ISO/IEC JTC1/SC2/WG2 N2713

Date: 2004-03-20

ISO/IEC JTC1/SC2/WG2

Universal Multiple-Octet Coded Character Set (UCS) - ISO/IEC 10646

Secretariat: ANSI

Title:	Revised Proposal for encoding A Supplemented Set of IPA Combining Marks, Modifier Letters & Five-Degree Contour Tone Marks in the BMP of the UCS
Source:	China National Body
Action:	For consideration by JTC1/SC2/WG2 and UTC
Distribution:	SC2/WG2 and UTC Experts

On WG2 #44 meeting in San Francisco, N2626 "Proposal on IPA Extensions & Combining Diacritical Marks" from China had widely attention. Most of arguments generally support the proposal. After reviewing the suggestions, the proposed proposal from China is revised as following:

- 1. In the proposed proposal, 24 more phonetic symbols or marks are in three categories whereas 320 phonetic symbols or marks are in the N2626 proposal.
- **2**. In the proposed proposal, 24 more phonetic symbols or marks are in three categories as the following:
 - 1) Keeping the No.A930 diacritic mark in N2626, a dot above right of the symbol [O], as in [O·], representing "greater openness" in vowels (lower vowel);
 - 2) A set of marks representing 8 tone categories in the Chinese languages; (A954~A95B)
 - 3) 15 tone value marks with a five-degree tone value measurement. (A95F \sim A963 \ AA36 \sim AA3A \ AA3B \sim AA3F)
- 3. The actual examples of three categories above mentioned are provided in the proposed proposal.

A. Administrative

1. Title: Revised Proposal for encoding A Supplemented Set of IPA Combining Marks, Modifier Letters & Five-Degree Contour Tone Marks in the BMP of the UCS						
marks, mounter Letters & Five-Degree Contour Tone marks	S III LITE DIVIF OF LITE OCS					
2. Requester's name: China						
3. Requester type (Member body/Liaison/Individual contribution):	Member body					
4. Submission date:	2004-03-19					
5. Requester's reference (if applicable): N2626						
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:	No
. b. The proposal is for addition of character(s) to an existing block: Name of the existing block:	Yes
IPA Extensions, Spacing Modifier Letters & Combining	Diacritical Marks
2. Number of characters in proposal:	24
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)	ection)
C-Major extinct D-Attested extinct E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage	symbols
4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):	1
Is a rationale provided for the choice?	No
If Yes, reference:	Yes
a. If YES, are the names in accordance with the "character naming guidelines"	165
in Annex L of P&P document?	Yes
b. Are the character shapes attached in a legible form suitable for review?	Yes
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostSc	cript format) for
publishing the standard? Institute of Linguistics, Chinese Academy of Sc	
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and	d indicate the tools
used: No.5 Jianguomennei Dajie, Beijing 100732, China	
Email: fy_yys@cass.org.cn; baichm@cass.org.cn 7. 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 of	
of proposed characters attached?	Yes
8. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) su	
presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose infor	
9. Additional Information:	No
Submitters are invited to provide any additional information about Properties of the proposed Cha	aracter(s) or Script
that will assist in correct understanding of and correct linguistic processing of the proposed chara	acter(s) or script.
Examples of such properties are: Casing information, Numeric information, Currency information	
information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode	
related information. See the Unicode standard at http://www.unicode.org for such information on	
see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Report	rts for information
needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Stand	lard.
C. Technical - Justification	
Has this proposal for addition of character(s) been submitted before?	Yes
If YES explain N2626 "Proposal on IPA Extensions & Combining Diacritical Marks for IS	
BMP"	
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?	National Body
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example:	<u> </u>
size, demographics, information technology use, or publishing use) is included?	Yes
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	Worldwide
2,	

6. After giving due considerations to the principles in the P&P document must the proposed characters in the BMP?	be entirely Yes
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)	? <u>No</u>
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 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 symborovided?	ols)
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 character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	

D. Proposal

Part I. Introduction

Designed by the Institute of Linguistics of Chinese Academy of Social Sciences and made public on *Dialect (Fangyan)* No. 2, 1979, the notational set of 494 IPA extensions and 106 tone marks has always served as the actual academic as well as industrial standard in China. At present, it is widely used by the communities of linguistics, education, and the publishing industry in language survey, linguistic research, language teaching, dictionary compilation, and publications in those fields. With the advent of the digital era, this notational set is also adopted in the typesetting integration platform by Founder Electronics, the largest pre-print software developer of Chinese electronic publishing.

