Arabic Quranic characters used to represent text in the minority orthographies appeared in proposals and documents from King Fahd Glorious Quran Printing Complex (L2/10-381), Mussa Abudena (L2/15-329, L2/16-044, and L2/17-004), Azzeddine Lazrek (L2/16-268, L2/17-252), Lateef Sagar Shaikh (L2/16-056, based on Abudena’s L2/16-044), as well as comments from Kamal Mansour, Lorna Evans, and Mussa Abudena (L2/16-102).

Some of the characters were considered eligible for encoding based on discussion in the Script Ad Hoc group (L2/17-037 and L2/17-384). This document includes those, as well as quite a few other characters deemed useful in further discussions between the authors and Marijn van Putten of Leiden University, an expert in Quranic Arabic, or found later by the authors. The orthographies we include are mostly used in Northwest Africa, and they have some shared symbols. It should be noted that this proposal may not be covering all needed characters in those orthographies. Instead, it should be considered a critical step to establish the most important characters and help future proposals advance.

Based on existing Quranic characters already encoded in Unicode, we have assumed a graphemic model for new characters. We consider modern Quranic orthographies comparable to technical writing: If a symbol is graphically quite different from an existing character used in another orthography for the same purpose or concept, we believe it should be encoded.

If the characters are encoded in the manner proposed below, they will be consistent with already encoded characters, and will be easier to add to existing Unicode Arabic fonts.

**New characters to be encoded**

Note that in the figures below, needed characters are typically circled in red, while existing characters are typically circled in blue.

1. **ARABIC SMALL HIGH WORD SAH**

   This is a pause/\textit{waqf} sign, and a combining character, like the other \textit{waqf} signs encoded for other Quranic orthographies (such as U+0615, U+0617, U+06D6..U+06DC, U+08D5..U+08D7, and U+08DD..U+08DF). It is the most common \textit{waqf} sign in modern Northwest African orthographies of the Quran.

   Graphically, this character is typically larger and wider than the mark above U+0671 ARABIC LETTER ALEF WASLA (as well as the ARABIC SYMBOL WASLA ABOVE proposed here (see #2 below) which is a spacing character for pedagogical purposes).
The proposed name is consistent with the name of existing characters U+08D4 and U+08DA..U+08DF. The name SAH comes from Abudena in L2/17-004, although he prefers other names for the character itself. The annotation “sign of waqf” should be added, with notes and references to make sure this character is not confused with a combining wasla.

Figure 2 from L2/16-044 (Abudena):

Figure 1 from L2/16-056 (Shaikh, based on Abudena):

From King Fahd Warsh, showing a narrower version, graphically more similar to a wasla:

2. ARABIC SYMBOL WASLA ABOVE

Most examples only show this character over alef. Azzeddine Lazrek in L2/17-252 recommended a combining version of this character (first proposed by Miikka-Markus Alhonen in L2/03-166), but that would cause two different ways to represent the existing Unicode character U+0671 ARABIC LETTER ALEF WASLA with no equivalence relationship between them.

There clearly is a need for a non-combining spacing version for pedagogical use. We think it should be encoded at U+FBC2 to complete the existing set of Arabic pedagogical symbols at U+FBB2..U+FBC1.

Evidence of stand-alone pedagogical usage (from the end notes of a King Fahd Hafs Quran):

Figure 4 from L2/03-166 (Alhonen):

3. ARABIC SMALL HIGH Farsi Yeh

This character is the al-Dani/Qaloon version of U+06E7 ARABIC SMALL HIGH YEH for perhaps one specific case in those editions and should have the same properties. According to Abudena, it always takes a FATHA above it (L2/17-004, p.3). These orthographies use it to distinguish a missing consonant /j/ from a missing vowel /i/. 
Contrastive usage with U+06E7 ARABIC SMALL HIGH YEH from Tripoli Qaloon:

From Tripoli Qaloon end notes:

4. ARABIC SMALL Farsi YEH

This is the spacing baseline version of the ARABIC SMALL HIGH Farsi YEH (#3 above) and is used in the Warsh orthography. Note that the Warsh orthography uses both this and U+06E6 ARABIC SMALL YEH contrastively. (Note that we already have a similar contrast in U+06E6/U+06E7 and U+06E5/U+08F3).

From King Fahd Warsh:

5. ARABIC SUPERSCRIPT ALEF MOKHASSAS

This symbol is used in Qaloon/al-Dani Qurans to specify cases where there are differences or doubts in the reading of a word.

From Tripoli Qaloon:
ومن ذلك الحذف المميز لهذه العلامة "\( \text{،} \) وهما تتماثلان به الخمسة، ويوجد في مئتان وستة وثلاثين كلمة (137)، ووجه تسميتها بهذا الاسم مأخوذة من قول أهل الرسم: هذه الكلمة أو الكلمات خمسة أو تسع وتسعون بالحذف.
وبد력 الإسناد كتب هذا النوع من الحذف في كلمات القرآن الكريم، وله أنه أحد أنواع الحذف الثلاثة المعروفة بالحذف "الإشارة" وهما يشيران إلى قراءة أخرى متناورة في الغالب والأكثرهما في: يبكون - أسى - تقعد - فمهم.

