Proposal to Encode ARABIC LETTER THREE QUARTER HIGH HAMZA for Jawi

Ahmad Ali A. Karim

(Tun) Suzana (Tun) Othman

Dr. Hasanuddin Yusof

11 January 2022

Abstract

The Jawi (Malay in Arabic script) letter, "*Three-Quarter High Hamza*" (otherwise simply known as "*Three-Quarter Hamza*" or locally known in Malay as "*Hamzah Tiga Suku*" or "همزة تيݢ سوكو" in the Jawi script) is currently not included in the latest version of Unicode. This makes it hard for users of the Malay language to properly write Jawi text as the character is one of the most common characters used in the Jawi alphabet. In this proposal, we provide a TrueType font named 'UnicodeJawiLetters' containing the glyph as specified by *Dewan Bahasa dan Pustaka*, together with the needed character properties for Unicode specification as well as evidence that this character have been used in Malay literature for over a century and is still actively in use to this day.

I. Acknowledgements

The authors would like to thank Akademi Jawi Malaysia for their feedback and support regarding this proposal, and ME'DIDIK Malaysia for providing the needed font as attached with the email.

The relevant document from Akademi Jawi Malaysia has been attached in Appendix A at the end of this proposal.

Akademi Jawi Malaysia has also kindly provided samples of usages of the character in question.

II. Proposal

The following character is proposed as an addition to the UCS repertoire. It could possibly be viewed as a variant of U+0621 or U+0674, but the guidelines on writing Jawi made by *Dewan Bahasa dan Pustaka*, the body responsible for coordinating the use of the Malay language including Jawi, clearly notes otherwise. In the case of U+0621, Jawi uses this character for a different purpose compared to the *Three-Quarter High Hamza* character. This means that the two letters are not interchangeable in Jawi.

It is also important to note that despite U+0674 being labelled as a Jawi character, this character is not the same as the one in use in Jawi. The Jawi letter does not have the same properties as this character which was originally intended for the Kazakh language.

The *Three-Quarter High Hamza* has the same exact size as U+0621, only with the vertical position being different, that is at three-quarters of the height of the character *Alef* (U+0627.)

Glyph	Code	Character name
ء	088F	ARABIC LETTER THREE QUARTER HIGH HAMZA

III. Rationale

1. Background

The Arabic letter *Hamza* is used to represent a /ʔ/ phoneme, otherwise known as a glottal sound. The letter in its default form stands on its own, however, it may also be 'seated' on top or beneath other letters, namely the letters *Alef* (U+0623 & U+0625,) *Waw* (U+0624,) and *Yeh* (U+0626.) In the Arabic language, these would depend on the position of the *Hamza* letter within a certain word.

$$ؤ ئ إ أ ء$$

Figure 1: Five forms of *Hamza* in the Arabic language.

As the Arabic script gets adopted by other languages, variations have been made towards the letter to represent different phonemes or simply to play a different role entirely in the writing of the language. Some variations were made to existing letter forms, such as in the Uyghur language where the character U+0626 (ئ) is instead utilised for a vocalic start of a certain phrase, such as "نئۇيغوُر" (/ʊjʁʊrˈtʃɛ/).

هەممە ئادەم زانىدنلا ئەركىن ، ئىززەت-ھۆرمەت ۋە ھوقۇقتا باپپاراۋەر بولۇپ تۇغۇلغان. ئۇلار ئەقىلگە ۋە ۋىجدانغا ئىگە ھەمدە بىر-بىرىگە قېرىنداشلىق مۇناسىۋەتىدە خاس روھ بىلەن مۇئامىلە قىلىشى كېرەك.

Figure 2: Article 1 of the Universal Declaration of Human Rights written in Uyghur; note the usage of Yeh With Hamza Above for the vocalic start of words.

However, in some languages there are also new variations of the letter *Hamza* made. This includes Pashto where another form of the seated *Hamza* is added, which is the *Hah With Hamza Above* (U+0681). Aside from seating the *Hamza* on top of another letter, there are also cases where the form of a standalone *Hamza* is changed, such as in Kazakh with their *High Hamza* (U+0674). The *High Hamza* is visually identical to the U+0621 by shape, but different by vertical position and size. Due to these differences, they have been encoded separately in Unicode.

