Unicode request for IPA modifier-letters (a), pulmonic

Kirk Miller, kirkmiller@gmail.com

Michael Ashby (President, IPA), m.ashby@ucl.ac.uk   2020 September 25

This proposal, officially supported by the International Phonetic Association (see letter on page 8), requests complete modifier support for current IPA letters, as well as for a few retired IPA letters, within the scope detailed below. Requested ⟨ʰ ʮ⟩ are also supported by the extIPA proposal recently accepted by the UTC in July 2020. The modifier lateral fricatives ⟨ɬ ɮ ꞎ ɭ̥⟩ were accepted by the UTC at that time and are not repeated here, but note that they have the support of the IPA in addition to the International Clinical Phonetics and Linguistics Association.

Due to the differing nature of the argument for acceptance by Unicode, letters for non-pulmonic consonants are requested in *Unicode request for IPA modifier-letter support (b), non-pulmonic*, for separate consideration by the UTC. In addition, the question of whether the IPA forms of modifier beta and chi should be Greek or Latin was only asked of the IPA in September, after SAH rejected Latin forms as not being sufficiently distinct for separate encoding. An adequate formal response from the IPA council may take some time, so evidence for those letters will be presented in the separate *Unicode request for IPA modifier-letter support (c), Greek letters*.

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

Background

This request expands on Peter Constable's 2003 'Proposal to Encode Additional Phonetic Modifier Letters in the UCS' (https://www.unicode.org/L2/L2003/03180-add-mod-ltr.pdf), and illustrates several characters that were requested in that proposal, but not illustrated and therefore not accepted at the time. Constable's notes (Section F of his proposal) includes a good summary of one of the two main uses of modifier IPA letters:

> In general, modifier letters are used in phonetic transcription to represent secondary aspects of articulation. Secondary articulations may involve aspects of simultaneous articulation that are considered to be in some sense less dominant to the basic sound (for instance, nasalized vowels are typically conceived in terms of their oral counterparts but with the additional secondary articulation of nasalization); or they may involve a transitional articulation of a type that might otherwise be considered a complete speech sound in its own right but for various reasons is interpreted by the linguist as a secondary element in a complex speech sound (for instance, diphthongs, or nasal onset of oral stop consonants). In some situations, the recommended transcription [by the International Phonetic Association] would not involve a modifier letter; thus, many of the proposed characters are not officially approved IPA notation. Nevertheless, the use of these modifier letters if fairly commonplace among linguists, even those that advocate the use of IPA.

It's notable that one of Constable's illustrations, of ⟨θ⟩, came from the IPA *Handbook* despite not being officially part of the IPA alphabet. Such usage goes back over a century, with ⟨ʃˤ⟩ given as an example in the IPA chart of 1900, and that the linguists that such usage is 'commonplace among' include Peter Ladefoged, president of the IPA from 1986 to 1991, organizer of the Kiel convention that overhauled the alphabet in 1989, and long-time editor of the Journal of the IPA (JIPA). An example of modifier letters for narrow phonetics in the official IPA *Handbook* is given as Figure 1.

A related but somewhat distinct use of modifier IPA letters is to indicate lightly or incompletely articulated sounds, for example in speech pathology but also in the transcription of running non-disordered speech. Several instances of such usage can be seen in the illustrations below. All consonant and vowel letters may be incompletely articulated as their target speech sounds, so all have potential modifier use.

Figure 1. Okada (1999: 118). Superscript IPA letters ⟨ᵈ ĭ ˢ⟩ used in the illustration of Japanese in the IPA *Handbook.* None of these characters are found in the set of superscripts provided by the IPA alphabet (according to the *Handbook*, the ⟨ˢ⟩ seen here was withdrawn a decade earlier), indicating continued sanction of the convention of superscripting IPA letters.

Of the IPA suprasegmentals, only the length marks are attested. As far as we can tell, tone letters and intonation-break symbols have not been used in this fashion, and we do not request them.

Superscript modifiers are semantically distinct from their base letters and may contrast with them, as illustrated by ⟨ɾ⟩ vs ⟨ᶿ⟩ in the narrow transcription ⟨baɾːɻ̩ndɔːəɾɻ̩z̩⟩ (*barn doors*) in Figure 18. The productivity of usage is illustrated by the unexpected appearance of ⟨fᵑ⟩ in Figure 33.

Precomposed letters such as ⟨ɚ⟩ in Figure 2 do not need support. (Compare custom precomposed ⟨ɚ ɝ⟩ with the combining diacritics in Gentium Plus font, ⟨ə˞ ɜ˞⟩.) The Script Ad Hoc group suggests that ⟨ɚ ɝ⟩ be accepted only if the IPA specifically requests them, and they do not.

Figure 2. An example of precomposed modifier ⟨ɚ⟩ in McDavid & O'Cain (1980: 121). The transcribed word is 'Maryland'.

For characters not illustrated in the following figures, see the attached letter of support from the International Phonetic Association. The unattested letters of the IPA chart are the vowels ⟨ɵ œ⟩ and the consonants ⟨ⱱ ɺ⟩, which are clearly accidental gaps: ⟨Œ ɞ⟩ are the rarest vowels in IPA transcription, and ⟨ⱱ ɞ⟩ are the two letters most recently added to the IPA alphabet.

In addition to the letters of the consonant and vowel blocks of the 2020 IPA chart, we request modifier support for the implicit IPA retroflex letter ⟨ɭ⟩ and the two length marks ⟨ː ˑ⟩, as well as for some retired IPA letters, namely the vowel ⟨ɷ⟩ and the affricates ⟨ʣ ʥ ʤ ʦ ʨ ʧ⟩. Although we have only attested to the two affricates that occur in English, ⟨ʤ ʧ⟩, this is a bias of our English-language sources, and the IPA supports including these characters as a set.

The SAH raised a concern about potential future demand to doubly superscript indefinite numbers

of modifier IPA letters. In most cases this would make no sense phonetically. For example, the transcription [tˢ] implies a lighter fricative release than the full affricate [t͡s]. If the entire consonant were lightly articulated, simple superscript should suffice for both elements: ⟨ᵗˢ⟩. Even for lightly articulated [pʰ], [tʲ] or [kʷ], there appears to be no demand for greater precision than currently supported ⟨ᵖʰ ᵗʲ ᵏʷ⟩. Although it is possible that a convention of doubly supercripted ⟨pʰ tʲ kʷ⟩ might develop in the future, it is unlikely that more characters will be needed than ⟨ʰ⟩ for aspiration, the four secondary articulations ⟨ʷ ʲ ˠ ˤ⟩ that are listed in the 'Diacritics' table of the IPA chart, and perhaps one or two unofficial IPA modifiers, such as voiced aspiration ⟨ʱ⟩ or the secondary articulation ⟨ᶣ⟩. (The 'no audible release' in the IPA diacritics table is a combining diacritic and so, like the rhotic vowels, should be handled by the font.) Beyond the dubiousness of extending such a convention to the entire alphabet, there is the practical problem that, at normal font sizes, doubly superscripted letters would be largely illegible, which should encourage other solutions for any perceived need.

Modifier letters

ᵛ   10781 MODIFIER LETTER SUPERSCRIPT TRIANGULAR COLON. Figures 54–55.

�   10782 MODIFIER LETTER SUPERSCRIPT HALF TRIANGULAR COLON. Figure 54.

ᵆ   10783 MODIFIER LETTER SMALL AE. Figures 45–48.

ᴮ   10784 MODIFIER LETTER SMALL CAPITAL B. Figure 16.

ᵈᶻ   10787 MODIFIER LETTER SMALL DZ DIGRAPH. IPA Letter of Support.

ᵈᶼ   10788 MODIFIER LETTER SMALL DZ DIGRAPH WITH RETROFLEX HOOK. IPA Letter of Support.

ᵈᶽ   10789 MODIFIER LETTER SMALL DZ DIGRAPH WITH CURL. IPA Letter of Support.

ᵈᶾ   1078A MODIFIER LETTER SMALL DEZH DIGRAPH. Figures 52–53.

ᵈ   1078B MODIFIER LETTER SMALL D WITH TAIL. Figure 25.

ᵊ   1078E MODIFIER LETTER SMALL REVERSED E. Figure 50.

ᵌ   1078F MODIFIER LETTER SMALL CLOSED REVERSED OPEN E. IPA Letter of Support.

ᵞ   10791 MODIFIER LETTER SMALL RAMS HORN. Figures 36–37.

ᴳ   10792 MODIFIER LETTER SMALL CAPITAL G. Figures 26–28.

ʰ   10795 MODIFIER LETTER SMALL H WITH STROKE. Figures 13–14.

