Before the

Federal Communications Commission

Washington, D.C. 20554

In the Matter of	)	
	)	
The Emergency Alert System	)	EB Docket No. 04-296
	)	
Frank W. Bell	)	

Re: NOTICE OF PROPOSED RULEMAKING 21st Century Video Access Act.

The ability of use EAS to alert disabled persons has a number of issues that relate to other EAS issues. So the following document is not restricted to just the disabled aspect. For example, to send alerts to the blind and deaf would provide for the use of Braille. Braille is included in the Unicode system of alphabets and is listed near the end. How to include Unicode in the EAS system is described elsewhere, but is described in outline next. Following is an adaptation of the ASCII ISO 8859-1 to significantly extend its capabilities. One inclusion is the provision of IPA (International Phonetic Alphabet), which can assist Text to Speech (TTS) capabilities, and Braille is included. The keyboard is also extended, based on the Japanese keyboard which has a few extra buttons. Also the status LED system is increased from 3 single color LEDs to 4 dual color LEDs, which uses all the bits of the LED register.

language; A language selection scheme is defined within the constraints of the EAS protocol, and translation between RFC 3066 is either achievable or it is a single nation language and can be handled by the Unicode standard. The details are in the EAS+ documentation. If this is not specified, then "en-US" SHALL be assumed

-JJJHHMM-. This header code block identifies the Julian calendar date and the time the message was originally disseminated in hours and minutes using the 24-hour Universal Time Coordinated (UTC) clock.

An implication of the New Orleans experience of EAS performance is the desirability to be able to carry different languages. Also an implication of specific message coding is to be able to select appropriate language messages by different users. This means that language identification SHOULD be in the EAS header. While the header has everything assigned, a redefinition is proposed for the first J of JJJ, the Julian calendar day of the year. This J at present can only have the ASCII values of 0, 1, 2 or 3. So the proposal is to keep this the same for English. The date only requires the last two bits. So use the first six bits as follows

Binary 000000 Octal 00 Use for National or local language, ASCII 8 bit. “Local1”

Binary 000001 Octal 01 Use for National or local language, Unicode extended data after Lat-Long. “Local1”

Binary 000010 Octal 02

To To be assigned to multi-country or major languages, 10 codes

Binary 001010 Octal 12

Binary 001011 Octal 13 Dutch

Binary 001100 Octal 14 English

Binary 001101 Octal 15 Spanish

Binary 001110 Octal 16 French

Binary 001111 Octal 17

To To be assigned to multi-country or major languages, 16 codes

Binary 001110 Octal 36

Binary 011111 Octal 37 Use of International Phonetic Alphabet (IPA) and others by bank switching extended ASCII.

Hexadecimal 0x80

To Reserved to keep 7 bit ASCII format

Hexadecimal 0xFE

Hexadecimal 0xFF Language is not selected for this message. For example, satellite use.

These characters will read as ASCII "0", "1", "2", "3" for English, "4", "5", "6", "7" for Spanish (i.e. subtract 4 for the date value). "8", "9", ":", ";" for French as the date hundreds change. The rest are more difficult and not a current concern for the U.S. EAS system. However a few examples of multi-country languages are:

German is the language of Germany, Austria and Switzerland, so it needs a code.

Korean is the language of the Republic of Korea and the Democratic Peoples' Republic of Korea, so it needs a code as it is multi-country.

Chinese has many languages/dialects with one writing system. It is used widely in Singapore for example. So to provide the local language option for another spoken language, Chinese needs a code.

Russian and Arabic are multi-country languages. Japanese is an important language, and they have an extensive alert system also.

Latin is the international language of botany and zoology, so it needs a code.

Esperanto is neither a national or local language, but it is an official language of the U.N. so it needs a code.

Greek is an historical language and is used in more than one country, so it needs a code.

As Unicode has been proposed, perhaps the languages can be grouped into those that would use extended ASCII and those that would use Unicode for the extended data. However U.S. ASCII SHALL be the basis for the header code e.g. event codes, originators, etc. unless otherwise specified later.

The first H SHALL have the first 6 bits defined with a default of 001100 and reserved with the last two bits used for tens hour value. The second H SHALL have the first 4 bits defined with a default value of 0011 and reserved with the last four bits used for units hour value. The first M SHALL have the first 5 bits defined with a default of 00110 and reserved with the last three bits used for the tens minutes value. The second M SHALL have the first 4 bits defined with a default of 0011 and reserved with the last four bits used for the units minutes value. If Unicode is mixed with ASCII in the text, the start delimiter SHALL be =?. The exit of the Unicode mode appears to be defined by ISO 2022 and 6049 as ?=. This SHALL be done before the pause and end of the message.

Multiple languages MAY be used on radio and TV. The bandwidth of TV permits a maximum of four AES streams, therefore with two, four mono languages to be transmitted simultaneously, or in parallel. As legacy TVs will not distinguish languages, they SHALL still be in serial for the interim. Dolby Digital MAY provide a solution for legacy TVs. The limited bandwidth of HD radio would only permit one language after the other, or in serial mode. The identification of the language 1 is as above. The other languages are indicated as below; **ISO 10646-1 Unicode**

The Use of ASCII and its Adaptation for Multiple Alphabets and Keyboards. Frank W. Bell

Introduction

The most extensive ASCII standard currently is ISO 8859-1. However it is only for Roman based alphabets. Byt the use of four characters, SO, SI, DC2 and DC4 a bank switching scheme is enabled. DC1 and DC3 are reserved for flow control. Also various keyboard layouts and other alphabets are considered so as to make a combination version. This is based on the Japanese keyboard which has a few more keys for other functions, which are also applied to other alphabets particularly Chinese. Braille is included at the end.

Contents

The Use of ASCII and its Adaptation for Multiple Alphabets and Keyboards. Frank W. Bell 4

Introduction 4

ASCII Code Table 6

Other European Table 19

IPA, the International Phonetic Alphabet 22

Notes on Keyboard Alphabet Switching and other items. 30

Keyboard Layout Chart following 33

Japanese Keyboard Layout Chart 47

Korean Hangul 56

Cyrillic 58

Russian, Belarusian and Ukranian standard keyboards 60

MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media 69

Code Table Basic Arabic 69

India, Devanagari 74

Special code points 76

Tibetan and Dzongkha (Bhutan) 81

Vowels 82

Consonants 84

Arpabet Symbols 86

Stress 86

Vowels 87

Consonants 88

Chinese 91
Finals 93
Rules given in terms of English pronunciation 95
Pronunciation of initials 96
Pronunciation of finals 98
Orthography 100
Letters 100
Word formation, capitalization, initialisms and punctuation 101
Tones 106
Numerals in place of tone marks 109
Rules for placing the tone mark 110
Using tone colors 111
Third tone exceptions 111
The *ü* sound 112
Pinyin in Taiwan 112
Comparison with other orthographies 113
Chart of comparison with other romanizations 114
Braille 118
Assignment 120
Form 121
Writing braille 122
Letters 122
Formatting 123
Punctuation 123
Contractions 124
Page dimensions 125
Braille transcription 125
Braille translation software 127
Braille-reading techniques 127
International uniformity 128
Other braille conventions 129
Uses 129

Braille phone 130
Unicode 131
Appendix J: Text to Speech Tips 133

ASCII Code Table

ASCII char		ISO 8859-1	ASCII uses, Greek, Hangul, IPA	Working Notes	Working Notes 2	The first ^ is CTRL-, C use
	&#00;	Unused	NUL null			^@ \0
	&#01;		SOH start of header			^A
	&#02;		STX start of text	Text mode	Note 45	^B
	&#03;		ETX end of text	Numeric mode	Note 45	^C
	&#04;		EOT end of transmission			^D
	&#05;		ENQ enquiry			^E
	&#06;		ACK Acknowledgement			^F
	&#07;		BEL Bell	Next item in row	e.g. matrix or	^G \a
	&#08;		BS backspace	Next row	determinant	^H \b
		Horizontal tab	Horizontal tab			^I \t
Line feed	Line feed			^J \n
	&#11;	Unused	VT vertical tab			^K \v
			NP or FF form feed			^L \f
Carriage Return	Carriage Return			^M \f
	&#14;	Unused	SO (ASCII BASE)			^N
	&#15;		SI shift in (IPA UC mostly)			^O
	&#16;		DLE data link escape			^P
	&#17;		DC1 XON			^Q
	&#18;		DC2 (shift in 2) (GREEK/HANGUL/IPA LC)			^R
	&#19;		DC3 XOFF			^S
	&#20;		DC4 (Unicode/other)			^T
	&#21;		NAK negative ACK			^U
	&#22;		SYN sync idle			^V

	&#23;		ETB end transmission block			^W
	&#24;		CAN cancel			^X
	&#25;		EM end of medium			^Y
	&#26;		SUBstitute			^Z
	&#27;		ESCape			^[
	&#28;		FS file separator			^\
	&#29;		GS group separator			^]
	&#30;		RS record separator			^^
	&#31;		US unit separator			^_
		Space (Em)		En space, non-breaking	Use CTRL-space	Note 52 for numbers
!	!	Exclamation mark		CTRL		
"	"	Quotation mark		ALT		
#	#	Number sign		DEL		
$	$	Dollar sign		Fn		
%	%	Percent sign		WIN/MAC		
&	&	Ampersand		0xA		
'	'	Apostrophe		0xB		
(	(	Left parenthesis		0xC		
)	)	Right parenthesis		0xD		
*	*	Asterisk		0xE		
+	+	Plus sign	**• Bullet**	0xF		
,	,	Comma		Cursor up	CSV field delimiter	May display as triple digit separator, but not included as data. Decimal point not required in value.
-	-	Hyphen		Cursor down		
.	.	Period (fullstop)		Cursor right		Decimal point.
/	/	Solidus (slash)		Cursor left		
0 - 9	0	Digit 0	0	reserved	reserved	reserved
1	1	Digit 1	1	reserved	reserved	reserved
2	2	Digit 2	2	reserved	reserved	reserved
3	3	Digit 3	3	reserved	reserved	reserved
4	4	Digit 4	4	reserved	reserved	reserved
5	5	Digit 5	5	reserved	reserved	reserved
6	6	Digit 6	6	reserved	reserved	reserved
7	7	Digit 7	7	reserved	reserved	reserved

8	8	Digit 8	8	reserved	reserved	reserved
9	9	Digit 9	9	reserved	reserved	reserved
:	:	Colon	Colon	Intersection		
;	;	Semi-colon	Semi-colon	Union		
<	<	Less than	Less than	Less than or equal to		
=	=	Equals sign	Equals sign	Equivalent to		
>	>	Greater than	Greater than	Greater than or equal to		
?	?	Question mark	Question mark	Right arrow		
@	@	Commercial at	Commercial at	Not equal to		
A	&#65	A	ㅁ (m)	↗ global rise		
B	&#66	B Voiced bilabial trill (b lower case)	ㅠ (yu)	↘ global fall		
C	&#67	C Voiceless palatal plosive	ㅊ (ch)	⊙	Bilabial click	
D	&#68	D	ㅇ (-/ng)	ǀ	Dental click	
E	&#69	E	ㄷ (d)	!	(Post) alveolar click	
F	&#70	F Voiceless labio-dental fricative	ㄹ (l/r)	ǂ	Palatal-alveolar	
G	&#71	G	ㅎ (h)	ǁ	Alveolar lateral click	
H	&#72	H	ㅗ (o)			
I	&#73	I	ㅑ (ya)	ɓ	Bilabial	Voiced implosive
J	&#74	J	ㅓ (eo)	ɓ	Bilabial	Voiced implosive
K	&#75	K	ㅏ (a)	ɗ	Dental / alveolar	Voiced implosive
L	&#76	L Voiced velar lateral approximant	ㅣ (i)	ɗ	Dental / alveolar	Voiced implosive

M	&#77	M	ㅡ (eu)	’	Ejective	
N	&#78	N	ㅜ (u)			
O	&#79	O	ㅐ (æ) e.g hay	ʃ	Palatal	Voiced implosive
P	&#80	P	ㅔ (e)	ʃ	Palatal	Voiced implosive
Q	&#81	Q Voiced uvular plosive,	ㅂ (b)	ɠ	Velar	Voiced implosive
R	&#82	R	ㄱ (g)	ɠ	Velar	Voiced implosive
S	&#83	S	ㄴ (n)	ɢ	Uvular	Voiced implosive
T	&#84	T	ㅅ (s)	ɢ	Uvular	Voiced implosive
U	&#85	U	ㅓ (yeo)	ɽ	Voiced retroflex tap/flap	
V	&#86	V Voiced labio-dental fricative	ㅍ (p)	ɽ	Voiced retroflex tap/flap	
W	&#87	W	ㅈ (j)	ɬ	Voiceless alveolar lateral fricative	
X	&#88	X Voiceless velar fricative	ㅌ (t)	ɬ	Voiceless alveolar lateral fricative	
Y	&#89	Y	ㅛ (yo)	ɻ	Voiced retroflex approximant	
Z	&#90	Z	ㅋ (k)	ɻ	Voiced retroflex approximant	
[	[	Left square bracket	Left square bracket			
\	\	Reverse solidus (backslash)	Reverse solidus (backslash)			\\
]	]	Right square bracket	Right square bracket		Left square bracket	
^	^	Caret	Caret			
_	_	Horizontal bar (underscore)	Horizontal bar			

			(underscore)			
`	`	Acute accent	Acute accent			
a - z	a - z	Letters a-z	Phonetic IPA		Hangul note	
a	&#97	a	m	M	ㅁ (m)	Voiced nasal bilabial
b	&#98	b Voiced bilabial trill	ʃ	ʃ Sh(oe)	ㅠ (yu) unused	ju
c	&#99	c Voiceless palative plosive, use C upper case	ʒ	Ʒ	ㅊ (ch)	Post-alveolar fricative
d	&#100	d	ŋ	Ŋ	ㅇ (-/ng)	Voiced nasal velar
e	&#101	e	d	D	ㄷ (d)	Voiced alveolar plosive
f	&#102	f Voiceless labio-dental fricative, use F upper case	r	R	ㄹ (l/r), r only	Voiced uvular trill L/l for l
g	&#103	g	h	H	ㅎ (h)	Glottal fricative
h	&#104	h	o	O	ㅗ (o)	
i	&#105	i	ɾ	ɾ	ㅑ (ya) unused	ja alveola lateral approximant
j	&#106	j	ə	Ə	ㅓ (eo)	
k	&#107	k	a	A	ㅏ (a) after /aːftə/	
l	&#108	l Voiced velar lateral approximant	i	I	ㅣ (i) slip /slɪp/	
m	&#109	m	ʊ	U	ㅡ (eu)	Voiced labio-dental approximant
n	&#110	n	y	Y	ㅜ (u)	
o	&#111	o	O æ	Æ	ㅐ (ae) cat /kæt/	
p	&#112	p	e	E	ㅔ (e) ten /ten/	
q	&#113	q Voiced uvular plosive, use Q upper case	b	B	ㅂ (b)	Voiced bilabial plosive

r	&#114	r	g	G	ㄱ (g)	Voiced uvular plosive
s	&#115	S	n	N	ㄴ (n)	Voiced alveolar nasal
t	&#116	t	s	S	ㅅ (s)	Alveolar fricative
u	&#117	U	z	Z Z(oo)	ㅕ (yeo) unused	Jo
v	&#118	v Voiced labio-dental fricative, use V upper case	p	P	ㅍ (p)	Bilabial plosive
w	&#119	w	ɟ	ʒ	ㅈ (j)	Voiced palatal fricative
x	&#120	x Voiceless velar fricative, use X upper case	t	T	ㅌ (t)	Alveolar plosive
y	&#121	y	j	J	ㅛ (yo) less o	Voiced palatal approximant
z	&#122	z	k	K	ㅋ (k)	Velar plosive
{	{	Left curly brace	Left curly brace	Stress 0	No stress	Default
|			Vertical bar	| Minor (foot) break	‖ Major (intonation) break		
}	}	Right curly brace	Right curly brace	Stress 1	Primary stress	
~	~	Tilde	Tilde	Stress 2	Secondary stress	
	&#127;	DEL		Stress 3	Soft or whisper	^? When not phonetic CTRL-
	€	€ EURO SIGN Left-ALT-4($)	‚	SINGLE LOW-9 QUOTATION MARK		
			ƒ	LATIN SMALL LETTER F WITH HOOK		
	‚	„	„	DOUBLE LOW-9 QUOTATION MARK		
	ƒ	Ƒ	…	HORIZONTAL ELLIPSIS		
	„	„	†	DAGGER		
	…	…	‡	DOUBLE		

				DAGGER		
	†	†	‰	PER MILLE SIGN		
	‡	‡	‹	SINGLE LEFT-POINTING ANGLE QUOTATION MARK		
	ˆ	ˆ	‘	LEFT SINGLE QUOTATION MARK		
	‰	‰	’	RIGHT SINGLE QUOTATION MARK		
	Š	Š	“	LEFT DOUBLE QUOTATION MARK		
	‹	‹	”	RIGHT DOUBLE QUOTATION MARK		
	Œ	Œ	•	BULLET		
		Start Superscript	–	EN DASH		
	Ž	Ž	—	EM DASH		
		Start Subscript	™	TRADE MARK SIGN		
		End Superscript or Subscript most recently set and not ended, i.e. nesting is possible. Double character for all clear or see note 3.	›	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK		
	‘	‘		NO-BREAK SPACE		
	’	’	¨	GREEK DIALYTIKA TONOS		
	“	“	Ά	GREEK CAPITAL LETTER ALPHA WITH TONOS	ɲ	Voiced retroflex nasal
	”	”	£	POUND SIGN		
	•	•	¤	CURRENCY SIGN	Multiply scalar	Vector dot product abCosθ
	–	–			Reserved for e.g. Arabic	
	—	—	¦	BROKEN BAR		

	˜	~	§	SECTION SIGN		
	™	™	¨	DIAERESIS	SGCI (XXX in Unicode)	Single Graphic Character Introducer
	š		©	COPYRIGHT SIGN		
	›	›	«	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK		
	œ	Unused	¬	NOT SIGN		
		Unused		SOFT HYPHEN	SHY or Optional Hyphen	A location to break a word at line end. CTRL- in Word.
	ž	Unused	®	REGISTERED SIGN	↻	- Rotate, clockwise
	Ÿ	Unused	¯	HORIZONTAL BAR	↺	+ rotatate
		Non-breaking Space	°	DEGREE SIGN		^BS
¡	¡	Inverted exclamation Right-ALT-1(!)	±	PLUS-MINUS SIGN		
¢	¢	Cent sign Right-ALT&shift-4($)	²	SUPERSCRIPT TWO	≡	Equivalent
£	£	Pound sterling Left&right-ALT-4($)	³	SUPERSCRIPT THREE	∵	Because
¤	¤	General currency sign Left-ALT&shift-4($)	´	GREEK TONOS	∴	Therefore
¥	¥	¥ YEN & Yuan/Renmimbi SIGN Right-ALT-4($)	µ	MICRO SIGN	∈	Includes
¦	¦	Broken vertical bar	¶	PILCROW SIGN		
§	§	Section sign	· MIDDLE DOT	. syllable break		
¨	¨	Umlaut (dieresis)	Έ	GREEK CAPITAL LETTER EPSILON WITH TONOS	Ʒ	voiced epiglottal fricative
©	©	Copyright	Ή	GREEK CAPITAL LETTER ETA WITH TONOS	ʕ	simultaneous X and ʃ
ª	ª	Feminine ordinal	Ί	GREEK CAPITAL LETTER IOTA	ɕ	Voiceless alveolo-palatal

				WITH TONOS		fricative
«	«	Left angle quote, guillemotleft	»	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	ɔ bored /bɔːd/	
¬	¬	Not sign Right-ALT-_	Ό	GREEK CAPITAL LETTER OMICRON WITH TONOS	ɕ	Voiceless alveolo-palatal fricative
		Soft hyphen ~- Or left-ALT--	½	VULGAR FRACTION ONE HALF		
®	®	Registered trademark Left-ALT-R	Ύ	GREEK CAPITAL LETTER UPSILON WITH TONOS	ɥ	Voiced labial-palatal approximant
¯	¯	Macron accent Left&right-ALT--	Ώ	GREEK CAPITAL LETTER OMEGA WITH TONOS	ʍ	Voiceless labial-velar approximant
°	°	Degree sign ^~ Or left-ALT-0	ΐ	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS	ɹ	Voiced alveolar approximant
±	±	Plus or minus Left-ALT-+	Α	GREEK CAPITAL LETTER ALPHA	ɑˑ	
²	²	Superscript two ^! Or right-ALT-2	Β	GREEK CAPITAL LETTER BETA	β Voiced bilabial fricative	German β
³	³	Superscript three Right-ALT-3	Γ	GREEK CAPITAL LETTER GAMMA	ɾ Voiced alveolar tap/flap	
´	´	Acute accent	Δ	GREEK CAPITAL LETTER DELTA	ð Voiced dental fricative	
µ	µ	Micro sign Left-Alt-m	Ε	GREEK CAPITAL LETTER EPSILON	sleep /sliːp/	
¶	¶	Paragraph sign	Ζ	GREEK CAPITAL LETTER ZETA	ʐ voiced retroflex fricative	

