). Reading Arabic: Legibility studies for the Arabic script. Doctoral Dissertation, Faculty of the Humanities, Leiden University.

Devroye, L. (2019a). Font design in Iran. Online index of font designers, http://luc.devroye.org/iran-index.html

Devroye, L. (2019b). IBM Selectric and Selectric II type fonts. Online type design information page. http://luc.devroye.org/fonts-44934.html

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. International Journal of Communication, Vol. 10, pp. 947-969.

Dunn, P. D. (1979). *Appropriate Technology: Technology with a Human Face*, New York: NY: Palgrave Macmillan.

Encyclopedia Iranica (2011). Computers in Persia. Vol. VI, Fasc. 2, pp. 118-119. http://www.iranicaonline.org/articles/computers-in-persia

Esfahbod, B. and Pournader, R. (2002). FarsiTeX and the Iranian TeX community. *TUGboat*, Vol. 23, No. 1, pp. 41-45.

Floor, W. M. (1990). ČĀP. In *Encyclopedia Iranica*, Vol. I/7, pp. 760-764. Available online at http://www.iranicaonline.org/articles/cap-print-printing-a-persian-word-probably-derived-from-hindichapna-to-print-see-turner-no

Green, N. (2010). Persian print and the Stanhope: Industrialization, evangelicalism, and the birth of printing in early Qajar Iran. *Comparative Studies of South Asia, Africa, and the Middle East*, Vol. 30, No. 3, pp. 473-490.

Heller, S. and Meggs, P. B. (2001). *Texts on type: Critical writings on typography*, New York, NY: Allworth Press.

Institute of Standards and Industrial Research of Iran (1976). *Character arrangement on keyboards of Persian typewriters* (in Persian), Report ISIRI 820.

International Business Machines Corp. (1959). 1620 data processing system. Online reference information. https://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP1620.html

Iran Plan and Budget Organization (1980). *Final proposal for the Iranian National Standard Information Code* (INSIC). Report, with Persian and English versions.

Kahan, B. (2000). *Ottmar Mergenthaler—The man and his machine*, New Castle, UK: Oak Knoll Press.

Keyghobadi, R. (2012). *Written spaces: Interpreting contemporary Iranian typography through the works of three Iranian artists*. Doctor of Education Dissertation, Columbia University.

Kia, M. (1998). Persian Nationalism and the campaign for language purification. *Middle Eastern Studies*, Vol. 34, No. 2, pp. 9-36.

Knuth, D. E. (1984). *The TEXbook*, Volume A in the 5-volume "Computers and Typesetting" series, Reading, MA: Addison-Wesley.

Lagally, K. (1992). ArabTEX, a system for typesetting Arabic. *Proc. 3rd Int'l Conf. Multi-lingual Computing: Arabic and Roman Script*, Vol. 9, No. 1.

Lazard, G. (2008). The rise of the new Persian language. *The Cambridge History of Iran,* Vol. 4 (Period from the Arab Invasion to the Saljuqs), pp. 566-594.

Lemire, D. (2019). Sorting strings properly is stupidly hard. Blog of a professor at University of Quebec. https://lemire.me/blog/2018/12/17/sorting-strings-properly-is-stupidly-hard/

Library of Congress (2014). A thousand years of the Persian book. Exhibition and Lecture Series.

Mavaddat, F. and Parhami, B. (1977). Informatics in Iran: Problems and prospects. *Proc. Int'l Conf. Computer Applications in Developing Countries,* Bangkok, Thailand, pp. 121-133.

Mohtaj, S., Roshanfekr, B., Zafarian, A., and Asghari, H. (2018). Parsivar: A language processing toolkit for Persian. *Proc. 11th Language Resources and Evaluation Conf.*, pp. 1112-1118.

Monk, T. (2013). GP IBM Selectric typeball catalog—Font styles. The Typosphere blog. https://munk.org/typecast/2013/11/03/gp-ibm-selectric-typeball-catalog-font-styles/

Moradi, F., Nanquette, L., Hosseinipour, M., Amiri, A., Rakhimov, D., and Biniaz, B. (2013). *Publishing in Persian language: In Iran, Afghanistan, Tajikistan, Uzbekistan, Europe and United States*. Paris: Int'l Alliance of Independent Publishers.

Nemeth, T. (2017). *Arabic type-making in the machine age: The influence of technology on the form of Arabic type, 1908-1993*. Leiden, The Netherlands: Brill, p. 288.

Parhami, B. and Mavaddat, F. (1977). Computers and the Farsi language: A survey of problem areas. *Information Processing 77* (Proc. IFIP World Congress), Amsterdam: North Holland, pp. 673-676.

Parhami, B. (1978a). Optically weighted dot-matrix Farsi and Arabic numerals. *Information Technology 78* (Proc. 3rd Jerusalem Conf.), Amsterdam: North-Holland, pp. 207-210.

Parhami, B. (1978b). On the use of Farsi and Arabic languages in computer -based information systems. *Proc. Symp. Linguistic Implications of Computer-Based Information Systems*, New Delhi, India.