ᴴ   10796 MODIFIER LETTER SMALL CAPITAL H. Figures 56–62.

ᶹ   10797 MODIFIER LETTER SMALL HENG WITH HOOK. Figure 33.

ᶷ   107A0 MODIFIER LETTER SMALL TURNED Y. Figures 29–32.

ᵒ   107A2 MODIFIER LETTER SMALL O WITH STROKE. Figure 49.

ᴼᴱ   107A3 MODIFIER LETTER SMALL CAPITAL OE. IPA Letter of Support.

ᵔ   107A4 MODIFIER LETTER SMALL CLOSED OMEGA. Figures 38–41.

ᑫ   107A5 MODIFIER LETTER SMALL Q. Figures 3–6.

ᶨ   107A6 MODIFIER LETTER SMALL TURNED R WITH LONG LEG. IPA Letter of Support.

ᶩ   107A7 MODIFIER LETTER SMALL TURNED R WITH LONG LEG AND RETROFLEX HOOK. IPA Letter of Support.

ᶉ   107A8 MODIFIER LETTER SMALL R WITH TAIL. Figures 17–19.

ꭉ 107A9 MODIFIER LETTER SMALL R WITH FISHHOOK. Figures 19–20.

ʀ 107AA MODIFIER LETTER SMALL CAPITAL R. Figures 34-35.

ʨ 107AB MODIFIER LETTER SMALL TC DIGRAPH WITH CURL. IPA Letter of Support.

ʦ 107AC MODIFIER LETTER SMALL TS DIGRAPH. IPA Letter of Support.

ʦ̢ 107AD MODIFIER LETTER SMALL TS DIGRAPH WITH RETROFLEX HOOK. IPA Letter of Support.

ʧ 107AE MODIFIER LETTER SMALL TESH DIGRAPH. Figures 51, 53.

ƫ 107AF MODIFIER LETTER SMALL T WITH RETROFLEX HOOK. Figures 21–23.

ᶹ 107B0 MODIFIER LETTER SMALL V WITH RIGHT HOOK. IPA Letter of Support.

ʸ 107B2 MODIFIER LETTER SMALL CAPITAL Y. Figures 42–44.

ˀ 107B3 MODIFIER LETTER LATIN GLOTTAL STOP WITH STROKE. Figure 12.

ˁ 107B4 MODIFIER LETTER REVERSED GLOTTAL STOP WITH STROKE. Figures 7–11.

Chart

Latin Extended-F (formerly Phonetic Extensions Supplement-A) was created for this series of proposals.

- Characters in magenta cells are illustrated in the figures.
- Characters in pink cells are accidental gaps, and requested as such by the International Phonetic Association.
- Characters in grey cells have already been accepted by the UTC, 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.

We propose moving the characters in Miller & Ball's extIPA request, which was accepted by the UTC in July, from Latin-D to Latin-F, so as to leave room in Latin-D for medieval manuscript letters and capitals of BMP characters that form casing pairs. We sort them together with the IPA letters of this proposal. Martin Ball of the International Clinical Phonetics and Linguistics Association said (p.c.) that he has no problem with this, and Marija Tabain, editor of JIPA, says that it is acceptable to place extIPA letters such as ⟨ʖ⟩ in the main IPA consonant chart in illustrations of the IPA in that journal. There would therefore seem to be no reason to segregate the two sets of modifier IPA letters in Unicode.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-F																
U+1078x	ꜳ	ː	ˈ	æ	ʙ	ɓ	β	dz	dʑ	dʐ	dʒ	ɖ	ɗ	ɗ	ə	ɞ
U+1079x	fŋ	ɣ	ɢ	ɠ	ɕ	ħ	ʜ	ɦ	ʄ	ls	ɮ	ɬ	ɺ	ʈ	ʐ	ʂ
U+107Ax	ʎ	ʎ	ø	œ	ɷ	q	ɻ	ɭ	ɾ	ɾ	ʀ	ʨ	ʦ	ʦ̢	ʧ	ʈ
U+107Bx	ʋ	x	ʏ	ʔ	ʕ	ʘ	ǀ	‖	ǂ	ǁ						

4

Properties

The cross-referenced number in red is the currently proposed codepoint for a new base letter. If that letter is placed elsewhere by the UTC, this number will need to be changed to match.

```
10781;MODIFIER LETTER SUPERSCRIPT TRIANGULAR COLON;Lm;0;L;<super> 02D0;
    ;;;N;;;;;
10782;MODIFIER LETTER SUPERSCRIPT HALF TRIANGULAR COLON;Lm;0;L;<super>
    02D1;;;;N;;;;;
10783;MODIFIER LETTER SMALL AE;Lm;0;L;<super> 00E6;;;;N;;;;;
10784;MODIFIER LETTER SMALL CAPITAL B;Lm;0;L;<super> 0299;;;;N;;;;;
10787;MODIFIER LETTER SMALL DZ DIGRAPH;Lm;0;L;<super> 02A3;;;;N;;;;;
10788;MODIFIER LETTER SMALL DZ DIGRAPH WITH RETROFLEX HOOK;
    Lm;0;L;<super> AB66;;;;N;;;;;
10789;MODIFIER LETTER SMALL DZ DIGRAPH WITH CURL;Lm;0;L;<super> 02A5;
    ;;;N;;;;;
1078A;MODIFIER LETTER SMALL DEZH DIGRAPH;Lm;0;L;<super> 02A4;;;;N;;;;;
1078B;MODIFIER LETTER SMALL D WITH TAIL;Lm;0;L;<super> 0256;;;;N;;;;;
1078E;MODIFIER LETTER SMALL REVERSED E;Lm;0;L;<super> 0258;;;;N;;;;;
1078F;MODIFIER LETTER SMALL CLOSED REVERSED OPEN E;Lm;0;L;<super> 025E;
    ;;;N;;;;;
10791;MODIFIER LETTER SMALL RAMS HORN;Lm;0;L;<super> 0264;;;;N;;;;;
10792;MODIFIER LETTER SMALL CAPITAL G;Lm;0;L;<super> 0262;;;;N;;;;;
10795;MODIFIER LETTER SMALL H WITH STROKE;Lm;0;L;<super> 0127;;;;N;;;;;
10796;MODIFIER LETTER SMALL CAPITAL H;Lm;0;L;<super> 029C;;;;N;;;;;
10797;MODIFIER LETTER SMALL HENG WITH HOOK;Lm;0;L;<super> 0267;;;;N;;;;;
107A0;MODIFIER LETTER SMALL TURNED Y;Lm;0;L;<super> 028E;;;;N;;;;;
107A2;MODIFIER LETTER SMALL O WITH STROKE;Lm;0;L;<super> 00F8;;;;N;;;;;
107A3;MODIFIER LETTER SMALL CAPITAL OE;Lm;0;L;<super> 0276;;;;N;;;;;
107A4;MODIFIER LETTER SMALL CLOSED OMEGA;Lm;0;L;<super> 0277;;;;N;;;;;
107A5;MODIFIER LETTER SMALL Q;Lm;0;L;<super> 0071;;;;N;;;;;
107A6;MODIFIER LETTER SMALL TURNED R WITH LONG LEG;Lm;0;L;<super> 027A;
    ;;;N;;;;;
107A7;MODIFIER LETTER SMALL TURNED R WITH LONG LEG AND RETROFLEX HOOK;
    Lm;0;L;<super> 1DF08;;;;N;;;;;
107A8;MODIFIER LETTER SMALL R WITH TAIL;Lm;0;L;<super> 027D;;;;N;;;;;
107A9;MODIFIER LETTER SMALL R WITH FISHHOOK;Lm;0;L;<super> 027E;
    ;;;N;;;;;
107AA;MODIFIER LETTER SMALL CAPITAL R;Lm;0;L;<super> 0280;;;;N;;;;;
107AB;MODIFIER LETTER SMALL TC DIGRAPH WITH CURL;Lm;0;L;<super> 02A8;
    ;;;N;;;;;
107AC;MODIFIER LETTER SMALL TS DIGRAPH;Lm;0;L;<super> 02A6;;;;N;;;;;
107AD;MODIFIER LETTER SMALL TS DIGRAPH WITH RETROFLEX HOOK;
    Lm;0;L;<super> AB67;;;;N;;;;;
107AE;MODIFIER LETTER SMALL TESH DIGRAPH;Lm;0;L;<super> 02A7;;;;N;;;;;
107AF;MODIFIER LETTER SMALL T WITH RETROFLEX HOOK;Lm;0;L;<super> 0288;
    ;;;N;;;;;
107B0;MODIFIER LETTER SMALL V WITH RIGHT HOOK;Lm;0;L;<super> 2C71;
    ;;;N;;;;;
```

```
107B2;MODIFIER LETTER SMALL CAPITAL Y;Lm;0;L;<super> 028F;;;;N;;;;;
107B3;MODIFIER LETTER GLOTTAL STOP WITH STROKE;Lm;0;L;<super> 02A1;
    ;;;N;;;;;
107B4;MODIFIER LETTER REVERSED GLOTTAL STOP WITH STROKE;Lm;0;L;<super>
    02A2;;;;N;;;;;
```

References

Roger Bagnall (2011) *The Oxford Handbook of Papyrology.*

William Bennett (forthcoming), 'Click Phonology'. In Sands, ed., *Click Consonants.* Brill, Leiden.