The present IPA Extensions, Combining Diacritical Marks for the Chinese languages in ISO/IEC 10646 cannot fully meet the actual needs of the Chinese communities of linguistics, culture and education, publication, and information technology. After careful examination and reviewing of the present IPA notational set, we find it urgent to add **24** more phonetic symbols in three categories as the following:

- 1. A combining dot above right representing open vowel featuring lower tongue position;
- 2. A set of modifier letters representing 8 tone categories in the Chinese languages;
- 3. 15 five-degree contour tone marks, including 10 five-level dotted forms for neutral tones and 5 five-degree contour tone marks.

Part II. Progress

The present 1979 *Dialect (Fangyan)* notational set was a meticulous design with reference to the rules and standards made by the International Phonetic Association on the basis of an extensive collection of the IPA symbols or marks ever used to represent the Chinese languages by Chinese linguists and publishers. In the recent years, the Institute of Linguistics of Chinese Academy of Social Sciences has called on for several times specialists in this area to discuss the writing standard and number of combining diacritical marks and modifier letters. It has been proposed to supplement and finally consummate the present set by determining the coding rules of IPA set of symbols and adding combining marks, modifier letters and five-degree contour tone marks to the 1979 set with ISO/IEC 10646 standards.

Part III. Principles

- 1. The proposed new IPA set for the Chinese languages is an extension of IPA Extensions, Spacing Modifier Letters & Combining Diacritical Marks in ISO/IEC 10646;
- 2. The proposed new set shall be capable of expressing and exchanging all data in the IPA documents or documents with IPA content, and meeting the needs of the Chinese linguists, educationists, publishers, and software developers;
- 3. The ordering of the proposed new set ought to be in accordance with that of the basic IPA set:
- 4. The proposed new set ought to be technically operable in information processing.

Part IV. Technical Demonstration

- 1. The ordering of the proposed new set ought to be in accordance with that of the basic IPA set;
- 2. The proposed new set ought to observe the principles of code writing and compositional features as issued by the International Phonetic Association, and capable of meeting the academic and industrial standards as much as possible;

3. Each one of the symbols and marks in the proposed new IPA set ought to be of one-one correspondence with one code, so that all the information with IPA marks as well as the Chinese characters represented could share the same plane of ISO / IEC 10646 to actualize the standardized information processing of the data with IPA all over the world.

Part V. Prospects

As a result of the active advocacy to establish a set of phonetic symbols and marks for the Chinese languages compatible with the internationally used IPA standards to help solve the related difficulties in linguistic research, language teaching, language engineering, and language information processing, the proposed new set has been repeatedly discussed by experts from all those fields. Should this proposed set be accepted as a part of the plane of ISO/IEC 10646, it will surely facilitate the globalization progress of information processing with IPA.

Part VI. Examples of the proposed new IPA symbols and marks

1. Combining mark No. A900, first used in some Western missionaries' works on Minnanhua, a sub-dialect of the Min group of the Chinese dialects, is a dot above right of the symbol [O], as in [O], representing "greater openness" in vowels (lower vowel). This mark was used in at least four early *pinyin* schemes, among more than five all together, to phonetically represent Minnanhua, and was also used later in most of the works on Minnanhua published in Taiwan. The following is some examples of its use as modifier letters and independent mark, respectively: The following example can be seen in Luo Changpei, the well-known modern linguist, 1999:42.

