An Overview of the New 8-Dots Arabic Braille Coding System

Oussama El Ghoul, Ikrami Ahmed, Achraf Othman, Dena A. Al-Thani, and Amani Al-Tamimi

1 Mada Center, Al Nasr Tower B, 8th Floor, Doha, Qatar
{oeelghoul,aothman,aaltamimi}@mada.org.qa, dalthani@hbku.edu.qa
2 Qatar Social and Cultural Center for Blind, Doha, Qatar
ikrami.ahmad@gmail.com

Abstract. Considering the rapid technological development and especially for assistive technology, the six-point Braille system has become insufficient to meet the needs of the blind and enable them to read, write content, and, to publish accessible documents. This system is not sufficient to write and produce scientific contents that contain several symbols. Despite this need, the Arabic language still lacks an eight-point coding system. In this context, this paper aims to present a unified eight-point Braille system and present it to Arab communities to get benefit from it in developing digital content for blind people. The Arabic language differs from the Latin and other languages in the number of letters and diacritics, which makes the coding system different from the one used in these languages. In this work, we studied the symbols used in the Arabic language and the current Braille system and looked for methods and recommendations regarding the design of the eight-point Braille system. A methodology and a set of principles have been identified that have been adopted in preparing the system, and rules for coding have been established.

Keywords: Braille · 8 dots Braille · Assistive technology · Accessibility

1 Introduction

Braille is the only and unique method that enables blind or deaf-blind people who have difficulties to access to printed materials to read and write. In fact, the ability to write and read in braille opens the door to knowledge, intellectual freedom, equal opportunity and personal security. Nowadays The cost of Braille electronic devices has begun to decrease significantly, which means that a larger segment of the blind - especially in developing countries - can access Braille electronic devices, but there is not enough content to support the Arab Braille.

Although Braille has become important for teaching literature and science for blind people, Braille’s coding of the Arabic language is still not developed to accommodate scientific symbols such as those used in mathematics, physics, or even chemistry. This is because the eight-point Braille coding system is not yet adopted.
In this context, this paper aims to present an eight-point braille coding system for the Arabic language. A scientific methodology for identifying symbols has been designed upon by relying on five rules that are subject to six principles that have been adjusted by following previous studies.

2 Related Work

The six-dots Braille systems allow encoding of a maximum of 63 characters, which is sufficient to encode letters, numbers, punctuation marks, and some signs of the approved Braille system. However, these systems do not provide enough coding capacity for coding symbols used in science such as mathematics, physics, chemistry or even music. Hence the importance of the eight-dots coding system, which enables coding up to 255 symbols. This gives sufficient scope for the inclusion of all the important symbols. Using this system allows people with visual disabilities to read and write scientific, literary and artistic contents. Adopting a Unified Arabic Braille coding system is become important to allow Arabic blinds to create and read Arabic content.

Worldwide there are two primarily unified braille systems that are mostly used: The Unified English Braille UEB and The Nemeth Uniform Braille System NUBS [2, 6]. Both are used to transcribe and represent all letters and formats used on English documents into braille. The UEB is more suitable for text representation, in fact, it’s derived from the classic six-dot Braille that encodes 26 small letters and 14 punctuation symbols with one cell. However, NUBS is mainly used for representing mathematical and scientific contents. Unfortunately, there are many mismatches between the two systems.

3 Motivation

According to many blind people, the current Braille coding system suffers from several limitations. In this context, the Mada Center for Assistive Technology conducted a questionnaire on the problems and limitations that users of the current Arab Braille see and want to fix. The questionnaire covered all Arab countries, and we got 80 answers from 17 countries. The questionnaire was directed to all persons who use this system from teachers, students and other users. We received 80 answers, as 67% of people who completed the questionnaire are using Braille to learn and/or read, which indicates the importance of this system in their daily lives (Fig. 1). The answers also confirmed that 96% participant are fluent in Arabic braille and 77% of them are fluent in a coding system other than Arabic, which means that their answers will be based on a comparison between the Arab Braille system and other systems. This makes their answers more realistic and accurate. In fact, 72% of the participants confirmed that they are daily users of Braille.

After analysing the results we concluded that 70% of the participants think that the current Arabic Braille contains many limitations. Furthermore, 83% of the participants believe that work should be done to develop the current Arab Braille system.
Concerning the limitations of the Arabic Braille we prepared a set of 8 options to be chosen by the participants. Each participant can choose one or many options and he can propose additional limitation.