وقد يوجد هذا الحذف في كلمات خاصة مثل: "ليقيق"، "كذبت"، "بالافاثة". 

وقد يأتي هذا الحذف في بعض الكلمات مشيا إلى التفرد أو عوض تكرارها في القرآن مثل: "تفرقكم".

وأما بالنسبة لكلمة "سيدة"، فهي تأتي على الحروف "التي أسلفناها ذكر كتب النحاة ولم يذكر، مما يدل على أن عدد هؤلاء البصرا سابق الإجاز.

في حسب القرآن الكريم، حيث أن القرآن الكريم كله مجزري وفظيه مجبر في سمعه.

ولعلي شبَّت هذه الكلمات الثلاثة بهذا النوع من الحذف سراح طيفا، فقد استمر القرآن الكريم اهتم به شيئا، وبعدها إلى نظر على ذلك الذي.

عُلموا علامة للحذف المعروفة، وكل أنواع الحذف الثلاثة دون تبيين

بقي أن نعلم أن أصل هذه الشمعة موجودة في كتاب كما كتاب، لكن هذه العلامة "\( \text{،} \) التي نستخدمها لم يبكون بتصور أو تكرر التف. ولكنهما يوجد في بعض المصاحف الليبية المخطوطة الموقعة، وكم كل الأوقاف طابع جزء.

"علامة الحذف "، تستخدم فيما حذف النحو مثل "الطلالين الرقيقين"، وتُصد بأحرف، "كل الف مطلق، أي يوجد مثل ذلك "، "أنازيم"، "ليقيق"، "وكلة المخصص."


6. ARABIC BASELINE ROUND DOT
First proposed by King Fahd Glorious Quran Printing Complex (L2/10-381) with different properties, this is a spacing character sitting on the baseline and can take harakat. It’s not the same dot as #7, which appears higher, at about the vertical center of an alef, and is sometimes smaller in size.

This is called a musahhala hamza. It’s a certain way of representing hamza, when it’s not formally written as a letter, while also hinting at a specific way of recitation. It may appear next to another hamza, but it’s not a modifier: traditionally, it’s considered a second “softened” hamza. (According to Marijn van Putten, it’s no longer a phonetical glottal stop, it becomes a syllabic break without an intervening consonant.)

Note that the size of the dot could vary a little in different editions, but there doesn’t seem to be contrastive usage in any edition. The identity of the character is defined by it sitting on the baseline and being round.

Another potential name is ARABIC INLINE ROUND DOT.

Figures 12b and 14b from L2/17-252 (Lazrek)

![Image](image1)

From King Fahd Warsh (note that it’s taken a Fatha):

![Image](image2)

7. ARABIC RAISED ROUND DOT
First proposed by King Fahd Glorious Quran Printing Complex (L2/10-381) with different properties, this is to be used as a choice in a group of three, with U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE (shown below with blue ovals) and U+065C ARABIC VOWEL SIGN DOT BELOW. While those two are combing and appear above and below an alef form, this follows the alef form and is vertically aligned to the center of its height. Compare with #6, which sits lower and may be larger in size.

Similar to #6, the size of the dot could vary a little in different editions, but there doesn’t seem to be contrastive usage in any edition. The identity of the character is defined by it sitting higher than the baseline, near the vertical center of an alef it always follows, as well as being round. There are editions of the Quran that use both of these contrastively.

Figure 6.a. L2/17-252 (Lazrek)

![Image](image3)
8. ARABIC LETTER ALEF WITH ATTACHED FATHA
9. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE
10. ARABIC LETTER ALEF WITH ATTACHED KASRA
11. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA
12. ARABIC LETTER ALEF WITH LEFT MIDDLE STROKE
13. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA

These are generally used to help recitation when there’s a choice for connection (wasl) or disconnection (ghat’).

Note that while one may not expect it, #13 is indeed right-joining, like a normal alef, as can be seen in this example from King Fahd al-Duri:

Typically, each orthography may use three of the above set of six. But both #9 and #12 exist in King Fahd Warsh in contrastive use, which uses four of them:

Discussion from King Fahd Warsh end notes discussing #8, #10, and #12:

Figure 14.b in L2/17-252 (Lazrek) showing #8 (followed by #6):
King Fahd Warsh showing #8, #9, and #10:

From Tripoli Qaloon showing #11 (with a *sukun* below):

From Tripoli Qaloon end notes, discussing #9, #11, and #13:

Figure 6d and 8d in L2/17-252 (Lazrek) showing #9 followed by #7:

From King Fahd Warsh showing #8 and #9:

From Tripoli Qaloon showing #9:

Figure 9b in L2/17-252 (Lazrek) showing #10:
من الملك فهد ورش يظهر رقم #8 و#10:

من الملك فهد الديري 1436 هـ نهاية التذكرة يظهر رقم #13:

من طرابلس القلوع يظهر رقم #13:

من تونس القلوع يظهر رقم #9 و#11 و#13:
هُمزة الوصل وكيفيّة ضبطها

وضع نقطّة فوق الألف، تدُل على أن الابتداء بها في اللُفظ يكون بحركة الفتح، وذلك نحو: (أَلْبَحَمْدُ)

وضع نقطّة تجاه منتصف يسار الألف، تدُل على أن الابتداء بها في اللُفظ يكون بحركة الضمة، وذلك نحو: (أَعِبَدُوا)

وضع نقطّة تحت الألف، تدُل على أن الابتداء بها في اللُفظ يكون بحركة الكسر، وذلك نحو: (إِهِدَأ)。

وضع جرّة الصلة فوق الألف، تدُل على أن الحركة التي قبل همزة الوصل هي فتحة، وذلك نحو: (رَبِّ الْأَلْفَ)

وضع جرّة الصلة في وسط الألف، تدُل على أن الحركة التي قبل همزة الوصل هي ضمة، وذلك نحو: (يُقَطَّلَ الْمُسْرِيٌّ)

وضع جرّة الصلة تحت الألف، تدُل على أن الحركة التي قبل همزة الوصل هي كسرة، وذلك نحو: (رَبِّ الْعَلِيمِيْنَ)
14. ARABIC LETTER ALEF WITH ATTACHED ROUND DOT ABOVE
15. ARABIC LETTER ALEF WITH ATTACHED LEFT ROUND DOT
16. ARABIC LETTER ALEF WITH ATTACHED RIGHT ROUND DOT
17. ARABIC LETTER ALEF WITH ATTACHED ROUND DOT BELOW

These are hints for various types of recitation of *hamza* in Warsh and al-Dani/Qaloon orthographies.

From Tripoli Qaloon, #14:

Description of #14, #16, and #17 from end notes of Tripoli Qaloon. Note the right circled example in the second line. It uses a fatha over #14:
Sample for #14 from King Fahd Warsh, in red (compare with the non-attached usage of the round dot over alef and under seen in the first line, in blue):

Discussion of #14, #15, and #17 from King Fahd Warsh end notes:

Sample for #15 from King Fahd Warsh:

Contrast the above with the disjoint version from the same edition:

Sample for #16 from Tripoli Qaloon:

Sample for #17 from Tripoli Qaloon:
Sample for #17 from King Fahd Warsh:

18. ARABIC LETTER ALEF WITH DOT ABOVE
19. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND DOT ABOVE
20. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND DOT ABOVE
21. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND DOT ABOVE

While other Northwest African orthographies tend to use the same graphical shape of a rounded dot in all three positions, the Tripoli Qaloon uses a normal “letter” dot above, an attached ring or sukun to the left, and a detached sukun below. Since Unicode encodes letters with modifier dots atomically, these characters are proposed atomically.

From the end notes of Tripoli Qaloon, discussing #18:

Figure 1 L2/16-056 (Shaikh, based on Abudena), showing #19:

From Tripoli Qaloon showing #19:

Note that this version, from King Fahd al-Duri 1436AH end notes, uses a round dot, so it’s best represented by #11, followed by U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE:
From Tripoli Qaloon showing #20:

Note that this version, from King Fahd al-Duri 1436AH end notes, uses a round dot, so it’s best represented by #9, followed by U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE:

A bismillah, from Tripoli Qaloon, showing #21 three times:

22. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND LEFT RING
23. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND LEFT RING
24. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND LEFT RING

These were originally proposed by Abudena in L2/15-329 with different names. While most other orthographies use a attached or detached round dot to represent these, the Tripoli Qaloon uses an attached ring or sukun. We use the name ring for consistency with existing character names.

L2/18-028 recommended character sequence to represent these, but in the Unicode Arabic model, when elements touch or intersect, they are encoded as a single precomposed character. Abudena originally proposed the precomposed characters in L2/15-329 and L2/16-044.

From Tripoli Qaloon, showing #22:

Figure 3 in L2/16-044 (Abudena), showing #23:

From Tripoli Qaloon end notes, showing #23:
25. ARABIC LETTER ALEF WITH ATTACHED RIGHT HAMZA

Some Northwest African editions use this when the *alef hamza* is followed by a *damma*, where the most common orthographies use U+0623 ALEF WITH HAMZA ABOVE instead.

26. ARABIC LETTER ALEF WITH ATTACHED LEFT HAMZA

This is similar to #25, but it’s used in Warsh orthographies, and the *hamza* is attached to the left of the *alef*. Depending on the style, the following *damma* may appear over the *alef* part or the *hamza* part.
27. ARABIC SUKUN BELOW

Abudena proposed this in L2/15-329 as a part of his precomposed proposed characters 25, 26, and 27. Since in all the examples he provides they appear detached from the alef form, and there may be a need for a sukun below in other orthographies, we think a combining sukun below is a better choice for encoding, as it’s more productive.

