

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации

**Doc Type:** Working Group Document  
**Title:** Proposal to encode three uppercase Latin letters used in early Pinyin  
**Source:** Andrew West (魏安), Eiso Chan (陈永聪), and Michael Everson (葉密豪)  
**Status:** Individual Contribution  
**Action:** For consideration by JTC1/SC2/WG2 and UTC  
**Date:** 2017-01-16

## 1. Introduction

This is a proposal to encode uppercase forms of three existing Latin characters which currently only exist in lowercase form:

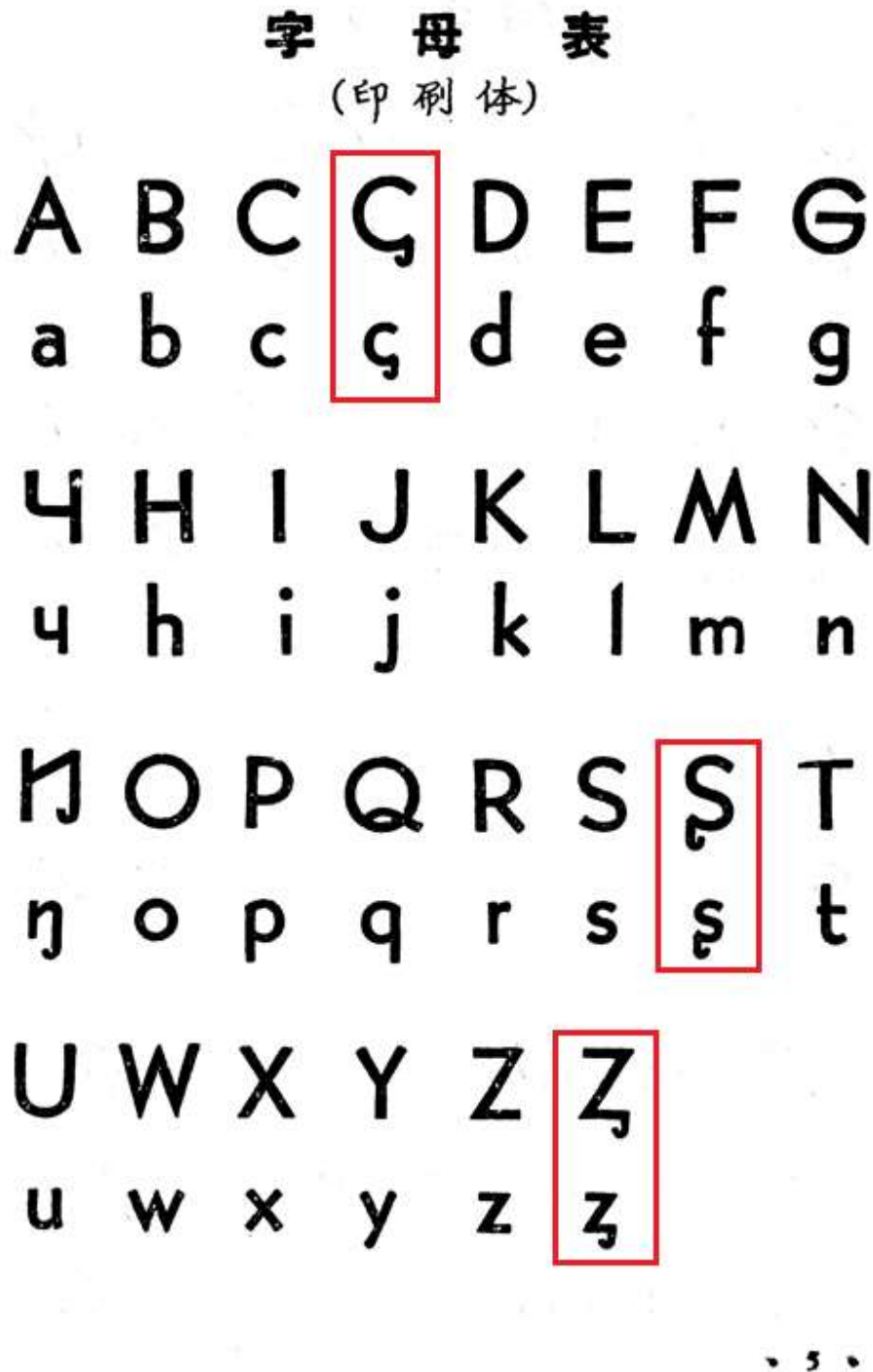
U+A794	ç	LATIN SMALL LETTER C WITH PALATAL HOOK
U+0282	ș	LATIN SMALL LETTER S WITH HOOK
U+1D8E	ȝ	LATIN SMALL LETTER Z WITH PALATAL HOOK