·	·	Middle dot ^+ Right-ALT-.	**Η**	GREEK CAPITAL LETTER ETA	ɦ glottal voiced fricative	
¸	¸	Cedilla	**Θ**	GREEK CAPITAL LETTER THETA	θ Voiceless dental fricative	
¹	¹	Superscript one Left-ALT-1	**Ι**	GREEK CAPITAL LETTER IOTA	ɭ Voiced retroflex lateral approximant	
º	º	Masculine ordinal	**Κ**	GREEK CAPITAL LETTER KAPPA	ɟ voiced palatal plosive	
»	»	Right angle quote, guillemotright Left-ALT-"	**Λ**	GREEK CAPITAL LETTER LAMDA	ʎ Voiced palatal lateral approximant	
¼	¼	Fraction one-fourth Left-ALT-1	**Μ**	GREEK CAPITAL LETTER MU	ɯ velar voiced approximant	
½	½	Fraction one-half Left-ALT-2	**Ν**	GREEK CAPITAL LETTER NU	ɳ retroflex voiced nasal	
¾	¾	Fraction three-fourths Left-ALT-3	**Ξ**	GREEK CAPITAL LETTER XI	Voiced labio-dental tap/flap	
¿	¿	Inverted question mark ~B Or right-ALT-?	**Ο**	GREEK CAPITAL LETTER OMICRON	ɬ	Voiced alveolar lateral fricative
À	À	Capital A, grave accent Left-ALT&shift-A	**Π**	GREEK CAPITAL LETTER PI	Aspirated voiced P	
Á	Á	Capital A, acute accent Right-ALT&shift-A	**Ρ**	GREEK CAPITAL LETTER RHO	ʁ Uvular voiced fricative	
Â	Â	Capital A, circumflex accent right-ALT&leftshift-A	**Σ**	GREEK CAPITAL LETTER SIGMA	ɣ Voiced velar fricative	
Ã	Ã	Capital A, tilde left-ALT&rightshift-A	**Τ**	GREEK CAPITAL LETTER TAU	ɖ voiced retroflex plosive	
Ä	Ä	Capital A, dieresis or umlaut mark. Left&right-ALT-A	**Υ**	GREEK CAPITAL LETTER UPSILON	boot /buːt/	

Å	Å	Capital A, ring right-ALT&shift+leftshift-A	**Φ**	GREEK CAPITAL LETTER PHI	Φ voiceless bilabial fricative	
Æ	Æ	Capital AE diphthong (ligature) left-ALT&shift+rightshift-A	**Χ**	GREEK CAPITAL LETTER CHI	χ Voiceless uvular fricative	
Ç	Ç	Capital C, cedilla Left-ALT-C	**Ψ**	GREEK CAPITAL LETTER PSI	ʂ voiceless retroflex fricative	
È	È	Capital E, grave accent Left-ALT-E	**Ω**	GREEK CAPITAL LETTER OMEGA	ɜ bird /bɜːd/	
É	É	Capital E, acute accent Right-ALT-E	**Ϊ**	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA	ʢ	Epiglottal plosive
Ê	Ê	Capital E, circumflex accent Right-ALT&leftshift-E	**Ϋ**	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA	ɥ	Voiced labial-palatal approximant
Ë	Ë	Capital E, dieresis or umlaut mark Left&right-ALT-E	**ά**	GREEK SMALL LETTER ALPHA WITH TONOS	ʑ	Voiced alveolo-palatal fricative
Ì	Ì	Capital I, grave accent Left-ALT-I	**έ**	GREEK SMALL LETTER EPSILON WITH TONOS	ʢ	voiced epiglottal fricative
Í	Í	Capital I, acute accent Right-ALT-I	**ή**	GREEK SMALL LETTER ETA WITH TONOS	ɧ	simultaneous X and ʃ
Î	Î	Capital I, circumflex accent Right-ALT&leftshift-I	**ί**	GREEK SMALL LETTER IOTA WITH TONOS	ɕ	Voiceless alveolo-palatal fricative
Ï	Ï	Capital I, dieresis or umlaut mark Left&right-ALT-I	**ϋ**	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS	ʑ	Voiced alveolo-palatal fricative
Ð	Ð	Capital Eth, Icelandic ESC left-ALT-D	**α**	GREEK SMALL LETTER ALPHA	ɑˑ car /	

					cɑːr/	
Ñ	Ñ	Capital N, tilde Left-ALT&rightshift-N	**β**	GREEK SMALL LETTER BETA	β Voiced bilabial fricative	German β
Ò	Ò	Capital O, grave accent Left-ALT-O	**γ**	GREEK SMALL LETTER GAMMA	ɾ Voiced alveolar tap/flap	
Ó	Ó	Capital O, acute accent Right-ALT-O	**δ**	GREEK SMALL LETTER DELTA	ð Voiced dental fricative	
Ô	Ô	Capital O, circumflex accent Right-ALT&leftshift-O	**ε**	GREEK SMALL LETTER EPSILON	sleep /sliːp/	
Õ	Õ	Capital O, tilde Left-ALT&rightshift-O	**ζ**	GREEK SMALL LETTER ZETA	ʐ voiced retroflex fricative	
Ö	Ö	Capital O, dieresis or umlaut mark Left&right-ALT-O	**η**	GREEK SMALL LETTER ETA	ɦ glottal voiced fricative	
×	×	× Multiply sign ^* Or left-ALT-X	**θ**	GREEK SMALL LETTER THETA	θ voiceless dental fricative	Vector cross product abSinθ
Ø	Ø	Capital O, slash Right-ALT&rightshift-O	**ι**	GREEK SMALL LETTER IOTA	ɭ Voiced retroflex lateral approximant	
Ù	Ù	Capital U, grave accent Left-ALT-U	**κ**	GREEK SMALL LETTER KAPPA	ɟ voiced palatal plosive	
Ú	Ú	Capital U, acute accent Right-ALT-U	**λ**	GREEK SMALL LETTER LAMDA	ʎ Voiced palatal lateral approximant	
Û	Û	Capital U, circumflex accent Right-ALT&leftshift-U	**μ**	GREEK SMALL LETTER MU	ɰ velar voiced approximant	
Ü	Ü	Capital U, dieresis or umlaut mark Left&right-ALT-U	**ν**	GREEK SMALL LETTER NU	ɳ retroflex voiced nasal	
Ý	Ý	Capital Y, acute accent Right-ALT-Y	**ξ**	GREEK SMALL LETTER XI	Voiced labio-dental tap/flap	
Þ	Þ	Capital THORN, Icelandic Left&right-ALT-B	**ο**	GREEK SMALL LETTER OMICRON	ɮ	Voiced alveolar lateral fricative
ß	ß	Small sharp s, German (sz	**π**	GREEK SMALL	Aspirated	

		ligature) Left&right-ALT-Z		LETTER PI	voiced P	
à	à	Small a, grave accent ~H	ρ	GREEK SMALL LETTER RHO	ʁ Uvular voiced fricative	
á	á	Small a, acute accent ~<	ς	GREEK SMALL LETTER FINAL SIGMA	ç Voiceless palatal fricative	
â	â	Small a, circumflex accent ~@	σ	GREEK SMALL LETTER SIGMA	ɣ Voiced velar fricative	
ã	ã	Small a, tilde ~!	τ	GREEK SMALL LETTER TAU	ɖ voiced retroflex plosive	
ä	ä	Small a, dieresis or umlaut mark ~S	υ	GREEK SMALL LETTER UPSILON	boot /buːt/	
å	å	Small a, ring ~W	φ	GREEK SMALL LETTER PHI	ɸ Voiceless bilabial fricative	
æ	æ	Small ae diphthong (ligature) ~Z	χ	GREEK SMALL LETTER CHI	χ Voiceless uvular fricative	
ç	ç	Small c, cedilla ~C	ψ	GREEK SMALL LETTER PSI	ʂ voiceless retroflex fricative	
è	è	Small e, grave accent ~>	ω	GREEK SMALL LETTER OMEGA	ʊ book /bʊk/	
é	é	Small e, acute accent ~(	ϊ	GREEK SMALL LETTER IOTA WITH DIALYTIKA	ʔ	Epiglottal plosive
ê	ê	Small e, circumflex accent ~D	ϋ	GREEK SMALL LETTER UPSILON WITH DIALYTIKA	ɲ	Voiced retroflex nasal
ë	ë	Small e, dieresis or umlaut mark ~+	ό	GREEK SMALL LETTER OMICRON WITH TONOS	ɕ	Voiceless alveolo-palatal fricative
ì	ì	Small i, grave accent ~I	ύ	GREEK SMALL LETTER UPSILON WITH TONOS	ɹ	Voiced alveolar approximant

í	í	Small i, acute accent ~V	ώ	GREEK SMALL LETTER OMEGA WITH TONOS	Ɱ	Voiceless labial-velar approximant
î	î	Small i, circumflex accent ^#	RESERVED for e.g. Arabic.	ĕ	Short duration example e	
ï	ï	Small i, dieresis or umlaut mark ~~	↑ upstep		Half long	
ð	ð	Small eth, Icelandic Left&right-ALT-D	ɜ bird /bɜːd/	ˑ	Long duration	
ñ	ñ	Small n, tilde ~N	↓ downstep	ȅ	Extra high example	
ò	ò	Small o, grave accent ~K	ʊ book /bʊk/	é	High example	
ó	ó	Small o, acute accent ~O	ɔ bored /bɔːd/	ē	Mid example	
ô	ô	Small o, circumflex accent ~A	√	è	Low example	
õ	õ	Small o, tilde ~Q	∞ infinity	ȅ̀	Extra low example	
ö	ö	Small o, dieresis or umlaut mark ~L	≈	ě	Rising example	
÷	÷	Division sign ~? Left-ALT-/	Division sign	ê	Falling example	
ø	ø	Small o, slash ~X	ɒ hot /hɒt/		High rising	
ù	ù	Small u, grave accent ~P	^ cup / k^p/		Low rising	
ú	ú	Small u, acute accent ~U	≥		High falling	
û	û	Small u, circumflex accent ~I	≤	^{2}√	Low falling	
ü	ü	Small u, dieresis or umlaut mark ~J	W	^{n}√	Rising falling	If n has value, enter alt-numlock-number-enter
ý	ý	Small y, acute accent ~Y	≠		Falling rising	
þ	þ	Small thorn, Icelandic Left&right-ALT-P	ħ	Ħ Th(em)		
ÿ	ÿ	Small y, dieresis or umlaut mark ~G	≈	‿	Linking, no break	

ASCII

Other European Table

128	80	€	EURO SIGN

130	82	‚	SINGLE LOW-9 QUOTATION MARK
131	83	ƒ	LATIN SMALL LETTER F WITH HOOK
132	84	„	DOUBLE LOW-9 QUOTATION MARK
133	85	…	HORIZONTAL ELLIPSIS
134	86	†	DAGGER
135	87	‡	DOUBLE DAGGER
137	89	‰	PER MILLE SIGN
139	8B	‹	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
145	91	‘	LEFT SINGLE QUOTATION MARK
146	92	’	RIGHT SINGLE QUOTATION MARK
147	93	“	LEFT DOUBLE QUOTATION MARK
148	94	”	RIGHT DOUBLE QUOTATION MARK
149	95	•	BULLET
150	96	–	EN DASH
151	97	—	EM DASH
153	99	™	TRADE MARK SIGN
155	9B	›	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
160	A0		NO-BREAK SPACE
161	A1	΄΅	GREEK DIALYTIKA TONOS
162	A2	Ά	GREEK CAPITAL LETTER ALPHA WITH TONOS
163	A3	£	POUND SIGN
164	A4	¤	CURRENCY SIGN
165	A5	¥	YEN SIGN
166	A6	¦	BROKEN BAR
167	A7	§	SECTION SIGN
168	A8	¨	DIAERESIS
169	A9	©	COPYRIGHT SIGN
171	AB	«	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
172	AC	¬	NOT SIGN
173	AD		SOFT HYPHEN
174	AE	®	REGISTERED SIGN
175	AF	―	HORIZONTAL BAR
176	B0	°	DEGREE SIGN
177	B1	±	PLUS-MINUS SIGN
178	B2	²	SUPERSCRIPT TWO
179	B3	³	SUPERSCRIPT THREE
180	B4	΄	GREEK TONOS
181	B5	µ	MICRO SIGN
182	B6	¶	PILCROW SIGN

183	B7	·	MIDDLE DOT
184	B8	Έ	GREEK CAPITAL LETTER EPSILON WITH TONOS
185	B9	Ή	GREEK CAPITAL LETTER ETA WITH TONOS
186	BA	Ί	GREEK CAPITAL LETTER IOTA WITH TONOS
187	BB	»	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
188	BC	Ό	GREEK CAPITAL LETTER OMICRON WITH TONOS
189	BD	½	VULGAR FRACTION ONE HALF
190	BE	Ύ	GREEK CAPITAL LETTER UPSILON WITH TONOS
191	BF	Ώ	GREEK CAPITAL LETTER OMEGA WITH TONOS
192	C0	ΐ	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS
193	C1	Α	GREEK CAPITAL LETTER ALPHA
194	C2	Β	GREEK CAPITAL LETTER BETA
195	C3	Γ	GREEK CAPITAL LETTER GAMMA
196	C4	Δ	GREEK CAPITAL LETTER DELTA
197	C5	Ε	GREEK CAPITAL LETTER EPSILON
198	C6	Ζ	GREEK CAPITAL LETTER ZETA
199	C7	Η	GREEK CAPITAL LETTER ETA
200	C8	Θ	GREEK CAPITAL LETTER THETA
201	C9	Ι	GREEK CAPITAL LETTER IOTA
202	CA	Κ	GREEK CAPITAL LETTER KAPPA
203	CB	Λ	GREEK CAPITAL LETTER LAMDA
204	CC	Μ	GREEK CAPITAL LETTER MU
205	CD	Ν	GREEK CAPITAL LETTER NU
206	CE	Ξ	GREEK CAPITAL LETTER XI
207	CF	Ο	GREEK CAPITAL LETTER OMICRON
208	D0	Π	GREEK CAPITAL LETTER PI
209	D1	Ρ	GREEK CAPITAL LETTER RHO
211	D3	Σ	GREEK CAPITAL LETTER SIGMA
212	D4	Τ	GREEK CAPITAL LETTER TAU
213	D5	Υ	GREEK CAPITAL LETTER UPSILON
214	D6	Φ	GREEK CAPITAL LETTER PHI
215	D7	Χ	GREEK CAPITAL LETTER CHI
216	D8	Ψ	GREEK CAPITAL LETTER PSI
217	D9	Ω	GREEK CAPITAL LETTER OMEGA
218	DA	Ϊ	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA
219	DB	Ϋ	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA
220	DC	ά	GREEK SMALL LETTER ALPHA WITH TONOS
221	DD	έ	GREEK SMALL LETTER EPSILON WITH TONOS
222	DE	ή	GREEK SMALL LETTER ETA WITH TONOS

223	DF	ΐ	GREEK SMALL LETTER IOTA WITH TONOS
224	E0	ΰ	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS
225	E1	α	GREEK SMALL LETTER ALPHA
226	E2	β	GREEK SMALL LETTER BETA
227	E3	γ	GREEK SMALL LETTER GAMMA
228	E4	δ	GREEK SMALL LETTER DELTA
229	E5	ε	GREEK SMALL LETTER EPSILON
230	E6	ζ	GREEK SMALL LETTER ZETA
231	E7	η	GREEK SMALL LETTER ETA
232	E8	θ	GREEK SMALL LETTER THETA
233	E9	ι	GREEK SMALL LETTER IOTA
234	EA	κ	GREEK SMALL LETTER KAPPA
235	EB	λ	GREEK SMALL LETTER LAMDA
236	EC	μ	GREEK SMALL LETTER MU
237	ED	ν	GREEK SMALL LETTER NU
238	EE	ξ	GREEK SMALL LETTER XI
239	EF	ο	GREEK SMALL LETTER OMICRON
240	F0	π	GREEK SMALL LETTER PI
241	F1	ρ	GREEK SMALL LETTER RHO
242	F2	ς	GREEK SMALL LETTER FINAL SIGMA
243	F3	σ	GREEK SMALL LETTER SIGMA
244	F4	τ	GREEK SMALL LETTER TAU
245	F5	υ	GREEK SMALL LETTER UPSILON
246	F6	φ	GREEK SMALL LETTER PHI
247	F7	χ	GREEK SMALL LETTER CHI
248	F8	ψ	GREEK SMALL LETTER PSI
249	F9	ω	GREEK SMALL LETTER OMEGA
250	FA	ϊ	GREEK SMALL LETTER IOTA WITH DIALYTIKA
251	FB	ϋ	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
252	FC	ό	GREEK SMALL LETTER OMICRON WITH TONOS
253	FD	ύ	GREEK SMALL LETTER UPSILON WITH TONOS
254	FE	ώ	GREEK SMALL LETTER OMEGA WITH TONOS

IPA, the International Phonetic Alphabet

Suprasegmentals				Tones and word accents											
ˈ	Primary stress			Level tones				Contour tones							
ˌ	Secondary stress			ȅ	or	˥	Extra high	ě	or	˩˥	Rising	ȅ̂	or	˥˩	High falling

e ːˑ	Long			é		˥	High	ê		˥˩	Falling	e᷄		˩˩	Low falling
eˑ	Half-long			ē		˧	Mid	e □		˥˧	High rising	e □		˩˧	Rising falling
ĕ	Extra-short			è		˨	Low	e □		˩˨	Low rising	e᷅		˩˨	Falling rising
|	Minor (foot) break			ȅ		˩	Extra low	↑	Upstep			↗	Global rise		
‖	Major (intonation) break							↓	Downstep			↘	Global fall		
.	Syllable break														
‿	Linking (no break)														

International Phonetic Alphabet (IPA) ˌɪntəˈnæʃn□l fəˈnɛtɪk ˈælfəˌbɛt

Consonants (pulmonic)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Consonants (non-pulmonic)

Clicks		Voiced implosives		Ejectives	
ʘ	bilabial	ɓ	Bilabial	ʼ	examples:
ǀ	Dental	ɗ	Dental / alveolar	pʼ	Bilabial
ǃ	(Post) alveolar	ʄ	Palatal	tʼ	Dental / alveolar
ǂ	Palatal-alveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

Other symbols

ʍ w	Voiceless / voiced labial-velar approximants
ɥ	Voiced labial-palatal approximant
ħ ʕ	Voiceless / voiced epiglottal fricatives
ʡ	Epiglottal plosive
ɕ ʑ	Voiceless / voiced alveolo-palatal fricatives
ɧ	simultaneous x and ʃ
k͡p t͡s	Affricatives and double articulations can be represented by two symbols and joined by a tie bar if necessary

Vowel				
	← Mouth wider horizontally		Mouth narrower horizontally →	
Mouth narrower vertically	iː sleep /sliːp/	ɪ slip /slɪp/	ʊ book /bʊk/	uː boot /buːt/
	e ten /ten/	ə after /aːftə/	ɜː bird /bɜːd/	ɔː bored /bɔːd/
Mouth wider vertically	æ cat /kæt/	^ cup / k^p/	ɑː car / cɑːr/	ɒ hot /hɒt/

ɜ

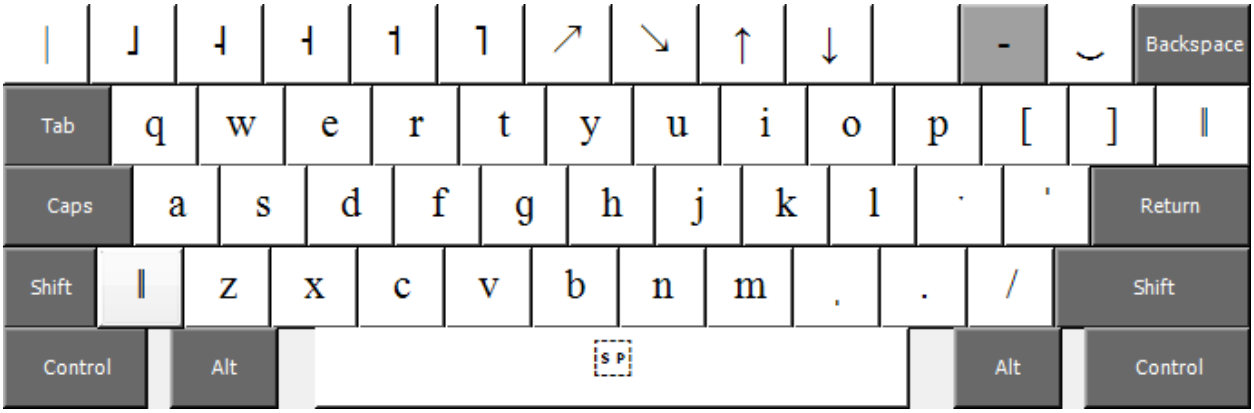
Diphthongs		
ɪə beer /bɪə/	eɪ say /seɪ/	

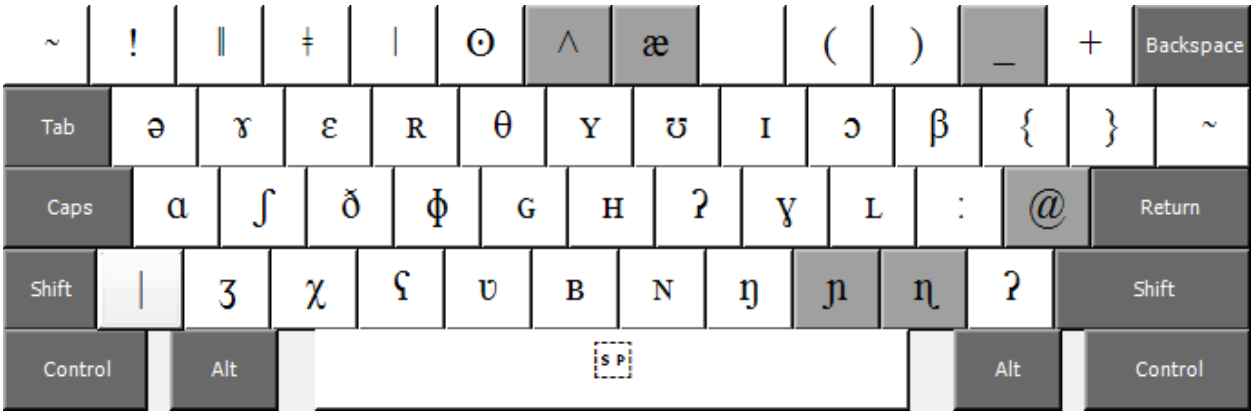
ʊə fewer /fjʊə/	ɔɪ boy /bɔɪ/	əʊ no /nəʊ/
eə bear /beə/	aɪ high /haɪ/	aʊ cow /kaʊ/