42 3	常常文集·第	一島・厦门音)	6		
a.)	n al				
新式	周辨明	Campbell	Douglas	Doty	Medhurst
Ь	Р	P	Р	þ	p
p	ph	ph	ph	p	ph
bb	ь	ь	ь	b	b
d	t	1	t	t	t
ι	th	th	th	ť.	ťh
g	k	k	k	k	k
k	kh	kh	kh	k	kh
98	g	g	g	g	g
ng	ng	ng	ng	ng	gn
	在四科	•厦门话拼 ○上角加水	音方案中 - 圆点表示	,都在	w.y.□
ts	量类	(舌位)) (カロ皮 	ch
ts	ch	chh	ehh	ch'	eh'h
dz	j	j	j	j	j
b.	前 母				
0	ő	oʻ	θ'	oʻ.	oe
e	e	e	e	е	ey
aí	ai	ai	ai	ai	ac
UAA	ati	811	au	au	aou
aft	a ⁶	a*	an	n ^o	9 _{uk}

A dot above right of the symbol [O], as in [O], representing "greater openness" in vowels (lower vowel) in four *pinyin* schemes on the Xiamen dialect.

This mark can also be seen on *Dialect (Fangyan)*, the well-known academic magazine, 1979, 2:160.

a*	•a	•a	`a	*8	*a
418	419	420	421	422	423
a ⁻	a'	a*	a'	a,	a*
436	437	438		440	441
a:	a*	a	a*	a.	*a
454	455	45	457	458	459
a†	a‡	a!	475	a 7	a)
472	473	474		476	477
η	2	Ž,	щ	a ⁷	a⊙
485	486	487	488	489	490

2. As first found four tones in the South-north Dynasty (420-589 A.D.) and recorded in the ancient Chinese phonological literature such as *Qieyun* published in 601 A.D., scholars termed the tones of Chinese as in its ancient form as *ping*, *shang*, *qu*, and *ru* which were represented using a half-circle at either of the four corners above or below a Chinese character. The number of tones of many Chinese dialects doubled to eight, respectively four for *Yang* tones and four for *Yin* tones, due to the influence of the dichotomy of voiced vs. unvoiced of the initials in the ancient times, and the former ones of which were represented using a low line combined with the half-circle. This method of marking the tones is the well-known Faquanfa, or the half-circle marking method. This method is convenient in language survey and research because it can not only represent tone categories of the contemporary Chinese dialects, but also reflect their ancient origins. The well-known modern linguists, first Swedish linguist B. Karlgren, and later Y-R Chao, Luo Changpei, Li Fanggui, and others all inherited this set of tone marks. Up to now this tone-marking method can still be found in the large number of works on Chinese languages and their phonologies.

The following example can be found on Karlgren (1940: 541):

	緒	論		541
專寫它們多	立一行,非	戊就只把因	調變而發生的	異讀放在註裏。
上啊,——	我們這裏	既然是做	的母變化的總則	展,——就把它放
延頭,並且有	小號碼跟	正文裏特	票的號碼對照信	着查。
上去符	虎:		<i>t</i> →	



员裏就可以看得出來了。 但在現代方言裏,調的陰陽不全能從

向來用這樣的辦法分辨:



《不大用。 要詳細研究現代方言的聲調,這工作要走得太遠

暫且放在一邊作為另一個專題的研究。

A half-circle can be at either of the four corners above or below an IPA symbol. A half-circle can be at either of the four corners above or below a Chinese character or other phonetic mark.

The following example can be found in Y-R Chao' (1948:17 General introduction.):



These modifier letters can also be seen on *Dialect (Fangyan)*, 1979, 2:157.

肆 :	本刊第一期用的 阴上	703—470 号 阴去	八个 阿	一亏太小,改 阳平	阳上	- 708 号调类 阳去	符号。 阳 A
。了 701	[°] 顶 703	订 [°] 705	滴。 707	ç亭 702	锭 704	定 ² 706	笛 ₂ 708
1979 年第	52期						. 157 .

A half-circle is at four corners above or below a Chinese character.

