ISO/IEC JTC1/SC2/WG2 N3081
L2/06-120
2006-04-09

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal for addition of Vai characters to the UCS**
Source: **Michael Everson (Evertype), Mohamed Nyei (New York University), Charles Riley (Yale University), Tombekai Sherman (Chair of the Vai New Testament Translation Committee)**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2006-04-09**

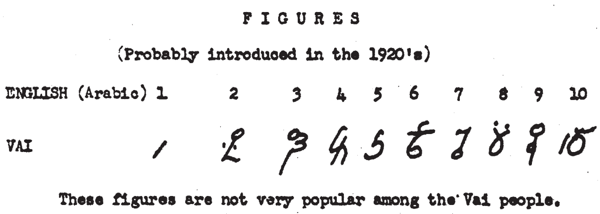
Request. This document asks for the addition of four nasal vowels and ten digits to the Vai character repertoire. Vai is a new script still under ballot, and the proposal here augments the character set which was proposed in N2948R. If this proposal is adopted, the following 14 characters would exist:

ꔁ	A501	VAI SYLLABLE EEN
ꔥ	A525	VAI SYLLABLE IN
ꕲ	A572	VAI SYLLABLE OON
ꖖ	A596	VAI SYLLABLE UN
꘠	A620	VAI DIGIT ZERO
꘡	A621	VAI DIGIT ONE
꘢	A622	VAI DIGIT TWO
꘣	A623	VAI DIGIT THREE
꘤	A624	VAI DIGIT FOUR
꘥	A625	VAI DIGIT FIVE
꘦	A626	VAI DIGIT SIX
꘧	A627	VAI DIGIT SEVEN
꘨	A628	VAI DIGIT EIGHT
꘩	A629	VAI DIGIT NINE

with the following properties:

```
A501;VAI SYLLABLE EEN;Lo;0;L;;;;;N;;;;;
A525;VAI SYLLABLE IN;Lo;0;L;;;;;N;;;;;
A572;VAI SYLLABLE OON;Lo;0;L;;;;;N;;;;;
A596;VAI SYLLABLE UN;Lo;0;L;;;;;N;;;;;
A620;VAI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
A621;VAI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
A622;VAI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
A623;VAI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
A624;VAI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
A625;VAI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
A626;VAI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
A627;VAI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
A628;VAI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
A629;VAI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

Digits. These were originally attested in S. Jangaba M. Johnson's *Traditional History, Customary Laws, Mores, Folkways, and Legends of the Vai Tribe*. (Monrovia: Department of the Interior):



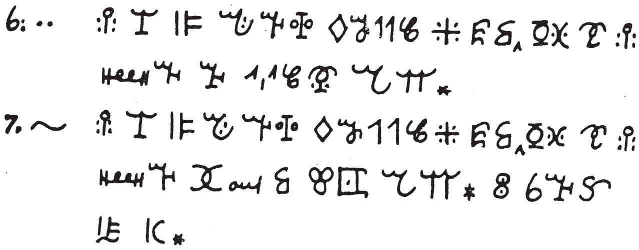
In N2948R, it was suggested that because the Vai digits were not popular and have since been replaced by European digits, their encoding was not urgent. Africanist Konrad Tuchscherer (Department of History, St. John's University, New York) responded to this, saying:

> The Vai numerals are attested to in sources earlier than Johnson. True, they were unpopular and I have no evidence to suggest that they were ever in widespread use. If you are encoding logograms that are only known from a single source, like the Book of Rora, then maybe these numerals should be encoded too.

Konrad is currently in the Cameroon working on Bamum, and is unable to provide further examples at present, but there seems to be little reason to dismiss Johnson's evidence as insufficient, and no reason not to finish the encoding of this facet of the Vai script at this time, while it is under ballot.

Nasal vowels. The addition of these four characters will complete the set of initial vowels, simplifying common tasks such as the representation of foreign vowel sounds, and in particular tasks related to keyboard input, such as typing tutorials and on-screen visual keyboard user feedback.

The Standard Vai Script published in Monrovia in 1962 makes explicit provision for the extension of the character set by the use of various marks. These are not "combining marks" in the UCS sense, but they are schemes for extending the character set.