Given that Unicode encodes these two forms of *Hamza*, ء and ٴ (U+0621 & U+0674) separately, it seems appropriate to treat the Jawi character *Three-Quarter High Hamza* in a similar manner.

2. Three-Quarter High Hamza

In the Jawi script, the *Three-Quarter High Hamza* character looks visually identical to U+0621 in its shape and size, but with one difference that is its vertical position, which is at three-quarters of the height of the letter *Alef*. This form of the letter does not replace the regular form of *Hamza* like in U+0621, and both forms are still used for different purposes.

3.10 Huruf *hamzah* ditulis tanpa "rumah" di tengah garisan apabila diapit oleh huruf *alif* dengan huruf *ya* (diftong [ai]) atau huruf *alif* dengan huruf *wau* (diftong [au]) atau huruf *alif-wau* dengan huruf *ya* (diftong [oi]) yang mewujudkan diftong sebagai satu suku kata. Contohnya:

Ejaan Jawi	Padanan Rumi
اءيسيغ	aising
اءيلرون	aileron
اءوديت	audit
اءوتوماتيک	automatik
اوءيديوم	oidium

3.11 Kata atau istilah serapan daripada bahasa Arab yang mengandungi huruf *hamzah* ditulis seperti kedudukannya dalam bahasa sumber. Contohnya:

Ejaan Jawi	Padanan Rumi
وضوء	wuduk
املاء	imlak
اخفاء	ikhfak
انبياء	anbia

Figure 3: From Ismail & Manshoor (2001), page 76 & 77; differences between *Three-Quarter High Hamza* (green) and the regular *Hamza* (red).

The Three-Quarter High Hamza is used for various purposes, most commonly for certain suffixes, diphthongs, and for connecting two vowels of certain combinations. It is also used for the first letter of non-Malay single-syllable loanwords and names starting with a vowel other than /a/, as well as certain Arabic loanwords which consists of any form of *Hamza*, that has been changed to use this form instead.

The standard form of *Hamza* is used for certain Arabic loanwords which has for a long time been spelt using the same spelling as in Arabic. This would generally be words used often in religious texts, as most Malay religious scholars and writers would use the Arabic spelling instead of Malay. Even so, there are still many words closely related to religion which had their *Hamza* changed to *Three-Quarter High Hamza*.

Unlike the Kazakh *High Hamza*, the *Three-Quarter High Hamza* does not differ in size from the original *Hamza* and remains with its full width and height.

Text	Character used	Comments
كأداءن	ARABIC LETTER THREE QUARTER HIGH HAMZA *(proposed)*	Correct position and size
كأداءن	ARABIC LETTER HAMZA (U+0621)	Position too low
كأدأن	ARABIC LETTER HIGH HAMZA (U+0674)	Size too small

Figure 4: Writing the Malay word 'keadaan' in Jawi using the three forms of *Hamza* mentioned.

For the record, during the reformation of the Jawi spelling in 1984 conducted by the Project Work Committee for the Reviewing of the Jawi Spelling under *Dewan Bahasa dan Pustaka*, the character was only mentioned to be positioned in the middle without any exact specifications on the exact position of the character. It would seem that before the publication of the *Daftar Kata Bahasa Melayu: Rumi-Sebutan-Jawi* (Malay Language Word Directory: Rumi-Pronunciation-Jawi) by *Dewan Bahasa dan Pustaka* in 1999, the position was not standardised and the name '*Hamzah Tiga Suku*' did not came to be until the late 2000s, where the earliest official use is seen in 2008 in the second edition of the Malay word directory published by *Dewan Bahasa dan Pustaka*. Prior to 1984, they would sometimes be written higher than the height of *Alef* or even slightly lower than the current standardised height. However, they would always be distinguishably higher than the standard form of *Hamza*.
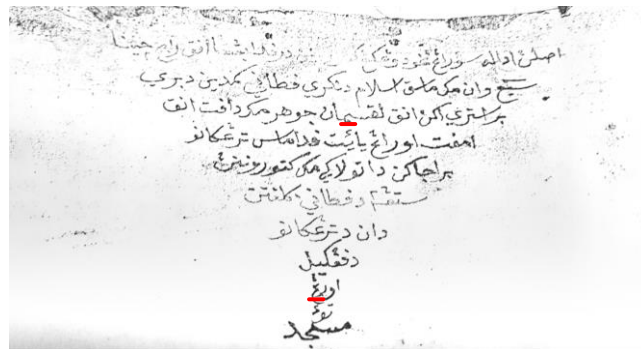

Figure 5: Heading of an old Malay family tree explaining the origins of the Tok Masjid (توﺉ مسجد) family.