Suprasegmentals			Tones and word accents											
ˈ	Primary stress		Level tones				Contour tones							
ˌ	Secondary stress		ḗ	or	˥	Extra high	ě	or	˩˥	Rising	ê	or	˥˩	High falling
eː	Long		é		˦	High	ê		˥˩	Falling	ê		˩˥˩	Low falling
eˑ	Half-long		ē		˧	Mid	e □		˧˦	High rising	e □		˧˦˧	Rising falling
ĕ	Extra-short		è		˨	Low	e □		˩˨	Low rising	ẽ		˩˨˩	Falling rising
ˈ	Minor (foot) break		ȅ		˩	Extra low	↑	Upstep			↗	Global rise		
‖	Major (intonation) break						↓	Downstep			↘	Global fall		
.	Syllable break													
‿	Linking (no break)													

- 14 simple consonant letters: 1. ㄱ (g); 2. ㄴ (n); 3. ㄷ (d); 4. ㄹ (l/r); 5. ㅁ (m); 6. ㅂ (b); 7. ㅅ (s); 8. ㅇ (-/ng); 9. ㅈ (j); 10. ㅊ (ch); 11. ㅋ (k); 12. ㅌ (t); 13.ㅍ (p); 14. ㅎ (h).
  - 5 double letters (glottalized): 1. ㄲ (kk); 2. ㄸ (tt); 3. ㅃ (pp); 4. ㅆ (ss); 5. ㅉ (jj).
  - 6 simple vowel letters: 1. ㅏ (a); 2. ㅓ (eo); 3. ㅗ (o); 4. ㅜ (u); 5. ㅡ (eu); 6. ㅣ (i)
  - 4 simple iotized vowel letters (semi consonant-semi vowel): 1. ㅑ (ya); 2. ㅕ (yeo); 3. ㅛ (yo); 4. ㅠ (yu)
  - 11 diphthongs: 1. ㅐ (ae); 2. ㅒ (yae); 3. ㅔ (e); 4. ㅖ (ye); 5. ㅘ (wa); 6. ㅙ (wae); 7. ㅚ (oe); 8. ㅝ (weo); 9. ㅞ (we); 10. ㅟ (wi); 11. ㅢ (yi).

From http://www.rejc2.co.uk/ipakeyboard/





However these layouts do not provide for a lower/upper case system and a possible adoption in Korea has the phonetic characters in a different layout from the Hangul layout.

Notes on Keyboard Alphabet Switching and other items.

An economical solution is proposed that uses all 8 bits of the keyboard display register and dual color LEDs.

1) Some phonetic characters are in the Greek alphabet, and are not separately listed in the phonetic alphabet. If IPA upper case is selected and a lower case Greek character is entered, then the corresponding Greek upper case character SHALL be transmitted.
2) Keyboard modification for multialphabet use. This requires dual color LEDs and additional circuitry;
   a) Modify the key caps to have multiple character sets, preferably color coded.
   b) Capslock, Numlock, Shift and ALT as below. A1, A2, A3 and A4 are Unicode alphabet sets.

Capslok	Off	Off	Off	Off	Green	Green	Green	Green
Numlok	Off	Off	Green	Green	Off	Off	Green	Green
Shift	Off	On	Off	On	Off	On	Off	On
select	lower	upper	Lwr&N	Up&N	upper	Lower	Up&N	Lwr&N
	(1)	(2)	(1) Numlok	(2) Numlok	(2)	(1)	(2) Numlok	(1) Numlok
Capslok	Red	Red	Red	Red	Red	Red	Red	Red
Numlok	Off	Off	Off	Off	Green	Green	Green	Green
Shift	Off	On	Off	On	Off	On	Off	On
select	XAscii	XAsciiS	XAsciiS	XAscii	XAsccii+	XAscci+S	XAscci+S	XAscii+
					Or ALT	Or ALT	Or ALT	Or ALT
Capslok	Red	Red	Red	Red	Red	Red	Red	Red
Numlok	Red	Red	Red	Red	Orange	Orange	Orange	Orange
Shift	Off	On	Off	On	Off	On	Off	On
select	Hangul	Hangul	Hangul	Hangul	Grk	GrkS	Grk	GrkS
Capslok	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange
Numlok	Off	Off	Off	Off	Green	Green	Green	Green
Shift	Off	On	Off	On	Off	On	Off	On

select	IPA	IPAS	IPA	IPAS	IPA+ Or ALT	IPA+S Or ALT	IPA+ Or ALT	IPA+S Or ALT
Capslok	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange
Numlok	Red	Red	Red	Red	Orange	Orange	Orange	Orange
Shift	Off	On	Off	On	Off	On	Off	On
select	A1	A2	A1	A2	A3 Or ALT	A4 Or ALT	A3 Or ALT	A4 Or ALT

c) Scroll Lock Off Green Red Orange
Normal Scroll Lock Alt function keyboard Scroll Lock Alt Function Keyboard
The Alt Function keyboard is for custom applications e.g. video or graphics editing systems, or music.
As laptops lack the number pad area separate for additional character usage, some alternative is desirable. This should be invisible to the operating system by BIOS implementation.
Unicode provides for other alphabets, perhaps as an overlay with holes. The character(s) after DC4 indicate the section of Unicode that is to be used.

d) The shift key shall operate as normal, if the capslock is red or orange, the shift shall exchange the character case.

e) The CTRL-Capslock, CTRL-Numlock and CTRL-Scroll lock shall reverse the selection one step back. Note that CTRL-Numlock- (i.e. with the – character) is reserved as an En dash.

f) The IPA Caps provides for future uppercase IPA characters not in the base IPA set. The IPA characters to the right of C to Z are accessed by using CTRL- when in the IPA mode. If uppercase of those IPA characters becomes defined, then they shall be accessed as CTRL- when in the IPA Caps mode. Also for font and formatting string start and end characters which are optional and the whole string will not print but will display on a text, xml or hexadecimal editor. The definitions of the strings are not provided here.

g) The IPA character ɘ also represents the number. The Greek character π also represents the number. The IPA character ħ also represents the positive or hole electric charge. As + is normally associated with red, the + shape in the character may be highlighted in red, and the rest of the character some other color preferably except black. The electrical negative charge is also represented by the IPA character **ə.** The – part of the character may be highlighted in black, The rest of the character some other character than red or black. The roman letters e, m, n, p, f, k, M, G, T, P, Y are with μ the magnitude characters where e is unity. E.g. 4e7 is a 4.7 value as the decimal point (or comma) is liable to be unreadable on component markings.

The IPA characters i , j, k are the three axis imaginary operators. The Roman characters A, B, C, D, E, F also represent 10, 11,12, 13, 14, 15 to provide hexadecimal characters. The use of 0x at the beginning means that the number is hexadecimal even if A, B, C, D, E or F are not in the number.

h) The Japanese keyboard key for Hirakana/Katakana, which is to the right of the spacebar, shall have a multicolor LED using the remaining two bits of the LED register. The assignment shall be OFF Normal ASCII/IPA/Hangul/Greek as noted above using extended ASCII.
   GREEN Hirakana (default) or green tagged overlay as assigned by user.
   ORANGE Katakana (default) or orange tagged overlay as assigned by user.
   RED Cyrillic using KOI8 or Devanagari ISCII in India (default) or red overlay as assigned by user, Arabic is recommended.

3) Arabic has a MARC 8 binary code following. Hebrew has Unicode characters, but does not seem to have an ASCII equivalent. Both of these scripts are right to left.
4) The rule for subscripts and superscripts is that subscripts SHALL be before superscripts and can only go one step deep. They shall be followed by the end subscript/superscript. A superscript may immediately follow. An additional superscript may follow so the depth is not limited until the end subscript/superscript is sent which then refers to the end of all subscript or superscript for the base character. Mathematical expressions are provided for.
5) Characters to right of C to G currently have no provision for an uppercase form. They are all clicks. While this is not a problem with Hangul or some other national alphabet, this may sometime be desirable with IPA. While IPA currently does not have an uppercase system currently, the letters in Roman and Greek that they line up with have both cases and so the provision is there for suitable fonts to be developed. Sometimes it is not apparent which instance is upper case and which is lower case, the first instance shall be the upper case.
6) Alternatives to Hangul (which has 26 characters) for other national languages are being researched, but the character set is limited. This means that the use of IPA for international use more desirable, and this is now provided for in this extended ASCII. Sanskrit has 36-48 phoneme letters, Arabic has 28 letters, Cyrillic has 33 letters but only 20 are non-roman and Thai has 44 consonants plus 21 vowels. So for some, a Hangul substitution would be possible, but for others, only IPA has sufficient characters. Numerals in other than modern European are not provided for. Chinese and Japanese have many more characters. Japanese has kanji, which are originally Chinese characters, hiragana, which has 48 characters which may be modified with dakuten or handakuten, and katakana which also has 48 characters. Romaji is the name for ASCII ISO 8859-1 characters.
7) The DC4 character is followed by a string defining which Unicode version or other standard is to be used. The DC1-DC3 selection defines the end of such usage.
8) The CTRL, ALT, DEL reboot is preceded by DC2, followed by DC1 then either CR LF which reboots all cascaded serial (not other connected) devices. Or a numeral 0 to 9 which is decremented each pass through and if it reaches zero (or already is), that device reboots.
9) The F is preceded by a DC2 then followed by a DC1 then a numeral 0 to F (which represents the function key number) then a CR LF. Digits 0 to 9 are assumed to be hexadecimal when used for

binary processing. The hexadecimal values 0xA to 0xF are not currently otherwise defined as font characters, but are always hexadecimal when used for binary processing.

10) The first character being ^ denotes CTRL- and being ~ denotes ALT- in character pairs for key entry.
11) Other standards such as KSC5601 (South Korea) and KPS9566-97 (North Korea) are two byte systems. This is basically incompatible with single byte systems but, like Unicode which is the preferred two byte system, they can be accommodated via the DC4 switch mechanism. Editing of text is more complex as the SO, SI, DC2 and DC4 characters must be found if present beforehand in the source and destination and incorporated with any switching required in the destination text. This proposal for a standard is assuming that there will be adoption of IPA as a worldwide alphabet. While two byte character sets are capable of large numbers of characters, it then becomes a problem of how to have keyboard entry for so many characters. The KPS and KSC sets includes scientific units. However, in the west, these are typed as individual characters.
12) The IPA ʊ in both lower and upper case are fitted to U because the IPA U key position in Hangul is used by ㅓ . Although there is no IPA equivalent, the Z key has another Hangul ㅋ which is IPA k or K. So instead, the ASCII U font shall be made to resemble IPA ʊ instead.
13) Although there is no required use of lower and upper case for all IPA characters, it is anticipated that font developers would be expected to extend in that direction. For this purpose at present, different size fonts are used instead.
14) The use of Unicode selections for other languages can be printed on the right of the key.

Keyboard Layout Chart following

ESC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12		
												˙		XAscii+
Ð	ɦ	ɭ	ɮ	ʁ	ʎ	ɯ	ɳ		ç	ɣ	ɖ	χ		IPAs
Ð	ɦ	ɭ	ɮ	ʁ	ʎ	ɯ	ɳ	×	ç	ɣ	ɖ	χ		IPA
` ~ ï	1 ! ã	2 @ â	3 # î	4 $ €	5 % £	6 ^ ¥	7 & ¤	8 *	9 (é	0)	- _ -	= + ë	BS	ALT
ɔ	**γ Γ**	**δ Δ**	**ζ Ζ**	**η Η**	**θ Θ**	**λ Λ**	**μ Μ**	**ξ Ξ**	**π Π**	**ρ Ρ**	ő	ȅ	ě	Grk/IPA
↑	B ɡ	ʒ ɬ	D !	G ɡ	S ɕ	J ɻ	Z ɾ	ʟ ɓ	Æ ʃ	E ʃ	ɕ	ɹ	ʍ	IPA+
↓	b ㅂ	ʝ ㅈ	d ㄷ	g ㄱ	s ㅅ	j ㅛ	z ㅓ	í ㅑ	æ ㅐ	e ㅔ	ɕ	ɹ	ʍ	Hangul
	q	W		R			ʊ					ASCII	not	IPA
TAB	Q õ	W å	E	R	T	Y ý	U ú	I û	O ó	P ù	[{	] }	\ |	ALT
	ς Σ	**σ Σ**	**τ Τ**	**φ Φ**	**ψ Ψ**	**ω Ω**	ɾ	ʒ	ɧ	ʊ	≈			Grk/IPA
	M ↗	N ɢ	ɲ |	R ǂ	H ‖	O	ə ɓ	A ɗ	I ɗ	ʔ ɒ	ɲ ^		IPAs	IPA+
	m ㅁ	n ㄴ	ŋ ㅇ	r ㄹ	h ㅎ	o ㅗ	ə ㅓ	a ㅏ	i ㅣ	ʔ ɒ	ɲ ^		IPA	Hangul
				f	G				L			ASCII	not	IPA
CAPS	A ô	S ä	D ê	F	G ÿ	H à	J ü	K ò	L ö	; :	‘ “	CRLF	Base	ALT
	±	≤	≥	≠	√	ⁿ√	²√	**°**	**•**	ʒ				XAscii+

	K ɻ	T ɬ	ʒ ʊ	P ɽ	ʃ ɹ	Y	U ʼ	Ħ	ː	‿			IPAs	IPA+
	k ㅋ	t ㄷ	ʒ ㅊ	p ㅍ	ʃ ㅠ	y ㅜ	u ㅡ	ħ		ʂ			IPA	Hangul
		x	c	v								ASCII	not	IPA
SHIFT	Z æ	X ø	C ç	V í	B ¿	N ñ	M	, < á	. > è	/ ? ÷	SHIFT		Base	ALT
	Alpha-ALT	Or XAscii											Norm ASCII	ALT/ XAscii

15) The yellow areas are information. The light green areas are information in a box of 4, of which the left half may be hidden by keys. Also the left and right apply to two characters on the same key. The light blue areas are information in a box of 2, of which the left half may be hidden by keys. Also the left and right apply to two characters on the same key and ALT may alternatively be entered in the XAscii mode.
16) The QWERTY layout was originally set to minimize typewriter keys jamming and is not ergonomic. The Hangul layout ignores this layout and is supposedly ergonomic. The IPA layout is initially based on the Hangul equivalent where applicable. The ergonomics of the IPA layout are unknown, and also presumably could have a different optimum layout varying by language. It may be an improvement to rearrange the non-Hangul IPA key assignments, but that is expected to be discussed as part of a standards definition process.
17) The Greek alphabet is only including characters that are not already in the ASCII alphabet. The primary purpose of including Greek is to provide for STEM (science, technology, engineering and math) usage.

18) An example key label would be as below. The underline of G indicates that it is also used as an IPA character. Space for Unicode assignment is to the right, Hirakana is bottom middle and katakana is top middle for Japanese use. Use A1 and A3 mode; Cyrillic is provided for later but not on the example below.

ᵩ ᵩ キ
H ǁ
h ɧ
G ÿ き

19) IPA characters without an extended ASCII upper case are some modifiers that are not base characters.

20) ツ and シ, ソ and ン may be confused.

21) The use of upper and lower case “l” and “L” for two different IPA characters is unsuitable for use of lower and upper case. So instead of the serif (Times New Roman) font, a script (Lucida Calligraphy) font has been substituted, ℓ $\mathcal{L}$.

22) Japanese Keyboard;

23) The Caps Lock text is an English/Hirakana switch that is provided by the Caps/Num color switching scheme where Hirakana and Katakana are the A1 and A3 alphabets. As existing software is using the Caps Lock, that mode is accepted, but user information needs to be supplied with the keyboard.
24) The key to the left of “1” is small/normal size characters. This would be a preset default percentage point change.
25) The key to the left of space means no change to Chinese characters. The key to the right of shift means change to Chinese characters. So pinyin can entered as roman, or katakana or hirakana.
26) The next key to the right of the last one means select Katakana on top and select Hirakana below, and add select roman for normal or pinyin entry for Chinese use.
27) The character 『 on the “P” key means start horizontal text.
28) The character 』 on the “+” key means end horizontal text.
29) The character 「 on the “[” key which is second right of the “P” key means start vertical text.
30) The character 」 on the “]” key which is third right of the “L” key means end vertical text.
31) The bullet on the “?” key is a space symbol.
32) A “£” like symbol with an extra dash is on the key to the right of the “0”
33) A long dash for punctuation is on the 3rd key to the right of the “0”.
34) The 、。
35) The characters noted from 22) to here are to be included in the charts below, but a printable source that works in Word has yet to be found.
36) The unprinted character above へ is not obvious, but may be the タ katakana character.

37) The "**¥**" Yen or Yuan character is used on the "6" key as well as of the character on the 3rd key to the right of "0" in kana mode.

38) The "**¤**" currency symbol may be replaced by the printing of other currency symbols as software may be configured. However such currency symbols shall be transmitted as "**¤**" when ASCII is used. The use of (ISO 10366-1 countryAbbreviation used by internet) after such a symbol shall be used.

39) *Command **ed**: Write LEDs*
This command is followed by a byte indicating the desired LEDs setting. Bits 7-3: unused, 0. Bit 2: 1: CapsLock LED on. Bit 1: 1: NumLock LED on. Bit 0: 1: ScrollLock LED on. When OK, both bytes are ACKed. If the second byte is recognized as a command, that command is ACKed and done instead. Otherwise a NACK is returned (and a keyboard enable may be needed). Based on this, assign bit 4 to be the red LED of capslock and bit 5 to be the red LED of numlock and bit 6 to be the red LED of scroll lock. Turning both red and green on produces orange. Bit 3 would be the green LED for the Normal/Katakana/Hirakana/Cyrillic key LED. Bit 7 would be the red LED for the Normal/Katakana/Hirakana/Cyrillic key LED.

40) One version of the Japanese Keyboard scancodes has "0" as 0C instead of 0B, "-" as 0D instead of 0C, "=" as 7D instead of 0D, an additional key to the right of "=" with a scancode of 0E. However here, as in a normal keyboard, the Backspace is 0E, and the added button is below on the end and is assigned a scancode of 7D.