28. ARABIC LARGE CIRCLE BELOW

This is clearly a “graphical” circle, intended to contrast with sukun and the filled round dot. It’s used in the al-Duri orthography to represent a certain kind of imala, in contrast with places where a filled circle would be used.
From King Fahd al-Duri 1436AH:

مُوسِي لْقُومِهِ

From King Fahd al-Duri 1436AH end notes, discussing the notation and calling the symbol “a large empty-centered dot”:

ووضع نقطة كبيرة خالية الوسط تحت الحرف بدلاً من الفتحة يدل على إشارة إملاء صغير نحو: (مُوسِي، وعَصِي، وَيْحِي) والإملاء الصغير هي ماربة الفتحة والإملاء المخصصة (الكبرى) ولذا يقال لها بين بين وبين الفظيرين، وسميت بالتقييد. واختار اللجنة هذه الكيفية في الإملاء الكبرى لضمان علماء الضبط. وأثرت هذه الكيفية في الإملاء الصغير للفرق بينهما، ولأن هذه الكيفية في الصغير أقرب إلى علامات ضبط الصحائف، ويفتق المقول عليه في الفرق بين الإماليين هو التلقى من أفواه المشايخ المتقيين.

29. ARABIC ROUND DOT INSIDE LARGE CIRCLE BELOW

This is similar to #28 but is used in the Tunis Qaloon. It used to differentiate imala types, but while the King Fahd al-Duri uses the empty circle for small *imala* and the filled circle for large *imala*, the Tunis Qaloon uses the filled circle for small *imala* and the two circles for large *imala*.

From Tunis Qaloon:
From Tunis Qaloon end notes, discussing the notation:

30. ARABIC DOUBLED MADDA
31. ARABIC HALF MADDA OVER MADDA

These graphically innovative characters are used in Tunis Qaloon in contrast with U+0653 ARABIC MADDAH ABOVE to represent different lengths for recitation of madda. #31 appears to be only used when followed by a pause (waqf).

From Tunis Qaloon, showing #30:

From Tunis Qaloon, showing #31:

Compare with normal madda, U+0653 ARABIC MADDAH ABOVE from Tunis Qaloon:
Discussion of the notation, from Tunis Qaloon end notes:

32. ARABIC SMALL HIGH ZAH

This is used to note that a *dad* should be pronounced as *zah* (compare to U+06DC ARABIC SMALL HIGH SEEN, which is used to note that a *sad* should be pronounced as *seen*). It is very rare and may only happen once in the few traditions that use it. Also note that since this is similar to U+06DC ARABIC SMALL HIGH SEEN in being closely tied to the letter it modifies, it should be listed in MCM characters in UTR #53.

Similar to U+06DC, we may later find that this symbol is also used as a *waqf* sign in some Quranic notation. In such a case, a CGJ would be used to make it appear after *harakat*.

From King Fahd al-Duri 1436AH:
From King Fahd al-Duri 1436AH end notes discussing the character, as well as U+06DC ARABIC SMALL HIGH SEEN:

33. ARABIC LETTER THIN YEH
This is a yeh written with a thinner pen, used contrastively with the normal yeh in the Warsh orthography to represent a yeh letter that is not present in early manuscripts but is added to ease reading. In other orthographies that try to remain true to the original letters, it may be written as U+06E6 ARABIC SMALL YEH.

From King Fahd Warsh, compare the first initial yeh (in blue), which is a normal yeh, with the thin yeh:

Compare the above text with a more common Middle Eastern orthography (from the website quranrasm.ir), which uses U+06E6 ARABIC SMALL YEH:

From King Fahd Warsh, showing the medial forms of both normal and thin yehs:

Compare the above text with a more common Middle Eastern orthography (the website quranrasm.ir), which takes the first yeh as a normal yeh and adds a stand-alone U+06E6 ARABIC SMALL YEH:
From the end notes of King Fahd Warsh, discussing the thin yeh:

34. ARABIC TATWEEL WITH TWO DOTS BELOW

The Tripoli Qaloon uses this instead of #33. Instead of a lighter pen, the notation doesn’t write the tooth starting the letter, practically making it a tatweel with two dots below.

From Tripoli Qaloon, used word-initially:

From Tripoli Qaloon, followed by U+06E7 ARABIC SMALL HIGH YEH as well as a shadda and kasra, used word-medially:

From Tripoli Qaloon end notes:

From another Qaloon edition, very similar in style to the Tripoli Qaloon, used word-initially:
From the same Qaloon edition, followed by U+06E7 ARABIC SMALL HIGH YEH as well as a shadda and kasra, used word-medially:

35. ARABIC TATWEEL WITH OVERSTRUCK WAW
36. ARABIC TATWEEL WITH OVERSTRUCK HAMZA

These represent an unwritten waw and an unwritten hamza in a specific word, “įlyasu’u” whose skeleton is written a unique way. Similar to tatweel, these are join-causing letters and connect to both sides, without changes in their general shape.