The uppercase and lowercase forms of these three letters were used for writing Mandarin Chinese using the early draft version of *pinyin* romanization during the mid-1950s. A chart showing the uppercase and lowercase forms of the letters used in this early version of *pinyin* is shown in Fig. 1 over the page. The letters C/c with palatal hook correspond to modern *pinyin* CH/ch, S/s with hook correspond to modern *pinyin* SH/sh, and Z/z with palatal hook correspond to modern *pinyin* ZH/zh.

The lowercase letters ç, ș, and ȝ do not have a decomposition mapping to c, s and z with (palatal) hook, so it is not appropriate to represent the corresponding uppercase forms as C, S and Z with (palatal) hook. In order to ensure correct casing behaviour, it is necessary to encode the uppercase letters as distinct characters, with case-mappings to the existing lowercase letters. We therefore propose encoding the three uppercase forms of ç, ș, and ȝ in the LATIN EXTENDED-D block at A7BA..A7BC.

U+A7BA	Ç	LATIN CAPITAL LETTER C WITH PALATAL HOOK
U+A7BB	Ș	LATIN CAPITAL LETTER S WITH HOOK
U+A7BC	Ț	LATIN CAPITAL LETTER Z WITH PALATAL HOOK

Fig. 1: *Hànyǔ pīnyīn fāng'àn (cǎo'àn)* (Běijīng, 1956) p. 5



Additional examples of these letters used in discussions of the draft scheme for *pinyin* published in 1956, as well as in Chinese literature (prose, poem, and song) written using this version of *pinyin*, are given in Figs. 2 through 7.

## 2. Unicode Properties

We suggest encoding the proposed characters at U+A7BA through U+A7BC in the LATIN EXTENDED-D block.

### Proposed Characters

Code Point	Glyph	Character Name
U+A7BA	Ç	LATIN CAPITAL LETTER C WITH PALATAL HOOK
U+A7BB	Ş	LATIN CAPITAL LETTER S WITH HOOK
U+A7BC	Ȥ	LATIN CAPITAL LETTER Z WITH PALATAL HOOK

### UCD Properties

A7BA;LATIN CAPITAL LETTER C WITH PALATAL HOOK;Lu;0;L;;;;N;;;A794;  
A7BB;LATIN CAPITAL LETTER S WITH HOOK;Lu;0;L;;;;N;;;0282;  
A7BC;LATIN CAPITAL LETTER Z WITH PALATAL HOOK;Lu;0;L;;;;N;;;1D8E;

Line break: AL

Script: Latin

The UCD properties for the corresponding lowercase letters should be modified to:

A794;LATIN SMALL LETTER C WITH PALATAL HOOK;Ll;0;L;;;;N;;;A7BA;;A7BA  
0282;LATIN SMALL LETTER S WITH HOOK;Ll;0;L;;;;N;;;A7BB;;A7BB  
1D8E;LATIN SMALL LETTER Z WITH PALATAL HOOK;Ll;0;L;;;;N;;;A7BC;;A7BC

### 3. Figures

Fig. 2: Hànyǔ pīnyīn fāng'àn cǎo'àn tāolùnjí (dì'èrjī) (Běijīng, 1957) p. 125

1. 把 **Zz** 改为 **ʒʒ**, 把 **ㄗ ㄘ** 改为 **ʒ ʒ**;
  2. 把 **Ss** 改为 **ʃʃ**, 把 **ㄙ ㄘ** 改为 **ʃ**, **ㄙ** (德文 **S** 在某种情形下读 **sch**, 在此借用亦说得过去, 不过 **ʃ** 原来用于字的末尾)。
  3. 把 **Cc** 改为 **ʈʈ**, 把 **ㄘ ㄘ** 改为 **ʈʈ**。
- 以上手写体小写 **ʒ**, **ʈ**, **ʃ** 在连写时稍有不便。
4. 把 **ㄌ** 改回 **ㄌ**。
  5. 把 **ㄗ ㄘ** 改回 **ㄗ ㄘ**。

This is a suggestion to replace the three highlighted pairs of letters with the letters shown to their right

Fig. 3: “Guānyú nǐdìng Hànyǔ pīnyīn fāng'àn (cǎo'àn) de jīdiǎn shuōmíng”; in *Rénmín Jiàoyù* 1956 no. 3 p. 31

