

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 to encode two missing modifier letters for Extended IPA****Source: Karl Pentzlin****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2010-04-30****Additions for Extended IPA**

ʰ	U+A7F8	MODIFIER LETTER CAPITAL H WITH STROKE · faucalized → 0126 latin capital letter h with stroke ≈ <super> 0126
œ	U+A7F9	MODIFIER LETTER SMALL LIGATURE OE · labialized: open-rounded → 0153 latin small ligature oe ≈ <super> 0153

Properties:

A7F8;MODIFIER LETTER CAPITAL H WITH STROKE;Lm;0;L;;;;N;;;;;
 A7F9;MODIFIER LETTER SMALL LIGATURE OE;Lm;0;L;;;;N;;;;;

The two proposed letters are the only two yet unencoded true characters, which are listed and given a number in the "Handbook of the International Phonetic Association" [1],

The following figure shows a scan from p. 190 of this handbook:

Superscript double-barred H	[ʰ]	Faucalized	677
Superscript O-E ligature	[œ]	Labialized: Open-rounded	678

Besides these two characters, unencoded entities given a number in [1] only are musical terms ("f", "ff", "p", "pp", "allegro", "lento"), which are written by subscript italic but otherwise common Latin letters (IPA numbers 634...639; [1] p. 189). These entities are special anyway and are not proposed at this time (which could be done e.g. by proposing the constituent letters in a way similar to the "Mathematical Alphanumeric Symbols").

The two proposed letters, like all IPA characters including the superscript letters, are for use in scientific texts and databases, and therefore must not depend on higher level protocols to preserve their intended meaning. Thus, they qualify as plain text elements suitable for encoding.

Having these two missing letters encoded would strengthen the view to the naive user that Unicode systematically gives him the whole IPA, including all superscript letters, in a uniform way, rather than having some single omissions which may seem arbitrary.

Also, having these letters, input methods (yielding characters without higher level protocol information like formatting instructions) for IPA relying on IPA numbers may claim completeness regarding to the whole IPA.

References:

[1] Handbook of the International Phonetic Association (without author), Cambridge 1999, Reprint 2003, ISBN 0 521 63751 1

**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:	<i>Proposal to encode two missing modifier letters for Extended IPA</i>
2. Requester's name:	<i>Karl Pentzlin</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2010-04-30</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input type="checkbox"/> Yes
(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"/> No	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> Yes	
Name of the existing block:	<i>Latin Extended-D</i>	
2. Number of characters in proposal:	<i>2</i>	
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?		
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input type="checkbox"/> Yes	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> Yes	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>The letters are derived from standard Latin letters, thus no specific font is necessary</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):		
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>see text</i>	
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 type="checkbox"/> No	

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	No
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:	No
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Scientific <i>see text</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>see text</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Listed in the IPA handbook</i>
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:	Yes Yes <i>like all IPA and Latin letters encoded so far; contemporary use</i>
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? If YES, is a rationale for its inclusion provided? If YES, reference:	No
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:	No
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:	No
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:	No
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)	No
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	No