List of limitations:

- **L1**: Some symbols and signs are not defined in Arabic Braille, such as new symbols that appeared in modern software and applications, non-Arabic characters that entered the Arabic language to encode local dialects, and so on.
- **L2**: There are ambiguities and inconsistencies between many letters and symbols, with more than one symbol sharing the same dot representation in Braille.
- **L3**: Relying on the reader’s ability to distinguish the symbol through the context, which makes the Arab braille inappropriate for all literary and artistic contexts and makes it difficult to learn.
- **L4**: Problems in writing Arabic texts via devices.
- **L5**: There are low-usage letters or symbols that can be replaced by used ones that are not present in the current system.
- **L6**: Inconsistent Braille usage rules for typing numbers and spaces before and after symbols.
- **L7**: Others.
- **L8**: No-Limitations.

As shown on the following Chart, 50% of participants confirms that the current Arabic Braille does not encode some important signs and symbols that appeared in modern software and applications or the new non-Arabic characters that entered the Arabic language to encode local dialects. Furthermore, many participants complain of confusion and inconsistencies between many letters and symbols, with more than one symbol sharing the same dot representation in Braille.

We asked participants about the use of 8-dot Braille to overcome the limitations of the current 6-dot Arabic Braille. 67% of them think that by adding 2 additional dots we...
can encode more symbols and decrease the confusion and the inconsistencies rate between many letters and symbols.

The results of the survey confirms that the development of an 8-dot for Arabic language is important and become a request from the Braille users communities (Fig. 2).

4 Specificities of the Arabic Language

The Arabic alphabet is the alphabet that uses the Arabic letters for writing, and it is described as the most complete writing system, it included most of the sounds that a person can speak; derived from it many of the alphabets and the alphabet remained the most used for many centuries and is currently the most used writing system after the alphabet Latin. Many languages rely on the Arabic alphabet for writing, such as Urdu, Ottoman, Kurdish, and Malay [3, 4].

Arabic letters are written from right to left, in a style that depends on joining the letters of one word to each other, and these alphabets include 28 basic letters, and some consider them 29 letters as Hamza is a separate letter.

Contextual adjustments can be made to some letters [6]. The following characters are not individual letters, but are contextual adjustments of some Arabic characters ٕ ٩ :

The Arabic language is distinguished from the rest of the other languages by the 8 diacritics that are drawn on its letters, and diacritics are signs and movements to set the letters to be pronounced correctly [5].

Some languages that use the Arabic alphabet to write in their language add some additional letters, in order to translate some foreign letters phonetically, here are some of those letters:
• ﻓ - which is used to denote the letter “V” in Latin characters when phonetic translation. This letter is also used in writing dialects that pronounce this sound. The letter “ﻓ” is usually used in the phonetic translation of the Latin “V”. This letter is used to express the sound “Pa”.

• ﺏ - used to indicate the letter “P” when transliterating. Therefore, the word “7up” can be copied in this way: ﺏﺃﻦﻔﺳ, or in this way: ﺏﺳ ﻓﺃﻦ ﻗ. The letter is also used in Persian and Urdu.

• ﺝ - Cha: The letter is used to clarify the sound [tʃ], in Persian, Urdu, and Kurdish. Sometimes the letter is used in phonetic translation, although the Arabic word “Cha” is used using the compound “تشا”.

• ﮟ - it refers to the non-thirsty “gym” in some languages.

• ﺝ - Zhi, (by breaking Zain). This letter is used to express a spoken frictional voice that comes out from behind the “palate”, in the Persian, Kurdish, Urdu and Uyghur languages. This letter is used to translate sounds that are of foreign origin. It is rare for the sound “[ʒ]” to appear in the Arabic language, and the shin “ش” is usually used in the phonetic translation of this sound “[ʒ]”. This letter is used in Persian and Urdu.

According to our study on the Arabic language, at least 45 symbols should be reserved for encoding the Arabic letters.

5 Methodology

An eight-dot Braille Unicode is defined by the Unicode Consortium. It offers an encoding capacity of 256 symbols using a single cell [1]. Based on the Unicode Braille we proposed a new mapping for the Arabic 8-dots Braille.