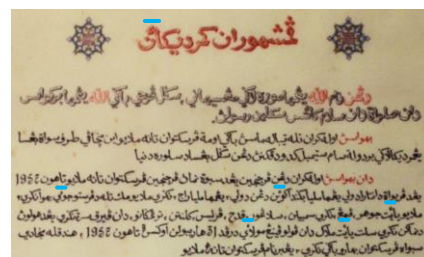Figure 6: Johore State Constitution (1895); position slightly higher than *Alef*.

Figure 7: From Malayan Declaration of Independence (1957); position exactly the same as today.

Figure 8: From Haris (2019), page 13; school textbook explaining the usage of this character.

The character serves various purposes which has changed over time as the Malay language itself develops into the modern era. It is the common-most form of the letter *Hamza*, contrary to its default form, known locally as *Hamzah Setara* or *Hamzah Sejajar* (lit. *Hamza that is on a par*, referring to the relativity of its base to the base of other letters), which is the second least common form out of the five *Hamza* forms in the Jawi alphabet.

Figure 9: From Haris (2019), page 1; the five forms of *Hamza* in the Jawi alphabet, which consists of, from right to left, U+0623, U+0625, *Three-Quarter High Hamza*, U+0621, U+0626.

It is clear from the visible usage of this letter as well as the explanation provided by *Dewan Bahasa dan Pustaka* being the official body which regulates the usage of the Malay language, that *Three-Quarter High Hamza* is not interchangeable with U+0621, and that U+0674 is not the appropriate representation of the letter in its present form, therefore it should not be recommended for usage in Jawi text.

Figure 10: From Aznan (2015), page 18; a guide to writing in Jawi explaining the proper position and size of the letter. The text underneath states: "the *Hamza* letter like above, must be written at the level of the middle or three-quarters of the height of the letter *Alef*. Its size must be big."

The problem right now is that users are resulted in using various techniques to attempt to display *Three-Quarter High Hamza* correctly. The most common way of doing so is using a superscript U+0621, which leads to the letter being smaller thus bringing back to the problem of using U+0674. The most precise technique would be to adjust the vertical position of U+0621 manually, which is only achievable through most word processors such as Microsoft Word like what is done in newspapers and other types of publications, but not possible for webpages without additional coding involved. In the Malay Wikipedia, the solution used would be to add an inline container with its vertical position altered, but such is tedious to execute and maintain in a long run especially if a change of font formatting occurs, since the position adjustments must use pixels. Furthermore, it is simply impossible to write the letter in plain text platforms which lack any text formatting feature, such as text messaging or social media platforms. This leads to many Jawi users to result in using U+0621 without any sort of alterations as it would be the closest and easiest way to achieve writing *Three-Quarter High Hamza*.

Figure 11: From Zakkina (2019-07-01); usage of Jawi in a local newspaper, *Utusan Malaysia.*

Figure 12:
Left: Screenshot from the Malay Wikipedia article on Richard Nixon, showing the usage of the *Three-Quarter High Hamza.*
Right: A look into the code of the page where an inline span container is added for a U+0621 to display it at a higher vertical position.

Given the clear distinction between these forms in users' minds, the fact that the forms ء and ٴ are not considered interchangeable, and that form ٴ is not applicable for Jawi usage, it is proposed that a new character ARABIC LETTER THREE QUARTER HIGH HAMZA should be encoded in the UCS to represent the Jawi *Three-Quarter High Hamza* letter.

Additionally, considering form ٴ is not applicable for Jawi usage, it is proposed that Jawi be dropped from the description of the character to avoid confusion among vendors and users.

3. Jawi user community information

The usage of the Jawi script can be traced back from as early as the 14th century CE, with the earliest form of Jawi writing found on the Terengganu Inscription Stone, a proclamation by the ruler of Terengganu on the Sharia (Islamic) law. Since then, Jawi has been widely used by most of the inhabitants of maritime Southeast Asia, especially those speaking Malay, Acehnese, Minangkabau, Banjarese, and various others.