41) To provide mitigation of errors, [CR][LF] SHALL be used in pairs, and if one character is an error, the pair is assumed and regenerated. The initial bank selection SHALL be S0 and after each [CR][LF] pair, SO SHALL be assumed UNLESS SI, DC2 or DC4 is the next character. Applications SHOULD provide for user ability to correct bank selection errors if they appear to have occurred. Error correction may be also provided by other means appropriate to the application

42) From www.acme.com, another version of the IPA. Unicode IPA Extensions (96, 250-2af):

ɐ ɐ	ɑ ɑ	ɒ ɒ	ɓ ɓ	ɔ ɔ	ɕ ɕ	ɖ ɖ	ɗ ɗ
ɘ ɘ	ə ə	ɚ ɚ	ɛ ɛ	ɜ ɜ	ɝ ɝ	ɞ ɞ	ɟ ɟ
ɠ ɠ	ɡ ɡ	ɢ ɢ	ɣ ɣ	ɤ ɤ	ɥ ɥ	ɦ ɦ	ɧ ɧ
ɨ ɨ	ɩ ɩ	ɪ ɪ	ɫ ɫ	ɬ ɬ	ɭ ɭ	ɮ ɮ	ɯ ɯ
ɰ ɰ	ɱ ɱ	ɲ ɲ	ɳ ɳ	ɴ ɴ	ɵ ɵ	ɶ ɶ	ɷ ɷ
ɸ ɸ	ɹ ɹ	ɺ ɺ	ɻ ɻ	ɼ ɼ	ɽ ɽ	ɾ ɾ	ɿ ɿ

ʀ ʀ	ʁ ʁ	ʂ ʂ	ʃ ʃ	ʄ ʄ	ʅ ʅ	ʆ ʆ	ʇ ʇ
ʈ ʈ	ʉ ʉ	ʊ ʊ	ʋ ʋ	ʌ ʌ	ʍ ʍ	ʎ ʎ	ʏ ʏ
ʐ ʐ	ʑ ʑ	ʒ ʒ	ʓ ʓ	ʔ ʔ	ʕ ʕ	ʖ ʖ	ʗ ʗ
ʘ ʘ	ʙ ʙ	ʚ ɚ	ʛ ɠ	ʜ ʜ	ʝ ʝ	ʞ ʞ	ʟ ʟ
ʠ ʠ	ʡ ʡ	ʢ ʢ	ʣ ʣ	ʤ ʤ	ʥ ʥ	ʦ ʦ	ʧ ʧ
ʨ ʨ	ʩ ʩ	ʪ ʪ	ʫ ʫ	ʬ ʬ	ʭ ʭ	ʮ ʮ	ʯ ʯ

43) IPA extensions. Phonetic Extensions (128, 1d00-1d7f):

ᴀ A	ᴁ Æ	ᴂ ᴂ	ᴃ ᴃ	ᴄ C	ᴅ D	ᴆ Ð	ᴇ E
ᴈ Ɜ	ᴉ ᴉ	ᴊ J	ᴋ K	ᴌ Ł	ᴍ M	ᴎ И	ᴏ O
ᴐ Ɔ	ᴑ ᴏ	ᴒ ᴐ	ᴓ ᴑ	ᴔ ᴔ	ᴕ ᴕ	ᴖ ᴖ	ᴗ ᴗ
ᴘ P	ᴙ Я	ᴚ ᴙ	ᴛ T	ᴜ U	ᴝ ᴝ	ᴞ ᴞ	ᴟ ᴟ
ᴠ V	ᴡ W	ᴢ Z	ᴣ Ʒ	ᴤ ᴤ	ᴥ ᴥ	ᴦ Γ	ᴧ Λ
ᴨ Π	ᴩ P	ᴪ Ψ	ᴫ Л	ᴬ A	ᴭ Æ	ᴮ B	ᴯ B
ᴰ D	ᴱ E	ᴲ Ǝ	ᴳ G	ᴴ H	ᴵ I	ᴶ J	ᴷ K
ᴸ L	ᴹ M	ᴺ N	ᴻ И	ᴼ O	ᴽ Ȣ	ᴾ P	ᴿ R
ᵀ T	ᵁ U	ᵂ W	ᵃ a	ᵄ ɐ	ᵅ ɑ	ᵆ ᴂ	ᵇ b
ᵈ d	ᵉ e	ᵊ ə	ᵋ ɛ	ᵌ ɜ	ᵍ g	ᵎ ᵢ	ᵏ k
ᵐ m	ᵑ ŋ	ᵒ o	ᵓ ɔ	ᵔ ᴖ	ᵕ ᴗ	ᵖ p	ᵗ t

ᵘ ᵘ	ᵙ ᵊ	ᵚ ᵚ	ᵛ ᵛ	ᵜ ᵜ	ᵝ ᵝ	ᵞ ᵞ	ᵟ ᵟ
ᵠ ᵠ	ᵡ ᵡ	ᵢ ᵢ	ᵣ ᵣ	ᵤ ᵤ	ᵥ ᵥ	ᵦ ᵦ	ᵧ ᵧ
ᵨ ᵨ	ᵩ ᵩ	ᵪ ᵪ	ᵫ ᵫ	ᵬ ᵬ	ᵭ ᵭ	ᵮ ᵮ	ᵯ ᵯ
ᵰ ᵰ	ᵱ ᵱ	ᵲ ᵲ	ᵳ ᵳ	ᵴ ᵴ	ᵵ ᵵ	ᵶ ᵶ	ᵷ ᵷ
ᵸ ᵸ	ᵹ ᵹ	ᵺ ᵺ	ᵻ ᵻ	ᵼ ᵼ	ᵽ ᵽ	ᵾ ᵾ	ᵿ ᵿ

44) More IPA extensions. Phonetic Extensions Supplement (64, 1d80-1dbf):

ᶀ ᶀ	ᶁ ᶁ	ᶂ ᶂ	ᶃ ᶃ	ᶄ ᶄ	ᶅ ᶅ	ᶆ ᶆ	ᶇ ᶇ
ᶈ ᶈ	ᶉ ᶉ	ᶊ ᶊ	ᶋ ᶋ	ᶌ ᶌ	ᶍ ᶍ	ᶎ ᶎ	ᶏ ᶏ
ᶐ ᶐ	ᶑ ᶑ	ᶒ ᶒ	ᶓ ᶓ	ᶔ ᶔ	ᶕ ᶕ	ᶖ ᶖ	ᶗ ᶗ
ᶘ ᶘ	ᶙ ᶙ	ᶚ ᶚ	ᶛ ᶛ	ᶜ ᶜ	ᶝ ᶝ	ᶞ ᶞ	ᶟ ᶟ
ᶠ ᶠ	ᶡ ᶡ	ᶢ ᶢ	ᶣ ᶣ	ᶤ ᶤ	ᶥ ᶥ	ᶦ ᶦ	ᶧ ᶧ
ᶨ ᶨ	ᶩ ᶩ	ᶪ ᶪ	ᶫ ᶫ	ᶬ ᶬ	ᶭ ᶭ	ᶮ ᶮ	ᶯ ᶯ
ᶰ ᶰ	ᶱ ᶱ	ᶲ ᶲ	ᶳ ᶳ	ᶴ ᶴ	ᶵ ᶵ	ᶶ ᶶ	ᶷ ᶷ
ᶸ ᶸ	ᶹ ᶹ	ᶺ ᶺ	ᶻ ᶻ	ᶼ ᶼ	ᶽ ᶽ	ᶾ ᶾ	ᶿ ᶿ

45) Diacritical Marks; Combining Diacritical Marks Supplement (64, 1dc0-1dff):

᷀ ᷀	᷁ ᷁	᷂ ᷂	᷃ ᷃	᷄ ᷄	᷅ ᷅	᷆ ᷆	᷇ ᷇
᷈ ᷈	᷉ ᷉	᷊ ᷊	᷋	᷌	᷍	᷎	᷏

᷐	᷑	᷒	ᷓ	ᷔ	ᷕ	ᷖ	ᷗ
ᷘ	ᷙ	ᷚ	ᷛ	ᷜ	ᷝ	ᷞ	ᷟ
ᷠ	ᷡ	ᷢ	ᷣ	ᷤ	ᷥ	ᷦ	ᷧ □
ᷨ □	ᷩ □	ᷪ □	ᷫ □	ᷬ □	ᷭ □	ᷮ □	ᷯ □
ᷰ □	ᷱ □	ᷲ □	ᷳ □	ᷴ □	᷵ □	᷶ □	᷷ □
᷸ □	᷹ □	᷺ □	᷻ □	᷼ □	᷽ □	᷾ ‹	᷿

46) Text mode is the default. It may be transmitted after SI in bank 2 and is resumed after NL. Numeric mode shall be entered after an SO in bank 2. In numeric mode, commas are field delimiters and are not stored or transmitted, but may be displayed as triple digit separators. Also in numeric mode, U means units count (in rack units it is 1.75”, 43.82 mm), D or da deca, h hecto, k kilo, M Mega, G Giga, T Tera, P Peta, E Exa (when followed by a letter e.g. EB is exabytes), Z zetta, Y Yoda (or yotta), d deci, m milli, u or **µ** or mc micro (medicines may use mc in the USA), n nano, p pico, f femto, a atto, z zepto, y yocto. The letter may replace a decimal point, e.g. 10k5. *e* the value of *e*, π has the numeric value, E means the exponent is following a sign for floating point values. I.e. 9M5 = 9,500,000=9E5+6. So exponents are normally integer, though with a decimal point roots are also permitted. This presents a problem when the decimal is not exact e.g. for cube roots so the rounding number of digits must then be written for a computer to calculate to the required accuracy. It may be better to express in the exponent form in such cases. Functions are referred to as the three or four character upper case letter, display is right justified. Commas in text of numbers are not copied when a number is copied into a numeric item space. A comma transmitted after a tab is interpreted as a field delimiter for csv data which may be including text. Normal scientific units notation shall apply e.g. m for metres when preceded by a value letter or decimal point, or the exponent is in the number. If different units from mksA are used, conversion shall be to the resolution of the largest number of digits plus one with rounding. A numeric mode character or character sequence shall be transmitted by entering ALT+NUMLOCK. A text mode character or character sequence shall be transmitted by entering ESC then NUMLOCK. A space may or may not be between the numeric value and the multiplier letter, which may be followed by the scientific units of measure if not a count of items.
47) The Cyrillic keyboard layout is so different, compatibility of layout was not attempted. Also it was missing Cyrillic characters.
48) The Cyrillic binary code is often different from ASCII. So by switching the key entry on a modified Japanese keyboard to Cyrillic, this indicates to the Operating System that the different binary is to be used. In addition, there is some provision for some individual characters to be selected using an ESC-CTRL-character (with or without shift) on non-modified Japanese keyboards.
49) Dvorak key layout appears to have de-facto lost in the market. However ergonomic study and improved layout for the new assignments is worth researching.

50) IF a [CR][LF] IS NOT found as a pair, BUT either a < or a bank switch character is found next, then an error is assumed in either the [CR] or the [LF], and is reinserted.
51) A line MAY begin with { and be followed by 0-9 or A,B,C,D,E,or F. If so, this is the start of a checksum and line number format as four characters for two bytes using ISO 1155. This SHALL be followed by L# and character pairs, with a leading 0 if needed for a line number starting at the second line. The first line is referred to as 00. The maximum line couint is 000000, which SHALL be a new start for error correction. If an error in the message checksum count is detected, a reply messageconsisting of {LRnnnnnn} SHALL initiate retransmission at that line number where nnnnnn is the line count number. IF there are not a set of four characters 0-9 or A,B,C,D,E,or F THEN this is not a checksum total. IF it is a line number, the character after the last digit SHALL be }.
52) The ISO recommended format for larger numbers is the period for the decimal point and a space between a 3 digit group and another digit or more. An illustration is “1 234.456 7” and commas are not used. However as Word 2010 defaults to ¼ Em space, this is what is used in the second space example as others display larger. When binary or hexadecimal are used, the digits are in groups of 4 with non-breaking space. Word uses CTRL-SHIFT-SPACE to enter the non-breaking space character, the first in the example. ISO 2022 defines the ESC usage, but ESC-SPACE may be available. When you add characters before to make the illustration above move to another line, the second space can be the break to the next line, but the first space does not. A normal space is an Em space which is larger and is a string delimiter and is a normal character size depending on the font although the WYSIWYG of the word processor usually makes it look smaller. See Wikipedia “Space (punctuation)”, “Non-breaking space” for the usage as a non-breaking space and use in SI, International System of Units. There SHALL be a non-breaking space between the last digit and the first character of the SI units designation. Then the whole value is a string. Because the non-breaking space would require a bank selection character before it, and normally would require a default ASCII character afterward, the bank switching SHALL take place back to default ASCII. In HTML the character is   and in Unicode is U+202F known as NNBSP. Excel 2010 has a symbol special character of non-breaking space, that can be manually inserted. Using CTRL-SHIFT-SPACE does not work. However it is not available as a menu format option for regular numbers or SI values. Also Excel arithmetic with such values that include non-breaking spaces yields a *VALUE error. So ISO SI is not implemented whereas the units should be able to be included in the math with non-breaking spaces between the numbers and the units. A non-breaking-space SHALL NOT be considered a string delimiter, unlike a normal space, comma, tab, null, quote marks or maybe others in computer languages. The arithmetic of interacting units to determine the appropriate units of the answer would also be desirable. The default units system SHALL be mksAC or mksAK when Kelvin temperature is used. Also Excel does not include as a menu option the ISO preferred date format of YYYY-MM-DD. Another limitation is the non-provision of frames as used in video or movie timecode, though the frame rate meeds specification. There is a macro available to address this limitation, which could be included in Excel. As can be seen below, other units can be expressed in terms of the base units, and in an equation, the units of everything on one side, when calculated out, equal the units of everything on the other side. Torque is also N.m. but vector multiplication, which is provided for as dot and cross product above. Fixed space may also be used in equations to keep them as a single string when parsing. Another code problem that Excel and various people have is international phone numbers. An example is +CCCC(AAA)EEEEE-LLLLxPPPPP where + means international access, C is 1-4 digits country code, AAA is 1-3 digits

area code (ocassionally more) EEEEE is 1-5 digits exchange code and LLLL is 4 digits local number (sometimes 3 or 2), PPPPP, if used is the 2-5 digit extension number. While this makes 15 digits for local number assignment, worldwide usage makes for less. Currently there is no definition as to when an area code is required prior to the EEEE-LLLL for local calls, but using {} instead of () may provide for this. Recognizing this pattern is not a fixed length is currently not in Excel. International caller ID does provide country code and national number does work from some countries. Wikipedia has more information. The U.S. and Canada have 1 as both the long distance access and the combined country code.

Named units derived from SI base units[31]:3

Name	Symbol	Quantity	Expressed in terms of other SI units	Expressed in terms of SI base units
radian	rad	angle		$m \cdot m^{-1}$
steradian	sr	solid angle		$m^2 \cdot m^{-2}$
hertz	Hz	frequency		s^{-1}
newton	N	force, weight		$kg \cdot m \cdot s^{-2}$
pascal	Pa	pressure, stress	N/m^2	$kg \cdot m^{-1} \cdot s^{-2}$
joule	J	energy, work, heat	$N \cdot m$	$kg \cdot m^2 \cdot s^{-2}$

watt	W	power, radiant flux	J/s	$kg \cdot m^2 \cdot s^{-3}$
coulomb	C	electric charge or quantity of electricity		$s \cdot A$
volt	V	voltage (electrical potential difference), electromotive force	W/A	$kg \cdot m^2 \cdot s^{-3} \cdot A^{-1}$
farad	F	electric capacitance	C/V	$kg^{-1} \cdot m^{-2} \cdot s^4 \cdot A^2$
ohm	Ω	electric resistance, impedance, reactance	V/A	$kg \cdot m^2 \cdot s^{-3} \cdot A^{-2}$
siemens	S	electrical conductance	A/V	$kg^{-1} \cdot m^{-2} \cdot s^3 \cdot A^2$
weber	Wb	magnetic flux	V·s	$kg \cdot m^2 \cdot s^{-2} \cdot A^{-1}$
tesla	T	magnetic field strength	Wb/m^2	$kg \cdot s^{-2} \cdot A^{-1}$
henry	H	inductance	Wb/A	$kg \cdot m^2 \cdot s^{-2} \cdot A^{-2}$
degree Celsius	°C	temperature relative to 273.15 K		K
lumen	lm	luminous flux	cd·sr	cd

lux	lx	illuminance	lm/m^2	m^{-2}·cd
becquerel	Bq	radioactivity (decays per unit time)		s^{-1}
gray	Gy	absorbed dose (of ionizing radiation)	J/kg	m^2·s^{-2}
sievert	Sv	equivalent dose (of ionizing radiation)	J/kg	m^2·s^{-2}
katal	kat	catalytic activity		s^{-1}·mol

Standard prefixes for the SI units of measure

Multiples	**Prefix name**		**deca**	**hecto**	**kilo**	**mega**	**giga**	**tera**	**peta**	**exa**	**zetta**	**yotta**
	Prefix symbol		da	h	k	M	G	T	P	E	Z	Y
	Factor	10^0	10^1	10^2	10^3	10^6	10^9	10^{12}	10^{15}	10^{18}	10^{21}	10^{24}

Fractions	**Prefix name**		**deci**	**centi**	**milli**	**micro**	**nano**	**pico**	**femto**	**atto**	**zepto**	**yocto**
	Prefix symbol		d	c	m	µ	n	p	f	a	z	y
	Factor	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-6}	10^{-9}	10^{-12}	10^{-15}	10^{-18}	10^{-21}	10^{-24}

53) To be Continued, this is an entry placeholder.

Japanese Keyboard Layout Chart

ESC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12		
~ `	ぬ	ふ	あ	う	え	お	や	ゆ	よ	わを	ほ=	·へ ~^		XAscii+
~ `	ɦ	ɭ	ɮ	ʁ	ʎ	ɯ	ɳ		ç	ɣ	ɖ	χ		IPAs
~ `	ɦ	ɭ	ɮ	ʁ	ʎ	ɯ	ɳ	×	ç	ɣ	ɖ	χ		IPA
~ `	1 ! ã	2 @ â	3 # î	4 $ €	5 % £	6 ^ ¥	7 & ¤	8 *	9 (é	0)	- _ -	= + ë	BS	ALT
ɔ Ɔ	ɣ Γ	ð Δ	ʒ Ζ	ŋ Η	θ Θ	ʎ Λ	μ Μ	ξ Ξ	n Π	ρ Ρ 『	ḗ 、	ȅ 。	ě 「	Grk/IPA
↑	B	Ʒ	D	G	S	J	Z	L	Æ	E	Ɛ	ɹ	Ɯ	IPA+
↓	b ㅂ	ʝ ㅈ	d ㄷ	g ㄱ	s ㅅ	j ㅛ	z ㅓ	i ㅣ	æ ㅐ	e ㅔ	ɛ	ɹ	ɯ	Hangul
	q た	W て	い	R す	か	ん	u な	に	ら	せ		ASCII	not	IPA
TAB	Q õ	W å	E	R	T	Y ý	U ú	I û	O ó	P ù	[{	] }	\ |	ALT
	ς Σ	σ Σ	τ Τ	φ Φ	ψ Ψ	ω Ω	ɾ	ʒ	ɦ	ʊ 』	≈	」 ¬		Grk/IPA
	M	N ɢ ɢ	ɲ |	R ǂ	H ‖	Oɡ ɡ	Ə ɓ ɓ	Ad ɗ	I ɟ ɟ	ʔ ɒ	ɲ^	ð– ¥|		IPAs IPA+
	m ▭	n ㄴ	ŋ ㅇ	r ㄹ	h ㅎ	o ㅗ	ə ㅓ	a ㅏ	i ㅣ	ʔ ɒ	ɲ ^	ð		IPA Hangul
Kana	ち	と	し	f は	G き	く	ま	の	L り	れ	け	む	not	IPA

CAPS	A ô	S ä	D ê	F	G ÿ	H à	J ü	K ò	L ö	; :	‘ “	` ~ ï	CRLF	ALT
	±	≤	≥	≠	√	ⁿ√	²√	ˋ	°	•	ɜ ʒ			XAscii+
	K ɻ ɻ	T ɨ ɨ	ʒ ⊙	P ɾ ɾ	∫ ↘ ↗	Y !	U ’	Ħ	ˑ	ʂ ‿	ʊ		IPAs	IPA+
	k ㅋ	t ㅌ	ʒ ㅊ	p ㅍ	∫ ㅠ	y ㅜ	u ㅡ	ħ	˚	ʂ	ʊ		IPA	Hangul
	つ	さ	そ	ひ	こ	み	よ	ね	る	め	ろ	ASCII	not	IPA
SHIFT	Z æ	X ø	C ç	V í	B ¿	N ñ	M	, < á	. > è	/ ? ÷		SHIFT	Base	ALT
	Alpha- ALT	Or XAscii											Norm ASCII	ALT/ XAscii

Keyboard Layout For 3270
Silicon Graphics
White = shift + key
Blue = alt + key
Attn
PF13 PF1
PF14 PF2
PF15 PF3
PF16 PF4
PF17 PF5
PF18 PF6
PF19 PF7
PF20 PF8
PF21 PF9
PF22 PF10
PF23 PF11
PF24 PF12
Print SysRq
Test
Clear
Dup Insert
FldMk Home
PA3 PA1
Delete
ErEOF ErInp
PA2
英 記号
カナ 記号
英数
カナ
Reset
Alt
Alt
Enter

01	3B	3C	3D	3E	3F	40	41	42	43	44	57	58		
ESC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12		
29	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	
	ぬ	ふ	あ	う	え	お	や	ゆ	よ	わ を	ほ=	·へ ~^		XAscii+
Jp.Small/ Normal Reg/ext Cyrillic	ɦ ヌ	ɭ フ	ɮ ア	ʁ ウ	ʎ エ	ɯ オ	ɳ ヤ	ユ	ç ヨ	ɣ ワ	ɖ ホ	χ ヘ		IPAs
	ɦ	ɭ	ɮ	ʁ	ʎ	ɯ	ɳ	×	ç	ɣ	ɖ	χ		IPA
	1 ! ã	2 @ â	3 # î	4 $ €	5 % £	6 ^ ¥	7 & ¤	8 *	9 (é	0)	- _ -	= + ë	BS	ALT
0F	10	11	12	13	14	15	16	17	18	19	1A	1B	7D	
ɔ Ɔ	ɣ Γ	δ Δ	ζ Ζ	η Η	θ Θ	λ Λ	μ Μ	ξ Ξ	π Π	ρ Ρ 』	ḗ 、	ȅ 。	ě 「	Grk/IPA
↑	B タ	ʒ テ	D イ	G ス	S カ	J ン	Z ナ	L ニ	Æ ラ	E セ	ɕ ロ	ɹ	ɱ ム	IPA+
↓	b ㅂ	ʝ ㅈ	d ㄷ	g ㄱ	s ㅅ	j ㅛ	z ㅓ	i ㅣ	æ ㅐ	e ㅔ	ɕ	ɹ	ɱ	Hangul
	q た	W て	い	R す	か	ん	υ な	に	ら	せ		ASCII	not	IPA
TAB	q õ	W å	E	R	T	Y ý	υ ú	I û	O ó	P ù	[{	] }	\ |	ALT