William Bright (1957) *The Karok Language.* University of California Press.

Terry Crowley & John Lynch (2006) *The Avava Language of Central Malakula (Vanuatu).*

Jerold Edmondson, John Esling, Jimmy Harris & Tung-chiou Huang. (2005) "A Laryngoscopic Study of Glottal and Epiglottal/Pharyngeal Stop and Continuant Articulations in Amis", *Language and Linguistics* 6: 3.

Jerold Edmondson, John Esling & Jimmy Harris (2008: ms p. 6) "Vocal register and other phonetic features of Somali". Data published in Edmondson & Esling (2006) "The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies," *Phonology* 23, 157–191.

John Esling (2010) "Phonetic Notation", in Hardcastle, Laver & Gibbon (eds.) *The Handbook of Phonetic Sciences*, 2nd ed.

Nick Evans (1996) "First – and last – notes on Wurrugu". A 'slightly revised' version of the publication in *University of Melbourne Working Papers in Linguistics* 16.

Mats Exter (2008) "Properties of the Anterior and Posterior Click Closures in N|uu." PhD dissertation, University of Cologne.

Alexandre François (2010) "Phonotactics and the prestopped velar lateral of Hiw," *Phonology* 27.3.

Barry Heselwood (2013) *Phonetic Transcription in Theory and Practice.* Edinburgh University Press.

Mark Harvey, Ben Davies, Susan Lin, Myf Turpin, Alison Ross & Katherine Demuth (2013) "Two types of pre-stopping in Kaytetye," *The 49th Annual Meeting of the Chicago Linguistic Society.* University of Chicago Press.

Theo Homan (1975), *Skíðaríma: An Inquiry into the Written and Printed Texts, References and Commentaries.* Amsterdamer Publikationen zur Sprache und Literatur.

International Phonetic Association. Letter of Support. (See following pages.)

Wolfgang Kehrein (2002) *Phonological Representation and Phonetic Phasing: Affricates and Laryngeals.* Walter de Gruyter.

Peter Ladefoged & Ian Maddieson (1996) *Sounds of the World's Languages.*

John Laver (1994) *Principles of Phonetics.* Cambridge University Press.

Raven McDavid & Raymond O'Cain (1980) *Linguistic Atlas of the Middle and South Atlantic States.* University of Chicago Press, fasc. 2.

Amanda Miller et al. (2007), "The Sounds of N|uu: Place and Airstream Contrasts", *Working Papers of the Cornell Phonetics Laboratory* 16.

Amanda Miller (2011) "The Representation of Clicks", in Oostendorp et al. (eds.) *The Blackwell Companion to Phonology.*

Miller-Ockhuizen (2010) "A prosodic account of Ju|'hoansi consonant distributional asymmetries," *Proceedings of the 1st International Symposium January 4-8, 2003, Riezlern/Kleinwalsertal.*

Hideo Okada (1999) "Japanese," *Handbook of the International Phonetic Association.*

Robert Penhallurick (1991) *The Anglo-Welsh Dialects of North Wales.*

Jill Perry (2000) *Phonological/phonetic assessment of an English-speaking adult with dysarthria.* Masters

thesis, Memorial University of Newfoundland.

Menán du Plessis (2009) *Unity Hypothesis for the Southern African Khoesan Languages*. PhD thesis, University of Cape Town.

Tomas Riad (2014) *The Phonology of Swedish*, Oxford University Press.

Anastasia Riehl & Abigail Cohn (2011) "Partially Nasal Segments," *The Blackwell Companion to Phonology*, vol. 1.

Bonny Sands (forthcoming) *Click Consonants*. Brill, Leiden.

Sands & Honken (2014) "ǂ'Amkoe body part terminology in comparative perspective," in Güldemann & Fehn (eds.) *Beyond 'Khoisan'*.

Deborah Tooker (2012) *Space and the production of cultural difference among the Akha prior to globalization*.

UPSID: UCLA Phonological Segment Inventory Database. Ian Maddieson (1981) "UPSID: Data and Index", *UCLA Working Papers in Phonetics* 53.

Clive Upton, David Parry & John Widdowson (1994) *Survey of English Dialects*. Routledge.

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: ⟨ʰ ʷ ʲ ˠ ˤ ⁿ ˡ⟩. (The IPA Chart classifies them as Diacritics; in Unicode, six are placed in the range Spacing Modifier Letters, but ⟨ⁿ⟩ 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 then 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 ⟨tʰ tʰ⟩ 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 ʃˢ means an **s**-like variety of ʃ ; aⁱ 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 ʃˢ and explained it as a 'variety of 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 (⟨ ɕ ʄ ʝ ǃ ⟩ and the ⟨ ɬ ɮ ⟩ that are also needed for the extensions to the IPA for disordered speech). Historical letter-like symbols should be supported if
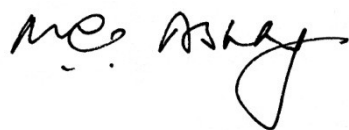
they are attested in the recent literature. A more methodical approach to historical and para-IPA symbols may be formulated after further consideration by the Council, but for now at least the affricate ligatures ⟨ ts dz tʃ dʒ tɕ dʑ ʈʂ ɖʐ ⟩ should be supported, as they remain in common use. If only the 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 ⟨ n̥ ⟩ and ⟨ tⁿ̥ ⟩. As for the data structure, the diacritic will be correctly taken as applying to the immediately preceding phonetic quality, ⟨ ⁿ ⟩. Three spacing diacritics play important roles in the IPA, the one for ejectives ⟨ ʼ ⟩ and the two length marks ⟨ ː ˑ ⟩. 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 ⟨ ʰ ⟩, since this yields, for instance, ⟨ tʰː ⟩ rather than ⟨ tʰˑ ⟩. In ⟨ tʰː ⟩ 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 q (ᑫ)

Used in romanized Akha for glottal stop and by Khoisanists for vowel pharyngealization. Also for an unreleased [q] in the conventions of Pike (1947) *Phonemics* (not illustrated).

probably pharyngealization. This is transliterated here by means of a superscript 'q', as in !gwaᑫ-xu 'sky'. Concerning Lloyd's annotation of 'accompaniments', Traill (1995) has determined that her use of a following letter 'k' almost certainly represented the 'plain' click, i.e. the click characterized by inaudible velar (i.e. posterior) release; while her

| 'ayin' (/ ̃) | |nú/ ̃ ŏ | 'old man' |
| | / ̃aoᑫ | 'cold, winter' |

Figure 3. du Plessis (2009: 30), with a description of the symbol as a superscript 'q'.

noisy dorsal burst which occurs after the anterior click burst. But, it is a convention which helps to distinguish the clicks in !Xóõ from the $/\widehat{!q}/$ click in N|uu that has different timing characteristics. In using diacritics to transcribe $/^k!/$ and $/^q!/$, I suggest that these do not pattern phonetically as contour segments or clusters.

Accompaniment	Ladefoged and Traill (1994)	Orthography (Traill 1994)	Transcription (Sands)
Voiceless unaspirated	k! q!	! !q	$^k!$ $^q!$
Voiced	g! ɢ!	!g !ɢ	$^g!$ $^{ɴ\!ɢ}!$
Voiced aspirated	g!h ɢ!h	g!qh[16] ɢ!qh	$^{gk}!h$ $^{ɴq}!h$

Figure 4. Sands (fc: 31), with the difference between superscript and baseline 'q'.

the anterior click burst, the dorsal burst and the vowel onset are very similar for the clicks shown in Figures 1.14 and 1.15. The center of gravity of the uvular-alveolar $/^q!/$ click shown in Figure 1.15 is lower than that of the velar-alveolar $/^k!/$ click shown in Figure 1.14 (1000 Hz vs. 1390 Hz). This is consistent with a uvular vs. velar place difference. Formant transitions also suggest a contrast of dorsal place.