Parhami, B. (1978c). Impact of Farsi language on computing in Iran. *Mideast Computer*, Vol. 1, pp. 6-7, September.

Parhami, B. (1981). "Language-Dependent Considerations for Computer Applications in Farsi and Arabic Speaking Countries," *System Approach for Development* (Proc. IFAC Conf.), North-Holland, 1981, pp. 507-513.

Parhami, B. and Daie, V. (1981). *Glossary of computers and informatics: English/Persian*, Tehran, Iran: Informatics Society of Iran. Reprinted many times up to 1991; revised and reissued by an ISI committee in 1994.

Parhami, B. and Taraghi, M. (1981). "Automatic Recognition of Printed Farsi Texts," *Pattern Recognition*, Vol. 14, Nos. 1-6, pp. 395-403.

Parhami, B. (1984a). *Computer Appreciation* (*Aashnaaee baa Computer*, in Persian), Tehran, 1984. Reprinted in 1985, 1987, and yearly thereafter. As of 2001, was still being used as a textbook.

Parhami, B. (1984b). Standard Farsi information interchange code and keyboard layout: A unified proposal. *J. Institution of Electrical and Telecommunications Engineers*, Vol. 30, No. 6, pp. 179-183.

Parhami, B. (1985). University education in computer science and technology: The new Iranian plan. *Proc. IFIP 4th World Conf. on Computers in Education*, Norfolk, VA, pp. 923-930.

Parhami, B. (1986). Computer science and engineering education in a developing country: The case of Iran. *Education and Computing*, Vol. 2, No. 4, pp. 231-242.

Parhami, B. (1995). On lower bounds for the dimensions of dot-matrix characters to represent Farsi and Arabic scripts. *Proc. 1st Annual CSI Computer Conf.*, Tehran, Iran, pp. 125-130.

Parhami, B. (2017). Computers and challenges of writing in Persian. Two lectures in the UCLA Bilingual Lecture Series on Iran. In Persian (November 19) and English (November 20).

Parhami, B. (2019a). Computers and challenges of writing in Persian (in Persian). *Iran Namag*, Vol. 4, No. 2 (Summer 1398, Persian calendar). pp. 3-32

Parhami, B. (2019b). Evolutionary changes in Persian and Arabic scripts to accommodate the printing press, typewriting, and computerized word processing. *Tugboat* (Proc. 40th Annual Meeting of the TeX Users Group, Palo Alto, CA, August), Vol. 40, No. 2, pp. 179-186.

Petroski, H. (1992). *To engineer is human: The role of failure in successful design*. New York, NY: Vintage.

Petroski, H. (1994). *The evolution of useful things: How everyday artifacts—from forks and pins to paper clips and zippers—came to be as they are*. New York, NY: Vintage.

Python.org (2019). Links to Python information in Persian/Iranian/Farsi. Online resource page: https://wiki.python.org/moin/PersianLanguage

Reisinger, D. (2011). IBM Celebrates 50 years with the Selectric typewriter. c|net.com article. https://www.cnet.com/news/ibm-celebrates-50-years-with-the-selectric-typewriter/

Renima, A., Tiliouine, H., and Estes, R. J. (2016). The Islamic Golden Age: A story of the triumph of the Islamic civilization. In: *The State of Social Progress of Islamic Societies*, ed. by Tiliouine, H. and Estes, R., pp. 25-52. Cham, Switzerland: Springer.

Ross, F. (2017). Non-Latin font design at linotype. Online article. https://tntypography.eu/resources-list/non-latin-type-design-linotype-fiona-ross/

Sameti, H., Veisi, H., Bahrani, M., Babaali, B., and Hosseinzadeh, K. (2008). Nevisa, a Persian continuous speech recognition system. *Proc. 13th Int'l CSI Computer Conf.*, Berlin: Springer, pp. 485-492.

Sanati, M. (2018). My recollections of desktop publishing (in Persian). *Computer Report*, Vol. 40, No. 239, pp. 53-60.

Twyman, M. (1999). *The British Library guide to printing: History and techniques*, Toronto: University of Toronto Press.

Unicode.org (2019). About the Unicode standard. Online information page with pertinent links. https://unicode.org/standard/standard.html

Wikipedia (2019). Non-English-based programming languages. https://en.wikipedia.org/wiki/Non-English-based_programming_languages

Williams, M. R. (1997). *A history of computer technology*, 2nd ed, Los Alamitos, CA: IEEE Computer Society Press.

Zamanian, M. and Heydari, P. (2012). Readability of texts: State of the art. *Theory & Practice in Language Studies*, Vol. 2, No. 1.

Image credits

Credit listed only for images that are not the author's own design and creation.

