

**Proposal to encode some combining Arabic character
for Quranic representation
in The Unicode Standard and ISO/IEC10646**

King Fahd Glorious Quran Printing Complex

P.O. Box 6262 - Al-Madinah Al-Munawwarah
Kingdom of Saudi Arabia
Tel.: 966-4-8615600 Fax: 966-4-8615495
it-dep@qurancomplex.gov.sa
<http://www.qurancomplex.org/>
2010-04-14

1. Introduction

The most important signs used in the largely met representation of Quranic text are already included in Unicode. However, still miss some others. This document proposes to encode two missing combining Arabic characters. Its encoding would be useful for representing correctly Quranic text.

2. Utilization

2.1. Facilitator Hamza

Some Arabic words begin with two Hamza. In some Quranic riwayaths well – known in the Muslim world, the second Hamza is a Facilitator Hamza. Facilitator Hamza is the pronunciation of a laxative Hamza out between its way and the way of the long vocalization letter congener its diacritical vocalization. It is not a veritable Hamza from its way nor is switched to the letter Alef or Waw or Yeh.

الهمزة المسهلة : هو لفظ الهمزة ملين يخرج من بين مخرجها ومخرج حرف المد المجانس لحركتها، أي بين بين، فلا هي همزة محققة من مخرجها ولا هي مبدلة إلى حرف الألف أو الواو أو الياء من مخرجه.

In order to distinguish between *veritable* Hamza and *facilitator* Hamza, a diacritical sign is used. The three cases of second Hamza are:

- The Hamza with Fatha pronounced between Hamza and Alef (See Figure 1). The second Hamza is transformed as Alef for Madda.
- The Hamza with Damma pronounced between Hamza and Waw (See Figure 2). The second Hamza is transformed to the diacritical dot.
- The Hamza with Kasra, pronounced between Hamza and Yeh (See Figure 3). The second Hamza is transformed to the diacritical dot.

In the two last cases, there is no Unicode character that could be used for this purpose. We propose to add one.

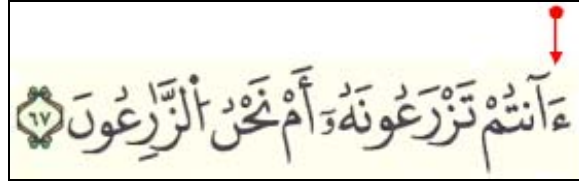


Figure 1.a - Hamza with Fatha (Sourat Al-Waqiat : Ayat 67) [1, pp. 489]



Figure 1.b - Hamza with Fatha (Sourat Al-Waqiat : Ayat 67) [2, pp. 452]

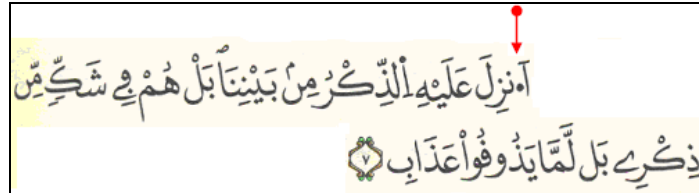


Figure 2.a - Hamza with Damma (Sourat Sâd : Ayat 7) [1, pp. 409]

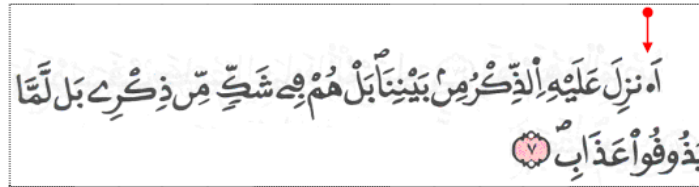


Figure 2.b - Hamza with Damma (Sourat Sâd : Ayat 7) [2, pp. 376]





Figure 3.a - Hamza with Kasra (Sourat Yûsuf : Ayat 90) [1, pp. 215]



Figure 3.b - Hamza with Kasra (Sourat Yûsuf : Ayat 90) [2, pp. 193]

2.2. Starting Hamza Wasl with Damma

In Arabic, Hamza Wasl is not pronounced if it is in middle of reading. However, it is pronounced if we begin with it in reading. In some Quranic riwayaths well – known in the Muslim world, Hamza Wasl is marked by a dot. The position of this dot indicates the vowel used when it is pronounced (See Figure 4). The three cases of vowel are:

- In the case of Fatha, the dot is on Alef. The Unicode character 06EC  “Arabic rounded high stop with filled center” could be used for this purpose.
- In the case of Kasra, the dot is under Alef. The Unicode character 065C  “Arabic vowel sign dot below” could be used for this purpose.
- In the case of Damma, the dot is in middle left of Alef (See Figures 5, 6, 7 and 8). There is no Unicode character that could be used for this purpose. We propose to add one.