3. It had been difficult to describe the tone values of the dialects and minority languages in China before the five-degree contour tone marking method was proposed by Y-R Chao. Tone values of the Chinese languages used to be described with explanatory words or be compared to intonation in English or French languages in the early survey reports and teaching textbooks by the Western missionary scholars. Obviously, it was difficult to describe precisely the tone values in their great number and complexity as in the Chinese languages using those methods. The well-known Linguists like Y-R Chao, Liu Fu, and Luo Changpei once tried to record the tone values of the Chinese languages in terms of stave as musicians recording the scores, which could be more precise yet very troublesome to operate. It was only after the creation of the five-degree contour tone marking method by Y-R Chao was the issue of describing the tone values of the Chinese languages fully resolved.

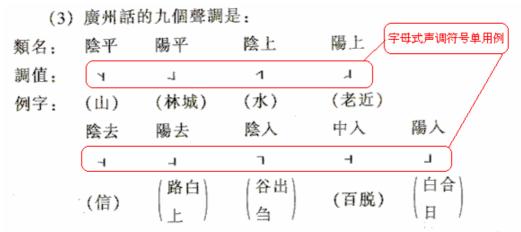
The creation of the set of five-degree contour tone marks is based on the five-degree method of marking the tone pitches of the Chinese languages, which divides the tone pitches into the five degrees of extra low, low, middle, high, and extra high, and divides the tone categories into five patterns of level, falling, rising, fall-rising and rise-falling. The set of five-degree contour tone marks has been widely used since its creation, because its precise grasp of the two key factors of tone pitch and tone pattern has enabled this method to describe perfectly the tone values of not only the Chinese languages, but also any other tonal language.

The set of five-degree contour tone marks is generally used as a set of diacritics, and only sometimes independently. The set of marks is widely used in all kinds of Chinese languages as well as other languages of East Asia. The following is an example of its use as diacritic and independent mark, respectively:

For Example 1 of describing Chinese dialects with five-degree contour tone marks, see *Dialect (Fangyan)* 1979, 2:157:

al	aJ	8.4	a.a	a1
	12:		14:	15:
			-,	
a.1	a.1	aı	a4	a4
21:	22:	23:	24:	25:
aJ	a4	ач	a-1	a1
31:	32:	33:	34:	35:
av	a٩	ач	ан	ап
41:	42:	43:	44:	45:
av	ач	ач	ar	ат
51:	52:	53:	54:	55:
aA	an	an		
131:	143:	153:		
a.ı	aJ	a.a	ал	aA
212:	213:	214:	215:	242:
av	ал	a-ı	an	an
313:	315:	325:	351:	353:
au	ач			
412:	424:			

For Example 2 of describing Chinese dialects with five-degree contour tone marks, see the well-known linguist Dong Tonghe (2001:47):



陰上與陽上大致在國語與蘇州都是上聲,只有一些如"近"官 話是去聲而吴語為陽去。

Examples of tone value marks.

The following is Example 3 of describing minority languages in China with five-degree contour tone marks:

```
    xa1 30 '液(動詞)'
    xa1 60 '寒*

    xa1 30 '下肢'
    xa1 60 '寒*

    xa1 30 '下肢'
    xa1 60 '寒*

    ma1 20 '猪笆'
    xa1 60 'ヌ草'

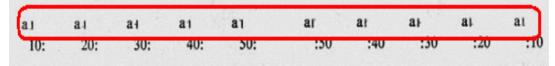
    ma1 20 '猪笆'
    xa1 60 'ヌ草'
```

The following is Example 4 of describing minority languages in China with five-degree contour tone marks:

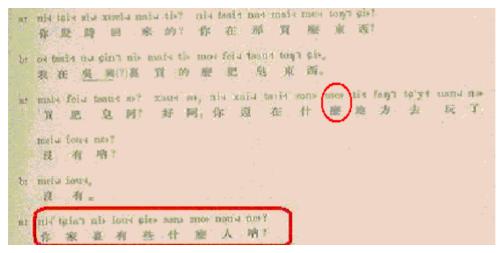
加查加	台	证	黄	275	篮	th	雷	th	劍	河
如	調值	例字	調佐	例字	网位	例字	测值	例字	测值	例字
1	433	ki+角	433	ka×炒	433	ta + 回答	433	ki + 19	433	tu-f带
2	√53	kin笙	755	kaフ拖	7 ⁵⁵	ta 7 殊	453	ki + 笙	₄ 5a	tu+步伐
8	444	ki + MC	744	ka 1 仮	744	ta 1長	144	tsa1房子	144	tu1便
4	422	ki ų 界	422	ka 1矮	422	tal遺失	122	ka 4 A	- 0	
5	735	kin族	A ¹³	ka / M	A13	ta /新	735	ki 1櫃	435	加力
6	423	ki a 🎹	421	ka u 側	A ²³	ta×死	A ²³	ki a III		
7	A ¹³	kiょ则	753	ki v ŋ	753	tan翅膀	A18	ki i 剪	A13	tu笑
8	J ²¹	kiょ揺				C-VISIT	₃ 21	maiょ砍	J ²¹	tus跨

Neutral tone is an important phonetic phenomenon of the tonal languages. The general neutral tone's feature of the Chinese Languages is lighter and shorter than the original syllable tone pitch. The neutral tone pitch can be largely changed on the different phonetic conditions, e.g. the neutral tones' pitches of the following three words, such as 奶奶(grandmother)[nai213 nai213-30], 妹妹(young sister)[mei51 mei51-10] and 哥哥(brother) [ke55 ke55-40] in Beijing Mandarin. The Chinese linguists have two descriptions for neutral tones: one is a middle-dot before the syllable, and the one is five-level dotted forms. The difference between the five-level dotted forms for neutral tones and the five-degree contour tone marks is that they are some high or dots rather than short lines. The following is an example of its use as modifier letter dotted tone bar, respectively:

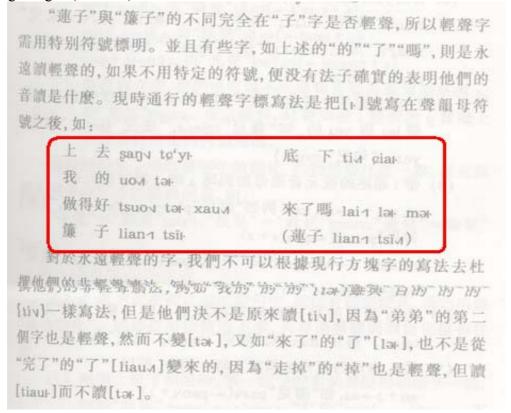
For Example 5 of describing Chinese dialects with 10 five-level dotted forms for neutral tones, half is modifier letter dotted tone bar and the other half is modifier letter left-stem tone bar, see *Dialect (Fangyan)* 1979, 2:157:



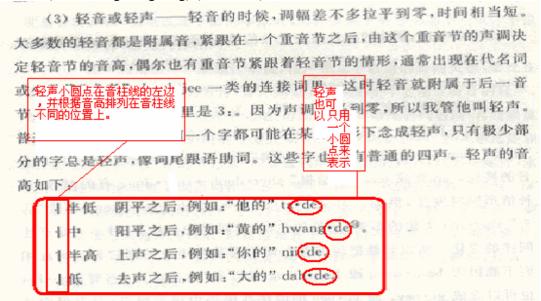
For Example 6 of describing Chinese dialects with five-level dotted forms for neutral tones, can be found in Y-R Chao' (1948:67):



For Example 7 of describing Chinese dialects with five-level dotted forms for neutral tones, see Dong Tonghe (2001:25):



For Example 8 of describing Chinese dialects with five-level dotted forms for neutral tones, see Y-R Chao' (Trans.Ting Pangtsing) (2002:225):



A different dotted bar.

A dot is before (left) the syllable.