Jawi became the default writing system for the Malay language and various sister languages. Being an adaptation of the foreign Arabic script, Jawi had to add several new letters to represent phonemes used by Malays but are non-existent in the Arabic language. A total of 6 new letters were added, with some alterations being done to existing Arabic letters including the letters *Kaf* and *Hamza*. The latest addition to the Jawi alphabet, Arabic Letter Waw With Dot Above (U+06CF; locally known as Va), was introduced in 1973 by a local Jawi magazine, *Majalah Dian* by its founder, Yusoff Zaky Yaacob, who also introduced new functions for *Three-Quarter High Hamza*. The letter was later included in the formalised Jawi spelling guide in 1986 developed by *Dewan Bahasa dan Pustaka.*

In 1963, the Latin or Rumi script has been adapted as the official writing script in Malaysia for the purpose of modernisation and standardisation upon unification with two Bornean states and Singapore, and attempts to standardise Malay in Brunei, Malaysia, and Indonesia. Despite this, Section 9 of the National Language Act of 1963/67 (Act 32) of Malaysia states that such is with the condition that the usage of the Rumi script is not restricting the usage of the Jawi script as the writing script of the national language. In Brunei, it is the official and formal writing script used for writing Malay alongside the Rumi script.

Today, although the Rumi script is more prevalent among the users due to modernisation and because it is used for formal purposes, the Jawi script is still widely used in daily writing, certain publications, road signages, and even in official government press releases for certain conditions or areas.
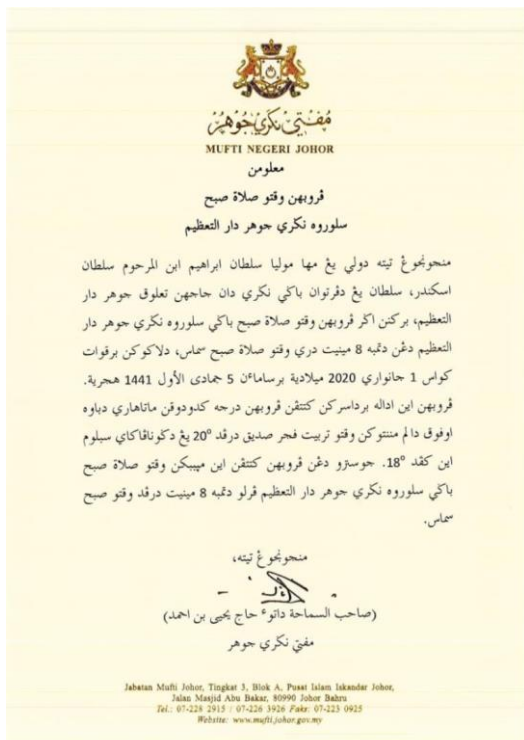

Figure 13: Official Johor (Malaysian state) Government statement from the State Mufti written in Jawi.

Figure 14: The Raja Permaisuri Agong (Queen) of Malaysia tweeting in Jawi as a Chinese Twitter user responds also in Jawi.

Figure 15: Street signage in Kelantan, Malaysia using the Jawi script.

Jawi is used mainly by Malays predominantly in Malaysia and Brunei, but also non-Malays within the area. It is also used in certain places or communities within Southern Thailand, Indonesia, and Singapore. It is estimated based on the Malay population in Malaysia and the population of Brunei that the amount of Jawi users is <u>at least 23 million people</u>, not including the non-Malays and those from outside Malaysia and Brunei.

IV. References

جوهر انداڠ٢ توبهكراجاٴن *[Johore State Constitution]*. (1895). Johor Bahru (Johore): Government of Johore.

ملايو تانه كاٴن كمرديقا ڤمشهوران *[Malayan Declaration of Independence]*. (1957).

(1988). *Daftar Ejaan Rumi-Jawi [Rumi-Jawi Spelling Directory]*. Kuala Lumpur (Malaysia): Dewan Bahasa dan Pustaka. ISBN 9789836200273.

Ismail bin Dahaman & Manshoor bin Haji Ahmad. (2001). *Daftar Kata Bahasa Melayu: Rumi-Sebutan-Jawi (Jilid 1) [Malay Language Word Directory: Rumi-Pronunciation-Jawi (Book 1)]*. Kuala Lumpur (Malaysia): Dewan Bahasa dan Pustaka. ISBN 9789836246721.