6. ..When this is placed under a character, one should utter a nasal sound.
7. ~When this is placed under a character, one should say it with a trilling tongue; like "[illegible]" ra.

Neither of these marks are implemented by combining characters per se. The r-series [illegible] is simply the l-series [illegible] with the "tilde" placed under the glyph; these are being encoded as unique characters in the UCS. Nasal characters are not really implemented by putting double dots *under* a character, despite the description quoted above. Instead a variety of mechanisms are used, about half using pairs of dots and half using horizontal or vertical bars incorporated into the glyphs. The four nasal vowels proposed here fill in gaps in the syllabic paradigm in conformance with the provision for such extension in the 1962 standard for Vai.

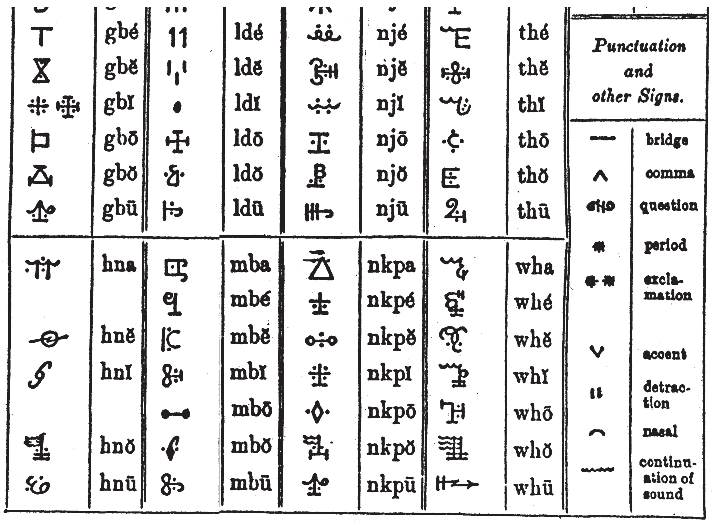
Reference: Zuke Kandakai, S. Jangaba Johnson, Bai Tamia Moore, & Fatima Massaquoi Fahnbulleh. 1962. [illegible] (*Vai kpolo saikilamaa mɛ*) = *The Standard Vai script*. Monrovia: University of Liberia African Studies Program.

Detailed discussion. The table on page 5 of this document shows the Vai character repertoire. Highlighted in green in that table are the four characters proposed here; highlighted in red are 40 characters created by Momolu Massaquoi in 1911 as additions to the traditional syllabary. The additions occur in two classes: series modifications and "gap fillers". The series modifications are simple: Massaquoi used additional dots or lines to extend the /t-/ series to a /θ-/ series, the /d-/ series to a /ð-/ series, the /s-/ series to a /ʃ-/ series, and the /z-/ series to a /ʒ-/ series. Of these, only the /ʃ-/ series has had any popularity, and that is fairly marginal. (Massaquoi made two other series innovations; his series for /ʍ/ was taken over in modern times to represent the nasal /w̃-/ series, and his series for /r/ has been adopted as part of the Standard syllabary.)

The "gap filler" modifications are quite interesting. Massaquoi filled out the nasal /w̃-/ series with forms for ꔄ wẽ, ꕵ wõ, ꖓ wũ, and ꖟ wɔ̃ (modifying ꔃ we, ꕴ wo, ꖒ wu, and ꖞ wɔ); he filled out the /mgb-/ series with forms for ꔣ mgbĩ and ꖌ mgbũ (modifying ꔛ gbe and ꖇ kpu); he filled out the nasal /m̃-/ series with ꔇ mẽ and ꕹ mõ (modifying ꔆ ɓe and ꕸ mɓo), the nasal /ñ-/ series with ꔌ nẽ and ꕻ nõ (modifying ꔋ nɗe and ꕺ nɗo), and the nasal /ɲ̃-/ series with ꔘ ɲẽ and ꖂ ɲõ (modifying ꔗ dʒe and ꖁ dʒo). None of these syllables actually occurs in Vai; they simply fill in "gaps" in the paradigm as Massaquoi conceived it. They would be useful for writing foreign words. In preparing a keyboard input method for Vai, however, we have encountered four gaps which, if filled, would be extremely useful.