[18] Nakagawa (2006: 197) notes timing differences between the anterior and dorsal bursts for uvular and velar accompaniments in G|ui that are greater than those I have seen in the corresponding accompaniments in !Xóõ but less than what is typically seen in N|uu clicks. G|ui clicks might be transcribed with a superscripted dorsal stop symbol after the click, e.g. $/!^q/$.

Figure 5. Sands (fc: 32, incl. fn 18)

above, at the flat[54] approach to the village
the *dzòma Làⁿ doᵍ doᵍ káⁿ káⁿ làⁿ* [= name of *dzòma*]
ruling at the village center[55]
friendly *Mìnm̀ kɔ́ tiᵍ ábəᵍ* [lit.-fast like the rolling *ábəᵍ* [56] seeds]
When *Kɔ́ tiᵍ ábəᵍ* strikes the flint for a fire,
if [she] doesn't strike it, [only] the space around her feet is lit [at night]
striking it, all the land and rivers light up.[57]
(*zoᵍla*) text, lines 313-319, my translation; see Tooker 1988. Appendix B).

Figure 6. Tooker (2012: 100)

Modifier reversed glottal stop with stroke, ʕ (ˤ)

Epiglottalization. Used in Khoisan and in Arabic dialectology. An additional source to those below is Miller-Ockhuizen (2003) *The Phonetics and Phonology of Gutturals: A Case Study from Juǀʼhoansi*, p. 72 ff., and Esling (p.c.) informs me that he's made extensive use of this character in his 2019 book *Voice Quality*, which I don't have access to.
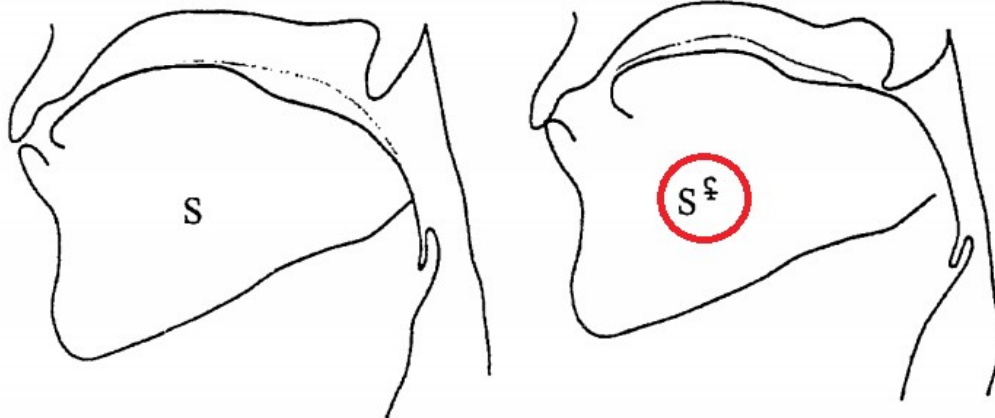

Figure 10.24 X-ray tracings of s and sˤ spoken by a Qatari Arabic speaker in the words sad 'to prevail' and sˤad (name of the letter).

Figure 7. Ladefoged & Maddieson (1996: 365)

Ejective stop. As was noted above, there is one ejective stop in N|uu (recently discovered in one word, *c'áˤè* 'be in pieces'), namely the *voiceless prepalatal ejective stop* /c'/ (3.9).

(3.9) *c'áˤè* 'be in pieces'

As for the VOT of aspirated plosives, no quantitative study has yet been conducted on the *glottal lag* (the duration between the release of the oral closure and the release of the glottal closure, i.e., the onset of voicing) of ejective stops in N|uu. From qualitative inspection of the tokens that are found in my data, an order of magnitude for the glottal lag of around 50 ms (similar to the estimate given for the VOT above) seems reasonable, though. Figure 3.7 shows spectrograms of typical realizations of *cáà* 'lie (recline)' and *c'áˤè* 'be in pieces'. Both /c/ and /c'/ show the slight affrication typical of prepalatal stops.
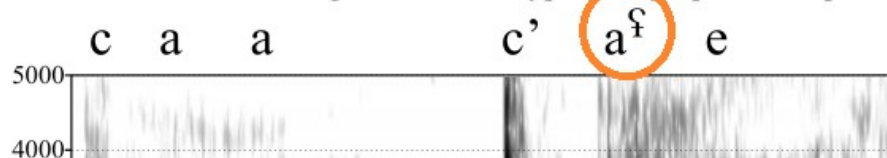

Figure 8. Exter (2008: 25)

Epiglottalized vowels. The epiglottalized vowel system consists of the segments /eˤ/, /aˤ/, /oˤ/, and /uˤ/ (3.25).

(3.25) *zéˤé* 'fly (verb)'
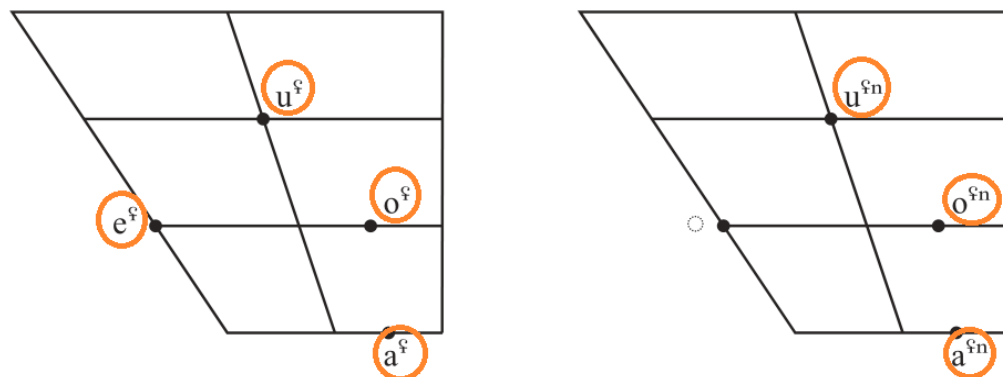ᵑǂàˤà 'kick'
kóˤó 'swell'
/ʰùˤá/ 'arrow poison'

Figure 3.30 Schematic diagrams of the 'plain' (left) and nasalized (right) epiglottalized vowel inventories of N|uu. The dotted circle indicates a presumed accidental gap.

Figure 9. Exter (2008: 51)

Nasalized		iⁿ		aⁿ			uⁿ
Epiglottalized			e^ˤ	a^ˤ		o^ˤ	u^ˤ
Nasal epiglottalized				a^{ˤn}		o^{ˤn}	

Table 1: Nǀuu vowels

Overall, there are fewer contrasts among the nasalized and epiglottalized vowels than their modal counterparts. Our lexicon contains examples of epiglottalization on all vowels except [i], but [e^ˤ] and [u^ˤ] are much less common than [a^ˤ] and [o^ˤ] and [u^ˤ] is most likely an allophone of [o^ˤ] Just three roots with [e^ˤ] have been elicited: [ze^{ˤ}e^ˤ] 'fly (v)', [je^{ˤ}βe] 'have arms crossed' and [ŋ‡^{h}e^{ˤ}βe] 'close your skirt'. To the best of our

Figure 10. Miller et al. (2007: 104)

Table 9. Some terms referring to the cheek		
Language	**"cheek"**	**"cheekbone"**
‡'Amkoe	[ŋ/ú^{ˤ}βí]	[ŋ|ú^{ˤ}βí (si) ɟà^{nˤ}]
Juǀ'hoan	ǃomm	ǀ'úbíkxáí

Figure 11. Sands & Honken (2014: 246)
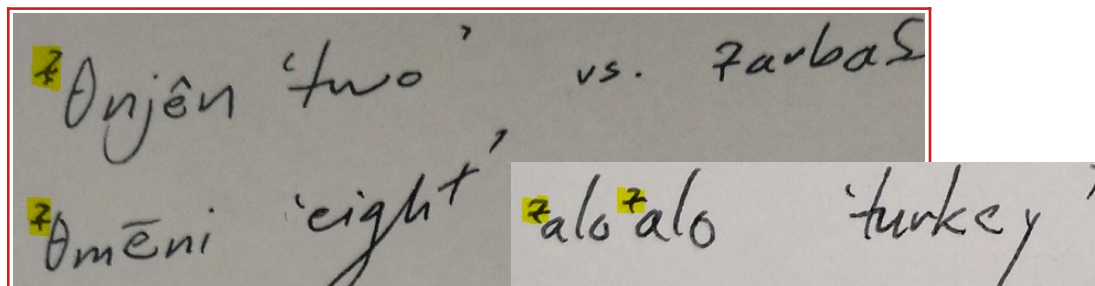
Modifier glottal stop with stroke, ʔ̵ (ˀ̵)

Figure 12. Timothy Henry, field notes for Moṣlawi Arabic from Oct 2015. Modifier letter for epiglottal constriction in [ˀθnjên] 'two' vs full letter for epiglottal stop in [ʔarbaʕ] 'rabbit'. Also [ˀθmēni] 'eight' and [ˀaloˀalo] 'turkey'.