From Tripoli Qaloon:

Compare these to the more common forms, from a Middle Eastern edition, that uses U+08F3 ARABIC SMALL HIGH WAW and U+0654 ARABIC HAMZA ABOVE:

From Tripoli Qaloon end notes:

From Al-Saleki 2017, mixing the formats, using #35 but not #36:

37. ARABIC SMALL HIGH YEH BARREE WITH TWO DOTS BELOW

This is a modifier symbol that specifies that an alef should be pronounced as yeh. The yeh is dotted, to hint that it should be pronounced as /j/ instead of /i:/ . It’s a rare character, and it seems that it appears
only once in the whole text of the Quran in Warsh orthography, to specify that a word that other traditions usually read as /li-ʔahaba/ should be read as /li-jahaba/.

From King Fahd Warsh (also mentioned in Lazrek L2/17-252), followed by a *fatha*:

![Image](image1)

From Morocco Warsh, followed by a *fatha*:

![Image](image2)

38. **ARABIC LARGE ROUND DOT ABOVE**
39. **ARABIC LARGE ROUND DOT BELOW**

While in most Quranic orthographies the size of a round dot does not appear to be significant, the al-Duri orthography makes a distinction between a large round dot and a small round dot. It uses a large round dot for *imala*, and a small round dot for *tassheel* of *hamza* and pronunciation of *alef* in case of a disconnect in pronunciation (*qat*).

Sample showing differentiation in King Fahd al-Duri 1429AH:

![Image](image3)
Description of each symbol from the end notes of King Fahd al-Duri 1429AH, where the difference is size is called out by calling these “large” and “small” round filled dots:
Discussion from the end notes of Tunis Qaloon, showing the differentiation in size:

While the King Fahd al-Dani uses a large dot for *imala* and a small dot for *ikhtilas*, the Tripoli Qaloon uses a large dot for both, which can go both above and below a letter:

Suggested codepoints and properties

There is not enough space in the existing Arabic Extended-A block, so a new block named Arabic Extended-B should be allocated at U+0870..U+089F.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0870</td>
<td>١</td>
</tr>
<tr>
<td>0871</td>
<td>١</td>
</tr>
<tr>
<td>0872</td>
<td>١</td>
</tr>
<tr>
<td>Unicode</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>0873</td>
<td>ARABIC LETTER ALEF WITH LEFT MIDDLE STROKE</td>
</tr>
<tr>
<td>0874</td>
<td>ARABIC LETTER ALEF WITH ATTACHED KASRA</td>
</tr>
<tr>
<td>0875</td>
<td>ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA</td>
</tr>
<tr>
<td>0876</td>
<td>ARABIC LETTER ALEF WITH ATTACHED ROUND DOT ABOVE</td>
</tr>
<tr>
<td>0877</td>
<td>ARABIC LETTER ALEF WITH ATTACHED RIGHT ROUND DOT</td>
</tr>
<tr>
<td>0878</td>
<td>ARABIC LETTER ALEF WITH ATTACHED LEFT ROUND DOT</td>
</tr>
<tr>
<td>0879</td>
<td>ARABIC LETTER ALEF WITH ATTACHED ROUND DOT BELOW</td>
</tr>
<tr>
<td>087A</td>
<td>ARABIC LETTER ALEF WITH DOT ABOVE</td>
</tr>
<tr>
<td>087B</td>
<td>ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND DOT ABOVE</td>
</tr>
<tr>
<td>087C</td>
<td>ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND DOT ABOVE</td>
</tr>
<tr>
<td>087D</td>
<td>ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND DOT ABOVE</td>
</tr>
<tr>
<td>087E</td>
<td>ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND LEFT RING</td>
</tr>
<tr>
<td>087F</td>
<td>ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND LEFT RING</td>
</tr>
<tr>
<td>0880</td>
<td>ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND LEFT RING</td>
</tr>
<tr>
<td>Unicode</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>0881</td>
<td>ARABIC LETTER ALEF WITH ATTACHED RIGHT HAMZA</td>
</tr>
<tr>
<td>0882</td>
<td>ARABIC LETTER ALEF WITH ATTACHED LEFT HAMZA</td>
</tr>
<tr>
<td>0883</td>
<td>ARABIC TATWEEL WITH OVERSTRUCK HAMZA</td>
</tr>
<tr>
<td>0884</td>
<td>ARABIC TATWEEL WITH OVERSTRUCK WAW</td>
</tr>
<tr>
<td>0885</td>
<td>ARABIC TATWEEL WITH TWO DOTS BELOW</td>
</tr>
<tr>
<td>0886</td>
<td>ARABIC LETTER THIN YEH</td>
</tr>
<tr>
<td>0887</td>
<td>ARABIC BASELINE ROUND DOT</td>
</tr>
<tr>
<td>0888</td>
<td>ARABIC RAISED ROUND DOT</td>
</tr>
<tr>
<td>089D</td>
<td>ARABIC SUPERSCRIPT ALEF MOKHASSAS</td>
</tr>
<tr>
<td>089E</td>
<td>ARABIC DOUBLED MADDA</td>
</tr>
<tr>
<td>089F</td>
<td>ARABIC HALF MADDA OVER MADDA</td>
</tr>
<tr>
<td>08C9</td>
<td>ARABIC SMALL Farsi Yeh</td>
</tr>
<tr>
<td>08CA</td>
<td>ARABIC SMALL HIGH Farsi Yeh</td>
</tr>
<tr>
<td>08CB</td>
<td>ARABIC SMALL HIGH YEH BARREE WITH TWO DOTS BELOW</td>
</tr>
<tr>
<td>Code</td>
<td>08CC</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>08CD</td>
</tr>
<tr>
<td></td>
<td>08CE</td>
</tr>
<tr>
<td></td>
<td>08CF</td>
</tr>
<tr>
<td></td>
<td>08D0</td>
</tr>
<tr>
<td></td>
<td>08D1</td>
</tr>
<tr>
<td></td>
<td>08D2</td>
</tr>
<tr>
<td></td>
<td>FBC2</td>
</tr>
</tbody>
</table>