Vai has seven oral vowels (ꔀ e, ꔤ i, ꕉ a, ꕱ o, ꖎ u, ꖙ ɔ, ꗋ ɛ); only three of these (ꕊ ã, ꖚ ɔ̃, ꗌ ɛ̃) have a traditional nasal representation, and it is proposed here to fill these four gaps. This will not only solve a particular implementation problem for Vai; it will make the script more flexible for the representation of foreign initial nasal vowels. Looking the set of oral and nasal syllables, it is easy to see that there are glyph relationships between ꔃ we and ꔄ wẽ; between ꔤ i, ꔦ hi, ꔨ wi, and ꔩ wĩ; between ꕊ ã and ꕋ ŋã; between ꕌ ha and ꕍ hã, between ꕱ o, ꕳ ho, ꕴ wo, and ꕵ wõ; between ꖎ u, ꖐ hu, ꖒ wu, and ꖓ wũ, between ꖙ ɔ, ꖜ hɔ, ꖞ wɔ, and ꖟ wɔ̃; between ꖚ ɔ̃ and ꖛ ŋɔ̃; and between ꗋ ɛ, ꗎ hɛ, and ꗏ hɛ̃. It can easily be seen that paired dots and horizontal bars are typically used as a device to distinguish a nasal vowel from an oral one.

Massaquoi's own chart gives a sort of inverted breve as a mark of nasalization (see the bottom right of the figure below) but there is no indication as to how it would be used, if it was intended to be productive.



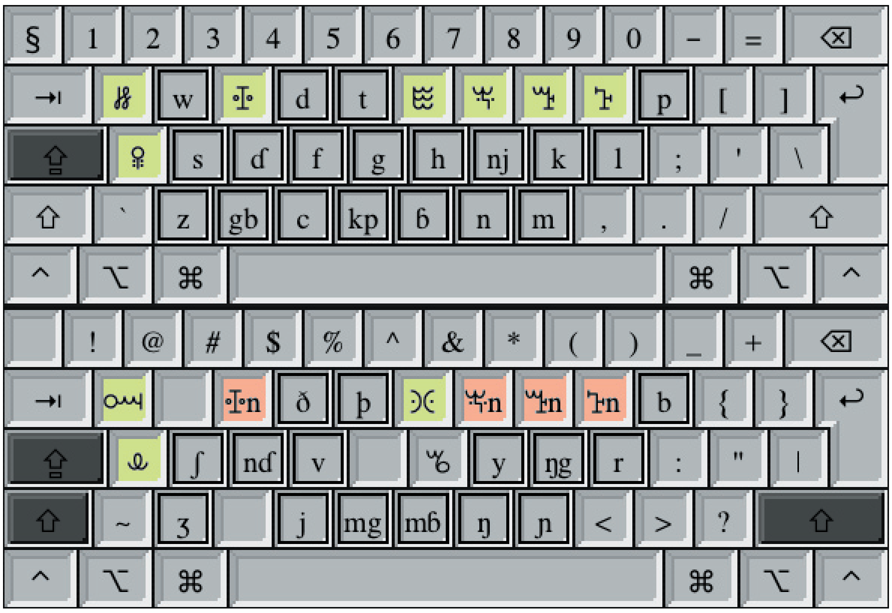
As noted above, the "nasalization dots" in the 1962 standard, like the tilde "r" modifier, were never realized as a productive combining character (in UCS terms), so if dots or strokes are to be used to extend the character set it has to be achieved via a unique syllable.