(4) 适当地注意系统化。为了表达汉语的语音特点, 为了便利教学和一般应用, 拼音方案应该尽可能地做到系统化。例如: **P** (资)、**ㄗ** (雌)、**ㄌ** (思) 用 **Z**、**C**、**S** 表示, **ㄗ** (知)、**ㄗ** (痴)、**P** (诗) 用 **Z**、**C**、**S** 表示, 系统地表明了现代汉语中这两组音的对比。这对方言区和少数民族学习普通话有很大方便。

以上这些原则都是相对的, 不是绝对的, 它们相互之间还有些矛盾。这些原则应该全面地加以考虑, 才能够设计出一个比较完善的方案。

Fig. 4: Je Şouşan, "WO RE'AI PINJIN ZIMU"; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* August 1956, p. 3

## WO RE'AI PINJIN ZIMU

5 nian qian, wo gangan man 7 sui de shihou, mama ba wo sundau Ly-Da Şrfan Fuşu Xiauxye 1 nian 1 ban li, qiu zai lauş de qiaudau xia xyexile Latinxua Sin Wenz. Lauş şouxian qiau womn lianxi bihua, ran womnd şouztou şulian, qieze qiu qiau womn du zimu he xie zimu, houlai hai qiau womn lianxi pinjin, renşırır he lanđu kewen. Bu dau lian ge ye de gunfu, wo he wod tunxyemn dou xyehuile zimu de dufa he xiefa, budan neŋ du Latinxua Sin Wenz heibanbau, erqie hai neŋ yŋ Latinxua Sin Wenz xie xin.

Wo he wod tunxyemn dou gauxiŋ qi le. Womn ye xye ye ai xye. Kexi daule 2 nianqi de shihou, lauş qiu bu qiau le.

Qinnian 1 ye, wo kandaule bauzi şan gunbu de "Han'y Pinjin Faŋan (cau'an)", huanxide liaubude!

Wo huixianqi zai 1 nianqi xyexi Latinxua Sin Wenz de qinquan lai le. Na shihou wo he tunxyemn dou xiwan zuguo kuai şıyŋ zezun wenzi. Xianzai zege xiwan, wo qyede qiu jau şıxian le. Wo je huixianqi zai 2 nianqi liyŋ Latinxua Sin Wenz xyexi Hanzi de hauçu lai le. Na shihou xyexi Hanzi jinwei jou Latinxua Sin Wenz zujin de banzu, rende kuai, dude ju zenqye. Ji huixianqi zexie qinquan, wo qiu xia qyexin kuaidianr xyehui xin pinjin zimu.

Zai ban-ge ye jiqian, wo qiu dau Xinhua Şudian li qy maaile ji ben "Han'y Pinjin Faŋan(cau'an)", mei-tian fanxye huiqia, qiu lianxi du, lianxi xie. Xianzai wo ju xyehuile pinjin zimu le. Wo xinli zen gauxiŋ a!

Ly-Da Şrfan Fuşu Xiauxye 5 nian 1 ban

Je Şouşan

## 我热爱拼音字母

5年前,我刚刚满7岁的时候,妈妈把我送到旅大师范附属小学1年1班里,就在老师的教导下学习了拉丁化新文字。老师首先教我们练习笔画,让我们的手指头熟练,接着就教我们读字母和写字母,后来还教我们练习拼音、认识词儿和朗读课文。不到两个月的工夫,我和我的同学们都学会了字母的读法和写法,不但能读拉丁化新文字黑板报,而且还能用拉丁化新文字写信。

我和我的同学们都高兴极了。我们越学越爱学。可惜到了2年级的时候,老师就不教了。

今年1月,我看到了报纸上公布的“汉语拼音方案(草案)”,欢喜的不得了!

我回想起在1年级学习拉丁化新文字的情况来了。那时候我和同学们都希望祖国快使用这种文字。现在这个希望,我觉得就要实现了。我也回想起在2年级利用拉丁化新文字学习汉字的好处来了。那时候学习汉字因为有拉丁化新文字注音的帮助,认得快,读得又正确。一回想起这些情况,我就下决心快点儿学会新拼音字母。

在半个月以前,我就到新华书店里去买了一本“汉语拼音方案(草案)”,每天放学回家,就练习读,练习写。现在我又学会了拼音字母了。我心里真高兴啊!