I have only found ⟨ˀ̵⟩ in field notes, not in print. John Esling (President of the IPA 2011–2015, p.c.) says "One of the only reasons epiglottal stop (ʔ̵) in general may not yet appear widely in the literature is because we only identified it phonetically (as a pharyngeal stop, a.k.a. aryepiglottic closure) 20 years ago." Indeed, he has used the transcriptions ⟨ʕ͡ʔ⟩ and ⟨ʜ͡ɦ⟩ for Amis, with a tie bar that he notes in print is not accurate, and says he could imagine using modifier letters (ʕ̢ʔ and ʜʰ) instead.

Timothy Henry of the Harrington Project, at CalState Fullerton, has used ⟨ˀ̵⟩ for Moslawi Arabic in

his field notes (see figure), and would like support for publication. Re. the image above, he says (p.c.), *the raised epiglottal stop symbol [is] useful in Maslawi as there is frequently strong epiglottal closure in the production of words, but its use isn't always phonemic (it doesn't always correspond to anything in MSA, either). I've illustrated this in the first example where the epiglottal constriction was audible (and visible) in the word for 'two', but not phonemic, and then a full epiglottal stop in the word for 'rabbit', where the epiglottal stop is a variant of the phonemic glottal stop.*

Modifier h with stroke, ħ (ᵸ)

Esling, a co-author of the article in the figures below, says "it would also be logical to use ħ as a superscript, since voiceless frication or trilling are common phonetic effects," and that a superscript ᵸ would be useful for transcribing the very light final [ħ] often produced in the acquisition of Arabic by infants at around 12 months.

Note also the extIPA request that all fricatives be supported as modifier letters for fricated release. Because ħ is a regular IPA letter, it is included here rather than with the extIPA request.

'other', and are distinct from /ʔ/, which is implemented as [ʔ] initially and medially, and as [͡ʕʔᵸ] finally, e.g. in *'icep* [ʔtsəpʰ] 'betel nut', *po'ot* [poʔɔtʰ] 'small knife', and *loma'* [ɾumɑ̰͡ʕʔᵸ] 'house'.

Figure 13. Edmondson et al (2005: 384)

Figure 5: The epiglotto-pharyngeal stop in Amis *riri'* [riri͡ʕʔᵸ] 'grasshopper'

Figure 14. Edmondson et al (2005: fig. 5, p. 389).

[ᵸ] faríngea fricativa sorda poco definida

[ᵸs] faríngeo-alveolar fricativa sorda

[xᵸ] velo-faríngea fricativa sorda

Figure 15. Comparison of the *Atlas Lingüístico de México* (Lope Blanch 1990–2000, summary from UMex? [from external link in Wikipedia-ES article on RFE]. In the second line, the description goes on to say that in [ᵸs], the [s] is the primary place of articulation, as opposed to [ħˢ]. [xᵸ] also contrasts with [ħˣ].

Modifier small capital ʙ (ᴮ)

Found for voiced and voiceless bilabial trilled release, such as [bᴮ] alongside [dʳ] for an alveolar trilled release. Attested from Avava (Vanuatu). Crowley & Lynch use superscript small-capital ᴘ (ᴾ) for a voiceless bilabial trilled release, equivalent to Americanist superscript p-with-tilde ⟨p̃⟩ in [tp̃] in

Everett & Kern (1997) *Wari'*. Brendon Yoder (p.c.) says a superscript ʙ would be useful for transcribing trilled release of Nias *mb*.

Figure 16. Crowley & Lynch (2006: 31). Red is the voiced trilled release, orange its voiceless allophone (also visible, alveolar *dʳ*). On p. 30 the consonant is described as a "prenasalized voiced bilabial stop with a bilabial trilled release, which typically involves some rounding – or at least thrusting forward – of the lips." The tilde in this transcription indicates that rounding rather than nasalization.

Modifier r with fishhook, ɾ, and r with tail, ɽ (ʴ ˞)

Used in Penhallurick and similar dialect atlases for weak allophones of English /r/. The first two are also found in descriptions of Iwaidjan languages, where superscript ⟨ɾ⟩ and ⟨ɽ⟩ function as diacritics modifying the articulations of l and ɭ. The resulting phonemic segments /lʴ/ and /ɭ˞/ are distinct from sequences /lɾ/ and /ɭɽ/. So far I have only been able to access papers with /ɭ˞/, though the languages that have that also have /lʴ/ (e.g. Robern Mailhammer (p.c.) said that "the symbol [ɭ] used has been an l with either a tap or a retroflex tap superscript.")

paddle	*miri*	[míɹ̪i]	*wijarldug*	Marrgu, Maung *miri*
place name: Cape Don	*Jamarldin.gi*	[ɟámaɭˤiŋgi]		
place name: Eastern Side of Popham Creek	*Wumarij*	[wúmaɹiɟ]		
place name: Western Side of Popham Creek	*Wagali*	[wagali]		
proper name (Cunningham	*Ngurldun*	[ŋuɭˤuɳn]		

Figure 17. ⟨ɭˤ⟩ in Evans (1996: 2)

Figure 18. Upton, Parry & Widdowson (1994: 26). Contrast between ⟨ɾ⟩ and ⟨ɹ⟩.

Figure 19. Penhallurick (1991: xviii). Modifier IPA diacritics used for weak articulation. The letters are ʴ ˞ ˔ ˕ ʵ ᴿ ʶ ˢ ˪ ˒ ᵘ ᵊ. The superscript rhotics ʴ ʵ ᴿ and ɹ-bar are not supported by Unicode.

Figure 20. Penhallurick (1991: 66). Modifier fishhook *r* (red) is handwritten, as opposed to typewritten superscript *r* (yellow).

Modifier t with retroflex hook, ʈ (ᵗ)

Used for pre-stopped retroflex nasals and laterals, for example in the Arandic languages.

	bilabial	dental	alveolar	retroflex	palatal
	ᵖm	ᵗn̪	ᵗn	ᵗɳ	ᶜɲ
p		t̪	t	ʈ	c
	m	n̪	n	ɳ	ɲ

Figure 21. Riehl & Cohn (2011: 555)

	Labial	Dental	Alveolar	Retroflex	Palatal	Dorsal
Stop	p	t̪	t	ʈ	c	k
Nasal	m	n̪	n	ɳ	ɲ	ŋ
Pre-stopped Nasal	ᵖm	ᵗn̪	ᵗn	ᵗɳ	ᶜɲ	ᵏŋ
Lateral		l̪	l	ɭ	ʎ	
Tap			ɾ			
Continuant	w			ɻ		ɰ

Table 1: Kaytetye consonantal inventory.

Figure 22. Harvey et al. (2013)

relevant subset of the recorded data are considered. The target words contained the coronal nasals /n, ɳ, n̪, ɲ/, pre-stopped nasals /ᵗn, ᵗɳ, ᵗn̪, ᶜɲ/, and laterals /l, ɭ, l̪, ʎ/, all in a /#V_V/ context. The target coronal segments were preceded by word-initial /a/

Figure 23. Harvey et al. (2013)

426 *Alexandre François*

Australia, an alveolar [d͡l] and a dental [d͡l̪]. Martuthunira, a now extinct language of Western Australia, also prestopped its four laterals syllable-finally: [t̪l̪], [ᵗl], [ᶜʎ], [ᵗɭ] (Dench 1995: 27).

Figure 24. François (2010: 426)

Modifier d with tail, ɖ (ᵈ)

Linguists differ on whether they transcribe the Australian stop series as voiceless ⟨p t̪ t ʈ c k⟩ or voiced ⟨b d̪ d ɖɟ g⟩. Thus both ⟨t̪⟩ and superscript ⟨ᵈ⟩ are needed. For example, Nukunu /ᵗn̪/ and /ᶜɲ/ have been transcribed phonetically as [ᵈn̪] and [ʲɲ] (Hercus 1992 *A Nukunu Dictionary*, cited in Butcher 1999 "what speakers of Australian Aboriginal languages do with their velums and why", *Proceedings of the XIVth International Congress of Phonetic Science*), though that particular language does not have a retroflex. Similarly, Acehnese has voiced post-stopped (orally released) nasals transcribed

⟨mᵇ nᵈ ɲ ŋᵍ⟩ in Long & Maddieson (1993: 142) "Consonantal evidence against Quantal Theory", *UCLA Working Papers in Phonetics* 83, but again lacks a retroflex series. Chan (1987) comments that such notation is quite common, and uses it herself ("Post-stopped nasals in Chinese", *UCLA Working Papers in Phonetics* 68). Lack of attested ⟨ᵈ⟩ is thus an accidental gap.

