

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Международная организация по стандартизации

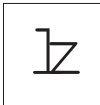
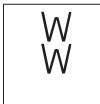

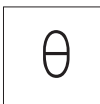
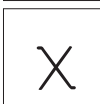


**Doc Type:** Working Group Document  
**Title:** Additional IPA characters for the UCS  
**Source:** Michael Everson  
**Status:** Working Document  
**Date:** 1998-05-25

This document proposes several IPA characters for inclusion in the UCS, and discusses the rationale for their inclusion.

Four of these characters are new characters, specified in “the extIPA Chart” (*Journal of the International Phonetic Association* (1994) 24.2, 95–98). These signs are used to describe disordered speech.

Three of these characters are Latin letters, borrowed from the Greek alphabet but used in the IPA. The International Phonetic Alphabet is the *Latin* alphabet with extensions. When the UCS was designed, *all* of the IPA characters derived from Greek letters should have been included, not just some of them. It makes no sense whatsoever for the UCS to contain, as it does for the IPA, Latin letters  $\alpha$ ,  $\gamma$ ,  $\epsilon$ ,  $\upsilon$ , and  $\varphi$ , but not  $\beta$ ,  $\theta$ , and  $\chi$ . The cost of this disunification is that it might affect some UCS coded IPA data. But UCS implementations are not that widespread, and the unification was a mistake in the first place. The benefit of the disunification is in the management and presentation of data. Unless LATIN LETTER THETA and GREEK LETTER THETA are distinguished, it would be impossible to correctly sort, for instance, a word list which contained Greek and Avestan words, or any other combination of IPA text and Greek text. Users expect (for instance) Avestan words to sort together and Greek words to sort together; the unification would force Avestan words beginning with the Greek-derived IPA letters to sort within the Greek list, not in the Latin-script list where they are required to sort.

The characters proposed to be added here are given below, with proposed code positions. The proposal summary form is appended thereafter.

02A9	LATIN SMALL LETTER FENG	02AB:	
02AA	LATIN SMALL LETTER LES		
02AB	LATIN SMALL LETTER LEZ	02AC:	
02AC	LATIN LETTER BILABIAL PERCUSSIVE		
02AD	LATIN SMALL LETTER BETA	02AD:	
02AE	LATIN SMALL LETTER THETA	02AE:	
02AF	LATIN SMALL LETTER CHI	02AF:	
02A9:			
02AA:			

## **A. Administrative**

### **1. Title**

Additional IPA characters for the UCS

### **2. Requester's name**

Michael Everson, EGT (WG2 member for Ireland)

### **3. Requester type**

Expert contribution

### **4. Submission date**

1998-05-25

### **5. Requester's reference**

### **6a. Completion**

This is a complete proposal.

### **6b. More information to be provided?**

No

## **B. Technical -- General**

### **1a. New script? Name?**

No.

### **1b. Addition of characters to existing block? Name?**

Yes. IPA (International Phonetic Alphabet) Extensions.

### **2. Number of characters**

7

### **3. Proposed category**

Category A

### **4. Proposed level of implementation and rationale**

Level 1 noncombining characters.

### **5a. Character names included in proposal?**

Yes

### **5b. Character names in accordance with guidelines?**

Yes

### **5c. Character shapes reviewable?**

Yes

### **6a. Who will provide computerized font?**

Michael Everson, Everson Gunn Teoranta

### **6b. Font currently available?**

Yes

### **6c. Font format?**

TrueType

### **7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?**

Yes. See *JIPA* (1994) 24.2 and *JIPA* (1995) 25.1.

### **7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?**

Hardcopy is provided to WG2 for distribution.

### **8. Does the proposal address other aspects of character data processing?**

Addition of three of the characters makes a multilingual sort including IPA and Greek data possible.

## **C. Technical -- Justification**

### **1. Contact with the user community?**

Yes: International Phonetic Association.

**2. Information on the user community?**

Linguists and publishers

**3a. The context of use for the proposed characters?**

Phonetic representations

**3b. Reference**

See above

**4a. Proposed characters in current use?**

Yes

**4b. Where?**

In many publications

**5a. Characters should be encoded entirely in BMP?**

Yes.

**5b. Rationale**

Contemporary use and keeping them with the other IPA characters.

**6. Should characters be kept in a continuous range?**

No.

**7a. Can the characters be considered a presentation form of an existing character or character sequence?**

No.

**7b. Where?**

**7c. Reference**

**8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?**

Some characters borrowed from the Greek alphabet into the Latin alphabet have a similar appearance as their Greek originals. They function differently, however, in terms of sorting, etc., as they are used in the IPA, which is comprised of the *Latin* alphabet with extensions.

**8b. Where?**

**8c. Reference**

**9a. Combining characters or use of composite sequences included?**

No

**9b. List of composite sequences and their corresponding glyph images provided?**

**10. Characters with any special properties such as control function, etc. included?**

No

## **D. SC2/WG2 Administrative**

To be completed by SC2/WG2

**1. Relevant SC 2/WG 2 document numbers:**

**2. Status (list of meeting number and corresponding action or disposition)**

**3. Additional contact to user communities, liaison organizations etc.**

**4. Assigned category and assigned priority/time frame**

**Other Comments**