

# Unicode request for IPA modifier letters (b), non-pulmonic

Kirk Miller, [kirkmiller@gmail.com](mailto:kirkmiller@gmail.com)

Michael Ashby (President, IPA), [m.ashby@ucl.ac.uk](mailto:m.ashby@ucl.ac.uk)

2020 November 08

This proposal is a continuation of the concurrent request, *Unicode request for IPA modifier-letter support (a), pulmonic*. The request for IPA modifier letters has been split into two sets for the convenience of the UTC, as the supporting argument for them is different. The pulmonic letters are widely attested in the literature, with only a few accidental gaps. The non-pulmonic letters are rare, presumably due to the limited amount of the relevant transcription conventions, such as for speech pathology, used with non-European languages. The argument for inclusion is the stated position of the International Phonetic Association (see letter on page 4), which requests modifier support for all current and implicit IPA letters for implosives and click releases. Support is not needed for ejectives.

Thanks to Deborah Anderson of the Universal Scripts Project for her assistance.

## Background

Superscripting indicates ‘shades of sound’. It is widely used in IPA transcription both phonemically, as for secondary articulation, and phonetically, as for incomplete articulation. It is easy to imagine productive use of non-pulmonic consonants parallel to pulmonic ones. For example, Constable’s illustration of <<sup>9</sup>b> for Igbo [gb̥] in his ‘[Proposal to Encode Additional Phonetic Modifier Letters in the UCS](#)’ could be extended to <<sup>9</sup>ḃ> for those Igbo dialects that realize gb as implosive [gb̥]. Likewise, realizations of click consonants described in the literature as having a lateral click release might be transcribed as e.g. <ɰ̥>, since \*ɰ̥ would suggest a lateral approximant release. Modifier click letters would of course also be used for partial articulation, such as <ɰ̥ḡḡ̥> for a lightly released bilabial nasal click, or the incidental clicks sometimes formed by the overlap of consonants in German and Kinyarwanda.

Beside the letters found in the non-pulmonic consonant block of the IPA chart, we request modifier support for the implicit IPA retroflex implosive and click release, <<sup>ɻ</sup>l̥>. The very large number of IPA ejective consonants, however, do not require additional Unicode support. In most cases, modifier ejectives are acceptably supported with the current Unicode diacritic for ejectives, U+2BC, as in <<sup>k</sup>> for modifier <<sup>k</sup>>. If a distinction needs to be made (say between a superscript ejective [s̥] and ejective /t̥/ with a superscript fricative release [s̥]), then the combining diacritic U+315 can be used to specify that the scope of the apostrophe is the superscript element rather than the segment or cluster as a whole: [s̥] vs [t̥s̥], with any difference in display to be handled by the font. (See Figure 1.) Not all typefaces will make a visible distinction, but the distinction will be encoded and thus preserved.

# Modifier letters, IPA

- 6 10785 MODIFIER LETTER SMALL B WITH HOOK.
- ɖ 1078C MODIFIER LETTER SMALL D WITH HOOK.
- ɗ 1078D MODIFIER LETTER SMALL D WITH HOOK AND TAIL.
- ɟ 10793 MODIFIER LETTER SMALL G WITH HOOK.
- ɠ 10794 MODIFIER LETTER SMALL CAPITAL G WITH HOOK.
- ɥ 10798 MODIFIER LETTER SMALL DOTLESS J WITH STROKE AND HOOK.
- ◌ 107B5 MODIFIER LETTER BILABIAL CLICK.
- ◌ 107B6 MODIFIER LETTER DENTAL CLICK.
- ◌ 107B7 MODIFIER LETTER LATERAL CLICK.
- ◌ 107B8 MODIFIER LETTER ALVEOLAR CLICK.
- ◌ 107B9 MODIFIER LETTER RETROFLEX CLICK WITH RETROFLEX HOOK.

## Comment on existing character

- ◌ 0315 COMBINING COMMA ABOVE RIGHT. Figure 1.

## Chart

Phonetic Extensions Supplement-A was created for the IPA modifier-letter proposal.

- Characters in light pink cells are requested by the International Phonetic Association.
- Characters in light grey cells have been accepted by the UTC for Unicode 14, though we moved them to the Latin-F block after discussion with the SAH.
- Characters in white cells are found in parts (b) and (c) of this proposal.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-F																
U+1078x	AA	:	·	ɥ	B	6	β	ɖ	ɗ	ɟ	ɠ	ɥ	ɟ	ɟ	ɟ	ɟ
U+1079x	ɥ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ
U+107Ax	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ
U+107Bx	v	x	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ	ɟ

# Properties

The **number in red** is a currently proposed codepoint for a base letter. If the UTC places that letter elsewhere, the number in this document will need to be changed to match.