Marija Trabain (editor of JIPA, p.c.) says that she's had trouble with soft-formatted ⟨ᵈ⟩ losing its formatting in Zoom for her online teaching. She notes that the prestopping in Australian languages is likely to be voiced, though since the languages don't have a phonemic voicing distinction, that is usually ignored in writing.

Figure 25. Marija Tabain (2020), from online course material.

Modifier small capital ɢ (ᴳ)

Used for pre-stopped uvular nasals and prevoicing of other uvular consonants. (A superscript is commonly used for prevoicing, e.g. in Laver 1994: 370.) For those who use superscripts for partial voicing, a /ɢ/ with final devoicing would be written ⟨ɢᶢ⟩. In addition, ⟨qᶜ⟩ is an old IPA convention for unaspirated [q] to be used when ⟨q⟩ transcribes aspirated [qʰ].

Figure 26. Wolfgang Kehrein (2002: p. 23 and fn 24). The two columns shown are 'stop' and 'affricate'.

Figure 27. Sands (fc: 31). Comparison of transcription conventions. (Excerpt from the table used to illustrate superscript q.)

Figure 28. Sands (fc: 31, fn 17)

Modifier small turned y, ʎ (ᶠ)

Modifier ʎ is used for a palatal lateral affricate, often with voiceless or raising diacritics. Even with the potential availability of a dedicated character from the extIPA, such usage could continue under adherence to the regular IPA.

Figure 29. **Kehrein (2002: 20).** A summary of attested laterals, lateral fricatives and lateral affricates. Modifier l ʎ ʟ are circled in yellow. (According to p. 6, the ring diacritics for voicelessness are omitted to simplify the transcription.)
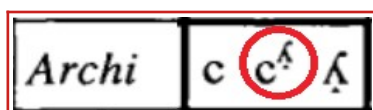
Figure 30. Kehrein (2002: 18) A palatal lateral affricate [cʎ].

Figure 31. Lateral affricated palatal clicks in the UPSID, p. 170.

Figure 32. Comparison of the *Atlas Lingüístico de México* (Lope Blanch 1990–2000, summary from UMex? from external link in Wikipedia-ES article on RFE[used the Greek letter by mistake, but intent is clear].

Modifier heng with hook, ɧ (ꚝ)

Used for theoretical allophones in Swedish. Note also extIPA request that all fricatives be supported as modifier letters for fricated release.

Figure 33. Riad (2014: 61, fn 11–12). ⟨ɟ̟ʲ⟩ and ⟨fᶣ⟩.

For the proposal to expand Unicode coverage of the extIPA, Script Ad Hoc asked for clarification of what exactly the convention of using modifier letters for fricated release was intended to cover. Martin Ball, primary author of the JIPA article in question, responded (2020 Feb 26),

> As regards the Unicode Committee's query concerning superscript symbols from the extIPA chart and the JIPA article describing them: My view is that the different fricative rows of the consonant chart should be supported in superscript form to allow the transcription of fricated release, as provided for in the 'other sounds' chart and summarized in the accompanying text.

The rows in the chart show additions to the IPA; other IPA letters are assumed but not shown explicitly.

Modifier small capital ʀ (ᴿ)

Used in Penhallurick and similar dialect atlases for weak allophones of English /r/. ɹ is used, along with a modifying superscript variant, for fricative [ɹ̝] in Penhallurick (1991). I have not found superscript ᶚ or ᴿ in that volume, apart from the list of superscript letters below, but they are used in the data it was drawn from, which is published in volume 3 of D.R. Parry (director, U. of Swansea) & Penhallurick (ed.) *Survey of Anglo-Welsh Dialects* [I do not have access to this]. ⟨ᴿ⟩ might also be found in Bright, if that is not a full capital (small cap for uvular trill vs full cap for voiceless alveolar trill).

The following superscript IPA letters in

The following superscript IPA letters indicate weak articulation :

Figure 34. Penhallurick (1991: xviii). Modifier IPA diacritics used for weak articulation. The letters are ʳ ɾ ᶚ ɹ ʈ ʀ ʁ s ɩ ɷ u ə.

šip "he flew away." Before front vowels, the velar friction of /x/ is often accompanied by a voiceless uvular trill: ʔa[xᴿxᴿ]i·č "child."

Figure 35. Bright (1957: 8)

Modifier rams horn, ɤ (ˠ)

Most instances of ˠ I've found are typos for ˠ (gamma). The following use ˠ for its intended value.

Similarly, Lanham and Macdonald (1979:40–1) identify fronted and glide-weakened MOUTH as a CE[15] variable and thus a defining feature of BrSAE. The most advanced values are [æ] for the onset, with a weak (unrounded) glide to [ə]. The vowel is corrected to a [äʊ]-like quality or even hypercorrected "towards backed [ɑʊ] or glide-weakened [ɑˠ] with hypercorrections most prominent among females in the Cape generally" (Lanham and Macdonald 1979:40). Importantly, and according to these

Figure 36. Bekker (2003: 165) *The Vowels of South African English*, PhD thesis. The source mistakenly used gamma (ɤ), which is why I refer to Bekker for its interpretation as ⟨ˠ⟩. P. 43 of Lanham & Macdonald (1979) supports Bekker, as the source used gamma for the full vowel letter ⟨ɤ⟩ as well, a not uncommon substitution in typewritten publications of the time. Regardless of whether the gamma was an intentional substitution in 1979, Bekker felt a need for superscript ram's horns in 2003.

stable, and invariably surfaces as back mid [o]. Lastly, back high /ɯ/ has the widest range of surface variants, ranging from high back to centralized mid front: /pipkɯp/ → [pipkɯˠp] or [pipkɪˠp] or [pipkɜˠp] 'nail'. In the processes of pre-vocalization and vocalization, other vowels may arise, as will be discussed in

Figure 37. Wetzels (2009: 244) "Nasal harmony and the representation of nasality in Maxacalí", in Calabrese & Wetzels (eds.) *Loan Phonology*. [ˠ] is a 'prevowel' derived from a coda /p/ or /m/ (p. 245).

Modifier closed omega, ɷ (ᵚ)

Former IPA usage. Used in English dialectology.

(b)   ME ŏ̤ before ld in old is represented as shown below.

old :   Gn 1 ô̤:²   2 o·ᵚ   3 ɔ:ɷ̣   4-7   o: 8 ô̤:   9 o:ᵚ
         10 o:¹, ɔu⁷

         Cl 1 o: 2 ô: 3 ɔ:u̥³   4 o:ᵚ   5 o:⁵   6 ŏ̤   7 o:¹
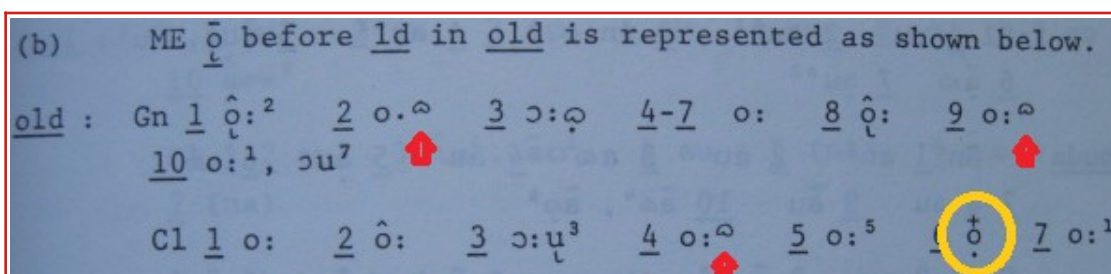
Figure 38. Penhallurick (1991: 79) *The Anglo-Welsh Dialects of North Wales*. [o:] with an offglide into [ɷ].

The long monophthongal forms [o:, o̤:, ọ:], [ɔ:] may be due to sound-substitution of Welsh /o:/ for RP /əɷ, oɷ/. The influence of this Welsh phoneme may also give rise to the diphthongal forms above with lengthened first elements – [ɔ:ɷ] – or weakened second elements – [o̤·ᵚ; oᵚ; ɔᵘ]. This Welsh influence is less likely to obtain at Cl 3, 6.

Figure 39. Penhallurick (1991: 55), explaining the use of superscript ɷ in diphthongs.