**Suggested properties**

No collation information is available. These characters should be included in the MCM set in UTR #53 since they can all take *fatḥas*:

U+08CA ARABIC SMALL HIGH Farsi Yeh
U+08CB ARABIC SMALL HIGH Yeh Barree With Two Dots Below
U+08CD ARABIC SMALL HIGH ZAH

The proposed character properties follow.

**UnicodeData.txt:**

0870;ARABIC LETTER ALEF WITH ATTACHED FATḥA;Lo;0;AL;;;;;N;;;;;
0871;ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATḥA;Lo;0;AL;;;;;N;;;;;
0872;ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKÉ;Lo;0;AL;;;;;N;;;;;
0873;ARABIC LETTER ALEF WITH LEFT MIDDLE STROKÉ;Lo;0;AL;;;;;N;;;;;
0874;ARABIC LETTER ALEF WITH ATTACHED KASRA;Lo;0;AL;;;;;N;;;;;
0875;ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA;Lo;0;AL;;;;;N;;;;;
0876;ARABIC LETTER ALEF WITH ATTACHED ROUND DOT ABOVE;Lo;0;AL;;;;;N;;;;;
0877;ARABIC LETTER ALEF WITH ATTACHED RIGHT ROUND DOT;Lo;0;AL;;;;;N;;;;;
0878;ARABIC LETTER ALEF WITH ATTACHED LEFT ROUND DOT;Lo;0;AL;;;;;N;;;;;
0879;ARABIC LETTER ALEF WITH ATTACHED ROUND DOT BELOW;Lo;0;AL;;;;;N;;;;;
ArabicShaping.txt:
0870; ALEF WITH ATTACHED FATHA; R; ALEF
0871; ALEF WITH ATTACHED TOP RIGHT FATHA; R; ALEF
0872; ALEF WITH RIGHT MIDDLE STROKE; R; ALEF
0873; ALEF WITH LEFT MIDDLE STROKE; R; ALEF
0874; ALEF WITH ATTACHED KASRA; R; ALEF
0875; ALEF WITH ATTACHED BOTTOM RIGHT KASRA; R; ALEF
0876; ALEF WITH ATTACHED ROUND DOT ABOVE; R; ALEF
0877; ALEF WITH ATTACHED ROUND DOT; R; ALEF
0878; ALEF WITH ATTACHED LEFT ROUND DOT; R; ALEF
0879; ALEF WITH ATTACHED ROUND DOT BELOW; R; ALEF
087A; ALEF WITH DOT ABOVE; R; ALEF
087B; ALEF WITH ATTACHED TOP RIGHT FATHA AND DOT ABOVE; R; ALEF
087C; ALEF WITH RIGHT MIDDLE STROKE AND DOT ABOVE; R; ALEF
087D; ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND DOT ABOVE; R; ALEF
087E; ALEF WITH ATTACHED TOP RIGHT FATHA AND LEFT RING; R; ALEF
087F; ALEF WITH RIGHT MIDDLE STROKE AND LEFT RING; R; ALEF
0880; ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND LEFT RING; R; ALEF
0881; ALEF WITH ATTACHED RIGHT HAMZA; R; ALEF
0882; ALEF WITH ATTACHED LEFT HAMZA; R; ALEF
0883; TATWEEL WITH OVERSTRUCK HAMZA; R; ALEF
0884; TATWEEL WITH OVERSTRUCK WA; R; ALEF
0885; TATWEEL WITH TWO DOTS BELOW; R; ALEF
0886; LETTER THIN YEH; R; ALEF
0887; BASELINE ROUND DOT; R; ALEF
0888; RAISED ROUND DOT; R; ALEF
0890; SUPERScript ALEF MOKHASSAS; Mn; 230; NSM; N; ; ; ;
0891; DOUBLED MADDA; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0892; HALF MADDA OVER MADDA; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0893; SMALL FARSI YEH; Lm; 0; AL; ; ; ; ; ; ; ;
0894; SMALL HIGH FARSI YEH; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0895; SMALL HIGH YEH BARREE WITH TWO DOTS BELOW; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0896; SMALL HIGH WORD SAH; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0897; SMALL HIGH ZAH; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0898; LARGE ROUND DOT ABOVE; Mn; 230; NSM; ; ; ; ; N; ; ; ;
0899; LARGE ROUND DOT BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ;
089A; LARGE CIRCLE BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ;
089B; ROUND DOT INSIDE LARGE CIRCLE BELOW; Mn; 220; NSM; ; ; ; ; N; ; ; ;
FBC2; SYMBOL WASLA ABOVE; Sk; 0; AL; ; ; ; ; N; ; ; ;

ArabicShaping.txt:
Annotations for existing round dot characters

New annotations should be added to the names list in order to help users of the standard identify the characters to use:

A1. To U+065C ARABIC VOWEL SIGN DOT BELOW, add an annotation “Also used in Quranic text in African and other orthographies“

A2. To U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE, add an annotation “Also used in Quranic text in African and other orthographies to represent wasla, ikhtilas, etc”.