The proposed Arabic eight-dot Braille maps most-used one-cell symbols – it contains:

- 29 small letters
- 3 contextual adjustment letters
- 8 diacritics \{ā ā ē ē ʻ ə ʃ ʒ \}
- 10 digits
- 35 punctuations and computer symbols
- 26 Greek symbols \{α β θ δ ε ϕ γ η ι χ κ λ μ ν ο π ρ σ τ υ Ω ξ ψ ζ\}
- 14 basic math symbols
- 10 trigonometrical, logarithmic, and calculus symbols
- 9 set theory & logic symbols
- 5 geometry symbols
- 8 formatting symbols
- 21 patterns towards new indictors/prefixes \{superscript, subscript, start-radical, endradical, start-radical-index, overlay-above, overlay-below, overlay-across, cancel-previous, endsubscript/superscript/radical-index/overlay, tally-markindicator, bullet-prefix, set-tactile-graphics-mode, set-braillegraphics-mode, set-math-mode, end-math/graphics-mode, setlanguage, set-currency, set-braille-code-system\}

Thus Braille-8 becomes an extension of classic six-dots braille.
In order to control the coding process through scientific rules, we have defined a set of principles that we will rely on when deciding which symbols to adopt. We adopted a set of principles proposed by [2] and we added one principle that we considered important (User friendly).

The adopted process was carried out in the following stages: 1- Defining the list of symbols used in the six-point system, including indicators and structuring symbols. 2- Defining the list of new symbols and indicators. 3- Classifying symbols into groups. 4- Applying the following rules [2] to encode the set of symbols of each group.

**Rule 1:** Retain the 6-dot representation. According to the principle of Intra-Similarity, the priority was given to the symbols mapped to one 6-dot cell.

**Rule 2:** Maintain as much as possible the system adopted in UEB and NUBS. It aims to maximize Inter-Similarity.

**Rule 3:** Use the dots 7 and 8 to encode the group’s identifier. In order to have a structured representation. It aims to have a user-friendly representation.

**Rule 4:** Use the dots 3 and 6 to encode the sub-groups identifier if needed.

After finishing the development of the new 8-Dots Braille system we implemented it on the opensource Liblouis library. The proposed tables are published on the last version of Liblouis (3.14). The tables can be now downloaded and used for free. We planned to conduct a new survey in order to evaluate the new library after weeks of uses.

| Principle          | Description                                                                 |
|--------------------|-----------------------------------------------------------------------------|
| Compression        | Map as many symbols as possible to a single braille cell. This principle aims to reduce the number of cells used to encode texts and facilitate reading. |
| Intra-Similarity   | Considering the existence of a 6-dot code. It is important to maintain a link with the 6-dot encoding system. |
| Inter-Similarity   | Maintain a minimum of similarity with transition rules adopted by other languages. |
| Unambiguity        | Ensure the coherence of the code when a combination of more that one cell is used to encode a character. |
| Consistency        | Use the same transition rules to encode characters of the same category. |
| User friendly      | Use a tactile based assignment/mapping for characters encoding to suit visually challenged users. |
| Coverage           | Unified Arabic Braille should cater to the needs of math, science, and computer science, as well as the transcribing codes. |
| Foresight          | Consider possible expansions in other areas (e.g. scientific braille, computer braille) by providing unbounded cells or sharing characters. |
6 Conclusion

In this paper, we presented a new Unified Arabic eight-dot braille system. The system is created according to a set of rules that we defined in order to follow the international requirements. The methodology of implementation and identification of rules was done by researchers and persons with visual impairments and blinds in all milestones through focus groups and one-to-one discussion. This approach helps us to get into consideration all challenges that face Braille-users and to understand the need to have a new 8-dots braille system.

References

1. ISO/TR 11548–2:2001, 1st ed., Part 2: Communication aids for blind persons identifiers, names and assignation to coded character sets for 8-dot Braille characters Latin alphabet based character sets. International Standards Organization. Zurich, Switzerland (2001)
2. Kacorri, H., Kouroupetroglou, G.: Design and Developing Methodology for 8-dot Braille Code Systems. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013. LNCS, vol. 8011, pp. 331–340. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39194-1_39
3. Martos, A., Kouroupetroglou, G., Argyropoulos, V., Deligiorgi, D.: Towards the 8-dot Nemeth braille code. Lect. Notes Comput. Sci. 8547, 533–536 (2014)
4. Abdallah, M.A., Khairuddin, O.: “Quranic Baille System”, World Academy of Science, Engineering and Technology,4–6,2008
5. Abualkishik, A., Omar, K.: Framework for translating the Holy Quran and its reciting rules to Braille code. In: International Conference on Research and Innovation in Information Systems (ICRIIS), 2013. IEEE (2013)
6. Vistrit, G.: Braille-8 – the Unified Braille Unicode System. In: IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), ISBN: 978–1-5090-2193-2, (2016)