旅大师范附属小学5年1班

叶寿山



Fig. 5: Li Qi, "ZOU XIAN ŞENHUO"; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* August 1956, p. 4

## ZOU XIAN ŞENHUO

Li Qi

Janguaŋ zaujauze piŋyan he ŝaŋaŋ,  
Geŝeŋ feiŋaŋ zai zuguo de dadi ŝaŋ,  
Tianje de maisui xian womn zauŝou,  
Guŋçaŋ de qidi wei womn huançaŋ.  
Womn ŝi lauduŋ renmin de erz,  
Yŋyan bu gufu zuguo de xiwaŋ;  
Womn xian ji qyn baised gez,  
Womn jou ji ŝuaŋ qianqiaŋd çıbaŋ.  
Çiaru zuguo xyjau womn qydau naali,  
Womn qiu zankai çıbaŋ xian naali feixian.  
Qin'aid xiauzaŋ qin'aid lauŝi,

Ganxie nimn xinqind peiŋaŋ,  
Muxiau de ruŋy womn jidiŋ jau bauçı,  
Nimnd hua yŋyan qi zai xinŝaŋ,  
Womn ŝenhuo zai xinqud guanhuai zuŋ,  
Womn qian zai douzeŋ li çeŋzaŋ,  
Womn xian ji qyn baised gez,  
Womn jou ji ŝuaŋ qianqiaŋd çıbaŋ,  
Çiaru zuguo xyjau womn qydau naali,  
Womn qiu zankai qianqiaŋd çıbaŋ xian  
naali feixian.

(Niu Yŋfeŋ pinxie)

## GUAIBUDE TINBUQIAN

Liy

Ji tian zauçen, şızen zeñ zı zai 8 dian, mou jiyan menzenbu li, kanbiñ de ren qian-qiandi duoqilai le. Jihuir, huş çuanhu biñren de miñzı, kanbiñ kaişı le. Jiqiñ deñde hen zauqi de xiau-Zaŋ haurunji tinđau le zıqıd miñzı, qiu qimañ zouqin wuz qy, zuo zai jişen de genqian. Jişen qiandandi wenle jixia biñkuan, qiu qiau ta qiekai kouz, yñ tinzenqi qiança tad feibu. Qiança de şıhou, xiau-Zaŋ kanqian jişen de lianse feiçañ qinzañ, qinqin zouze meitou, hen bu naifan di ba xiau-Zaŋ de şenz banlai-banqy, jihuir qiau ta zejañ zuo, jihuir qiau ta neijañ zuo, tinzenqi jihuir tie zai zeli, jihuir tie zai nali. Jişen xianran şı huanle şouqiau, biz li "hu-qi hu-qi" zı çuqi, meimau niñçen ji tiaxian, janqin li liuluçu ji zuñ kunbud şenqi. Jişen zeñun fançañd biauqin, ba xiau-Zaŋ xiahuaile, ta xian: "Ke zaule, jişen zejañ zauqi, şı-buşı wo dele şenme guai-biñ le? Jexy şı wo zuo de zışı bu hau ba? Jexy şı wo bu hui zuo şenhuxi, şıde ta zeme şenqi ba?" Zejañ ji lai, xiau-Zaŋ de şenqing je qinzañ qilai le, xinli putuñ-putuñdi luantiau. Jişen ne? Zai nar ju şı jautou, ju şı tanqi,

şouli naze ji kuai bai-şoupa buduandi caze lianşañ de han. Zeñ zai zege şıhou, huşı cuñ waimian qinlai le, kanqian zeñun qinqin, qiu renbu zu daxiau qilai. Jişen bu zıdau taa xiau şenme, qiu banqi lian wen taa şuo: "Ni xiau şenme?" Huşı jibian xiau, jibian zıze jişen şuo: "Ni kan nid tinzenqi, hai meijou ça zai erduo şañ nel" Jişen yñ şou ji mo, faxian tinzenqi de lian ge gougou, dou gou zai boz şañ, ze cai huan-ran-da-wudi



şuo: "Ol Guaibude tinbuqian!"

Ze şıhou, xiau-Zaŋ je hauxian cuñ emen li xinguolai jijañ, qinqin tule ji kou qi.

(Sun Guibai pinxie)

**Fig. 7:** Waŋ Xiŋian, “FAŊKUAIRZI HE PINJINZI”; in *SHIJAN: PINJIN WENZI SHIJAN DUWU*  
August 1956, p. 3

## FAŊKUAIRZI HE PINJINZI

**Waŋ Xiŋian**

Faŋkuairzi, zen qiguai,  
Jige zir ji faŋkuair,  
**Zuanqilai** neŋ zuan ji koudai.  
Neŋ zuanman ji koudai,  
Ke nan zuanman ji naudai,  
Jau zuanman ji naudai,  
Dei xyexi 3-nian 5-zai.