From King Fahd Warsh, as wasla marks:

From King Fahd al-Duri 1436AH end notes, explaining how it’s used for tassheel of hamza:

Annotations for existing arrowhead characters

Abudena, in L2/15-329, proposes Al-Dani versions of damma and dammatan. While others thought the normal damma character should be used for these, we found that African dammas are already encoded in Unicode! U+08F8 and U+08FB (proposed by Lorna A. Priest and Martin Hosken in L2/10-288R) are indeed African forms of dammas, also used in the Warsh orthography.

The end notes of King Fahd Qaloon visually describe these as “a small headless wow”:
A3. To U+08F8 ARABIC RIGHT ARROWHEAD ABOVE, add an annotation “Also used in Quranic text in African and other orthographies to represent damma”.

A4. To U+08FB ARABIC DOUBLE RIGHT ARROWHEAD ABOVE, add an annotation “Also used in Quranic text in African and other orthographies to represent dammatan”.

Symbols not proposed

In reviewing the above-mentions proposals and various modern published Qurans, we ran into a few other symbols that we don’t properly understand yet. The symbols found in these samples are still not analyzed to the authors’ satisfaction, so they are not proposed yet.

X1-X4. Abudena, in L2/15-329, proposes several tanween forms, which are also used in other African orthographies. As explained in “Annotations for existing arrowhead characters”, above, one of 2 or 3 is already encoded at U+08FB ARABIC DOUBLE RIGHT ARROWHEAD ABOVE. The four others should be encoded for consistency, but proper names and potential unification for them depends on better understanding of alternate tanween forms across the different orthographies.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>ALDANI ARABIC DAMMATAN IN CASE OF ETHHAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>ALDANI ARABIC DAMMATAN IN CASE OF EDGAHM AND EKHPA</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>ALDANI ARABIC DAMMATAN IN CASE OF EQLAB</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>ALDANI ARABIC FATHATAN IN CASE OF EQLAB</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>ALDANI ARABIC KASRATAN IN CASE OF EQLAB</td>
</tr>
</tbody>
</table>

X5. From King Fahd al-Duri 1636AH, perhaps a baseline small alef, similar to U+06E5 ARABIC SMALL WAW and U+06E6 ARABIC SMALL YEH:
X6. Abudena proposes a HIGH YEH WITH DOT ABOVE form in L2/15-329 (his character 16), which actually appears to be a noon. It’s unclear if this should be unified with U+06E8 ARABIC SMALL HIGH NOON or not. Here is a sample from the Tripoli Qaloon:

X7. Abudena, in L2/15-329, proposes two lam-alef ligature forms common is Northwest African orthographies (his characters 28 and 29), which are actually a ligature of lam with a superscript alef. Encoding these as an atomic ligature could be problematic, since each part can take a different harakat. But it’s unclear if we should encode another combining superscript alef form or treat the whole thing as a ligature of U+0644 ARABIC LETTER LAM and U+0670 ARABIC LETTER SUPERSCRIPT ALEF.
Here’s a sample from Tripoli Qaloon, where the lam is taking a fatha:

Here’s another sample from Tripoli Qaloon, where the lam is taking a fatha and the alef part is taking a madda:

X8-X9. In L2/15-329, Abudena proposes a toothless sad and toothless dad. At first, they appear to be a different style, but there is indeed potentially intentional contrastive usage of normal sad and toothless sad in Tripoli Qaloon.

Here is contrastive evidence (both from page 380 of Tripoli Qaloon):

We need to understand if there’s a meaning or an intentional pattern, especially since this doesn’t exist in the King Fahd Qaloon:

... or Tunis Qaloon:

We also need to understand if a final or isolated form of such a toothless sad or dad exist.
X10. A baseline dash that indicated deletion of a hamza sound. This may be representable by either U+06D4 ARABIC FULL STOP or <tawteel, ZWNJ>. Discussion from end notes of King Fahd Warsh:

Acknowledgments

Marijn van Putten was key to us deciphering many of the symbols and helped the authors with understanding their semantic and behavior is several communications, as well as by finding sources. Lorna Evans kindly designed the glyphs used in the chart based on SIL’s Scheherazade font. Kamal Mansour and members of the Unicode script ad hoc reviewed the proposal and provided us with comments that helped improve it.