3A	1E	1F	20	21	22	23	24	25	26	27	28	2B	1C	
See Table	ς Σ チ	σ Σ ト	τ Τ シ	φ Φ ハ	ψ Ψ キ	ω Ω ク	ɾ ɾ マ	ʒ ʒ ノ	ɦ ɦ リ	ʊ ʊ レ ‖	≈ ケ	⌋ ¬	CRLF Enter	Grk/IPA
	M	N ɢ ɢ	ɲ |	R ǂ	H ‖	O ɡ ɡ	Ə ɓ ɓ	A ɗ ɗ	I ɟ ɟ	ʔ ɒ	ɲ ^	ð – ¥ |	IPAs	IPA+
	m ▭	n ∟	ŋ ⬭	r ⌠	h ⩲	o ⊥	ə ┤	a ├	i |	ʔ ɒ	ɲ ^	ð	IPA	Hangul
Kana	ち	と	し	f は	G き	く	ま	の	L り	れ	け	む	not	IPA
CAPS	A ô	S ä	D ê	f	G ÿ	H à	J ü	K ò	L ö	; :	‘ “	` ~ ï	CRLF	ALT
2A	2C	2D	2E	2F	30	31	32	33	34	35	73	36		
	± ツ	≤ サ	≥ ン	≠ ヒ	√ コ	ⁿ√	²√ モ	ネ	ル	•	ʒ Ʒ			XAscii+
	K ɻ ɻ	T ɬ ɬ	Ʒ ⊙	P ʈ ʈ	∫ ↘ ↗	Y ! ミ	U ’	Ħ	˸	ʂ ‿ メ	ʊ		IPAs	IPA+
	k ㅋ	t ㅌ	ʒ ㅊ	p ㅍ	ʃ ㅠ	y ㅜ	u ㅡ	ħ、	˚	ʂ	ʊ		IPA	Hangul
	つ	x さ	c そ	v ひ	B こ	み	よ	ね	る	め	ろ	ASCII	not	IPA
SHIFT	Z æ	X ø	C ç	V í	B ¿	N ñ	M	, < á	. > è	/ ? ÷		SHIFT	Base	ALT

Ctl 1D	Alpha-ALT	Or XAscii	WIN E0 5B	Alt 38	7B No change Chinese	SP 39	79 Change Chinese	70 kata /hira/ Cyrillic	Alt E0 38	WIN E0 5c	Menu E0 5d	Ctl E0 1D	Norm ASCII	ALT/ XAscii

WITH VIDEO EDITING FOR ADOBE CS6

01	3B	3C	3D	3E	3F	40	41	42	43	44	57	58		
					Capture	Batch capture			Title design Shift, Edit Work space	Shift Effects Work Space	Shift Audio Work space			2xLeft Click on project wksp, import
ESC	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12		
29	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	
	ぬ	ふ	あ	う	え	お	や	ゆ	よ	わ を	ほ=	·へ~^		XAscii+
Jp.small/ Normal Reg/ext Cyrillic	ɦ ヌ	ɭ フ	ʧ ア	ʁ ウ	ʏ エ	ɰ オ	ɳ ヤ	ユ	ç ヨ	ɣ ワ	ɖ ホ	χ へ		IPAs
	ɦ	ɭ	ʧ	ʁ	ʏ	ɰ	ɳ	×	ç	ɣ	ɖ	χ		IPA
											Zoom in	Zoom out		

	1 ! ã	2 @ â	3 # î	4 $ €	5 % £	6 ^ ¥	7 & ¤	8 *	9 (é	0)	- _ -	= + ë	BS	ALT
								Un-numbered marker						
0F	10	11	12	13	14	15	16	17	18	19	1A	1B	7D	
ɔ Ɔ	γ Γ	δ Δ	ζ Ζ	η Η	θ Θ	λ Λ	μ Μ	ξ Ξ	π Π	ρ Ρ 『	ḗ 、	ḕ 。	ḗ 「	Grk/IPA
↑	B タ	ʒ テ	D イ	G ス	S カ	J ン	Z ナ	L ニ	Æ ラ	E セ	ɕ ロ	ɟ	ʍ ム	IPA+
↓	b ㅂ	ʝ ㅈ	d ㄷ	g ㄱ	s ㅅ	j ㅛ	z ㅑ	i ㅏ	æ ㅐ	e ㅔ	ɕ	ɟ	ʍ	Hangul
	q た	W て	い	R す	か	ん	ʊ な	に	ら	せ		ASCII	not	IPA
						Slip edit tool	Slide edit tool			Pen tool key frame				
TAB	q õ	W å	E	R	T	Y ý	ʊ ú	I û	O ó	P ù	[{	] }	\ |	ALT
3A	1E	1F	20	21	22	23	24	25	26	27	28	2B	1C	
See table	ς Σ チ	σ Σ ト	τ Τ シ	φ Φ ハ	ψ Ψ キ	ω Ω ク	ɾ ɾ マ	ʐ ʐ ノ	ɦ ɦ リ	ʊ ʊ レ	≈ ケ	⌋ ¬	CRLF Enter	Grk/IPA

										『				
	M	N ɢ ɢ	ɲ |	R ǂ	H ‖	Oɡ ɡ	Ə ɓ ɓ	Ad ɗ	I ɟ ɟ	ʔ ɒ	ɲ^	ð–¥ |	IPAs	IPA+
	m ▭	n ∟	ŋ ○	r ⌐	h ⌂	o ⊥	ə ┤	a ├	i |	ʔ ɒ	ɲ ^	ð	IPA	Hangul
Kana	ち	と	し	f は	G き	く	ま	の	L り	れ	け	む	not	IPA
	Ctl SelAll CtlSft UnSel All		Ctl Add defaul t trans				Rev	Play	Fwd				Render	Left arrow, frame -
CAPS	A ô	S ä	D ê	f	G ÿ	H à	J ü	K ò	L ö	; :	‘ “	` ~ ī	CRLF	ALT
2A	2C	2D	2E	2F	30	31	32	33	34	35	73	36		
End, GoTo End	±ツ	≤サ	≥ン	≠ヒ	√コ	ⁿ√	²√モ	ネ	ル	•	ʒ 3		Home, GoTo Beginning	XAscii+
PageUp Next Cut	K ɻ ɻ	T ɬ ɬ	3 ⊙	P ʈ ʈ	∫ ↘ ↗	Y ! ミ	U ’	Ħ	ˑ	ʂ ‿ メ	ʊ		IPAs	IPA+
	k ㅋ	t ㅌ	3 ㅊ	p ㅍ	∫ ㅠ	y ㅜ	u ㅡ	ħ、	◦	ʂ	ʊ		IPA	Hangul
PgeDwn	つ	x さ	c そ	v	B こ	み	よ	ね	る	め	ろ	ASCII	Not	IPA

Prev. Cut				ひ										
	Ctl Undo CtlSft Redo	Rate stretch tool Ctl Cut	Razor tool Ctl Copy	Sel Clip. Shift, add to sel. Ctl Paste	Ripple edit tool			Insert source	Overlay source				Right arrow, frame+	
SHIFT	Z æ	X ø	C ç	V í	B ¿	N ñ	M	, < á	. > è	/ ? ÷		SHIFT	Base	ALT
Ctl 1D Drag ripple	Alpha-ALT	Or XAscii	WIN E0 5B	Alt 38	7B No change Chinese	SP 39 Play /Stop	79 Change Chinese	70 kata /hira/ Cyrillic	Alt E0 38 Drag, unlink A/V	WIN E0 5c	Menu E0 5d	Ctl E0 1D	Norm ASCII	ALT/ XAscii

Korean Hangul

Sound like ae as in hay, from ㅐ

ㅐ

	ka	kya	ko	kyo	kō	kyo	ku	kyu	kū	ki		
	가	갸	거	겨	고	교	구	규	그	기	ㄲ	ㄱ
	1	27	27	35	46	70	73	87	88	97	106	23 74
	1	85	86	114	152	220	229	269	271	301	333	
na	나	냐	너	녀	노	뇨	누	뉴	느	니		ㄴ
	113	127	127	129	129	–	136	139	139	140		123 405
	365	425	425	435	435		450	473	474	484		
ta	다	댜	더	뎌	도	됴	두	듀	드	디	따	ㄷ
	141	–	161	–	164	–	177	181	181	187	187	153 526
	485		552		565		610	627	627	647	649	
ra	라	랴	러	려	로	료	루	류	르	리		ㄹ
	194	–	195	–	195	–	196	196	196	196		195 686
	680	687	687	690	691		695	695	696	696		
ma	마	먀	머	며	모	묘	무	뮤	므	미		ㅁ
	197	–	212	214	218	226	226	–	238	238		208 749
	698		761	774	787	817	818	863	864	864		
pa	바	뱌	버	벼	보	뵤	부	뷰	브	비	빠	ㅂ
	245	–	263	269	274	–	282	–	299	299	305	258 937
	885		951	969	985		1013	1067	1067	1068	1092	
sa	사	샤	서	셔	소	쇼	수	슈	스	시	싸	ㅅ
	308	331	331	345	345	355	356	367	367	370	387	327 1178
	1104	1196	1196	1247	1248	1282	1282	1321	1321	1333	1391	
a	아	야	어	여	오	요	우	유	으	이		ㅇ
	390	404	411	421	437	449	453	467	476	484		402 1452
	1406	1458	1476	1520	1573	1610	1621	1662	1684	1710		
cha	자	쟈	저	져	조	죠	주	쥬	즈	지	짜	ㅈ
	508	–	526	–	553	–	563	–	573	575	590	522 1854
	1746		1867		1947		1977		2012	2019	2068	
ccha	차	챠	처	쳐	초	쵸	추	츄	츠	치		ㅊ
	593	–	599	608	608	–	613	–	621	621		598 2078
	2078		2103	2134	2134		2153		2181	2183		
kka	카	캬	커	켜	코	쿄	쿠	큐	크	키		ㅋ
	626	–	627	628	628	–	630	630	630	631		--- 2203
	2200		2204	2207	2208		2214	2216	2216	2231		
tta	타	탸	터	텨	토	툐	투	튜	트	티		ㅌ
	632	–	638	–	640	–	645	646	646	649		637 2244
	2222		2249		2259		2276	2282	2282	2291		
ppa	파	퍄	퍼	펴	포	표	푸	퓨	프	피		ㅍ
	649	655	655	656	660	662	664	667	667	668		653 2308
	2293	2312	2312	2314	2331	2342	2347	2359	2359	2362		
ha	하	햐	허	혀	호	효	후	휴	호	히		ㅎ
	671	686	686	690	694	706	707	711	712	716		682 2410
	2373	2427	2429	2442	2459	2518	2520	2538	2543	2563		

the numbers are the pages in dictionaries, ignore them

Cyrillic

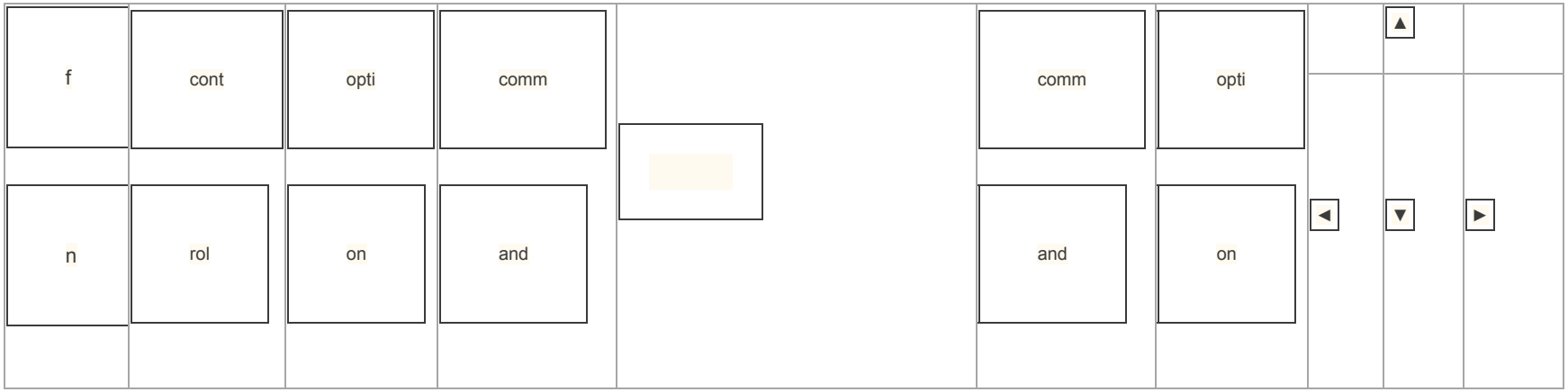
Letters of the Cyrillic alphabet (see also Cyrillic digraphs)											
А A	Б Be	В Ve	Г Ge	Ґ Ge upturn	Д De	Ђ Dje	Ѓ Gje	Е Ye	Ё Yo	Є Yest	Ж Zhe
З Ze	Ѕ Zje	S Dze	И I	І Dotted I	Ї Yi	Й Short I	Ј Je	К Ka	Л El	Љ Lje	М Em
Н En	Њ Nje	О O	П Pe	Р Er	С Es	Ć [ɕ]	Т Te	Ћ Tshe	Ќ Kje	У U	Ў Short U
Ф Ef	Х Kha	Ц Tse	Ч Che	Џ Dzhe	Ш Sha	Щ Shcha	Ъ Hard sign (Yer)	Ы Yery	Ь Soft sign (Yeri)	Э E	Ю Yu
Я Ya											
Cyrillic non-Slavic letters											
Ӏ Palochka	Ә Cyrillic Schwa	Ғ Ayn	Ҙ Bashkir Dhe	Ҫ Bashkir The	Ҡ Bashkir Qa	Ҟ Ka with descender	Ӈ Ng	Ө Barred O	Ү Straight U	Ұ Straight U with stroke	Ҥ He
Cyrillic letters used in the past											
□	Ѥ	Ѧ	Ѫ	Ѩ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ҁ

A iotified | E iotified | Yus small | Yus big | Yus small iotified | Yus big iotified | Ksi | Psi | Fita | Izhitsa | Izhitsa okovy | Koppa

Оу Uk | ꙍ Omega | ѿ Ot | Ѣ

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Ё	1	2	3	4	5	6	7	8	9	0	-	+	dele te	
tab	Я	Ж	Е	Р	Т	Ю	У	И	О	П	Ь	Ъ	.:	
caps lock	А	С	Д	Ф	Г	Х	Й	К	Л	Ш	Щ	return		
shift	З	Ч	Ц	В	Б	Н	М	Э	Ы	,;	shift			



Russian, Belarusian and Ukranian standard keyboards

Green color denotes differences between standard Russian and Belarusian layouts; blue color denotes differences between standard Russian layout and “Ukranian (Enhanced)” one (available in Windows Vista and above).

[Alt Gr] key is used in the Ukranian layout only, being responsible for the single letter “ge with upturn” [Ґґ], denoted here with red color.

These layouts are so incompatible with the ASCII layout that layout similarity is not attempted. Instead, non-ASCII characters are code entered for users and the use of binary converted in software when required. Not all characters

in the list above are shown on the keyboard layouts. The backwards S (top left in Belorussian) appears to be infrequently used, so is not provided for.

KOI8-U (see also KO17 following for 7 bit caps only)

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
0_																
1_																
2_	SP 0020 ***32***	! 0021 ***33***	" 0022 ***34***	# 0023 ***35***	$ 0024 ***36***	% 0025 ***37***	& 0026 ***38***	' 0027 ***39***	(0028 ***40***	) 0029 ***41***	* 002A ***42***	+ 002B ***43***	, 002C ***44***	- 002D ***45***	. 002E ***46***	/ 002F ***47***
3_	0 0030 ***48***	1 0031 ***49***	2 0032 ***50***	3 0033 ***51***	4 0034 ***52***	5 0035 ***53***	6 0036 ***54***	7 0037 ***55***	8 0038 ***56***	9 0039 ***57***	: 003A ***58***	; 003B ***59***	< 003C ***60***	= 003D ***61***	>>
003E ***62***	? 003F ***63***
4_	@ 0040 ***64***	A 0041 ***65***	B 0042 ***66***	C 0043 ***67***	D 0044 ***68***	E 0045 ***69***	F 0046 ***70***	G 0047 ***71***	H 0048 ***72***	I 0049 ***73***	J 004A ***74***	K 004B ***75***	L 004C ***76***	M 004D ***77***	N 004E ***78***	O 004F ***79***
5_	P 0050 ***80***	Q 0051 ***81***	R 0052 ***82***	S 0053 ***83***	T 0054 ***84***	U 0055 ***85***	V 0056 ***86***	W 0057 ***87***	X 0058 ***88***	Y 0059 ***89***	Z 005A ***90***	[005B ***91***	\ 005C ***92***	] 005D ***93***	^ 005E ***94***	_ 005F ***95***
	` ``	a a	b b	c c	d d	e e	f f	g g	h h	i i	j j	k k	l l	m m	n n	o o

6_	0060 ***96***	0061 ***97***	0062 ***98***	0063 ***99***	0064 ***100***	0065 ***101***	0066 ***102***	0067 ***103***	0068 ***104***	0069 ***105***	006A ***106***	006B ***107***	006C ***108***	006D ***109***	006E ***110***	006F ***111***
7_	p 0070 ***112***	q 0071 ***113***	r 0072 ***114***	s 0073 ***115***	t 0074 ***116***	u 0075 ***117***	v 0076 ***118***	w 0077 ***119***	x 0078 ***120***	y 0079 ***121***	z 007A ***122***	{ 007B ***123***	| 007C ***124***	} 007D ***125***	~ 007E ***126***	
8_	─ 2500 ***128***	│ 2502 ***129***	┌ 250C ***130***	┐ 2510 ***131***	└ 2514 ***132***	┘ 2518 ***133***	├ 251C ***134***	┤ 2524 ***135***	┬ 252C ***136***	┴ 2534 ***137***	┼ 253C ***138***	▀ 2580 ***139***	▄ 2584 ***140***	▀ 2588 ***141***	▌ 258C ***142***	▐ 2590 ***143***
9_	▒ 2591 ***144***	▒ 2592 ***145***	▒ 2593 ***146***	⌠ 2320 ***147***	■ 25A0 ***148***	· 2219 ***149***	√ 221A ***150***	≈ 2248 ***151***	≤ 2264 ***152***	≥ 2265 ***153***	NBSP 00A0 ***154***	⌡ 2321 ***155***	° 00B0 ***156***	² 00B2 ***157***	· 00B7 ***158***	÷ 00F7 ***159***
A_	═ 2550 ***160***	║ 2551 ***161***	╒ 2552 ***162***	ё 0451 ***163***	є 0454 ***164***	╗ 2554 ***165***	і 0456 ***166***	ї 0457 ***167***	╕ 2557 ***168***	╘ 2558 ***169***	╙ 2559 ***170***	╚ 255A ***171***	╛ 255B ***172***	ґ 0491 ***173***	╝ 255D ***174***	╞ 255E ***175***
B_	╟ 255F ***176***	╠ 2560 ***177***	╡ 2561 ***178***	Ё 0401 ***179***	Є 0404 ***180***	╣ 2563 ***181***	І 0406 ***182***	Ї 0407 ***183***	╦ 2566 ***184***	╧ 2567 ***185***	╨ 2568 ***186***	╩ 2569 ***187***	╪ 256A ***188***	Ґ 0490 ***189***	╬ 256C ***190***	© 00A9 ***191***
C_	ю 044E ***192***	а 0430 ***193***	б 0431 ***194***	ц 0446 ***195***	д 0434 ***196***	е 0435 ***197***	ф 0444 ***198***	г 0433 ***199***	х 0445 ***200***	и 0438 ***201***	й 0439 ***202***	к 043A ***203***	л 043B ***204***	м 043C ***205***	н 043D ***206***	о 043E ***207***
D_	п 043F ***208***	я 044F ***209***	р 0440 ***210***	с 0441 ***211***	т 0442 ***212***	у 0443 ***213***	ж 0436 ***214***	в 0432 ***215***	ь 044C ***216***	ы 044B ***217***	з 0437 ***218***	ш 0448 ***219***	э 044D ***220***	щ 0449 ***221***	ч 0447 ***222***	ъ 044A ***223***

E_	Ю 042E ***224***	А 0410 ***225***	Б 0411 ***226***	Ц 0426 ***227***	Д 0414 ***228***	Е 0415 ***229***	Ф 0424 ***230***	Г 0413 ***231***	Х 0425 ***232***	И 0418 ***233***	Й 0419 ***234***	К 041A ***235***	Л 041B ***236***	М 041C ***237***	Н 041D ***238***	О 041E ***239***
F_	П 041F ***240***	Я 042F ***241***	Р 0420 ***242***	С 0421 ***243***	Т 0422 ***244***	У 0423 ***245***	Ж 0416 ***246***	В 0412 ***247***	Ь 042C ***248***	Ы 042B ***249***	З 0417 ***250***	Ш 0428 ***251***	Э 042D ***252***	Щ 0429 ***253***	Ч 0427 ***254***	Ъ 042A ***255***
	_0	**_1**	**_2**	**_3**	**_4**	**_5**	**_6**	**_7**	**_8**	**_9**	**_A**	**_B**	**_C**	**_D**	**_E**	**_F**

In the table above, 20 is the regular SPACE character, and 9A is the NO-BREAK SPACE.

The difference with KOI8-R consists of the positions 0xA4; 0xA6; 0xA7; 0xAD; and 0xB4; 0xB6; 0xB7; 0xBD; which consist of extra letters that don't exist in Russian.