(?) GWŶDD MAIN   Gn 5   gâwîð 'maiṇ,

Figure 40. Penhallurick (1991: 258). Excrescent superscript *â*. The circumflex indicates pharyngealization.

Ribchester	/ʏ:/	/ɷə/
Eccleston	/ᵚu: ~ ᵚü:/	/ɷə/·
		/o:/
Skelmanthorpe	/ɷɪ/	/ɷə/
Youlgreave	/ɛᵚ/	/ᵚü:/
		/o:ᵚ/
Keelby	/u:/	/ɷə/

Figure 41. Anderson (1987: 102) *A Structural Atlas of the English Dialects*. The superscript is a bit faint, due to the thinness of the bottom curves, but is clearer in Note 8 on p. 119; the segments are /ᵚu: ~ ᵚü:/, /ɛᵚ/, /ᵚü:/, /o:ᵚ/.

Modifier small capital ʏ (ᵞ)

Used when measuring the formants leading into a consonant adjacent to [ʏ] (here an allophone /ɪ/ preceding /u/). Also for diphthongs like [øᵞ] in Swedish and [ɛᵞ] in some English dialects.

(71)

	pᵉ	pʲ	pᵒ	pʷ	tᴵ	tʲ	tᵞ	tʷ
trans F2	4	5	1	1	4	5	3	1
trans F3	3	5	1	1	5	5	1	1
burst NF	2	2	2	2	6	6	5	5

(72)

		MINDIST = V F2:3 & trans F2:4 (V)	or burst NF:2 (C)	MINDIST = V F2:4 & trans F2:4 (V)	or burst NF:2 & (trans F2:1 or trans F3:1) (C)	MAX CON-TRASTS
a.	pᵉe-pᵒo tᴵe-tᵞo	*!*		**		✓✓✓✓
b.	☞ pʲe-pʷo tʲe-tʷo				**	✓✓✓✓

Figure 42. Edward Flemming (2002: 103) *Auditory Representations in Phonology*. Routledge Outstanding Dissertations in Linguistics.

26

(6) American English: MINDIST = trans F2:3 >> MAX CONTRASTS		MINDIST = trans F2:2	MINDIST = trans F2:3	MAXIMIZE CONTRASTS
a.	bᵁu-dᴵu-dju	*	**	✓✓✓
b.	bᵁu-dᵞu-dju		**	✓✓✓
c. ☞	bᵁu-dᴵu			✓✓
d. ☞	bᵁu-dju			✓✓
e.	bᵁu-dᵞu		*!	✓✓

Figure 43. (Flemming 2002: 123)

while others are rising; some are closing, while others are centering: /eː/ > [eːⁱ], /ɛː/ > [ᵉɛː], /ʉː/ > [ᵋʉ], /øː/ > [øːᵞ], /uː/ > [ᵒuː], /oː/ > [oːᵋ] (/iː/, /yː/ and /aː/ are monophthongs) (Elert, 2000, 42).

Figure 44. Therese Leinonen (2010: 20) *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.* Doctoral thesis, University of Groningen.

Modifier æ (ᵆ)

Appears in diphthongal offglides and ms abbreviations.

fonetiske typer: en æ-agtig vokal og en diftong *ā·ᵋ, ā·ᵈ*. Nu forholder det sig ifølge Poul Andersen (FØF p. 138 ff.) således i østfynsk, at diftongiske manifestationer som *ɛ·ᵆ, ɛ·ᵆ* o.l. modsigelsesfrit kan tolkes som repræsenterende *æ*-fonemet, strukturelt *ɛᵆ*, idet der ikke findes sproglig opposition mellem et dobbelt *æ*-fonem med

Figure 45. Akademisk Forlag. (Denmark, 1974: 199) *Dialektstudier*, vol. 3. Contrast between [ɛ·æ] and [ɛ·ᵆ].

ɑᵉ:	*fje* s. Fjed	*meᵆ* præp. med (KA)
	le s. Led	*sdeᵆ* s. Sted (KA)
	sbe s. Sped	*seᵆ* s. Sæd, Vognbund

Figure 46. Andersen (1958: 138) *Fonemsystement i østfynsk: på grundlag af dialekten i revninge sogn.*

ɑ:	*ve·ᵊd* præp. ved det	*sde·ᵆd* s.bn. Stedet
		me·æd præp. med det
	sde·l·d s.bn. Stellet	*sdeᵆl·ᵊd* s.bn. Stellet
		dæl·ᵊd s.bn. Dællet

Figure 47. Andersen (1958: 139). Contrast between [ᵆ] (red) and [æ] (yellow).

27

> æ: superscript-æ represents:
> ær.   Ex. þær 3³, kiærleikz 3⁴, kiærn 123². written in full: kiærn 42².
>   As for the occurrence of superscript-æ in the older MSS, the remarks given above about superscript-ö hold here too, except that superscript-æ seems to represent ræ exclusively.

Figure 48. Homan (1975: 151), explaining ᵃᵉ in Icelandic mss. (The italics are a convention for the abbreviations spelled out in full.)

Modifier o with stroke, ø (ᵒ̷)

Appears in diphthongs in Scanian.

> The picture is similar in the rounded counterpart /y/. Ingers has in 1962 in Hjärsås noted an older form [by:] besides a younger [bᵒ̷y:] for *by* 'village'. From Loshult, Sigvard Bengtsson in 1931 reports three diphthongised realisations, which in IPA would be approximately [bᵒ̷yⁱ] 'village', [²nᵒ̷yʲa] 'new.PLUR' and [sᵒ̷yⁱn] *syn* (a certain legal procedure). An etymologically near-lying [²sᵒ̷yⁱnas] 'be visible' is found in Vankiva in the neighbouring hundred Västra Göinge, where it has been noted by Nils Nilsson in 1932.

Figure 49. Martin Persson (forthcoming thesis on Scanian dialectology).

Modifier reversed e, ǝ (ᵊ)

Appears in diphthongal offglides.

> (48) Eᵊhɔt də kuǝxə gɛᵊsǝ
> Il avoir-Pfkt.3.sg. le gâteau mangé.
> « Il mangeait / a mangé le gâteau… »

Figure 50. Caudal (2011: 27) "Contribution aspectuelle des temps et de la phrase", in Hancil (ed.) *Perspectives théoriques et empiriques: Sur l'aspect en anglais.*

Modifier tesh and dezh, ʧ ʤ (ᵗʃ ᵈʒ)

Tesh and dezh are former IPA usage, retired in 1989 with the Kiel convention. The examples below date from a decade later, and ligatures for affricates continue to be used. Because affricates are unlikely candidates for the releases or onsets of consonants, the most likely use of these letters as superscripts is for weak or incompletely articulated pronunciations, as in narrow transcription in speech pathology. That literature is heavily biased toward English, and our scouring of the literature even more heavily biased toward English, so it is unsurprising that we have only found instances of the two affricates that occur in English.

P	g		*	ʝ					r
AU		ɪ			ə	n	ʧ	ə	
			*					*	
P			*		*			*	
IU	kʰ	ɪ	ʧ		ə	n	ʧ	ɛ	r

Figure 51. Perry (2000: 90). The top line, ⟨ɡɪʝəntʃərⱼ⟩, is the recorded utterance; below it is the target phrase *kitchen chair*.

P				*	ʤ		w	ʤ	*		
AU	ð	ə	b	æ		ə			ʌ	m	d
				*					*		
				*					*		
P		*		*		*			*		
IU	ð	ə	b	æ	ʤ	ə	r	ʤ	ʌ̃	m	p	t

Figure 52. Perry (2000: 91). ⟨ðəbæʤəᵂʤʌmd⟩; the target phrase is *the badger jumped.*

C: SLP: (1) exaggerates speech. Subject: (1) [oʷ] instead of [ɔ] in *your*. (2) omits [ɪ] in *your*. (3) [ʃ] instead of [ʧ] in *watch*. (4) [n] instead of [ð] in *the*. (5) [ɡ] instead of [kʰ] in *kitchen*. (6) [ʝ] instead of [ʧ] in *kitchen*. (7) [ə] instead of [ɛ] in *chair*. (8) [ʲ] instead of [r] in *chair*.

C: SLP: (1) exaggerates speech. Subject: (1) [ʤ] instead of [ʤ] in *badger*. (2) [ʷ] instead of [r] in *badger*. (3) omits [p] in *jumped*. (4) [d] instead of [t] in *jumped*. (5) [v] instead of [f] in *off*. (6) [o] instead of [ə] in *the*.

Figure 53. Perry (2000: 90, 91), specifying the contrast of [ʝ ʤ] and [ʧ ʤ].