方块儿字，真奇怪，  
一个字儿一方块儿，  
装起来能装一口袋，  
能装满一口袋，  
可难装满一脑袋，  
要装满一脑袋，  
得学习三年五载。

Pinjinzi, xyede kuai,  
30-ge zimu ji han pai,  
Xyehuile neŋ nian ju hui xie,  
**Şuŋian** ynbuliau ŋige libai,  
Daudi naajan hau, naajan huai?  
Bai zai janqian bu yn cai,  
Xiauhaiz je neŋ kançulai.

拼音字，学得快，  
三十个字母一行排，  
学会了能念又会写，  
时间用不了几个礼拜，  
到底哪样好，哪样坏？  
摆在眼前不用猜，  
小孩子也能看出来。



#### 4. Bibliography

Committee for Language Reform of China (中國文字改革委員會), *Hànyǔ pīnyīn fāng'àn (cǎo'àn)* 漢語拼音方案 (草案) [Chinese pinyin scheme (draft)]. Běijīng: Rénmín jiàoyù chūbǎnshè (人民教育出版社), 1956.

Committee for Language Reform of China (中國文字改革委員會), “Guānyú nǐdìng Hànyǔ pīnyīn fāng'àn (cǎo'àn) de jǐdiǎn shuōmíng” 关于拟定汉语拼音方案 (草案) 的几点说明 [A few points of explanation regarding the draft pinyin scheme]; in *Rénmín Jiàoyù* 人民教育 [Peoples' Education] 1956 no. 3.

Script Reform Publishing House (文字改革出版社), *Hànyǔ pīnyīn fāng'àn cǎo'àn tāolùnjí (dì'èrjí)* 漢語拼音方案草案討論集 (第二輯) [Collection of discussions on the Chinese pinyin scheme draft]. Běijīng: Wénzì gǎigé chūbǎnshè (文字改革出版社), 1957.

Je Şouşan 叶寿山, “WO RE'AI PINJIN ZIMU” 我热爱拼音字母 [I deeply love pinyin letters]; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* 实验：拼音文字实验读物 August 1956, p. 3.

Li Īi (Li Ji), “ZOU XIAN ŞENHUO” (走向生活) [Moving towards life]; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* 实验：拼音文字实验读物 August 1956, p. 4.

Liy (Liyu), “GUAIBUDE TINBUĀIAN” (怪不得听不见) [No wonder I can't understand what's said]; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* 实验：拼音文字实验读物 August 1956, p. 4.

Wan Xiqian (Wang Xijian), “FANĀKUAIRZI HE PINJINZI” (方块儿字和拼音字) [Chinese characters and pinyin letters]; in *SHIJAN: PINJIN WENZI SHIJAN DUWU* 实验：拼音文字实验读物 August 1956, p. 3.

## 5. Proposal Summary Form

<b>ISO/IEC JTC 1/SC 2/WG 2</b> <b>PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS</b> <b>FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup></b> <b>Please fill all the sections A, B and C below.</b> Please read Principles and Procedures Document (P & P) from <a href="http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html">http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html</a> for guidelines and details before filling this form. Please ensure you are using the latest Form from <a href="http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html">http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html</a> . See also <a href="http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html">http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html</a> for latest Roadmaps.
--

### A. Administrative

1. Title:	<b>Proposal to encode three uppercase Latin letters used in early Pinyin</b>		
2. Requester's name:	<i>Andrew West, Eiso Chan, and Michael Everson</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:	<i>2017-01-16</i>		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:	<b>YES</b>		
(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):			<b>NO</b>
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			<b>YES</b>
Name of the existing block:	<i>Latin Extended-D</i>		
2. Number of characters in proposal:			<b>3</b>
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	<b>X</b>	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?			<b>YES</b>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			<b>YES</b>
b. Are the character shapes attached in a legible form suitable for review?			<b>YES</b>
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Michael Everson</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Michael Everson</i>		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			<b>YES</b>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			<b>YES</b>
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)?			<b>NO</b>
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 <a href="http://www.unicode.org">http://www.unicode.org</a> for such information on other scripts. Also see Unicode Character Database ( <a href="http://www.unicode.org/reports/tr44/">http://www.unicode.org/reports/tr44/</a> ) 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: N4102-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?	NO
If YES explain	
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.)?	NO
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	NO
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	NO
If YES, where? Reference:	
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:	