Although RFC 2319 says that character 95 should be U+2219 (·), it may also be U+2022 (•) to match the bullet character in Windows-1251.

KOI7																
	–0	**–1**	**–2**	**–3**	**–4**	**–5**	**–6**	**–7**	**–8**	**–9**	**–A**	**–B**	**–C**	**–D**	**–E**	**–F**
0_	NUL 0000 ***0***	SOH 0001 ***1***	STX 0002 ***2***	ETX 0003 ***3***	EOT 0004 ***4***	ENQ 0005 ***5***	ACK 0006 ***6***	BEL 0007 ***7***	BS 0008 ***8***	HT 0009 ***9***	LF 000A ***10***	VT 000B ***11***	FF 000C ***12***	CR 000D ***13***	SO 000E ***14***	SI 000F ***15***
1_	DLE 0010 ***16***	DC1 0011 ***17***	DC2 0012 ***18***	DC3 0013 ***19***	DC4 0014 ***20***	NAK 0015 ***21***	SYN 0016 ***22***	ETB 0017 ***23***	CAN 0018 ***24***	EM 0019 ***25***	SUB 001A ***26***	ESC 001B ***27***	FS 001C ***28***	GS 001D ***29***	RS 001E ***30***	US 001F ***31***
2_	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	– 002D	. 002E	/ 002F

	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
3_	0 0030 48	1 0031 49	2 0032 50	3 0033 51	4 0034 52	5 0035 53	6 0036 54	7 0037 55	8 0038 56	9 0039 57	: 003A 58	; 003B 59	< 003C 60	= 003D 61	> 003E 62	? 003F 63
4_	@ 0040 64	A 0041 65	B 0042 66	C 0043 67	D 0044 68	E 0045 69	F 0046 70	G 0047 71	H 0048 72	I 0049 73	J 004A 74	K 004B 75	L 004C 76	M 004D 77	N 004E 78	O 004F 79
5_	P 0050 80	Q 0051 81	R 0052 82	S 0053 83	T 0054 84	U 0055 85	V 0056 86	W 0057 87	X 0058 88	Y 0059 89	Z 005A 90	[005B 91	\ 005C 92	] 005D 93	^ 005E 94	_ 005F 95
6_	Ю 042E 96	А 0410 97	Б 0411 98	Ц 0426 99	Д 0414 100	Е 0415 101	Ф 0424 102	Г 0413 103	Х 0425 104	И 0418 105	Й 0419 106	К 041A 107	Л 041B 108	М 041C 109	Н 041D 110	О 041E 111
7_	П 041F 112	Я 042F 113	Р 0420 114	С 0421 115	Т 0422 116	У 0423 117	Ж 0416 118	В 0412 119	Ь 042C 120	Ы 042B 121	З 0417 122	Ш 0428 123	Э 042D 124	Щ 0429 125	Ч 0427 126	DEL 007F 127

The dollar sign character ("$" hex 24), however, was often replaced with the universal currency sign "¤" in Soviet computers. A popular legend has it that this was to demonstrate independence of the American-dominated computer industry — which was ironic, since a significant number of Soviet computers were actually implementations of various American designs.

ASCII NOT CYRILLIC	ASSIGN			unassigned	secondary		
	a						a

	А	А			Ђ	Dje	А
	в				Б		b
	В	В			Б	Б	В
	с				Ч		с
	С	С			Ч	Ч	С
d	Ь				Д		
D	Ь	Ь			Д	Д	
	е				Э		е
	Е	Е			Э	Э	Е
f	Є						
F	Є	Є			Ə	schwa	
g	Г						
G	Г	Г			Ѓ	Gje	
h	Н						

H	Н	Н			Һ	He	
	І	Or palochka					І
	ӏ	Dotted I			S	Dze	І
j	Ґ						
J	Ґ	Ґ			З́	Zje	
k	к	Or К					
K	К	К			Ќ	Kje	
l	п						
L	П	П			Љ	Lje	
m	м						
M	М	М			Ӊ	Ng	
n	и						
N	И	И			Њ	Nje	
	О						о

	О	О			Ө	Barred O	о
	Р				Л		р
	Р	Р			Л	Л	Р
Q	Й				Ъ		
Q	Й	Й			Ъ	Ъ	
R	Я						
R	Я	Я			Ҳ	Tshe	
s	Ф				Щ		
S	Ф	Ф			Щ	Щ	
t	Т						
T	Т	Т			Ҫ	Bashkir the	
u	Ц				Ю		
U	Ц	Ц			Ю	Ю	
v	Ы				Ӣ		

V	Ы	Ы			Ӣ	Ӣ	
w	Ш						
W	Ш	Ш			Short u	ў	
x	Х	Or kha					
X	Х	Х			Ҭ	Straight U With stroke	
y	У				Ӭ		
Y	У	У			Ӭ	Ӭ	
z	З				Ж		
Z	З	З			Ж	Ж	
-					Ҿ	[ҿ]	
^ “= on US”					Ԥ	dzhe	

[					Ғ	ayn	
]					Ҙ	Bashkir dhe	
\					Ҡ	Bashkir oa	
/ “right of /? On US”					Ҟ	Ka with descender	
A quirk in Word is to	convert	the lower		letter	to	uppercase	

Arabic5

MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media

Code Table Basic Arabic

December 2007

The first column in this table contains the MARC-8 code (in hex) for the character as coming from the G0 graphic set, the second column contains the MARC-8 code (in hex) for the character as coming from the G1 graphic set, the third column contains the

UCS/Unicode 16-bit code (in hex), the fourth column contains the UTF-8 code (in hex) for the UCS characters, the fifth column contains a representation of the character (where possible), the sixth column contains the MARC character name, followed by the UCS name. If the MARC name is the same as or very similar to the UCS name, only the UCS name is given.

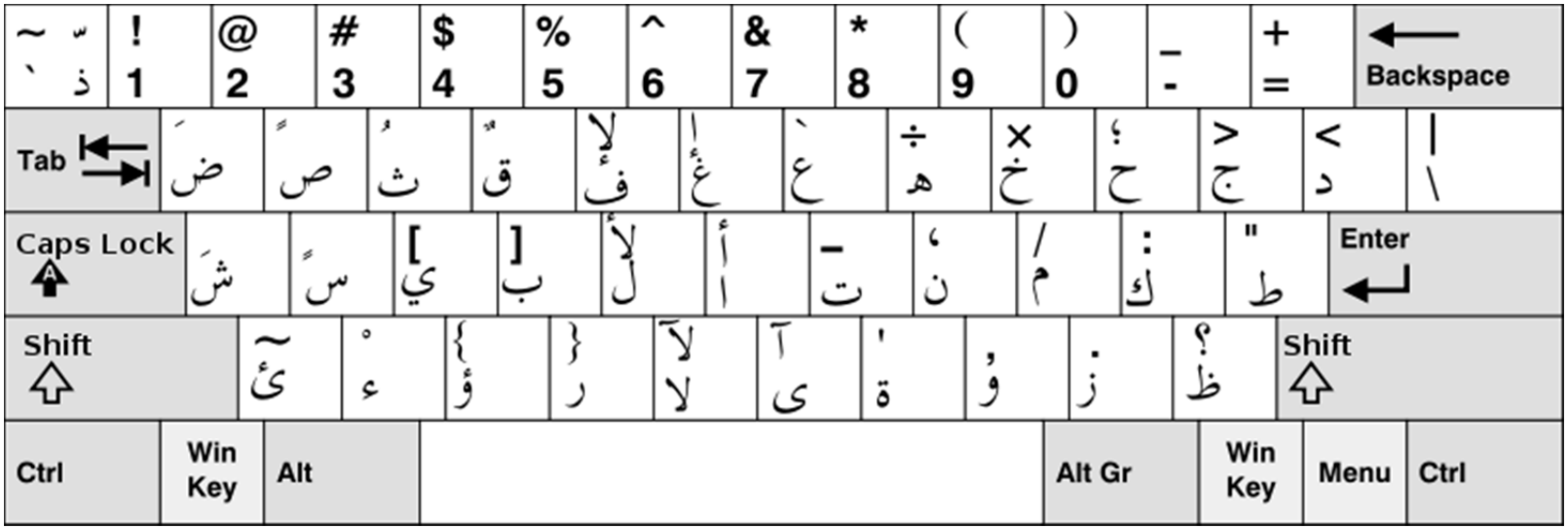
Not all characters display in all browsers. We have attempted to allow for font families that show each character set, but you must have one of these fonts on your computer. See the W3C site for a discussion of fonts: http://www.w3.org/TR/REC-CSS2/fonts.html#generic-font-families.

MARC-8 as G0	MARC-8 as G1	UCS	UTF-8	CHAR	C?	NAME
21	A1	0021	21	!		EXCLAMATION MARK
22	A2	0022	22	"		QUOTATION MARK
23	A3	0023	23	#		NUMBER SIGN
24	A4	0024	24	$		DOLLAR SIGN
25	A5	066A	D9AA	٪		PERCENT SIGN / ARABIC PERCENT SIGN
26	A6	0026	26	&		AMPERSAND
27	A7	0027	27	'		APOSTROPHE
28	A8	0028	28	(		OPENING PARENTHESIS / LEFT PARENTHESIS
29	A9	0029	29	)		CLOSING PARENTHESIS / RIGHT PARENTHESIS
2A	AA	066D	D9AD	*		ASTERISK / ARABIC FIVE POINTED STAR
2B	AB	002B	2B	+		PLUS SIGN
2C	AC	060C	D88C	،		ARABIC COMMA
2D	AD	002D	2D	–		HYPHEN-MINUS

2E	AE	002E	2E	.		PERIOD, DECIMAL POINT / FULL STOP
2F	AF	002F	2F	/		SLASH / SOLIDUS
30	B0	0660	D9A0	٠		ARABIC-INDIC DIGIT ZERO
31	B1	0661	D9A1	١		ARABIC-INDIC DIGIT ONE
32	B2	0662	D9A2	٢		ARABIC-INDIC DIGIT TWO
33	B3	0663	D9A3	٣		ARABIC-INDIC DIGIT THREE
34	B4	0664	D9A4	٤		ARABIC-INDIC DIGIT FOUR
35	B5	0665	D9A5	٥		ARABIC-INDIC DIGIT FIVE
36	B6	0666	D9A6	٦		ARABIC-INDIC DIGIT SIX
37	B7	0667	D9A7	٧		ARABIC-INDIC DIGIT SEVEN
38	B8	0668	D9A8	٨		ARABIC-INDIC DIGIT EIGHT
39	B9	0669	D9A9	٩		ARABIC-INDIC DIGIT NINE
3A	BA	003A	3A	:		COLON
3B	BB	061B	D89B	؛		ARABIC SEMICOLON
3C	BC	003C	3C	<		LESS-THAN SIGN
3D	BD	003D	3D	=		EQUALS SIGN
3E	BE	003E	3E	>		GREATER-THAN SIGN
3F	BF	061F	D89F	؟		ARABIC QUESTION MARK
41	C1	0621	D8A1	ء		HAMZAH / ARABIC LETTER HAMZA
42	C2	0622	D8A2	آ		ARABIC LETTER ALEF WITH MADDA ABOVE
43	C3	0623	D8A3	أ		ARABIC LETTER ALEF WITH HAMZA ABOVE
44	C4	0624	D8A4	ؤ		ARABIC LETTER WAW WITH HAMZA ABOVE
45	C5	0625	D8A5	إ		ARABIC LETTER ALEF WITH HAMZA BELOW

46	C6	0626	D8A6	ئ		ARABIC LETTER YEH WITH HAMZA ABOVE
47	C7	0627	D8A7	ا		ARABIC LETTER ALEF
48	C8	0628	D8A8	ب		ARABIC LETTER BEH
49	C9	0629	D8A9	ة		ARABIC LETTER TEH MARBUTA
4A	CA	062A	D8AA	ت		ARABIC LETTER THE
4B	CB	062B	D8AB	ث		ARABIC LETTER THEH
4C	CC	062C	D8AC	ج		ARABIC LETTER JEEM
4D	CD	062D	D8AD	ح		ARABIC LETTER HAH
4E	CE	062E	D8AE	خ		ARABIC LETTER KHAH
4F	CF	062F	D8AF	د		ARABIC LETTER DAL
50	D0	0630	D8B0	ذ		ARABIC LETTER THAL
51	D1	0631	D8B1	ر		ARABIC LETTER REH
52	D2	0632	D8B2	ز		ARABIC LETTER ZAIN
53	D3	0633	D8B3	س		ARABIC LETTER SEEN
54	D4	0634	D8B4	ش		ARABIC LETTER SHEEN
55	D5	0635	D8B5	ص		ARABIC LETTER SAD
56	D6	0636	D8B6	ض		ARABIC LETTER DAD
57	D7	0637	D8B7	ط		ARABIC LETTER TAH
58	D8	0638	D8B8	ظ		ARABIC LETTER ZAH
59	D9	0639	D8B9	ع		ARABIC LETTER AIN
5A	DA	063A	D8BA	غ		ARABIC LETTER GHAIN
5B	DB	005B	5B	[		OPENING SQUARE BRACKET / LEFT SQUARE BRACKET
5D	DD	005D	5D	]		CLOSING SQUARE BRACKET / RIGHT SQUARE BRACKET
60	E0	0640	D980	ـ		ARABIC TATWEEL

61	E1	0641	D981	ف		ARABIC LETTER FEH
62	E2	0642	D982	ق		ARABIC LETTER QAF
63	E3	0643	D983	ك		ARABIC LETTER KAF
64	E4	0644	D984	ل		ARABIC LETTER LAM
65	E5	0645	D985	م		ARABIC LETTER MEEM
66	E6	0646	D986	ن		ARABIC LETTER NOON
67	E7	0647	D987	ه		ARABIC LETTER HEH
68	E8	0648	D988	و		ARABIC LETTER WAW
69	E9	0649	D989	ى		ARABIC LETTER ALEF MAKSURA
6A	EA	064A	D98A	ي		ARABIC LETTER YEH
6B	EB	064B	D98B	ً	C	ARABIC FATHATAN
6C	EC	064C	D98C	ٌ	C	ARABIC DAMMATAN
6D	ED	064D	D98D	ٍ	C	ARABIC KASRATAN
6E	EE	064E	D98E	َ	C	ARABIC FATHA
6F	EF	064F	D98F	ُ	C	ARABIC DAMMA
70	F0	0650	D990	ِ	C	ARABIC KASRA
71	F1	0651	D991	ّ	C	ARABIC SHADDA
72	F2	0652	D992	ْ	C	ARABIC SUKUN
73	F3	0671	D9B1	ٱ		ARABIC LETTER ALEF WASLA
74	F4	0670	D9B0	ٰ		ARABIC LETTER SUPERSCRIPT ALEF
78	F8	066C	D9AC	٬		ARABIC THOUSANDS SEPARATOR
79	F9	201D	E2809D	”		RIGHT DOUBLE QUOTATION MARK
7A	FA	201C	E2809C	“		LEFT DOUBLE QUOTATION MARK



India, Devanagari

Indian languages have different keyboards, but ISCII is most common.

ISCII Devanagari																
	–0	–1	–2	–3	–4	–5	–6	–7	–8	–9	–A	–B	–C	–D	–E	–F
0_	NUL 0000 ***0***	SOH 0001 ***1***	STX 0002 ***2***	ETX 0003 ***3***	EOT 0004 ***4***	ENQ 0005 ***5***	ACK 0006 ***6***	BEL 0007 ***7***	BS 0008 ***8***	HT 0009 ***9***	LF 000A ***10***	VT 000B ***11***	FF 000C ***12***	CR 000D ***13***	SO 000E ***14***	SI 000F ***15***
1_	DLE 0010 ***16***	DC1 0011 ***17***	DC2 0012 ***18***	DC3 0013 ***19***	DC4 0014 ***20***	NAK 0015 ***21***	SYN 0016 ***22***	ETB 0017 ***23***	CAN 0018 ***24***	EM 0019 ***25***	SUB 001A ***26***	ESC 001B ***27***	FS 001C ***28***	GS 001D ***29***	RS 001E ***30***	US 001F ***31***
2_	SP 0020 ***32***	! 0021 ***33***	" 0022 ***34***	# 0023 ***35***	$ 0024 ***36***	% 0025 ***37***	& 0026 ***38***	' 0027 ***39***	(0028 ***40***	) 0029 ***41***	* 002A ***42***	+ 002B ***43***	, 002C ***44***	– 002D ***45***	. 002E ***46***	/ 002F ***47***

3_	0 0030 **48**	1 0031 **49**	2 0032 **50**	3 0033 **51**	4 0034 **52**	5 0035 **53**	6 0036 **54**	7 0037 **55**	8 0038 **56**	9 0039 **57**	: 003A **58**	; 003B **59**	< 003C **60**	= 003D **61**	> 003E **62**	? 003F **63**
4_	@ 0040 **64**	A 0041 **65**	B 0042 **66**	C 0043 **67**	D 0044 **68**	E 0045 **69**	F 0046 **70**	G 0047 **71**	H 0048 **72**	I 0049 **73**	J 004A **74**	K 004B **75**	L 004C **76**	M 004D **77**	N 004E **78**	O 004F **79**
5_	P 0050 **80**	Q 0051 **81**	R 0052 **82**	S 0053 **83**	T 0054 **84**	U 0055 **85**	V 0056 **86**	W 0057 **87**	X 0058 **88**	Y 0059 **89**	Z 005A **90**	[005B **91**	\ 005C **92**	] 005D **93**	^ 005E **94**	_ 005F **95**
6_	` 0060 **96**	a 0061 **97**	b 0062 **98**	c 0063 **99**	d 0064 **100**	e 0065 **101**	f 0066 **102**	g 0067 **103**	h 0068 **104**	i 0069 **105**	j 006A **106**	k 006B **107**	l 006C **108**	m 006D **109**	n 006E **110**	o 006F **111**
7_	p 0070 **112**	q 0071 **113**	r 0072 **114**	s 0073 **115**	t 0074 **116**	u 0075 **117**	v 0076 **118**	w 0077 **119**	x 0078 **120**	y 0079 **121**	z 007A **122**	{ 007B **123**	| 007C **124**	} 007D **125**	~ 007E **126**	DEL 007F **127**
8_																
9_																
A_		ँ 0901	ं 0902	ः 0903	अ 0905	आ 0906	इ 0907	ई 0908	उ 0909	ऊ 090A	ऋ 090B	ऐ 090E	ए 090F	ऐ 0910	ऍ 090D	ओ 0912

		161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
B_	ओ 0913 176	औ 0914 177	ऑ 0911 178	क 0915 179	ख 0916 180	ग 0917 181	घ 0918 182	ङ 0919 183	च 091A 184	छ 091B 185	ज 091C 186	झ 091D 187	ञ 091E 188	ट 091F 189	ठ 0920 190	ड 0921 191
C_	ढ 0922 192	ण 0923 193	त 0924 194	थ 0925 195	द 0926 196	ध 0927 197	न 0928 198	ऩ 0929 199	प 092A 200	फ 092B 201	ब 092C 202	भ 092D 203	म 092E 204	य 092F 207	य़ 095F 206	र 0930 205
D_	ऱ 0931 208	ल 0932 209	ळ 0933 210	ऴ 0934 211	व 0935 212	श 0936 213	ष 0937 214	स 0938 215	ह 0939 216	INV 217	ा 093E 218	ि 093F 219	ी 0940 220	ु 0941 221	ू 0942 222	ृ 0943 223
E_	े 0946 224	े 0947 225	ै 0948 226	ॅ 0945 227	ॊ 094A 228	ो 094B 229	ौ 094C 230	ॉ 0949 231	् 094D 232	़ 093C 233	। 0964 234					ATR 239
F_	EXT 240	० 0966 241	१ 0967 242	२ 0968 243	३ 0969 244	४ 096A 245	५ 096B 246	६ 096C 247	७ 096D 248	८ 096E 249	९ 096F 250					

Special code points

INV character—code point D9 (217)

The INV character is used as a pseudo-consonant to display combining elements in isolation. For example, क (ka) + ् (halant) + INV = क् (half ka). The Unicode equivalent is no break space 00A0 or dotted circle ◌ 25CC.

ATR character—code point EF (239)

The ATR character followed by a byte code is used to switch to a different font attribute (such as bold) or language (such as Bengali), up to the next ATR sequence or the end of the line. This has no direct Unicode equivalent, as font attributes are not part of Unicode, and each script has a distinct set of code points.

EXT character—code point F0 (240)

The EXT character followed by a byte code indicates a Vedic accent. This has no direct Unicode equivalent, as Vedic accents are assigned to distinct code points.

Halant character ्—code point E8 (232)

The halant character removes the implicit vowel from a consonant and is used between consonants to represent conjunct consonants. For example, क (ka) + ् (halant) + त (ta) = क्त (kta). The sequence ् (halant) + ् (halant) displays a conjunct with an explicit halant, for example क (ka) + ् (halant) + ् (halant) + त (ta) = क्त. The sequence ् (halant) + ़ (nukta) displays a conjunct with half consonants, if available, for example क (ka) + ् (halant) + ़ (nukta) + त (ta) = क्त.