Modifier triangular colon, ˑ, and half triangular colon, ꞏ (ˑ ꞏ)

Used for narrow transcription of voice-onset time (length of aspiration).

this is systematic cross-linguistically. Cho and Ladefoged (1999) have shown that languages exploit different ranges of voice onset time (VOT) values in their aspirated plosives depending on place of articulation, which could be indicated in narrower systematic transcriptions with superscript length marks, for example [pʰ, tʰˑ, kʰꞏ], but supplied by conventions in broader ones. See Section 4.6 below

Figure 54. Heselwood (2013: 146), with half-long ⟨ˑ⟩ and long ⟨ꞏ⟩. Heselwood distinguishes [ʰ], [ʰˑ] and [ʰꞏ]. If the length signs were to lose their superscript formatting, they would imply the consonant was geminate rather than that the aspiration was extended.

tions. She will produce them as the surface forms /.tʰæ.kɨ./ and /.tɛ.kʰɨ./ and as the auditory-phonetic forms [_t̪ʰꞏ̪ɛ̞ːɡ˺_ɡ̰ɨ] and [_d̪ɛ̀k˺_kʰꞏɨ], as the following four

Figure 55. Boersma & Hamann (2009: 34). [ʰꞏ] in the second line is distinguished from [ʰ] in the first line.

Modifier small capital ʜ (ᴴ)

The clicks transcribed [!kx'] and [g!kx'] in Juǀ'hoansi are transcribed by Miller-Ockhuizen (2003) and Miller (2007) as having epiglottalized releases, [!ᴴ] and [!ˤ]. Nakagawa (2006) transcribes the voiceless consonant of the pair as a cluster con-

Figure 56. Miller (2011: 421). A full-cap ᴴ has been substituted, but the description indicates that a small-capital was intended, as seen below.

last row of the table, each click may have several secondary accompaniments, such as in the series [ǀ ǀʰ ⁿ ǀ ᶢǀ ǀˀ ǀˣ ǀᴴ] (Miller, 2007; Miller et al., 2009), but only the basic primary values are represented here. This qualification can be seen

Figure 57. Esling (2010: 700). The ᴴ has x height.

GLOTTALIZED	ts' tʃ'	ds' dʃ' dz' dʒ'		ǀˀ ǂˀ !ˀ ‖ˀ	
UVULARIZED	tˣ tʃˣ	dˣ dzˣ dʒˣ		ǀˣ ǂˣ !ˣ ‖ˣ	gǀʶ gǂʶ g!ʶ g‖ʶ
EPIGLOTTALIZED	tᴴ kᴴ	ᶢˤ		ǀᴴ ǂᴴ!ᴴ ‖ᴴ	gǀˤ gǂˤg!ˤ g‖ˤ

Figure 58. Miller-Ockhuizen (2010: 44). Modifier ʜ (red) and ʕ (yellow).

in narrow phonetic transcriptions, I have sometimes debated whether a syllable such as [!ɑˤ] with an epiglottalized vowel might not be better considered as containing a click with an epiglottalized accompaniment [!ᴴɑ]. Accompaniments are important because the number of click consonants (or click clusters) in the inventory of a language

Figure 59. Sands (fc: 23). The ᴴ clearly has x height.

vary even when the clicks described are the same. For instance, Miller (2007) has described Juǀ'hoan as having voiced and voiceless epiglottalized clicks /ǀᴴ, gǀˤ/ where others have transcribed them as ejected velar affricated clicks /!x', g!x'/ (e.g. Snyman

Figure 60. Sands (fc: 34).

The orthographic **q** represents a voiceless dorso-uvular stop, accompanied by epiglottal components depending upon the word position. The peculiar thing about this stop is that it might be called a voiceless epiglottalized uvular stop [q͡ʔᴴ] (often with faucalization), because it is accompanied by an epiglottal stop. Finally, in careful, overly correct speech it represents a voiceless dorso-uvular stop with a schwa offglide [qᵊ]. In more normal conversational style it will be an epiglottalized dorso-uvular stop that can be released as an ejective. Examples are: *qod* [ˈq͡ʔᴴɔ̠ʔt] 'to dig', *maqal* [ˈmá q͡ʔᴴà̠l] 'to hear' *aqal* [ˈʔá̠ q͡ʔᴴà̠l] 'a house', *liq* [ˈlì̠q͡ʔᴴ] 'to swallow'.

Figure 61. Edmondson, Esling & Harris (2008: ms p. 6)

The SAH had originally suggested annotating U+1D78 ᵸ (modifier Cyrillic ⁿ) for IPA usage. Michael Everson argued (next figure) that the behaviour of the two characters is distinct enough to warrant a

separate code for IPA usage, and his argument was accepted by the SAH meeting of 2020 Sept 25. Note that U+1D34 ᴴ is a full capital and so is inappropriate.

It has been suggested that MODIFIER LETTER SMALL CAPITAL H be unified with 1D78 MODIFIER LETTER CYRILLIC EN. This would be a false unification. The letter was proposed to be MODIFIER LETTER SMALL CYRILLIC EN, as can be seen in this figure from N1962 "Optimizing Georgian representation in the BMP of the UCS" (1999-02-26).

I.Чеченская			2. Ингушская			2.Бацбийская	
Латин-ския	Грузин-ская	Лит. чеч.	Латин-ская	Грузин-ская	Лит. ингуш.	Латинская	Грузинская
a	ა	a	a	ა	a	a	ა
a	ა	–	a	ა	–	a	ა
ä	�	აჼ	–	–	–	–	–
ä	ჴ	აჼ	ä	ჴ	აჼ	–	–
aⁿ	ავჼ	(aᴴ)	–	–	–	aⁿ	ავჼ
aⁿ	ჴჼ	–	–	–	–	–	–
b	ბ	ბ	b	ბ	ბ	b	ბ
g d	გ	г д	g d	გ	г д	g d	გ
e	ჯ	e	e	ჯ	e	e	ჯ
eⁿ	ჯჼ	–	–	–	–	eⁿ	ავჼ

Figure 6. Sample of Latin and Georgian transcriptions of Chechen, Ingush, and Bats, with Cyrillic orthogrpahy for Chechen and Ingush, showing 207F SUPERSCRIPT LATIN SMALL LETTER N, MODIFIER LETTER SMALL CYRILLIC N, and MODIFIER LETTER GEORGIAN NAR, in Имнайшвили 1977.

Clearly both Latin and Cyrillic are lowercase, and in addition to the fact that CYRILLIC EN does not sort where SMALL CAPITAL H should sort, the glyph behaviour in italic is not the same. Below is shown a followed by MODIFIER LETTER SMALL CAPITAL H, by SUPERSCRIPT LATIN LETTER N, and by MODIFIER LETTER CYRILLIC EN in Roman and Italic.

aᴴ aⁿ aᴴ *a*ᴴ *a*ⁿ *a*ᴴ

Figure 62. Everson's argument against unification with U+1D78. Note that the italic Cyrillic letter at bottom right is graphically distinct from the italic Latin letter. In some Slavic fonts, the two can be even more different, with the bar of the Cyrillic *n* not being horizontal.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:** *Modifier IPA letters (a), pulmonic*
2. Requester's name: *Kirk Miller, Michael Ashby*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *2020 September 25*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *yes*
   (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: *yes*
      Name of the existing block: *Latin Extended-F*
2. Number of characters in proposal: *35*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   *√*   B.1-Specialized (small collection)   B.2-Specialized (large collection)
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document? *yes*
   b. Are the character shapes attached in a legible form suitable for review? *yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Kirk Miller
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      SIL (Gentium release)
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other
      sources) of proposed characters attached? *yes*
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)? *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 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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *yes for some*

    If YES explain *some repetition from Constable (2003) 'Proposal to Encode Additional Phonetic Modifier Letters in the UCS' (https://www.unicode.org/L2/L2003/03180-add-mod-ltr.pdf)*

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.)? *yes*

    If YES, with whom? *Marija Tabain, Editor of JIPA; the 30 elected members of the Council of the IPA; various other phoneticians*

    If YES, available relevant documents: *Letter of IPA support from Ashby, acting as President of the IPA*

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *IT, publishing*

    Reference:

4. The context of use for the proposed characters (type of use; common or rare) *phonetic*

    Reference:

5. Are the proposed characters in current use by the user community? *yes*

    If YES, where? Reference: *see References section, IPA letter of support*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *no*

    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)? *yes*

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *no*

    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? *no*

    If YES, is a rationale for its inclusion provided?

    If YES, reference:

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? *no*

    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? *no*

    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:

12. Does the proposal contain characters with any special properties such as control function or similar semantics? *no*

    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? *no*

    If YES, are the equivalent corresponding unified ideographic characters identified?

    If YES, reference: