

Unicode request for Greek letters with palatal hook

Kirk Miller, kirkmiller, gmail.com

2024 June 04

While reviewing [L2/24-050 Letters with Palatal Hook](#), the SAH recommended that the Greek-derived phonetic symbols <β̣ θ̣ χ̣> be deferred pending a decision as to whether they should be encoded as Latin or Greek script, as that decision would impact their properties and naming. Subsequent review by SAH decided that they should be encoded as Greek script, though added to a Latin extended block as other Greek-derived phonetic characters are.

There is some discussion among phoneticians over whether the IPA letters beta, theta and chi should be Latin or Greek in form, but IPA letters with a palatal hook were retired with the Kiel convention in 1989, and it's unlikely that such a distinction will need to be made for them. James Kirby contacted the Alphabets, Charts and Fonts committee of the IPA, and their response of 2 May was, "There is no protest against having Latin-block versions of these three letters, so we can support Kirk's proposal for Latin-block versions of the three base characters."

Characters

IPA letters with palatal hook (subheading continued)

- β̣ 1DF3B GREEK SMALL LETTER BETA WITH PALATAL HOOK. Figure 1 *ff*.
- θ̣ 1DF3C GREEK SMALL LETTER THETA WITH PALATAL HOOK. Figure 3.
- χ̣ 1DF3D GREEK SMALL LETTER CHI WITH PALATAL HOOK. Figure 4 *ff*.

Properties

1DF3B;GREEK SMALL LETTER BETA WITH PALATAL HOOK;Ll;0;L;;;;N;;;;;
 1DF3C;GREEK SMALL LETTER THETA WITH PALATAL HOOK;Ll;0;L;;;;N;;;;;
 1DF3D;GREEK SMALL LETTER CHI WITH PALATAL HOOK;Ll;0;L;;;;N;;;;;

The script properties of these characters is Greek, similar to U+1D5D MODIFIER LETTER SMALL BETA (β̣), U+1DBF MODIFIER LETTER SMALL THETA (θ̣), U+1D61 MODIFIER LETTER SMALL CHI (χ̣), U+1D66 GREEK SUBSCRIPT SMALL LETTER BETA (β̵), and U+1D6A GREEK SUBSCRIPT SMALL LETTER CHI (χ̵).

DoNotEmit data

For historical reasons, IPA letters with palatal hook are not canonically equivalent to the letter plus the palatal hook diacritic. They should thus be listed in DoNotEmit.txt.

03B2 0321; 1DF3B; Precomposed_Form # GREEK SMALL LETTER BETA, COMBINING PALATALIZED HOOK BELOW; GREEK SMALL LETTER BETA WITH PALATAL HOOK

03B8 0321; 1DF3C; Precomposed_Form # GREEK SMALL LETTER THETA, COMBINING PALATALIZED HOOK BELOW; GREEK SMALL LETTER THETA WITH PALATAL HOOK

03C7 0321; 1DF3D; Precomposed_Form # GREEK SMALL LETTER CHI, COMBINING PALATALIZED HOOK BELOW; GREEK SMALL LETTER CHI WITH PALATAL HOOK

References

- John Catford (1970) *abk_word-list_1970_02.jpg*. Transcription of *abk_word-list_1970_02.wav*.
Accessed from: archive.phonetics.ucla.edu/Language/ABK/abk_record_details.html.
- Eugénie Henderson (1949) A Phonetic Study of Western Ossetic (Digoron). *Bulletin of the School of Oriental and African Studies*, 13(1), 36–79.
- Michael Job (1981) Grammatischer Wechsel im Lesgischen. *Bedi Kartlisa* 39, 279–296.
- Jeffrey Kallen (2013) *Irish English*. Volume 2: *The Republic of Ireland*. Walter de Gruyter.
- John Kelly & John Local (1989) *Doing Phonology*. Manchester University Press.
- Kretzschmar *et al.* (1993/1994) *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. University of Chicago Press. [LAMSAS is a project of the University of Georgia]
- Katerina Nicolaidis (2024) ‘Unicode support for historical and para-IPA letters.’ Letter submitted to the Unicode Technical Committee, 01 January 2024. [L2/24-049](#).
- Jamin Pelkey (2011) *Dialectology as Dialectic: Interpreting Phula Variation*. De Gruyter Mouton.
- John Stonham (1999) *Aspects of Tsishaath Nootka Phonetics and Phonology*. Lincom Europa.

Chart

Greyed out cells are assigned (medium grey) or proposed elsewhere (light grey).

Latin Extended-G

1DF00

1DFFF

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ƒŋ	ɣ	ɔ̥	ɟ												
1	ɸ	ɮ	ɔ̥	ɟ												
2	ɔ̥	ɔ̥	ɮ	ɮ												
3	ɣ	ɮ	ɮ	ɮ												
4	ɣ	ɮ	ɮ	ɮ												
5	ɮ	ɮ	ɮ	ɮ												ɮ
6	ɮ	ɮ	ɮ	ɮ												ɮ
7	ɮ	ɮ	ɮ	ɮ												ɮ
8	ɮ	ɮ	ɮ	ɮ												ɮ
9	ɮ	ɮ	ɮ	ɮ												ɮ
A	ɮ	ɮ	ɮ	ɮ												ɮ
B	ɮ	ɮ	ɮ	ɮ												ɮ
C	ɮ	ɮ	ɮ	ɮ												ɮ
D	ɮ	ɮ	ɮ	ɮ												ɮ
E	ɮ	ɮ	ɮ	ɮ												ɮ
F	ɮ	ɮ	ɮ	ɮ												ɮ

Figures

Most speakers of contemporary Irish English produce sounds of the /p, b, f, v, w, m, m/ group in ways that are not significantly different from general patterns found elsewhere. Traditional dialect, however, presents a different picture. Henry (1957: 59), whose study of traditional dialect in Roscommon is based on fieldwork done in the 1940s, states quite simply that “in good dialect usage the variants of the *f*- and *v*-phonemes are bi-labial fricatives (ϕ , ϕ , β , β ...) as in Ir[ish]”. The use of ϕ extends to /m/ in this data as well. Examples cited by Henry (1957:

Figure 1. Kallen (2013: 49). $[\phi]$ and $[\beta]$ in Irish English.

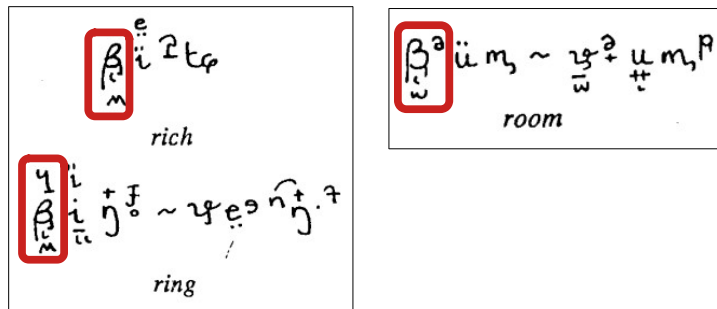


Figure 2. Kelly & Local (1989: 154, 245). $[\betḁ]$ in English in the speech of a 5-year-old. The diacritic $\langle \underset{\circ}{\rangle}$ under the letter is the old IPA diacritic for ‘open.’

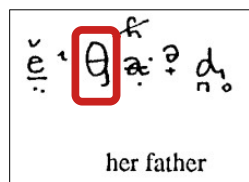


Figure 3. Kelly & Local (1989: 164). $[\thetḁ]$ in Welsh.

167	ἀχῆ	πράξι μυστα	
168	ἀχῆ	κράδι	
169	ἀχῆ	καλὸν	} οὐκ ἔστι ἀλλ' ἔστι ἄλλ' ἔστι ἄλλ' ἔστι
170	ἀχῆ	ἐόν	
1	ἀχῆ	ἡμέ	
2	ἀχῆ	ἡμέ	
3	ἀχῆ	ἡμέ	

Figure 4. Catford (1970). Entries for [χ̥] in Abkhaz.

χ is a voiceless fricative, usually uvular, but articulation may sometimes be pre-uvular, with accompanying action of the front part of the tongue sufficient to justify its description as palatalized.

<i>Absolute</i>		<i>Junction</i>	
χ̥ist	<i>xist</i>	raχ̥es	<i>razes</i>
χ̥ets ^w on	<i>xecon</i>		

Figure 5. Henderson (1949: 51). [χ̥] in Digor Ossetian. The illustrated glyph would not be a good shape for a digital font because it would leave little room for diacritics under the letter.

/k ^h	k'	x	γ	κ	ḡ	κ ^h	κ'	q	ḡ	q ^h	q'	χ	κ	ḡ	ḡ ^h	ḡ'	χ̥	κ	?	h/	
												[χat ^h]	<i>хат бусина</i>				[χtar]	<i>хтар</i>			
												[χ̇at ^h]	<i>хват слива</i>				[χ̇utar]	<i>хутар</i>			

Figure 6. Job (1981: 280–281). [χ̥] in Lezgin. The glyph is rather crude, and the diacritic is placed above letters with a descender like χ.

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 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 std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. Title:	<u>Greek letters with palatal hook</u>
2. Requester's name:	<u>Kirk Miller</u>
3. Requester type (Member body/Liaison/Individual contribution):	<u>individual</u>
4. Submission date:	<u>2024 June 6</u>
5. Requester's reference (if applicable):	<u></u>
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical - General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>	
Proposed name of script:	<u></u>	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>	
Name of the existing block:	<u>Latin Extended-G</u>	
2. Number of characters in proposal:	<u>3</u>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/>	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<u>Kirk Miller</u>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<u>SIL (Gentium Release)</u>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/>	
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)?	<input checked="" type="checkbox"/>	

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 www.unicode.org for such information on other scripts. Also see Unicode Character Database (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?	<i>yes</i>
If YES explain	<i>deferred pending decision on whether they should be identified as Latin or Greek</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?	<i>author is a member of the user community</i>
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	<i>phonetic</i>
Reference:	
5. Are the proposed characters in current use by the user community?	<i>yes</i>
If YES, where? Reference:	<i>see illustrations</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>no</i>
If YES, is a rationale provided?	
If YES, reference:	
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?	<i>no</i>
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?	<i>yes</i>
If YES, is a rationale for its inclusion provided?	<i>yes</i>
If YES, reference:	<i>dynamic generation of characters with U+0321 COMBINING PALATALIZED HOOK BELOW should be avoided; atomic Unicode characters are preferable</i>
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?	<i>no</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>no</i>
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?	<i>no</i>
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	<i>no</i>
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	