إِنَّ الصَّلَاةَ تَابِعَةٌ لِلْحَرَكَةِ الَّتِي قَبْلَ أَلِفِ الْوَصْلِ (هَمْزَةُ الْوَصْلِ) سَوَاءَ أَكَانَتِ الْحَرَكَةُ لَازِمَةً أَمْ عَارِضَةً ، (وَأَلِفُ الْوَصْلِ هِيَ الَّتِي تَسْقُطُ وَصَلًا وَتَثْبُتُ ابْتِدَاءً) فَإِنْ كَانَتِ الْحَرَكَةُ فَحَةً جُعِلَتْ حِرَّةُ الصَّلَاةِ فَوْقَ الْأَلِفِ نَحْوُ : (هُوَ اللَّهُ) وَإِنْ كَانَتْ كَثْرَةً جُعِلَتْ تَحْتَهَا نَحْوُ : (إِنِّي إِصْطَفَيْتُكَ) وَإِنْ كَانَتْ ضَمَّةً جُعِلَتْ فِي وَسْطِهَا نَحْوُ : (أَنْ تُشْكِرَ) .
وَالنَّقْطَةُ الْمُسْتَدِيرَةُ الشَّكْلِي الْمَطْمُوسَةُ الْوَسْطَى تَدُلُّ عَلَى كَيْفِيَّةِ الْإِبْتِدَاءِ بِأَلِفِ الْوَصْلِ فَإِنْ وُضِعَتْ فَوْقَ الْأَلِفِ ابْتِدَاءً بِهَا مَفْتُوحَةً ، وَإِنْ وُضِعَتْ تَحْتَهَا ابْتِدَاءً بِهَا مَكْسُورَةً وَإِنْ وُضِعَتْ فِي وَسْطِهَا ابْتِدَاءً بِهَا مَضْمُومَةً كَمَا رَأَيْتَ فِي الْأَمْثَلَةِ الثَّلَاثَةِ السَّابِقَةِ .

Figure 4.a - Position of Arabic starting point rule [1, pp. w]

بَعْضُ الْمُضَمَّاتِ فِي الْمَبْدِ

العلامة	دلالتها	أمثلتها
◌ِ	النفخة، إن وضعت فوق ألف الوصل يبتدأ بها مفتوحة	الْتَبَّأُ الْعَظِيمُ
◌ِ	وإن وضعت تحت الألف ابتداءً بها مكسورة	أَقْرَأُ - أَهْدِيْنَا
◌ِ	وإن وضعت في وسط الألف ابتداءً بها مضمومة	أَمْكُرُوا - أَمْخُلُومًا

Figure 4.b - Position of Arabic starting point rule [3, pp. 29]

لَقَدْ كَفَرَ الَّذِينَ قَالُوا إِنَّ اللَّهَ هُوَ الْمَسِيحُ
ابْنُ مَرْيَمَ وَقَالَ الْمَسِيحُ بَنِي إِسْرَائِيلَ اعْبُدُوا اللَّهَ رَبِّي وَرَبَّكُمْ
إِنَّهُ مَنْ يُشْرِكْ بِاللَّهِ فَقَدْ حَرَّمَ اللَّهُ عَلَيْهِ الْجَنَّةَ وَمَأْوَاهُ النَّارُ
وَمَا لِلظَّالِمِينَ مِنْ أَنْصَارٍ

Figure 5.a - Hamza with Damma after a letter with Fatha
(Sourat Al-Mâ'idah : Ayat 74) [1, pp. 104]