10785;MODIFIER LETTER SMALL B WITH HOOK;Lm;0;L;<super> 0253;;;N;;;;  
1078C;MODIFIER LETTER SMALL D WITH HOOK;Lm;0;L;<super> 0257;;;N;;;;  
1078D;MODIFIER LETTER SMALL D WITH HOOK AND TAIL;Lm;0;L;<super>  
1D91;;;N;;;;  
10793;MODIFIER LETTER SMALL G WITH HOOK;Lm;0;L;<super> 0260;;;N;;;;  
10794;MODIFIER LETTER SMALL CAPITAL G WITH HOOK;Lm;0;L;<super>  
029B;;;N;;;;  
10798;MODIFIER LETTER SMALL DOTLESS J WITH STROKE AND  
HOOK;Lm;0;L;<super> 0284;;;N;;;;  
107B5;MODIFIER LETTER BILABIAL CLICK;Lm;0;L;<super> 0298;;;N;;;;  
107B6;MODIFIER LETTER DENTAL CLICK;Lm;0;L;<super> 01C0;;;N;;;;  
107B7;MODIFIER LETTER LATERAL CLICK;Lm;0;L;<super> 01C1;;;N;;;;  
107B8;MODIFIER LETTER ALVEOLAR CLICK;Lm;0;L;<super> 01C2;;;N;;;;  
107B9;MODIFIER LETTER RETROFLEX CLICK WITH RETROFLEX TAIL;Lm;0;L;<super>  
**1DF0A**;;;N;;;;

# References

William Bennett (forthcoming), 'Click Phonology'. In Sands, ed., *Click Consonants*. Brill, Leiden. International Phonetic Association (see following pages)



# THE INTERNATIONAL PHONETIC ASSOCIATION

*Registered Office:* INTERNATIONAL HOUSE, 24 HOLBORN VIADUCT, LONDON EC1A 2BN, UK

*from The President*

23 September 2020

## Unicode support for IPA letters as superscript modifiers

### Introduction

The IPA wishes to lodge a formal request for general Unicode support for all IPA letters to be used as superscript modifiers. This document outlines the consultation process that has taken place prior to this request, explains the general motivation for the proposal, and provides some details on the intended scope of the expression “all IPA letters”.

### Consultation

Prompted by an initial approach from K. Miller, the President wrote an internal discussion document dealing with the issue, which was circulated to the 30 members of the Council on 13 May 2020.

Opinions were also sought from a number of senior phoneticians and advisors who are not currently Council members. All replies received supported the proposal, and some of the more detailed responses have helped refine the explanation of the motivation given below. The documents which were then prepared for submission to the Unicode script ad hoc group on 30 August 2020 were shared with the IPA Council with a further call for discussion and debate. Again no objections were registered. The present proposal can therefore be regarded as coming jointly and severally from the

entire Council of 30 elected members, who are listed below as individual signatories.

### Motivation

A small set ( $n = 7$ ) of superscript modifiers is already defined on the IPA Chart:  $\langle^h \ ^w \ ^j \ ^\text{v} \ ^\text{r} \ ^\text{n} \ ^l \rangle$ . (The IPA Chart classifies them as Diacritics; in Unicode, six are placed in the range Spacing Modifier Letters, but  $\langle^n \rangle$  is in Superscripts and Subscripts.)

But as the work of Kirk Miller amply illustrates, a survey of the phonetic and linguistic literature reveals many more than 7 superscripts in actual use, for the most part with interpretations that are immediately obvious—and indeed natural—to the phonetically-aware reader, even though no official IPA meaning has been explicitly attached to them. Although modern word-processing or typesetting applications may allow users to obtain a reasonable approximation to any superficial effect that may be required, it is now clear that a more important consideration in the long term is the adequate preservation of the underlying data structure. Superscripting is a crucially meaningful aspect of IPA coding: it is an indicator that the superscript element has “modifier” status. If we remain content to allow superscripting to be achieved by superficial means, relying upon application-specific markup, this dimension of the meaning is lost.

The main concern may be with data structure, but that is not to say that legibility and appearance are of no importance—indeed they have always been major considerations for the IPA. It is worth noting, therefore, that the definition of separate codepoints for superscripts gives the font designer the opportunity to produce much more satisfactory forms than those which result from superscripting in a word processor, especially if the IPA is invited to advise on the design. This has already been exploited to good effect in existing fonts: compare for instance  $\langle t^h \ t^h \rangle$  where the first

has the dedicated modifier for aspiration (the font is SIL Gentium Plus) and the second is the result of applying the superscript effect in Microsoft Word to the same font's ordinary < h >.

Kirk Miller has assembled evidence of a large number of further IPA letters in use as superscripts—so many, in fact, that it would already begin to appear unwise to restrict the proposal to those letters hitherto attested, since the gaps are arguably mere accidents.