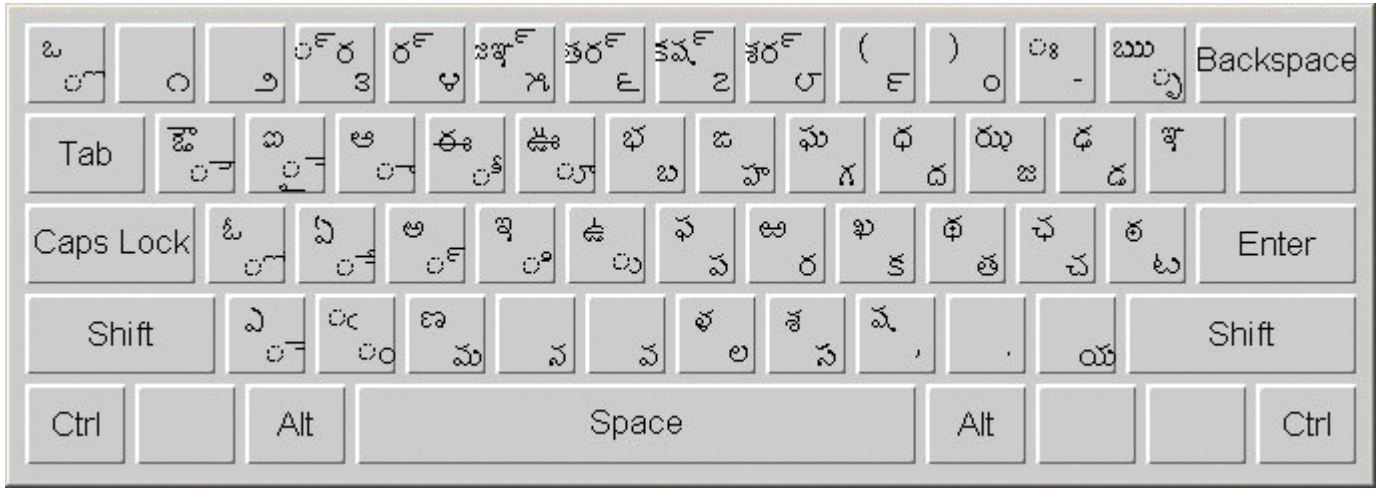
	ISCII		**Unicode**	
single halant	E8	halant	094D	
halant + halant	E8 E8	halant + ZWNJ	094D 200C	
halant + nukta	E8 E9	halant + ZWJ	094D 200D	

Nukta character ़—code point E9 (233)

The nukta character after another ISCII character is used for a number of rarer characters which don't exist in the main ISCII set. For example क (ka) + ़ (nukta) = क़ (qa). These characters have precomposed forms in Unicode, as shown in the following table.

ISCII code point	Original character	Character with nukta	Unicode code point
A1 (161)	ॐ	ॐ	0950
A6 (166)	इ	ऌ	090C
A7 (167)	ई	ॡ	0961
AA (176)	ऋ	ॠ	0960
B3 (179)	क	क़	0958
B4 (180)	ख	ख़	0959
B5 (181)	ग	ग़	095A
BA (186)	ज	ज़	095B
BF (191)	ड	ड़	095C

C0 (192)	ढ	ढ़	095D
C9 (201)	फ	फ़	095E
DB (219)	ि	ृ	0962
DC (220)	ी	ॄ	0963
DF (223)	्	ॢ	0944
EA (234)	l	s	093D



There are other keyboards for Benagli, Gujarati and others.

The use of IPA can be further researched e.g. google “IPA speech recognition”. One item there is the CMU dictionary which has a limited character set known as the ARPA phonetic alphabet. This is reproduced below courtesy of Wikipedia. The ARPA phonetic alphabet “Arpabet”, used in the CMU Dictionary is reproduced below. It has limited IPA character use.

Tibetan and Dzongkha (Bhutan)

ཀ ka /ká/	ཁ kha /kʰá/	ག ga /kà, kʰà/	ང nga /ŋà/
ཚ ca /tʃá/	ཆ cha /tʃʰá/	ཇ ja /tʃà/	ཉ nya /ɲà/
ཏ ta /tá/	ཐ tha /tʰá/	ད da /tà, tʰà/	ན na /nà/
པ pa /pá/	ཕ pha /pʰá/	བ ba /pà, pʰà/	མ ma /mà/
ཙ tsa /tsá/	ཚ tsha /tsʰá/	ཛ dza /tsà/	ཝ wa /wà/ (not originally part of the alphabet)[5]
ཞ zha /ʃà/[6]	ཟ za /sà/	འ 'a /hà/[7]	
ཡ ya /jà/	ར ra /rà/	ལ la /là/	
ཤ sha /ʃá/[6]	ས sa /sá/	ཧ ha /há/[8]	

ཨ a /á/

		UNASPIRATED - HIGH		ASPIRATED - MID		VOICED - LOW		NASALS - LOW	
		I		II		III		IV	
GUTTURAL	①	ཀ	ka	ཁ	kha	ག	ga	ང	nga
PALATAL	②	ཚ	ca	ཆ	cha	ཇ	ja	ཉ	nya
DENTAL	③	ཏ	ta	ཐ	tha	ད	da	ན	na
LABIAL	④	པ	pa	ཕ	pha	བ	ba	མ	ma
DENTAL	⑤	ཙ	tsa	ཚ	tsha	ཛ	dza	ཝ	wa
LOW	⑥	ཞ	zha	ཟ	za	འ	a	ཡ	ya
MID	⑦	ར	ra	ལ	la	ཤ	sha	ས	sa
HIGH	⑧	ཧ	ha	ཨ	a				

Vowels

Devanagari	IAST	Tibetan	Dependent vowel signs		Devanagari	IAST	Tibetan	Dependent vowel signs
अ	a	ཨ			औ	au	ཨཱུ	ཱུ

आ	ā	ཨཱ	ཱ		ऋ	ṛ	རྀ	ྲྀ
इ	i	ཨི	ི		ॠ	ṝ	རཱྀ	ཷ
ई	ī	ཨཱི	ཱི		ऌ	ḷ	ལྀ	ཹ
उ	u	ཨུ	ུ		ॡ	ḹ	ལཱྀ	ཻ
ऊ	ū	ཨཱུ	ཱུ		अं	aṃ	ཨཾ	ཾ
ए	e	ཨེ	ེ		अँ		ཨྀ	ྀ
ऐ	ai	ཨཻ	ཻ		अः	aḥ	ཨཿ	ཿ
ओ	o	ཨོ	ོ					

Consonants

Devanagari	IAST	Tibetan		Devanagari	IAST	Tibetan
क	ka	ཀ		द	da	ད
ख	kha	ཁ		ध	dha	ཌྷ
ग	ga	ག		न	na	ན
घ	gha	ཎྷ		प	pa	པ
ङ	ṅa	ང		फ	pha	ཕ
च	ca	ཚ		ब	ba	བ

छ	cha	ཆ		भ	bha	བྷ
ज	ja	ཇ		म	ma	མ
झ	jha	ཛྷ		य	ya	ཡ
ञ	ña	ཉ		र	ra	ར
ट	ṭa	ཏ		ल	la	ལ
ठ	ṭha	ཐ		व	va	ཝ
ड	ḍa	ད		श	śa	ཤ
ढ	ḍha	ཌྷ		ष	ṣa	ཥ

ण	ṇa	ཎ		स	sa	ས
त	ta	ཏ		ह	ha	ཧ
थ	tha	ཐ		क्ष	kṣa	ཀྵ

The Sanskrit "cerebral" (retroflex) consonants ट ठ ड ण ष (ṭa, ṭha, ḍa, ṇa, ṣa) are represented by reversing the letters ཏ ཐ ད ན ཤ (ta, tha, da, na, sha) to give ཊ ཋ ཌ ཎ ཥ(Ta, Tha, Da, Na, Sa).

It is a classic rule to transliterate च छ ज झ (ca cha ja jha) to ཙ ཚ ཛ ཛྷ (tsa tsha dza dzha), respectively. Nowadays, ཅ ཆ ཇ ཈ (ca cha ja jha) can also be used.

Arpabet Symbols

In Arpabet, every phoneme is represented by one or two capital letters. Digits are used as stress indicators and are placed at the end of the stressed syllabic vowel. Punctuation marks are used like in the written language, to represent intonation changes at the end of clauses and sentences. The stress values are:

Stress

Value	Description
0	No stress
1	Primary stress

2 Secondary stress

Vowels

Monophthongs

Arpabet	IPA	Word examples
AO	ɔ	off (AO1 F); fall (F AO1 L); frost (F R AO1 S T)
AA	ɑ	father (F AA1 DH ER), cot (K AA1 T)
IY	i	bee (B IY1); she (SH IY1)
UW	u	you (Y UW1); new (N UW1); food (F UW1 D)
EH	ɛ	red (R EH1 D); men (M EH1 N)
IH	ɪ	big (B IH1 G); win (W IH1 N)
UH	ʊ	should (SH UH1 D), could (K UH1 D)
AH	ʌ	but (B AH1 T), sun (S AH1 N)
AX	ə	sofa (S OW1 F AH0), alone (AH0 L OW1 N) discus (D IH1 S K AX0 S); note distinction from discuss (D IH0 S K AH1 S)
AE	æ	at (AE1 T); fast (F AE1 S T)

Diphthongs

Arpabet	IPA	Word Examples

EY	eɪ	say (S EY1); eight (EY1 T)
AY	aɪ	my (M AY1); why (W AY1); ride (R AY1 D)
OW	oʊ	show (SH OW1); coat (K OW1 T)
AW	aʊ	how (HH AW1); now (N AW1)
OY	ɔɪ	boy (B OY1); toy (T OY1)

R-colored vowels

Arpabet	IPA	Word Examples
ER	ɝ	her (HH ER0); bird (B ER1 D); hurt (HH ER1 T), nurse (N ER1 S)
AXR	ɚ	father (F AA1 DH ER); coward (K AW1 ER D)
EH R	ɛr	air (EH1 R); where (W EH1 R); hair (HH EH1 R)
UH R	ʊr	cure (K Y UH1 R); bureau (B Y UH1 R OW0), detour (D IH0 T UH1 R)
AO R	ɔr	more (M AO1 R); bored (B AO1 R D); chord (K AO1 R D)
AA R	ɑr	large (L AA1 R JH); hard (HH AA1 R D)
IH R *or* IY R	ɪr	ear (IY1 R); near (N IH1 R)
AW R	aʊr	*This seems to be a rarely used r-controlled vowel. In some dialects* flower (F L AW1 R; *in other dialects* F L AW1 ER0)

Consonants

Stops

Arpabet	IPA	Word Examples
P	p	pay (P EY1)
B	b	buy (B AY1)
T	t	take (T EY1 K)
D	d	day (D EY1)
K	k	key (K IY1)
G	g	go (G OW1)

Affricates

Arpabet	IPA	Word Examples
CH	tʃ	chair (CH EH1 R)
JH	dʒ	just (JH AH1 S T); gym (JH IH1 M)

Fricatives

Arpabet	IPA	Word Examples
F	f	for (F AO1 R)
V	v	very (V EH1 R IY0)
TH	θ	thanks (TH AE1 NG K S); Thursday (TH ER1 Z D EY2)
DH	ð	that (DH AE1 T); the (DH AH0); them (DH EH1 M)

S	s	say (S EY1)
Z	z	zoo (Z UW1)
SH	ʃ	show (SH OW1)
ZH	ʒ	measure (M EH1 ZH ER0); pleasure (P L EH1 ZH ER)
HH	h	house (HH AW1 S)

Nasals

Arpabet	IPA	Word Examples
M	m	man (M AE1 N)
EM	m̩	keep 'em (K IY1 P EM)
N	n	no (N OW1)
EN	n̩	button (B AH1 T EN)
NG	ŋ	sing (S IH1 NG)
ENG	ŋ̩	Washington (W AO1 SH ENG T EN)

Liquids

Arpabet	IPA	Word Examples
L	ɫ	late (L EY1 T)
EL	ɫ̩	bottle (B AO1 DX EL)

R	r *or* ɹ	run (R AH1 N)
DX	ɾ	wetter (W EH1 DX AXR)
NX	ɾ̃	wintergreen (W IY2 NX AXR G R IY1 N)

Semivowels

Arpabet	**IPA**	
Y	j	yes (Y EH1 S)
W	w	way (W EY1)
Q	ʔ	glottal stop (uh-oh - ʔʌʔoʊ)
(missing)	hw *or* ʍ	"when" etc. in some dialects

Chinese (in Wikipedia)

In each cell below, the bold letters indicate pinyin and the brackets enclose the symbol in the International Phonetic Alphabet.

	Bilabial		**Labiodental**	**Alveolar**		**Retroflex**		**Alveolo-palatal**	**Velar**
	Voiceless	**Voiced**	**Voiceless**	**Voiceless**	**Voiced**	**Voiceless**	**Voiced**	**Voiceless**	**Voiceless**
Nasal		**m** [m]			**n** [n]				

Plosive	**Unaspirated**	**b** [p]			**d** [t]					**g** [k]
	Aspirated	**p** [pʰ]			**t** [tʰ]					**k** [kʰ]
Affricate	**Unaspirated**				**z** [ts]		**zh** [ʈʂ]		**j** [tɕ]	
	Aspirated				**c** [tsʰ]		**ch** [ʈʂʰ]		**q** [tɕʰ]	
Fricative				**f** [f]	**s** [s]		**sh** [ʂ]	**r** [ʐ~ɻ][1]	**x** [ɕ]	**h** [x]
Lateral						**l** [l]				
Approximant		**y**[3] [j]/[ɥ][2] and **w**[3] [w]								

[1] *r* may phonetically be [ʐ] (a voiced retroflex fricative) or [ɻ] (a retroflex approximant). This pronunciation varies among different speakers, and is not two different phonemes.

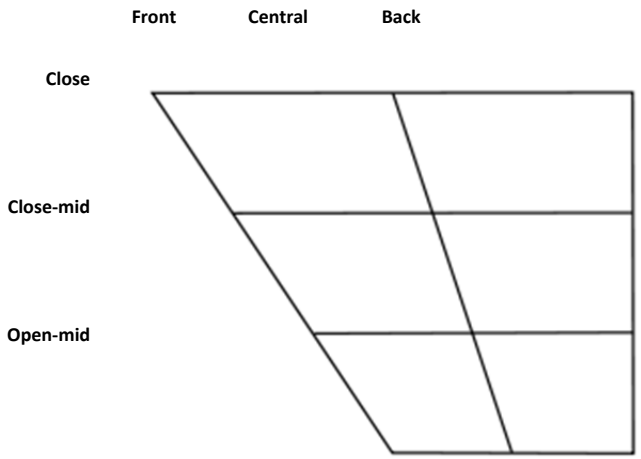
[2] *y* is pronounced [ɥ] (a labial-palatal approximant) before *u*.

[3] the letters *w* and *y* are not included in the table of initials in the official pinyin system. They are an orthographic convention for the medials *i, u* and *ü* when no initial is present. When *i, u* or *ü* are finals and no initial is present, they are spelled *yi*, *wu*, and *yu*, respectively. The conventional order (excluding *w* and *y*), derived from the zhuyin system, is:

b	**p**	**m**	**f**	**d**	**t**	**n**	**l**	**g**	**k**	**h**	**j**	**q**	**x**	**zh**	**ch**	**sh**	**r**	**z**	**c**	**s**

Finals

Standard Chinese vowels (with IPA and Pinyin)



i • y (ü)

ɨ (i)

u

ɤ (e) • o

ɚ (er)

ɛ (ê)

a

The following chart gives the combinations of medials and finals based on an analysis that assumes just two vowel nuclei, /a/ and /ə/;[28] various allophones result depending on phonetic context.

In each cell below, the first line indicates IPA, the second indicates pinyin for a standalone (no-initial) form, and the third indicates pinyin for a combination with an initial. Other than finals modified by an *-r*, which are omitted, the following is an exhaustive table of all possible finals.[1][29]

The only syllable-final consonants in Standard Chinese are *-n* and *-ng*, and *-r*, which is attached as a grammatical suffix. A Chinese syllable ending with any other consonant either is from a non-Mandarin language (a southern Chinese language such as Cantonese, or a minority language of China), or indicates the use of a non-pinyin romanization system (where final consonants may be used to indicate tones).

Final	**Nucleus**	/a/					/ə/					∅
	Coda	∅	/i/	/u/	/n/	/ŋ/	∅	/i/	/u/	/n/	/ŋ/	
Medial	∅	[ä] a	[aɪ̯] ai	[ɑʊ̯] ao	[än] an	[ɑŋ] ang	[ɯ̩ʌ] e	[eɪ̯] ei	[oʊ̯] ou	[ən] en	[ɤŋ] eng	[ɨ]

		-a	-ai	-ao	-an	-ang	-e	-ei	-ou	-en	-eng	-i
	/i/	[i̯ä] ya -ia		[i̯ɑʊ̯] yao -iao	[i̯ɛn] yan -ian	[i̯ɑŋ] yang -iang	[i̯ɛ] ye -ie		[i̯oʊ̯] you -iu	[in] yin -in	[iŋ] ying -ing	[i] yi -i
	/u/	[u̯ä] wa -ua	[u̯aɪ̯] wai -uai		[u̯än] wan -uan	[u̯ɑŋ] wang -uang	[u̯ɔ] [3] wo -uo/-o	[u̯eɪ̯] wei -ui		[u̯ən] wen -un	[u̯ɤŋ], [ʊŋ] [4] weng -ong	[u] wu -u
	/y/				[yɛn] yuan -üan [2]		[yœ] yue -üe [2]			[yn] yun -ün [2]	[i̯ʊŋ] yong -iong	[y] yu -ü [2]

[1] [ɑɻ] is written *er*. For other finals formed by the suffix *-r,* pinyin does not use special orthography; one simply appends *r* to the final that it is added to, without regard for any sound changes that may take place along the way. For information on sound changes related to final *r*, please see Erhua#Rules.

[2] *ü* is written as *u* after *j, q, x,* or *y.*

[3] *uo* is written as *o* after *b, p, m,* or *f.*

[4] *weng* is pronounced [ʊŋ] (written as *ong*) when it follows an initial.

Technically, *i, u, ü* without a following vowel are finals, not medials, and therefore take the tone marks, but they are more concisely displayed as above. In addition, *ê*[ɛ] (欸; 誒) and syllabic nasals *m* (呒, 呣), *n* (嗯, 唔), *ng* (嗯, 吼) are used as interjections.

Rules given in terms of English pronunciation

This section includes inline links to audio files. If you have trouble playing the files, see Wikipedia Media help.

All rules given here in terms of English pronunciation are approximations, as several of these sounds do not correspond directly to sounds in English.

Pronunciation of initials

Pinyin	IPA	English approximation[30]	Explanation
b	[p]	spit	unaspirated **p**, as in s**p**it
p	[pʰ]	pay	strongly aspirated **p**, as in **p**it
m	[m]	may	as in English **mummy**
f	[f]	fair	as in English **f**un
d	[t]	stop	unaspirated **t**, as in s**t**op
t	[tʰ]	take	strongly aspirated **t**, as in **t**op
n	[n]	nay	as in English **n**it
l	[l]	lay	as in English **l**ove
g	[k]	skill	unaspirated **k**, as in s**k**ill
k	[kʰ]	kay	strongly aspirated **k**, as in **k**ill
h	[x]	loch	roughly like the Scots **ch**. English **h** as in **h**ay or, more closely in some American dialects, **h**ero is an acceptable a The best way to produce this sound is by very slowly making a "k" sound, pausing at the point where there is just flowing over the back of your tongue (after the release at the beginning of a "k")
j	[tɕ]	churchyard	No equivalent in English, but similar to an unaspirated "-chy-" sound when said quickly. Like *q,* but unaspirated. N despite the common English pronunciation of "Beijing". The sequence "ji" word-initially is the same as the Japane of じ(ジ) *ji*.
q	[tɕʰ]	punch yourself	No equivalent in English. Like pun**ch y**ourself, with the lips spread wide with *ee*. Curl the tip of the tongue downw

			the back of the teeth and strongly aspirate. The sequence "qi" word-initially is the same as the Japanese pronunciation of チ (チ) *chi*.
x	[ɕ]	push yourself	No equivalent in English. Like **-sh y-**, with the lips spread and the tip of your tongue curled downwards and stuck to the bac teeth when you say *ee.* The sequence "xi" is the same as the Japanese pronunciation of し(シ) *shi.*
zh	[ʈʂ]	junk	Rather like **ch** (a sound between **ch**oke, **j**oke, **tr**ue, and **dr**ew, tongue tip curled more upwards). Voiced in a toneless syllab
ch	[ʈʂʰ]	church	as in **ch**in, but with the tongue curled upwards; very similar to nur**tu**re in American English, but strongly aspirated.
sh	[ʂ]	shirt	as in **sh**oe, but with the tongue curled upwards; very similar to mar**sh** in American English
r	[ʐ],[ɻ]	ray	Similar to the English **z** in a**z**ure and **r** in **r**educe, but with the tongue curled upwards, like a cross between English "r" and "j". InCyrillised Chinese the sound is rendered with the letter "ж".
z	[ʦ]	reads	unaspirated **c**, similar to something between su**ds** and ca**ts**; as in su**ds** in a toneless syllable
c	[ʦʰ]	hats	like the English **ts** in ca**ts**, but strongly aspirated, very similar to the Czech, Polish and Slovak **c**.
s	[s]	say	as in **s**un
w	[w]	way	as in **w**ater.*
y	[j],[ɥ]	yea	as in **y**es. Before a *u*, pronounce it with rounded lips.*

*** Note on *y* and *w***

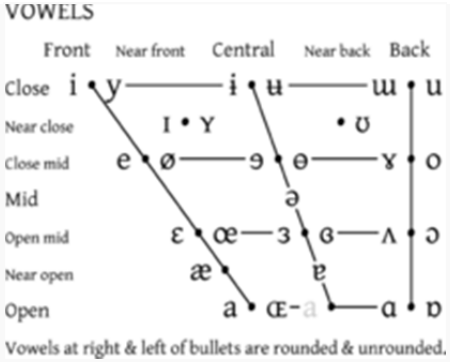
Y and *w* are equivalent to the semivowel medials *i, u,* and *ü* (see below). They are spelled differently when there is no initial consonant in order to mark a new syllable: *fanguan* is *fan-guan,* while *fangwan* is *fang-wan* (and equivalent to **fang-uan).* With this convention, an apostrophe only needs to be used to mark an initiala*, e,* or *o: Xi'an* (two syllables: [ɕi.an]) vs. *xian* (one syllable: [ɕi̯ɛn]). In addition, *y* and *w* are added to fully vocalic *i, u,* and *ü* when these occur without an initial consonant, so that they are written *yi, wu,* and *yu.* Some Mandarin speakers do pronounce a [j] or [w] sound at the beginning of such words—that is, *yi* [i] or [ji], *wu*[u] or [wu], *yu* [y] or [ɥy],—so this is an intuitive convention. See below for a few finals which are abbreviated after a consonant plus *w/u* or *y/i* medial: *wen* → C+*un,wei* → C+*ui, weng* → C+*ong,* and *you* → C+*iu.*

**** Note on the apostrophe**

The apostrophe (') is used before a syllable starting with a vowel (*a*, *o*, or *e*) in a multiple-syllable word when the syllable does not start the word (which is most commonly realized as [ɥ]), unless the syllable immediately follows a hyphen or other dash.[31] This is done to remove ambiguity that could arise, as in *Xi'an*, which consists of the two syllables *xi* ("西") *an* ("安"), compared to such words as *xian* ("先"). (This

ambiguity does not occur when tone marks are used: The two tone marks in *Xīān* unambiguously show that the word consists of two syllables. However, even with tone marks, the city is usually spelled with an apostrophe as *Xī'ān*.)