References


The following editions of the Quran have also been used:

1. Tripoli Qaloon:

2. King Fahd al-Duri 1429AH:


MUSHAF MADINAH AL-NABIYAH, BRAWYA AL-DARYAH MIN AYUB UMRO AL-BESYRI, MAQAM AL-MALIK FAHD AL-DURI, AL-SHIFA, 1429 H.
3. King Fahd al-Duri 1436AH:

مصحف المدينة النبوية، برواية الدوري عن أبي عمرو البصري، مجمع الملك فهد لطباعة المصحف الشريف، 1436 هـ.

4. Tunis Qaloon:

المصحف المعلم، برواية قالون عن الإمام نافع المدني، مع إبراز أهم قواعد الترتيل، حنبل، تونس.

ISBN 9789973690029

5. King Fahd Warsh:

مصحف المدينة النبوية، وفق رواية ورش عن الإمام نافع، مجمع الملك فهد لطباعة المصحف الشريف، 1428 هـ.

6. King Fahd Qaloon:

مصحف المدينة النبوية، وفق رواية قالون عن الإمام نافع، مجمع الملك فهد لطباعة المصحف الشريف، 1431 هـ.

7. Morocco Warsh:

قرآن كريم برواية ورش، المصحف الحسن المبسوط، وزارة الأوقاف والشؤون الإسلامية، المملكة المغربية، 1417 هـ.
A. Administrative

1. **Title:** *Arabic Additions for Quranic orthographies*

2. Requester’s name: **Roobeh Pournader and Deborah Anderson**

3. Requester type (Member body/Liaison/Individual contribution): **Expert Contribution**

4. Submission date: **September 29, 2019**

5. Requester’s reference (if applicable): 

6. Choose one of the following:
   - This is a complete proposal: **Yes**
   - (or) More information will be provided later: **No**

B. Technical – General

1. Choose one of the following:
   - a. This proposal is for a new script (set of characters): **No**
   - Proposed name of script: **N/A**
   - b. The proposal is for addition of character(s) to an existing block: **Yes**
   - Name of the existing block: *Arabic Extended-A and Arabic Presentation Forms-A (as well as new block Arabic Extended-B)*

2. Number of characters in proposal: **39**

3. Proposed category (select one from below - see section 2.2 of P&P document):
   - A-Contemporary
   - B.1-Specialized (small collection)
   - B.1
   - B.2-Specialized (large collection)
   - C-Major extinct
   - D-Attested extinct
   - E-Minor extinct
   - F-Archaic Hieroglyphic or Ideographic
   - G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided? **Yes**
   - a. If YES, are the names in accordance with the “character naming guidelines”
   - in Annex L of P&P document? **Yes**
   - b. Are the character shapes attached in a legible form suitable for review? **Yes**

5. Fonts related:

---

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

| Lorna Evans (SIL International) |

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

| SIL International |

6. References:

| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | Yes |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | Yes |

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? Yes

| UTR #53 information is provided for rendering order of combining marks. |

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
C. Technical - Justification

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this proposal for addition of character(s) been submitted before?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES explain</td>
<td></td>
</tr>
<tr>
<td></td>
<td>See various proposals mentioned on page 1</td>
</tr>
<tr>
<td>2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, with whom?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Several Quranic experts, including Marijn van Putten of Leiden University</td>
</tr>
<tr>
<td>If YES, available relevant documents:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Offline communications</td>
</tr>
<tr>
<td>3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?</td>
<td>Yes</td>
</tr>
<tr>
<td>Reference:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Several Muslim communities with minority interpretations of the Quran</td>
</tr>
<tr>
<td>4. The context of use for the proposed characters (type of use; common or rare)</td>
<td></td>
</tr>
<tr>
<td>Reference:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical</td>
</tr>
<tr>
<td>5. Are the proposed characters in current use by the user community?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, where? Reference:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mostly based in Northwest Africa (aka Maghrib)</td>
</tr>
<tr>
<td>6. After giving due considerations to the principles in the P&amp;P document must the proposed characters be entirely in the BMP?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, is a rationale provided?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Existing similar characters are in the BMP, since some of these are combining or have contextual shaping, it’s better to keep them in the same plane as the characters they interact with.</td>
</tr>
<tr>
<td>7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?</td>
<td>No</td>
</tr>
<tr>
<td>8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>N/A</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>N/A</td>
</tr>
<tr>
<td>9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>N/A</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>N/A</td>
</tr>
<tr>
<td>10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?</td>
<td>Yes</td>
</tr>
<tr>
<td>Question</td>
<td>Yes/No</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>11. Does the proposal include use of combining characters and/or use of composite sequences?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, is a rationale for such use provided?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?</td>
<td>N/A</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
<tr>
<td>12. Does the proposal contain characters with any special properties such as control function or similar semantics?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, describe in detail (include attachment if necessary)</td>
<td>N/A</td>
</tr>
<tr>
<td>13. Does the proposal contain any Ideographic compatibility characters?</td>
<td>No</td>
</tr>
<tr>
<td>If YES, are the equivalent corresponding unified ideographic characters identified?</td>
<td>N/A</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td></td>
</tr>
</tbody>
</table>