In fact, however, the IPA has a principle—long in place and never revoked—which permits *any* IPA symbol to be superscripted with modifier status. This provision is not currently mentioned on the IPA Chart, but it is important to distinguish between the Chart and the Alphabet. The Chart is a convenient one-page summary of the main provisions, but not an exhaustive enumeration of the Alphabet. The IPA's *Handbook* (1999) says (p.160, section 4 (c) (ii)) that one justifiable use of diacritics is 'For representing minute shades of sounds'. This phrase is taken over directly from the 1949 *Principles* (page 2, section 3. (d) (iv)). Reference to the *Principles* reveals a more extensive discussion of the matter, including specifically the means by which 'shades of sounds' are to be symbolised (see Figure 1).

**Small index letters may be used to indicate shades of sound.  
For instance  $\text{ʃ}^{\text{s}}$  means an s-like variety of  $\text{ʃ}$ ;  $\text{a}^{\text{r}}$  is another  
method of representing an r-coloured a.**

**Figure 1.** From the IPA *Principles* (1949: 17).

The formulation there seems to provide for essentially unlimited extrapolation in the use of 'small index letters'. And as late as 1979 the chart retained the example  $\text{ʃ}^{\text{s}}$  and explained it as a 'variety of  $\text{ʃ}$  resembling **s**, etc.', where that 'etc.' seems to offer unlimited licence.

It should not be assumed that 1999 *Handbook* was intended entirely to supersede earlier formulations.

In fact, there are no explicit revocation clauses in the *Handbook*. The view of the Council is therefore that the long-established provision remains in force allowing for limitless use of superscript symbols ('small index letters') as modifiers.

The conclusion, then, is that a codepoint should be allocated for the superscript version of every IPA letter-like symbol, and indeed that a Unicode implementation of the IPA which lacks this is incomplete.

Some may argue that it doesn't really matter whether a code point is allocated, even to existing superscripts such as the aspiration diacritic, since an approximation to the required appearance can always be obtained by means of text effects (in MS Word, for instance). That may have been the thinking when IPA symbols were first introduced into Unicode. But there is a fundamental difference in the type of data structure which results. This is seen clearly if we consider the construction of cross-language phonetic segment databases, or imagine future attempts at data-mining from texts containing IPA symbols. As has been shown, superscripting is not a merely superficial text effect, but a meaningful dimension of phonetic representation. Ideally Unicode codepoints should be so allocated that every meaningfully distinct IPA transcription corresponds to a unique code string, even after any markup is discarded.

#### Scope of the proposal

By 'letter-like symbols', we mean the symbols in the pulmonic and non-pulmonic consonant tables, in the vowel table, and in the 'other symbols' section of the current IPA chart, along with the retroflex consonants that are implicit in the alphabet (<ɖ ɗ ʎ!> and the <ʈ ʂ> that are also needed for the extensions to the IPA for disordered speech). Historical letter-like symbols should be supported if

they are attested in the recent literature. A more methodical approach to historical and para-IPA symbols may be formulated after further consideration by the Council, but for now at least the affricate ligatures < ts dz tʃ dʒ tɕ dʑ tʂ dʐ > should be supported, as they remain in common use. If only the subset that have hitherto been documented as superscripts were to be included this would bias Unicode toward particular languages.

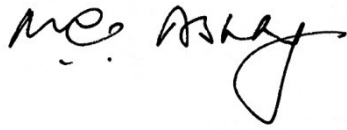
Finally we give some consideration to the superscripting of letters which carry diacritics. With a combining diacritic, no particular issues arise. Existing well-designed fonts intelligently move the combining diacritic when applied to a superscript; compare < ɲ̥ > and < tʰ̥ >. As for the data structure, the diacritic will be correctly taken as applying to the immediately preceding phonetic quality, < <sup>n</sup> >. Three spacing diacritics play important roles in the IPA, the one for ejectives < ' > and the two length marks < : · >. It seems that no new provision is required to cope with the ejective letters. The two length marks, however, do require Unicode support for use in superscripts, since length may be a feature of a modification rather than of a base segment. Extra length of aspiration, for example, cannot be indicated by simply adding the existing length mark after the superscript < <sup>h</sup> >, since this yields, for instance, < t<sup>h:</sup> > rather than < t<sup>h:</sup> >. In < t<sup>h:</sup> > the length mark will be taken as applying to < t >, and the representation interpreted as a long (geminate) [t] which is aspirated (to an unspecified degree).

## References

- IPA. (1949). *The principles of the International Phonetic Association*. London: International Phonetic Association.
- IPA. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.