Keyboard design and user interface. Vai's 284 characters do not fit onto a standard 48-key keyboard; even plain, shift, alt, and alt-shift only yield 192 positions. A deadkey keyboard based on consonant + rhyme, however, is simple to design and implement. (Apple ships such keyboards for Cherokee and Inuktitut in Mac OS X.) In such an implementation, each of the consonant keys is a deadkey (conventionally displayed with one or more Latin consonants), and each of the vowel keys (displayed with the Vai vowels) completes the deadkey and inserts a Vai character into the text, so "w" + ꗡ e = ꕃ we. The oral vowels are on the unshifted keys: ꗡ e is on the E key, ꔤ i is on the I key, ꕉ a is on the A key, ꗣ o is on the O key, ꖕ u is on the U key, ꖺ ɔ is on the Y key, and ꔀ ɛ is on the Q key. The nasal vowels are on the shifted keys: ꕊ ã is on the shift-A key, ꖻ ɔ̃ is on the shift-Y key, ꔁ ɛ̃ is on the shift-Q key. For the gap of four nasal vowels, needed as keystroke sequences to complete a number of encoded syllables, there is nothing adequate to display to the user on these keys.

One possibility for working around this would be to add the Latin letter "n" to each of these four keys to indicate the nasality of the vowel, as we did in the alpha version of the keyboard drivers. Another possibility would be to use a combining diaeresis below for ꗡ̤ ẽ, ꔤ̤ ĩ, ꗣ̤ õ, and ꖕ̤ ũ. We reject both of these as unsatisfactory, the first for a host of reasons, not the least of which is aesthetic, and the second because no other nasal syllable uses dots in such a way. We favour an alternate solution that makes use of intuitive modifications of existing glyphs.

The next question which we had to answer, then, taking the provisions of the 1962 standard into accounts, was "What should these characters look like?" For two of them, the answer is quite straightforward. On the analogy of ꗇ wo, ꗈ wõ, we should have ꔤ i, ꔤ ĩ, and ꗣ o, ꗣ õ. For the other two, there should be some discussion. The similarities between ꖕ u, ꖘ hu, ꖙ wu, and ꖚ wũ are quite suggestive: ꖕ ũ would be a strong candidate, on the analogy of ꖙ/ꖚ wu/wũ. Another option would be to take the ꖒ hnu given in Massaquoi in the position for modern ꖗ hũ; it has the glyph for ꖒ ng with two dots. For ꗡ e there are also two choices. The first is ꗡ ẽ, on the analogy of ꕃ/ꕄ we/wẽ. The second would be ꗡ, a glyph variant of ꗡ given by Massaquoi which we have not seen elsewhere.

To sum up: following analogies based on other existing glyph transformations, we are certain that ꗡ/ꗡ e/ẽ (cf ꕃ/ꕄ we/wẽ), ꔤ/ꔤ i/ĩ (cf ꗇ/ꗈ wo/wõ), ꗣ/ꗣ o/õ (cf ꗇ/ꗈ wo/wõ), and ꖕ/ꖕ u/ũ (cf ꖙ/ꖚ wu/wũ) are the correct forms to use to fill these gaps.



In the sample here our rather unsatisfactory temporary workaround is given; the Latin letter "n" is given following the oral vowel.

When these characters are added to the UCS, the interface difficulty we have identified will no longer be a problem, as can be seen in the sample below:

The fact that there has been very few Vai computer keyboards before now lends weight to the argument that this innovation is needed. The SIL keyboard "Vai Script version 0.1 beta" does not attempt to implement the full Vai character set as implemented in the UCS, and so was not designed with this sort of deadkey arrangement. The arrangement here has already been implemented for Mac OS and a Tavultesoft Keyman implementation is in hand.


-e -i -a -o -u -ɔ -ɛ
– -
˜ -N
ŋ˜ NG-N
h– H-
h˜ H-N
w– W-
w˜ H-N
p– P-
b– BH-
ɓ– B-
mɓ– MB-
kp– KP-
kp˜ KP-N
mgb– MGB-
gb– GB-
gb˜ GB-N
f– F-
v– V-
t– T-
þ– TH-
d– DH-
ð– DHH-
l– L-
r– R-
ɗ– D-
nɗ– ND-
s– S-
ʃ– SH-
z– Z-
ʒ– ZH-
c– C-
j– J-
nj– NJ-
y– Y-
k– K-
k˜ K-N
ŋg– NGG-
ŋg˜ NGG-N
g– G-
g˜ G-N
m˜ M-
n˜ N-
ɲ˜ NY-
-EE -I -A -OO -U -O -E
ɷ = * ʊ
ŋ : , . ?
NG