Part VII. Reference

- 1. 方言编辑部,1979,音标及其他记音符号。《方言》第2期157页、160页。 Editors 1979. Yinbiao ji qita jiyinfuhao (List of IPA phonetic symbols and other phonetic marks). *Dialect (Fangyan)* 2:157 and 160.
- 2. 董同龢,1965,《汉语音韵学》台湾:学生书局。第47页。2001 重版,北京:中华书局。
 - Dong, Tonghe. 1965. *Zhongguo Yinyunxue* (*Chinese Phonology*). Taiwan: Student Press. P.47; also see reprint. 2001. Beijing: Zhonghua Shuju. P47.
- 3. 高本汉(Karlgren, Bernhard), 1915, Etudes sur la Phonologie Chinoise. 瑞典乌普萨拉大学博士论文。中译本: 赵元任、罗常培、李方桂译, 《中国音韵学研究》1940, 北京: 商务印书馆。541页。
 - Karlgren, Bernhard. 1915. *Etudes sur la Phonologie Chinoise*. Ph.D, Uppsala University, Sweden. Trans. Y-R Chao, Luo Changpei, Li Fang-Kuei. *Zhongguo Yinyunxue Yanjiu*. 1940. Beijing: Commercial Press. P.541.
- 4. 罗常培, 1956, 《厦门音系》, 上海: 科学出版社重印(原刊于《中央研究院历史语言研究所集刊》1930分册)。又见: 《罗常培文集》1999, 济南: 山东教育出版社。Luo, Changpei. 1956. Xiamen Yinxi (The Phonology of Xiamen Dialect)(Reprint), Shanghai: Science Press. First published in 1930, Zhongyang Yanjiuyuan Lishi Yuyan Yanjiusuo Jikan (Bulletin of the Institute of History and Philology, Academia Sinica); also see: 1999. Luo Changpei Wenji (Collection of Luo Changpei). Ji'nan: Shandong Educational Press.
- 5. 赵元任等, 1948, 《湖北方言调查报告》北京: 商务印书馆。 Chao, Y-R, etc. al. 1948. General introduction. Hubei Fangyan Diaocha Baogao (A Survey Report on the Dialects in North Hubei Province). Beijing: Commercial Press. P.17 and P.67.
- 6. 赵元任,1968, 《中国话的文法》,加利福尼亚大学出版社。中译本:丁邦新译,《赵元任全集·中国话的文法》2002,北京:商务印书馆。Chao, Y-R. 1968. *A Grammar of Spoken Chinese*. University of California Press. Trans.

Table 1 – Row A9: IPA Combining Diacritical Marks & Tone-letter System

	A90	A91
	ं	ŀ
0	000	016
	c	Į.
1	001	017
•	<u>c</u>	l.
2	002	018
	C	Γ
3	003	019
	<u>S</u>	⊢
4	004	020
	>	⊢
5	005	021
	3	⊢
6	006	022
7	>	L
′	007	023
	5	
8	008	
_	1	
9	009	
	1	
A	010	
	-{	
В	011	
~	-1	
С	012	
D	.I	
ט	013	
E	ľ	
Ľ	014	
T-7	ľ	
F	015	

Table 2 – Row A9: IPA Combining Diacritical Marks & Tone-letter System

HEX	NAME
A900	combining dot above right
A901	modifier letter Chinese tone Yin Ping
A902	modifier letter Chinese tone Yang Ping
A903	modifier letter Chinese tone Yin Shang
A904	modifier letter Chinese tone Yang Shang
A905	modifier letter Chinese tone Yin Qu
A906	modifier letter Chinese tone Yang Qu
A907	modifier letter Chinese tone Yin Ru
A908	modifier letter Chinese tone Yang Ru
A909	modifier letter extra-high dotted tone bar
A90A	modifier letter high dotted tone bar
A90B	modifier letter mid dotted tone bar
A90C	modifier letter low dotted tone bar
A90D	modifier letter extra-low dotted tone bar
A90E	modifier letter extra-high dotted left-stem tone bar
A90F	modifier letter high dotted left-stem tone bar
A910	modifier letter mid dotted left-stem tone bar
A911	modifier letter low dotted left-stem tone bar
A912	modifier letter extra-low dotted left-stem tone bar
A913	modifier letter extra-high left-stem tone bar
A914	modifier letter high left-stem tone bar
A915	modifier letter mid left-stem tone bar
A916	modifier letter low left-stem tone bar
A917	modifier letter extra-low left-stem tone bar