Signed on behalf of the Council,



Michael Ashby MA DPhil, (President)

The full list of Council members for 2019–2023 is:

Prof. Dr Amalia Arvaniti	NETHERLANDS
Dr Michael Ashby	UNITED KINGDOM
Dr Patricia D S Ashby	UNITED KINGDOM
Prof. Plinio Almeida Barbosa	BRAZIL
Dr Patrice Speeter Beddor	UNITED STATES
Prof. Sonia Marise de Campos Frota	PORTUGAL
Prof. Gerry Docherty	AUSTRALIA
Prof. Emer. John H Esling	CANADA
Prof. Janet Fletcher	AUSTRALIA
Prof. Emer. William J Hardcastle	UNITED KINGDOM
Keith Johnson	UNITED STATES
Prof. Patricia Keating	UNITED STATES
Dr Ghada Khattab	UNITED KINGDOM
Dr James Kirby	UNITED KINGDOM
Prof. Alexei Kochetov	CANADA
Prof. Asher Laufer	ISRAEL
Dr Wai-Sum Lee	HONG KONG
Prof. Kikuo Maekawa	JAPAN
Dr Alexis Michaud	FRANCE
Prof. Ailbhe N Ni Chasaide	IRELAND
Prof. Katerina Nicolaidis	GREECE
Prof. Francis J Nolan	UNITED KINGDOM
Prof. Daniel Recasens	SPAIN
Prof. Jane E Setter	UNITED KINGDOM
Prof. Dr Adrian P Simpson	GERMANY
Prof. Maria-Josep Solé	SPAIN
Prof. Jane Stuart-Smith	UNITED KINGDOM
Prof. Marija Tabain	AUSTRALIA
Prof. Emer. Masaki Taniguchi	JAPAN
Prof. Dr Weijing ZHOU	CHINA

# Figures

## Modifier ejectives

A distinction might need to be made between an ejective consonant that has a superscript component, such as an allophone of /t'/ with fricative release, [t<sup>s'</sup>], and an entirely superscript consonant such as weakly articulated [k<sup>'</sup>]. Existing Unicode remedies should be sufficient for this.

(1985: 89–92) for !Xóǀ. Traill defines this constraint as an implication: in any CV sequence, if C is [+back], then V is also [+back]. The [+back] consonants of the language are velars and uvulars {k, k<sup>h</sup>, g, kx<sup>'</sup>, g<sup>kx'</sup>, q, q<sup>h</sup>, ɠ, q', x}, all coronals with velar accompaniments {C<sup>x</sup>, C<sup>kx'</sup>}, as well as bilabial, lateral, and alveolar clicks {⊙ || !} – to the exclusion of dental and palatal clicks {! †}.<sup>16</sup> Thus, we find cer-

Figure 1. Bennett (forthcoming: 86). Another convention for <g<sup>kx'</sup>> is <ɠkx'>. <kx'> is a unit here; the apostrophe modifies <kx'> specifically and not the consonant as a whole.

The !Xoon series of consonants {C<sup>kx'</sup>} mentioned in the figure above includes clicks with affricated ejective releases, such as the voiced series /g<sup>⊙kx'</sup> g<sup>||kx'</sup> g<sup>||kx'</sup> g<sup>!kx'</sup> g<sup>!kx'</sup>/, but also ejectives with the same release, namely /p<sup>'kx'</sup> t<sup>'kx'</sup> ts<sup>'kx'</sup>/. When asked, Bennett (p.c., 2020 August 24) replied,

*!Xoon /p<sup>'kx'</sup>/ should have the first ' modify the inline p, and the second ' should modify the superscript kx (and therefore could be super-superscripted where feasible?). [...] as far as I remember, there is no contrast between p<sup>kx'</sup> and p<sup>'kx'</sup>, so one could argue that we can make due using just one ' per segment. But it would be nice to have both options.*

Regardless of contrast or lack thereof, linguists working with !Xoon consonants have typically transcribed these segments with two apostrophes each.

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

**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:	<i>Modifier IPA letters (b), non-pulmonic</i>
2. Requester's name:	<i>Kirk Miller, Michael Ashby</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual</i>
4. Submission date:	<i>2020 November 08</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(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:	<i>yes</i>	
Name of the existing block:	<i>Phonetic Extensions Supplement-A</i>	
2. Number of characters in proposal:	<i>11</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?	<i>yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Kirk Miller</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>SIL (Gentium release)</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</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)?	<i>no</i>	

**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.

<sup>1</sup> 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? If YES explain	<i>no</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? If YES, available relevant documents:	<i>yes</i> <i>the 30 elected members of the Council of the IPA; various other phoneticians</i> <i>Letter of support from the International Phonetic Association</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>IT, publishing</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>phonetic</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>see IPA letter of support</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:	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>perhaps best to sort in alphabetic order with the other proposed modifier IPA characters</i>
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:	<i>no</i>
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:	<i>no</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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
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:	<i>no</i>
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)	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>no</i>