Aznan Haji Ahmad, Haji. (2015). *Kaedah Pembelajaran Jawi : Peringkat Asas [Method of Learning Jawi : Basic Level]*. Kuala Lumpur (Malaysia): Perpustakaan Negara Malaysia. ISBN 9789679312607.

Haris bin Kasim, Haji. (2019). بوکو تيکس جاٴيس جاوي تاهون 4 [*Year 4 Jawi JAIS Text Book]*. Seri Kembangan (Malaysia): Jabatan Agama Islam Selangor. ISBN 9789671650004.

Zakkina Wati Ahmad Tarmizi. (2019-07-01). "برسندر باكت" [Dependent on Talent]. *Utusan Malaysia*.

Akta 32 Akta Bahasa Kebangsaan 1963/67 Malaysia [Act 32 National Language Act 1963/67 of Malaysia]

V. Submitters

A work committee known as the Three-Quarter High Hamza Proposal Committee (Jawatankuasa Usul Hamzah Tiga Suku / جاوتنكواس اصول حمزة تيڬ سوکو) was formed between researchers and experts to motion this proposal. The committee members are:

- Ahmad Ali A. Karim / احمد علي عبد الكريم
  Author, student, & columnist on local newspaper *Utusan Malaysia*.
  Email: ahmadalikarimomar@hotmail.com
- (Tun) Suzana (Tun) Othman / (تون) سوزانا (تون) عثمان
  Author, historical researcher on Malay & Islamic civilisation.
  Email: akademiwatan2021@gmail.com
- Dr. Hasanuddin Yusof / دوكتور فلسفه حسن الدين يوسف
  Author, historical researcher on Malay & Islamic civilisation.
  Email: mohdhasanuddin@yahoo.com

Appendix A

Unicode Technical Committee,
The Unicode Consortium
P.O. Box 391476
Mountain View, CA 94039-1476
U.S.A.

10th January 2022

Dear members of the Unicode Technical Committee,

Proposal to Encode ARABIC LETTER THREE QUARTER HIGH HAMZA

We are writing in referring to the proposal document titled "Proposal to Encode ARABIC LETTER THREE QUARTER HIGH HAMZA for Jawi" dated 11th January 2022, proposing the addition of the character Three-quarter High Hamza or *Hamzah Tiga Suku* (حمزة تيݢ سوكو) which is used in daily Jawi writing.

As an organisation which publishes Jawi publications, provides training and education regarding the Jawi writing system and researches on the usage and history of Jawi, we can confirm that the letter is indeed in use by Jawi users.

We have also attached to this document additional samples of the usage of the character in question. Please refer to the attachment

Ahmad, Suzana & Hasanuddin. (2022). *Proposal to Encode ARABIC LETTER THREE-QUARTER HIGH HAMZA for Jawi*.

Your consideration is highly appreciated.

Thank you.

Regards,

Muhammad Syukri Rosli
Chairman
Akademi Jawi Malaysia

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html **.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode ARABIC LETTER THREE QUARTER HIGH HAMZA for Jawi*
2. Requester's name: *Ahmad Ali A. Karim, (Tun) Suzana (Tun) Othman, Dr. Hasanuddin Yusof*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *11 January 2022*
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-B*
2. Number of characters in proposal: *1*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary **X**   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?
   ME'DIDIK Malaysia
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   ME'DIDIK Malaysia (email: medidik.malaysia@gmail.com)
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*
   Right-to-left directionality

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.

---

[1] Form number: N4502-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, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	*No*
If YES explain	
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.)?	*Yes*
If YES, with whom?	*Akademi Jawi Malaysia*
If YES, available relevant documents:	*Appendix*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	*Yes*
Reference:	*Proposal document*
4. The context of use for the proposed characters (type of use; common or rare)	*Common*
Reference:	*Proposal document*
5. Are the proposed characters in current use by the user community?	*Yes*
If YES, where? Reference:	*Proposal document*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	*Yes*
If YES, is a rationale provided?	*Yes*
If YES, reference:	*Contemporary characters in current use; see proposal document*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	*Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
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?	*Yes*
If YES, is a rationale for its inclusion provided?	*Yes*
If YES, reference:	*Proposal document*
11. Does the proposal include use of combining characters and/or use of composite sequences?	*No*
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:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	*Yes*
If YES, describe in detail (include attachment if necessary)	
Characters have right-to-left (AL) directionality.	
13. Does the proposal contain any Ideographic compatibility characters?	*No*
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	