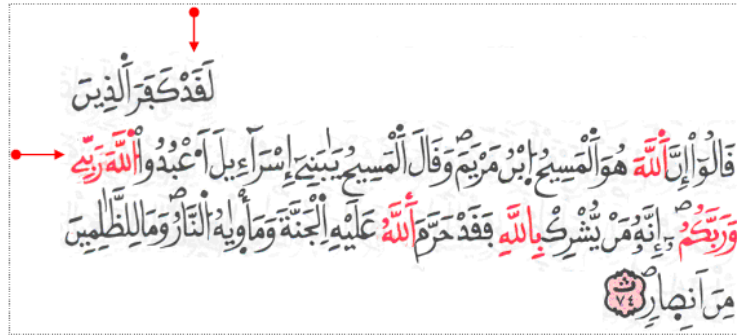


Figure 5.b - Hamza with Damma after a letter with Fatha
(Sourat Al-Mâ'idah : Ayat 74) [2, pp. 88]

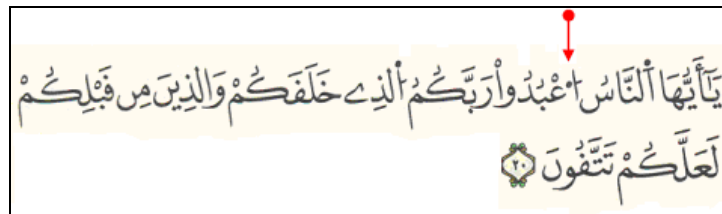


Figure 6.a - Hamza with Damma after a letter with Damma
(Sourat Al-Baqarah : Ayat 20) [1, pp. 4]

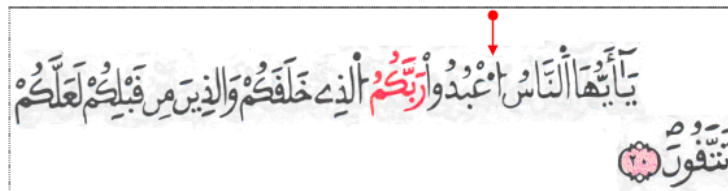


Figure 6.b - Hamza with Damma after a letter with Damma
(Sourat Al-Baqarah : Ayat 20) [2, pp. 5]

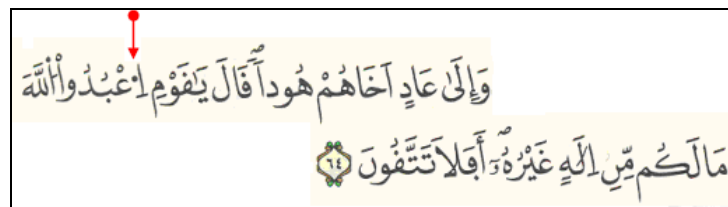


Figure 7.a - Hamza with Damma after a letter with Kasra
(Sourat Al-A'râf : Ayat 64) [1, pp. 138]

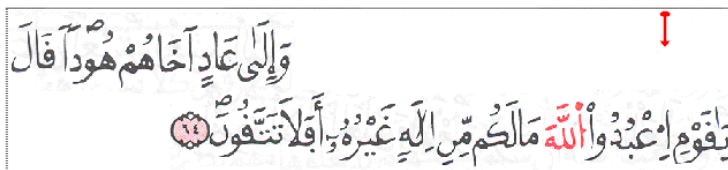


Figure 7.b - Hamza with Damma after a letter with Kasra
(Sourat Al-A'râf : Ayat 64) [2, pp. 122]

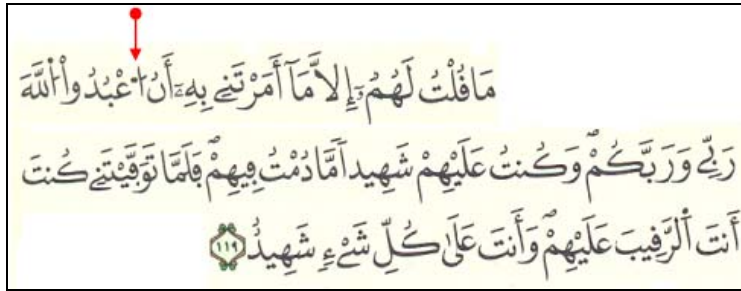


Figure 8.a - Hamza with Damma after a letter with Sukun transformed to Damma (Sourat Al-Mâ'idah : Ayat 119) [1, pp. 110-111]



Figure 8.b - Hamza with Damma after a letter with Sukun transformed to Damma (Sourat Al-Mâ'idah : Ayat 119) [2, pp. 93]

3. Name

The proposed name for the facilitator Hamza sign is: ARABIC VOWEL SIGN DOT BELOW LEFT. However, the name could be COMBINING DOT BELOW LEFT.