2 (left): Art by Farrokh Mahjoubi. https://festiveart.com/persian-calligraphy-farrokh-mahjoubi.html#.XS1ut-hKiUk

2 (right): Pinterest https://i.pinimg.com/originals/cc/0c/f6/cc0cf6edb8bb6851d91c8ce2edbc2c76.jpg

3 (left): Art Point, School of Calligraphy and Fine Arts. http://www.handwriting.pk/calligraphy.html

3 (right, top): Arabic Calligraphy Fonts Naskh. https://www.rock-cafe.info/posts/arabic-calligraphy-fonts-naskh-617261626963.html

4 (left, top/bottom): Old Internet stock images whose sources could not be identified via a recent Google image search

4 (right, top): El Nuevo nacimiento de la impression. https://franciscojaviertostado.files.wordpress.com/2013/12/metal_movable_type.jpg

4 (right, bottom): ABC South East NSW. https://www.abc.net.au/news/image/7189966-3x2-460x307.jpg

5 (left, top/bottom; right, top): Old Internet stock images whose sources could not be identified via a recent Google image search

5 (right, bottom): Coloring Book Printing and Binding

https://neilpaints.com/wp-content/uploads/2018/09/coloring-book-printing-and-binding-marvelous-manhattan-perfect-bound-book-of-coloring-book-printing-and-binding.jpg

6: 29LT Blog. https://blog.29lt.com/

7 (left): Old Internet stock image whose source could not be identified via a recent Google image search

7 (center): Arabic Tattoo Designs. https://arabictattoo.wordpress.com/2009/02/03/spoon-and-fork-in-naskh/

7 (right): Susan Leat. https://www.researchgate.net/figure/Demonstration-of-Arabic-typeface-characteristics-A-A-sample-of-a-five-letter-word_fig1_302059929

8: Own design, except for the single Arabic word, which is from: https://ya-webdesign.com/transparent250_/arabic-alphabet-png-4.png

9: ***The Atlantic***, July 18, 2017 (Atta Kenare / Getty Images)

14 (left): Modified from ***PC Magazine*** Encyclopedia image

14 (right): IBM

15 (left): Computer Sc IT & Management Web site

15 (right): ***PC Magazine*** Encyclopedia

16 (right): 14-Segment Display. https://upload.wikimedia.org/wikipedia/en/thumb/0/0a/14_Segment_LCD_characters.jpg/220px-14_Segment_LCD_characters.jpg

18 (left): Dot-matrix Wimbledon scoreboard in use from 1982 to 2008 (Wikipedia)

18 (right): Computronics Products. http://computronics.biz/productimages/prodairport3.jpg

19 (left): ***PC Magazine*** Encyclopedia

19 (right): Glazed-tile-based writings at Isfahan's Jam'eh Mosque; uncredited photo

20 (left): Shutterstock.com, stock world map with round pixels

20 (center): Japanese Computers—Still Living It 8-Bit. https://blog.gatunka.com/2009/11/03/japanese-computers-still-living-it-8-bit/

25 (right): Screenshot of the science and technology section from an unidentified news site

31: Texts generated using the tools of the Web site NastaliqOnline.ir

32: Texts generated using the tools of the Web site NastaliqOnline.ir

Author

Behrooz Parhami (PhD, UCLA 1973) is Professor of Electrical and Computer Engineering, and former Associate Dean for Academic Personnel, College of Engineering, at University of California, Santa Barbara, where he teaches and does research in computer hardware architecture. A Life Fellow of IEEE, a Fellow of IET, and British Computer Society, and recipient of several other awards (including a most-cited paper award from J. Parallel & Distributed Computing), he has written six textbooks and more than 300 peer-reviewed technical papers. Most relevant to this paper is his 12 years of experience at Iran's Sharif (formerly Arya-Mehr) University of Technology, spanning 1974 to 1988, the final two years on sabbatical leave in Canada, where he studied the Persian language in connection with computers and their applications, participated in technology transfer and standardization projects, helped establish the Informatics Society of Iran (and its technical journal, Computer Report), and was a leader in developing computer engineering educational programs and associated curricula. Professionally, he serves on journal editorial boards and conference program committees and is also active in technical consulting.

Subscription Rates

United States	Individual	Institutional
1 year	$35.00	$65.00
2 year	$65.00	$124.00

Canadian*	Individual	Institutional
1 year	$44.00	$ 74.00
2 year	$83.00	$142.00

Foreign**	Individual	Institutional
1 year	$ 56.00	$ 86.00
2 year	$107.00	$166.00

Prepayment is required. Make checks payable to University of Cincinnati *Visible Language* in U.S. currency only, foreign banks need a U.S. correspondent bank.

* Canadian subscriptions include additional postage ($9.00 per year).
**Foreign subscriptions include additional postage ($21.00 per year).

ISSN 0022-2224
Published continuously since 1967.

Back Copies

A limited number of nearly all back numbers is available. The journal website at http://visiblelanguagejournal.com is searchable and lists all issues, past article PDFs, contents, and abstracts.

Copyright Information

Authorization to photocopy items for internal or personal use, or for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of $1.00 per article, plus .10 per page is paid directly to:

CCC
21 Congress Street
Salem, Massachusetts 01970
Telephone 508.744.3350
0022-22244/86 $1.00 plus .10

UNIVERSITY OF Cincinnati