Proposal to Encode Phonetic Symbols with Middle Tilde in the UCS

Date: 2003-09-30

Author: Peter Constable, Microsoft

Address: One Microsoft Way
             Redmond, WA 98052
             USA

Tel: +1 425 722 1867

Email: petercon@microsoft.com

L2/03-174R2

A.   Administrative

1.	Title	Proposal to Encode Phonetic Symbols with Middle Tilde in the UCS
2.	Requester's name	INCITS/L2, UTC, SIL International (contact: Peter Constable)
3.	Requester type	Expert contribution
4.	Submission date	2003-09-30
5.	Requester's reference	L2/03-174r
6a.	Completion	This is a complete proposal
6b.	More information to be provided?	Only as required for clarification.

B.   Technical—General

1a.	New Script? Name?	No
1b.	Addition of characters to existing block? Name?	Yes — Phonetic Extensions
2.	Number of characters in proposal	11
3.	Proposed category	A
4.	Proposed level of implementation and rationale	1 (no combining marks or jamo)
5a.	Character names included in proposal?	Yes
5b.	Character names in accordance with guidelines?	Yes
5c.	Character shapes reviewable?	Yes
6a.	Who will provide computerized font?	SIL International
6b.	Font currently available?	Yes
6c.	Font format?	TrueType
7a.	Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes

7b.	Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8.	Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included (see section E).

C.   Technical—Justification

1.	Has this proposal for addition of character(s) been submitted before?	No
2a.	Has contact been made to members of the user community?	Yes
2b.	With whom?	Linguists
3.	Information on the user community for the proposed characters is included?	Linguists
4.	The context of use for the proposed characters	Linguistics text books, linguistic descriptions (books, journal publications, etc.); dictionaries.
5.	Are the proposed characters in current use by the user community?	Yes
6a.	Must the proposed characters be entirely in the BMP?	Preferably
6b.	Rationale?	If possible, should be kept with other phonetic symbols in the BMP.
7.	Should the proposed characters be kept together in a contiguous range?	Preferably
8a.	Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	Possibly (see discussion in section F below)
8b.	Rationale for inclusion?	See discussion in section F below.
9a.	Can any of the proposed characters be considered to be similar (in appearance or function) to an existing character?	No
9b.	Rationale for inclusion?	n/a
10.	Does the proposal include the use of combining characters and/or use of composite sequences?	No.
11.	Does the proposal contain characters with any special properties?	No.

D. SC2/WG2 Administrative

1. Relevant SC2/WG2 document numbers
2. Status (list of meeting number and corresponding action or disposition)
3. Additional contact to user communities, liaison organizations, etc.
4. Assigned category and assigned priority/time frame

Other comments

E. Proposed Characters

A code chart and list of character names are shown on a new page.

E.1   Code Chart

	1D6	1D7
0		n̴
1		p̴
2		r̴
3		f̴
4		s̴
5		t̴
6		z̴
7		
8		
9		
A		
B		
C	b̴	
D	d̴	
E	f̴	
F	m̴	

E.2   Character Names

1D6C   LATIN SMALL LETTER B WITH MIDDLE TILDE
1D6D   LATIN SMALL LETTER D WITH MIDDLE TILDE
1D6E   LATIN SMALL LETTER F WITH MIDDLE TILDE
1D6F   LATIN SMALL LETTER M WITH MIDDLE TILDE
1D70   LATIN SMALL LETTER N WITH MIDDLE TILDE
1D71   LATIN SMALL LETTER P WITH MIDDLE TILDE
1D72   LATIN SMALL LETTER R WITH MIDDLE TILDE
1D73   LATIN SMALL LETTER R WITH FISHHOOK AND MIDDLE TILDE
1D74   LATIN SMALL LETTER S WITH MIDDLE TILDE
1D75   LATIN SMALL LETTER T WITH MIDDLE TILDE
1D76   LATIN SMALL LETTER Z WITH MIDDLE TILDE

E.3 Unicode Character Properties

All of these characters should have a general category of Ll; no case mappings for these characters are proposed. Other properties should match those of similar characters (e.g. U+026B LATIN SMALL LETTER L WITH MIDDLE TILDE).

F. Other Information

F.1 Rationale

In phonetic transcription, consonant symbols with overlaid tilde are generally used to represent consonants with velarized or pharyngealized articulation. In the Americanist phonetic tradition, overlaid tilde has been used only for simultaneous velar articulation. In the IPA tradition, overlaid tilde has been used for either velarization or pharyngealization. There is a newer IPA recommendation for transcribing velarization and pharyngealization in distinct ways using modifier letters:
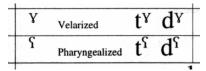

Figure 1. Distinct IPA notation for velarized and pharyngealized articulation (IPA 1999, p. 204).

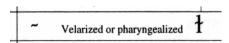Overlaid tilde is still approved IPA usage, however:

Figure 2. Overlaid tilde: current IPA usage (IPA 1999, p. 204).

Figure 3. Overlaid tilde: current IPA usage (IPA 1999, p. 173).

It is probably the case that the former notation is becoming the more common, both because it is more explicit, and because it poses fewer problems typographically. Nevertheless, symbols with overlaid tilde are still encountered in linguistics reference books and phonetics textbooks, and also in language descriptions:

Figure 4. From IPA (1949), p. 14.

Figure 5. From Heffner (1950), p. 134.

Production:
- Concentrate on production of the velarized lateral [ɫ]. Say [ɫ] with the tongue tip up.
- Produce syllables with [ɫ] syllable-initial: [ɫa] [ɫe]...
- Produce other voiced continuants alternating normal and velarized articulations: n - n̴   m - m̴   z - z̴.
- Contrast: za   z̴a
           na   n̴a
           ma   m̴a
- Produce voiceless continuants alternating normal and velarized articulations: s - s̴   f - f̴
- Contrast: sa   s̴a
           fa   f̴a

Figure 6. From Floyd (1981), p. 102.

We can now recapitulate the principal types of secondary articulation:

labialization, e.g. [tʷ] [dʷ] [kʷ] [sʷ] [zʷ] [lʷ] etc.
palatalization, e.g. [tʲ] [dʲ] [kʲ] [sʲ] [zʲ] [lʲ] etc.
velarization, e.g. [t̴] [d̴] [s̴] [z̴] [ɫ] etc.
pharyngalization, e.g. [t̴] or [tˤ], [s̴] or [sˤ] etc.
nasalization, e.g. [ĩ] [ɛ̃] [ã] [õ] etc.

Figure 7. From Catford (1988), p. 110.

Scottish Gaelic (Skye)

[kʲaʃiç] 'of an old woman'

[kʰõːl̴ʑfi ʃ] 'accompanying him/along with him'

[ɲiᵊd̥ θrʌxˑkˑ] '(family) washing'

Scottish Gaelic (Islay)

[kʰʷɵɥˑxʷt̴ʷfʲ] 'cat'

[n̴dᶻɒʔə̣] 'the day'

Figure 8. From Laver (1994), p. 559.

Figure 9. Overlaid tilde used in African-language descriptions (Mann and Dalby 1987, p. 218).

Figure 10. Overlaid tilde used in African-language descriptions (Tucker 1971, p. 647).

While it is clear that characters with overlaid tilde are used, what is more difficult to assess is the inventory of such characters needed. After "ɫ" (for the "dark l" of English), these characters appear to be most frequently used in relation to Arabic (for pharyngealized "emphatic" consonants) and other Semitic languages; and following that, for Russian (some describe Russian palatalized consonants as contrasting with velarized counterparts). Velarization and pharyngealization do occur in other languages, yet not all languages have been described, and not all existing descriptions can be examined. Therefore, while the possibilities for base characters that may require a modifier to indicate these secondary articulations are limited to at most the phonetic symbols for consonants, it is not clear just how many of those are actually required.

Because velarization is said to contrast with palatalization in Russian, and because Russian has a large inventory of palatalized consonants, it may constitute a language requiring one of the larger inventories of consonants with velarized modification. Such an inventory for Russian is described in IPA 1949:

Russian

[Broad transcription except that the retracted variety of i is shown by ɨ. Digraphic representation of affricates and palatalised consonants. tj, nj, etc., stand for ţ, ŋ, etc. snj, tnj, etc., stand for ʂŋ, ţŋ, j being used as a sign denoting palatalisation and being held to affect all preceding palatalisable consonants in a sequence unless a hyphen is inserted (the hyphen indicating that consonants preceding it are not palatalised, e.g. 'v-rjemja). lj is, however, written in full before another palatalised consonant on account of the frequent occurrence of velarised l in this position. p, b, m, f, v, t, d, n, l, r, s, z, without j are velarised, l particularly so (= ł). ts, tʃ are affricates, tʃ being rather palatal. Sequences of t and s, t

Figure 11. Inventory of Russian velarized phones (IPA 1949, p. 28).

This corresponds almost exactly with the inventory of overlaid-tilde characters I have encountered in the linguistics literature I have surveyed, and that are evidenced in the examples shown above, as summarized in Table 1:

Base Character	Samples
p	Figure 5
b	Figure 5
m	Figure 6
f	Figure 6
v	(not encountered)
t	Figure 4, Figure 5, Figure 7, Figure 9, Figure 10
d	Figure 3, Figure 4, Figure 5, Figure 7, Figure 9
n	Figure 3, Figure 6, Figure 8
l	Figure 2, Figure 6, Figure 7, Figure 8, Figure 9
r	Figure 9
ɾ	Figure 8
s	Figure 4, Figure 6, Figure 7, Figure 9, Figure 10
z	Figure 4, Figure 6, Figure 7, Figure 9

Table 1. Base characters and samples in which they appear with overlaid tilde[1]

Of the base characters listed in IPA 1949 in relation to Russian velarized consonants (Figure 11), all are attested in the literature I surveyed except "v".

I have encountered descriptions of Hebrew and Arabic that cite pharyngealization of other consonants — ð, k, and ʔ — but these descriptions did not use the overlaid-tilde notation. It is my impression that these consonants are not mentioned as often in Arabic and Hebrew descriptions in relation to pharyngealization as are others, particularly t, d and s.

It appears, then, that those characters with overlaid tilde that are illustrated in the samples above are probably the inventory that are most likely to be used by linguists. Therefore, I am limiting the proposed inventory to just those for which attestation can be demonstrated at this time. It is possible that a need for others may come to light at some point in the future, though that is not anticipated at present.

---

[1] Note: Some of the sources from which these samples were taken included overlaid tildes with other base characters than what the given sample reflects. The overall inventory encountered is that shown in the table, however.

F.2   Representation as sequences with U+0334

Question 8a of section C above asks whether these characters can be considered presentation forms of existing characters or character sequences. They could possibly be viewed as sequences involving U+0334 COMBINING TILDE OVERLAY, but I suggest that this would be inappropriate and is irrelevant. While combining marks in general are assumed to be applicable to arbitrary characters in a generative manner, allowing dynamic representation of text elements such as *Latin small a with bridge below*, there are certain combining marks for which this is not appropriate. I suggest that U+0334 COMBINING TILDE OVERLAY is among them.

Linguistically, there simply are only certain base characters that have any likelihood of being modified with a tilde overlay. There are far more characters, even when limiting discussion to phonetic symbols, for which there would be little motivation to add a tilde overlay. For instance, it would be silly (for typographic reasons as well as linguistic) to encode a character sequence < U+026B LATIN SMALL LETTER L WITH MIDDLE TILDE, U+0334 COMBINING TILDE OVERLAY >. In practice, there is a very limited inventory of characters that are used with tilde-overlay modification, because there are very few speech sounds that can have velarized or pharyngealized secondary articulation.

Also, whereas it is feasible to create font/rendering implementations that can productively display sequences involving arbitrary base characters followed by a combining mark such as U+0300 COMBINING GRAVE ACCENT using mechanisms such as glyph attachment points, this is not as readily feasible for U+0334 COMBINING TILDE OVERLAY: the way in which a base character is modified using a tilde overlay is dependent on the particular base character involved. It may be necessary to have different attachment positions, tildes of varying sizes and shapes, and possibly even adjustments to the design of the outline for the base character. Such issues cannot be dealt with by a font developer in a generalized manner but must be handled on an individual basis.

Thus, in terms of usage requirements and the realities of implementation, dynamic composition using U+0334 COMBINING TILDE OVERLAY is not a good choice, and should be avoided.

Note that this view is corroborated by existing characters in Unicode itself in that the characters U+026B LATIN SMALL LETTER L WITH MIDDLE TILDE does not have a decomposition. The combining mark U+0334 COMBINING TILDE OVERLAY is not currently used in any decomposition.

Therefore, since there are good reasons why productive use of U+0334 COMBINING TILDE OVERLAY is not recommended, and insofar as existing characters with tilde overlay are not considered presentation forms of existing sequences, it is suggested that the characters proposed here are likewise not to be considered presentation forms of existing sequences.

G.   References

Catford, J.C. 1988. *A practical introduction to phonetics.* Oxford: Clarendon Press.

Floyd, Rick. 1981. *Manual for articulatory phonetics.* Dallas: Summer Institute of Linguistics

Heffner, R-M. S. 1950. *General phonetics.* Madison, WI: University of Wisconsin Press.

International Phonetic Association (L'association Phonétique Internationale). 1946. "parti administratiːv." *Le maître phonétique* 85:1, p. 16–7.

——. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Laver, John. 1994. *Principles of phonetics.* (*Cambridge textbooks in linguistics.*) Cambridge: Cambridge University Press.

Mann, Michael, and David Dalby. 1987. *A thesaurus of African languages: a classified and annotated inventory of the spoken languages of Africa with an appendix on their written representation.* London: Hans Zell Publishers.

Tucker, A.N. 1971. "Orthographic systems and conventions in Sub-Saharan Africa." *Current trends in linguistics, volume 7: Linguistics in Sub-Saharan Africa*, ed. by Thomas A. Sebeok, 618–53. The Hague: Mouton.