The proposed name for the starting Hamza Wasl with Damma is: ARABIC VOWEL SIGN DOT LEFT. However, the name could be COMBINING DOT LEFT.

As are diacritical signs, which could be combined with an Arabic letter, especially Hamza, we could propose to add the term “combining” in its name. However, we could propose to not add the term “Arabic” in its name in order to not be confused with Arabic diacritical mark dot, which is used for another purpose and have another shape. In that case, it could be used for other purpose.

The chosen names is descriptive not functional.
















The chosen names is in order to be consistence with names of existing Diacritical marks and Arabic characters in ISO/IEC10646, and to help making the Unicode Standard easier to use.

4. Shape

The shape of the dot (full circle) in these signs differs from dot, “Nuqta” (full inclined lozenge), used in some Arabic letters.

5. Notes

We suggest the following annotations for these characters:

	<p>ARABIC VOWEL SIGN DOT BELOW LEFT = Arabic facilitator Hamza → 0358 </p>
	<p>ARABIC VOWEL SIGN DOT LEFT = Arabic starting Hamza Wasl with Damma → 302E  → 065C  Arabic vowel sign dot below → 06EC  Arabic rounded high stop with filled center → 0323  combining dot below → 0307  combining dot above → 02D9  dot above → 0A02  Gurmukhi sign bindi → 0A3C  Gurmukhi sign nukta → 11081  Kaithi sign anusvara → 110BA  Kaithi sign nukta → 082D  Samaritan mark nequdaa → 0830  Samaritan punctuation nequdaa</p>

The character "HANGUL SINGLE DOT TONE MARK", already encoding in CJK Symbols and Punctuation Block as a Nonspacing mark, has the same shape as Arabic starting Hamza Wasl with Damma. However, the actual technologies font and rendering don't allow using characters from a script in another.

6. Range

The proposed diacritical marks could be encoded in preference at:

- The Arabic block in 0600-06FF range;
- The Arabic Presentation Forms-B block in FE70- FEFF range;

7. Proprieties

The data for UnicodeData.txt file is as follows:

```
nnnn; ARABIC VOWEL SIGN DOT BELOW LEFT;Sk;0;AL;;;;;N;;;;;
nnnn; ARABIC VOWEL SIGN DOT LEFT;Sk;0;AL;;;;;N;;;;;
```

8. Font

We have design a font in TrueType containing glyphs of the proposed diacritical marks.

9. Reference

- [1] Mushaf Almadinah Anabawiyat, Warsh Riwayah, King Fahd Glorious Quran Printing Complex, 2005.
- [2] Mushaf Warsh, Dar AlMushaf, Beirut.
- [3] Juza Eamma, Adar Alealamia lilkitab, Casablanca, 2009.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. **Title:** *Proposal to encode some combining Arabic character for Quranic representation*

2. Requester's name: *King Fahd Glorious Quran Printing Complex*

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

4. Submission date: *2010-04-14*

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:

B. Technical – General

1. Choose one of the following:
a. This proposal is for a new script (set of characters):
Proposed name of script: _____
b. The proposal is for addition of character(s) to an existing block: *Yes*
Name of the existing block: *Arabic Extended-A*

2. Number of characters in proposal: *2*

3. Proposed category (select one from below - see section 2.2 of P&P document):
A-Contemporary B.1-Specialized (small collection) 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?
King Fahd Glorious Quran Printing Complex
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
King Fahd Glorious Quran Printing Complex
http://fonts.qurancomplex.gov.sa/unicode/signs.ttf
it-dep@qurancomplex.gov.sa

6. References:
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes (See above)*
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes (See above)*
http://fonts.qurancomplex.gov.sa/unicode/QuranMarkEncode.pdf

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*

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain		<i>No</i>
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.)? If YES, with whom? If YES, available relevant documents:	<i>Professor Azzeddine LAZREK lazrek@ucam.ac.ma http://www.ucam.ac.ma/fssm/rydarab/unicode.htm</i>	<i>Yes</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>All Muslims (about 1.5 milliard in the World) reciting correctly the Holy Quran</i>	<i>Yes</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:		
5. Are the proposed characters in current use by the user community? If YES, where? Reference:		
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:		<i>Yes</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>Yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>No</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>No</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>See above</i>	<i>Yes combining Yes No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		<i>No</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:		<i>No</i>