Pronunciation of finals



This table may be a useful reference for IPA vowel symbols

The following is a list of finals in Standard Chinese, excepting most of those ending with *r*.

To find a given final:

1. Remove the initial consonant. *Zh, ch,* and *sh* count as initial consonants.
2. Change initial *w* to *u* and initial *y* to *i.* For *weng, wen, wei, you,* look under *ong, un, ui, iu.*
3. For *u* after *j, q, x,* or *y,* look under *ü.*

Pinyin	IPA	Form with zero initial	Explanation
-i	[ɨ], [ɯ]	(*n/a*)	-i is a buzzed continuation of the consonant following z-, c-, s-, zh-, ch-, sh- or r-. (In all other cases, -i has the sound of b**ee**; this is listed below.)
a	[ä]	**a**	like English "f**a**ther", but a bit more fronted

e	[u̯ʌ], [ə]	**e**	a diphthong consisting first of a back, unrounded semivowel (which can be formed by first pronouncing "w" and then spread lips without changing the position of the tongue) followed by a vowel similar to English "d**uh**". Many unstressed syllables in C use theschwa [ə] (ide**a**), and this is also written as *e*.
ai	[aɪ̯]	**ai**	like English "eye", but a bit lighter
ei	[eɪ̯]	**ei**	as in "h**ey**"
ao	[ɑʊ̯]	**ao**	approximately as in "c**ow**"; the *a* is much more audible than the *o*
ou	[oʊ̯]	**ou**	as in "s**o**"
an	[än]	**an**	like British English "b**an**", but more central
en	[ən]	**en**	as in "tak**en**"
ang	[ɑŋ]	**ang**	as in German *Angst* (starts with the vowel sound in f**a**ther and ends in the velar nasal; like s**ong** in some dialects of America English)
eng	[ɤŋ]	**eng**	like *e* in *en* above but with ng added to it at the back
er	[ɑɻ]	**er**	similar to the sound in "b**ar**" in American English
Finals beginning with i- (y-)			
i	[i]	**yi**	like English b**ee**.
ia	[i̯ä]	**ya**	as **i** + **a**; like English "**ya**rd"
ie	[i̯ɛ]	**ye**	as **i** + **ê**; but is very short; *e* (pronounced like *ê*) is pronounced longer and carries the main stress (similar to the initial sound **ye** in **yet**)
iao	[i̯ɑʊ̯]	**yao**	as **i** + **ao**
iu	[i̯oʊ̯]	**you**	as **i** + **ou**
ian	[i̯ɛn]	**yan**	as **i** + **ê** + **n**; like English **yen**
in	[in]	**yin**	as **i** + **n**
iang	[i̯ɑŋ]	**yang**	as **i** + **ang**
ing	[iŋ]	**ying**	as **i** + **ng**
Finals beginning with u- (w-)			
u	[u]	**wu**	like English "oo"

ua	[u̯ä]	**wa**	as **u** + **a**
uo, *o*	[u̯ɔ]	**wo**	as **u** + **o** where the *o* (compare with the *o* interjection) is pronounced shorter and lighter (spelled as **o** after b, p, m or f).
uai	[u̯aɪ̯]	**wai**	as **u** + **ai** like as in **why**
ui	[u̯eɪ̯]	**wei**	as **u** + **ei**;
uan	[u̯än]	**wan**	as **u** + **an**;
un	[u̯ən]	**wen**	as **u** + **en**; like the *on* in the English *won*;
uang	[u̯ɑŋ]	**wang**	as **u** + **ang**;
ong	[ʊŋ], [u̯ɤŋ]	**weng**	starts with the vowel sound in b**oo**k and ends with the velar nasal sound in si**ng**; as **u** + **eng** in zero initial.
Finals beginning with ü- (yu-)			
u, *ü*	[y] (listen)	**yu**	as in German "**ü**ber" or French "l**u**ne" (To pronounce this sound, say "ee" with rounded lips)
ue, üe	[y͡œ]	**yue**	as **ü** + **ê**; the *ü* is short and light
uan	[y͡ɛn]	**yuan**	as **ü** + **ê** + **n**;
un	[yn]	**yun**	as **ü** + **n**;
iong	[i̯ʊŋ]	**yong**	as **i** + **ong**
Interjections			
ê	[ɛ]	(*n/a*)	as in "b**e**t".
o	[ɔ]	(*n/a*)	Approximately as in "**o**ffice" in British accent; the lips are much more rounded.
io	[i̯ɔ]	**yo**	as **i** + plain continental[clarification needed] "o".

Orthography

Letters

Pinyin differs from other romanizations in several aspects, such as the following:

- Syllables starting with *u* are written as *w* in place of *u* (e.g., *uan* is written as *wan*). Standalone *u* is written as *wu*.
- Syllables starting with *i* are written as *y* in place of *i* (e.g., *ian* is written as *yan*). Standalone *i* is written as *yi*.

- Syllables starting with *ü* are written as *yu* in place of *ü* (e.g., *üe* is written as *yue*).
- *ü* is written as *u* when there is no ambiguity (such as *ju*, *qu*, and *xu*), but written as *ü* when there are corresponding *u* syllables (such as *lü* and *nü*). In such situations where there are corresponding *u* syllables, it is often replaced with *v* on a computer, making it easier to type on a standard keyboard.
- When preceded by a consonant, *iou*, *uei*, and *uen* are simplified as *iu*, *ui*, and *un* (which do not represent the actual pronunciation).
- As in zhuyin, what are actually pronounced as *buo*, *puo*, *muo*, and *fuo* are given a separate representation: *bo*, *po*, *mo*, and *fo*.
- The apostrophe (') is used before a syllable starting with a vowel (*a*, *o*, or *e*) in a multiple-syllable word when the syllable does not start the word (which is most commonly realized as [ɥ]), unless the syllable immediately follows a hyphen or other dash.[31] This is done to remove ambiguity that could arise, as in *Xi'an*, which consists of the two syllables *xi* ("西") *an* ("安"), compared to such words as *xian* ("先"). (This ambiguity does not occur when tone marks are used: The two tone marks in "Xīān" unambiguously show that the word consists of two syllables. However, even with tone marks, the city is usually spelled with an apostrophe as "Xī'ān".)
- *Eh* alone is written as *ê*; elsewhere as *e*. Schwa is always written as *e*.
- *zh*, *ch*, and *sh* can be abbreviated as *ẑ*, *ĉ*, and *ŝ* (*z*, *c*, *s* with a circumflex). However, the shorthands are rarely used due to difficulty of entering them on computers, and are confined mainly to Esperanto keyboard layouts.
- *ng* has the uncommon shorthand of *ŋ*.
- The letter *v* is unused (except in spelling foreign languages, languages of minority nationalities, and some dialects), despite a conscious effort to distribute letters more evenly than in Western languages. However, sometimes, for ease of typing into a computer, the *v* is used to replace a *ü*.

Most of the above are used to avoid ambiguity when writing words of more than one syllable in pinyin. For example *uenian* is written as *wenyan* because it is not clear which syllables make up *uenian*; *uen-ian*, *uen-i-an* and *u-en-i-an* are all possible combinations whereas *wenyan* is unambiguous because *we*, *nya*, etc. do not exist in pinyin. See the pinyin table article for a summary of possible pinyin syllables (not including tones).

Word formation, capitalization, initialisms and punctuation

Although Chinese characters represent single syllables, Mandarin Chinese is a polysyllabic language. Spacing in pinyin is based on whole words, not single syllables. However, there are often ambiguities in partitioning a word. *The Basic Rules of the Chinese Phonetic Alphabet Orthography* (汉语拼音正词法基本规则; 漢語拼音正詞法基本規則; *Hànyǔ Pīnyīn Zhèngcífǎ Jīběn Guīzé*) were put into effect in 1988 by the

National Educational Commission (国家教育委员会; 國家教育委員會;*Guójiā Jiàoyù Wěiyuánhuì*) and the National Language Commission (国家语言文字工作委员会; 國家語言文字工作委員會; *Guójiā Yǔyán Wénzì Gōngzuò Wěiyuánhuì*).[32] These rules became a Guobiao standard in 1996[32] and were updated in 2012.[33]

1. **General**
    1. **Single meaning:** Words with a single meaning, which are usually set up of two characters (sometimes one, seldom three), are written together and not capitalized: *rén* (人, person); *péngyou* (朋友, friend); *qiǎokèlì* (巧克力, chocolate)
    2. **Combined meaning (2 or 3 characters):** Same goes for words combined of two words to one meaning: *hǎifēng* (海风; 海風, sea breeze); *wèndá* (问答; 問答, question and answer); *quánguó* (全国; 全國, nationwide); *chángyòngcí* (常用词; 常用詞)
    3. **Combined meaning (4 or more characters):** Words with four or more characters having one meaning are split up with their original meaning if possible: *wúfèng gāngguǎn* (无缝钢管; 無縫鋼管, seamless steel-tube); *huánjìng bǎohù guīhuà* (环境保护规划; 環境保護規劃, environmental protection planning); *gāoměngsuānjiǎ* (高锰酸钾; 高錳酸鉀, potassium permanganate)
2. **Duplicated words**
    1. **AA:** Duplicated characters (AA) are written together: *rénrén* (人人, everybody), *kànkan* (看看, to have a look), *niánnián* (年年, every year)
    2. **ABAB:** Two characters duplicated (ABAB) are written separated: *yánjiū yánjiū* (研究研究, to study, to research), *xuěbái xuěbái* (雪白雪白, white as snow)
    3. **AABB:** Characters in the AABB schema are written together: *láiláiwǎngwǎng* (来来往往; 來來往往, come and go), *qiānqiānwànwàn* (千千万万; 千千萬萬, numerous)
3. **Prefixes** (前附成分; *qiánfù chéngfèn*) **and Suffixes** (后附成分; 後附成分; *hòufù chéngfèn*): Words accompanied by prefixes such as *fù* (副, vice), *zǒng* (总; 總, chief), *fēi* (非, non-), *fǎn* (反, anti-), *chāo* (超, ultra-), *lǎo* (老, old), *ā* (阿, used before names to indicate familiarity), *kě* (可, -able), *wú* (无; 無, -less) and *bàn*(半, semi-) and suffixes such as *zi* (子, noun suffix), *r* (儿; 兒, diminutive suffix), *tou* (头; 頭, noun suffix), *xìng* (性, -ness, -ity), *zhě* (者, -er, -ist), *yuán* (员; 員, person), *jiā* (家, -er, -ist), *shǒu* (手, person skilled in a field), *huà* (化, -ize) and *men* (总; 們, plural marker) are written together: *fùbùzhǎng* (副部长; 副部長, vice minister), *chéngwùyuán* (乘务员; 乘務員, conductor), *háizimen* (孩子们; 孩子們, children)
4. **Nouns and names** (名词; 名詞; *míngcí*)
    1. Words of position are separated: *mén wài* (门外; 門外, outdoor), *hé li* (河里; 河裏, under the river), *huǒchē shàngmian* (火车上面; 火車上面, on the train), *Huáng Hé yǐnán* (黄河以南; 黃河以南, south of the Yellow River)
        1. Exceptions are words traditionally connected: *tiānshang* (天上, in the sky or outerspace), *dìxia* (地下, on the ground), *kōngzhōng* (空中, in the air), *hǎiwài* (海外, overseas)
    2. Surnames are separated from the given names, each capitalized: *Lǐ Huá* (李华; 李華), *Zhāng Sān* (张三; 張三). If the

surname and/or given name consists of two syllables, it should be written as one: *Zhūgě Kǒngmíng* (诸葛孔明; 諸葛孔明).

3. Titles following the name are separated and are not capitalized: *Wáng bùzhǎng* (王部长; 王部長, Minister Wang), *Lǐ xiānsheng* (李先生, Mr. Li), *Tián zhǔrèn* (田主任, Director Tian), *Zhào tóngzhì* (赵同志; 趙同志, Comrade Zhao).
4. The forms of addressing people with suffixes such as *Lǎo* (老), *Xiǎo* (小), *Dà* (大) and *Ā* (阿) are capitalized: *Xiǎo Liú* (小刘; 小劉, [young] Ms./Mr. Liu),*Dà Lǐ* (大李, [great; elder] Mr. Li), *Ā Sān* (阿三, Ah San), *Lǎo Qián* (老钱; 老錢, [senior] Mr. Qian), *Lǎo Wú* (老吴; 老吳, [senior] Mr. Wu)
    1. Exceptions are: *Kǒngzǐ* (孔子, Confucius), *Bāogōng* (包公, Judge Bao), *Xīshī* (西施, Xishi), *Mèngchángjūn* (孟尝君; 孟嘗君, Lord Mengchang), among others
5. Geographical names of China: *Běijīng Shì* (北京市, city of Beijing), *Héběi Shěng* (河北省, province of Hebei), *Yālù Jiāng* (鸭绿江; 鴨綠江, Yalu River),*Tài Shān* (泰山, Mount Tai), *Dòngtíng Hú* (洞庭湖, Dongting Lake), *Táiwān Hǎixiá* (台湾海峡; 臺灣海峽, Taiwan Strait)
    1. Monosyllabic prefixes and suffixes are written together with their related part: *Dōngsì Shítiáo* (东四十条; 東四十條, Dongsi 10th Alley)
    2. Common geographical nouns that have become part of proper nouns are written together: *Hēilóngjiāng* (黑龙江; 黑龍江, Heilongjiang)
6. Chinese names of Non-Chinese names are written in Hanyu Pinyin: *Āpèi Āwàngjìnměi* (阿沛·阿旺晋美; 阿沛·阿旺晉美, Ngapoi Ngawang Jigme);*Dōngjīng* (东京; 東京, Tokyo)

5. **Verbs** (动词; 動詞; *dòngcí*): Verbs and their suffixes *-zhe* (着; 著), *-le* (了) or *-guo* ((过; 過) are written as one: *kànzhe* (看着; 看著, seeing), *jìnxíngguo* (进行过;進行過, have been implemented). *Le* as it appears in the end of a sentence is separated though: *Huǒchē dào le.* (火车到了; 火車到了, The train [has] arrived).
    1. Verbs and their objects are separated: *kàn xìn* (看信, read a letter), *chī yú* (吃鱼; 吃魚, eat fish), *kāi wánxiào* (开玩笑; 開玩笑, to be kidding).
    2. If verbs and their complements are each monosyllabic, they are written together; if not, they are separated: *gǎohuài* (搞坏; 搞壞, to make broken), *dǎsǐ*(打死, hit to death), *huàwéi* (化为; 化為, to become), *zhěnglǐ hǎo* (整理好, to sort out), *gǎixiě wéi* (改写为; 改寫為, to rewrite as)
6. **Adjectives** (动词; 動詞; *xíngróngcí*): A monosyllabic adjective and its reduplication are written as one: *mēngmēngliàng* (矇矇亮, dim), *liàngtángtáng* (亮堂堂, shining bright)
    1. Complements of size or degree such as *xiē* (些), *yīxiē* (一些), *diǎnr* (点儿; 點兒) and *yīdiǎnr* (一点儿; 一點兒) are written separated: *dà xiē* (大些), a little bigger), *kuài yīdiǎnr* (快一点儿; 快一點兒, a bit faster)
7. **Pronouns** (代词; 代詞; *dàicí*)
    1. Personal pronouns and interrogative pronouns are separated from other words: *Wǒ ài Zhōngguó.* (我爱中国。; 我愛中國。, I love China); *Shéi shuō de?* (谁说的？; 誰說的？, Who said it?)

2. The demonstrative pronoun *zhè* (这; 這, this), *nà* (那, that) and the question pronoun *nǎ* (哪, which) are separated: *zhè rén* (这人; 這人, this person),*nà cì huìyì* (那次会议; 那次會議, that meeting), *nǎ zhāng bàozhǐ* (哪张报纸; 哪張報紙, which newspaper)
    1. Exception—If *zhè*, *nà* or *nǎ* are followed by *diǎnr* (点儿; 點兒), *bān* (般), *biān* (边; 邊), *shí* (时; 時), *huìr* (会儿; 會兒), *lǐ* (里; 裏), *me* (么; 麼) or the general classifier*ge* (个; 個), they are written together: *nàlǐ* (那里; 那裏, there), *zhèbiān* (这边; 這邊, over here), *zhège* (这个; 這個, this)

8. **Numerals** (数词; 數詞; *shùcí*) **and measure words** (量词; 量詞; *liàngcí*)
    1. Numbers and words like *gè* (各, each), *měi* (每, each), *mǒu* (某, any), *běn* (本, this), *gāi* (该; 該, that), *wǒ* (我, my, our) and *nǐ* (你, your) are separated from the measure words following them: *liǎng ge rén* (两个人; 兩個人, two people), *gè guó* (各国; 各國, every nation), *měi nián* (每年, every year), *mǒu gōngchǎng* (某工厂; 某工廠, a certain factory), *wǒ xiào* (我校, our school)
    2. Numbers up to 100 are written as single words: *sānshísān* (三十三, thirty-three). Above that, the hundreds, thousands, etc. are written as separate words: *jiǔyì qīwàn èrqiān sānbǎi wǔshíliù* (九亿七万二千三百五十六; 九億七萬二千三百五十六, nine hundred million, seventy-two thousand, three hundred fifty-six). Arabic numerals are kept as Arabic numerals: *635 fēnjī* (635 分机; 635 分機, extension 635)
    3. The *dì* (第) of ordinal numerals is hyphenated: *dì-yī* (第一, first), *dì-356* (第 356, 356th). The *chū* (初) in front of numbers one to ten is written together with the number: *chūshí* (初十, tenth day)
    4. Numbers representing month and day are hyphenated: *wǔ-sì* (五四, May fourth), *yīèr-jiǔ* (一二·九, December ninth)
    5. Words of approximations such as *duō* (多), *lái* (来; 來) and *jǐ* (几; 幾) are separated from numerals and measure words: *yībǎi duō gè* (一百多个; 一百多個, around a hundred); *shí lái wàn gè* (十来万个; 十來萬個, around a hundred thousand); *jǐ jiā rén* (几家人; 幾家人, a few families)
        1. *Shíjǐ* (十几; 十幾, more than ten) and *jǐshí* (几十; 幾十, tens) are written together: *shíjǐ gè rén* (十几个人; 十幾個人, more than ten people); *jǐshí*(几十根钢管; 幾十根鋼管, tens of steel pipes)
    6. Approximations with numbers or units that are close together are hyphenated: *sān-wǔ tiān* (三五天, three to five days), *qiān-bǎi cì* (千百次, thousands of times)

9. Other **function words** (虚词; 虛詞; *xūcí*) are separated from other words, including:
    1. Adverbs (副词; 副詞; *fùcí*): *hěn hǎo* (很好, very good), *zuì kuài* (最快, fastest), *fēicháng dà* (非常大, extremely big)
    2. Prepositions (介词; 介詞; *jiècí*): *zài qiánmiàn* (在前面, in front)
    3. Conjunctions (连词; 連詞; *liáncí*): *nǐ hé wǒ* (你和我, you and I/me), *Nǐ lái háishi bù lái?* (你来还是不来？; 你來還是不來？, Are you coming or not?)
    4. "Constructive auxiliaries" (结构助词; 結構助詞; *jiégòu zhùcí*) such as *de* (的/地/得), *zhī* (之) and *suǒ* (所): *mànmàn de zou* (慢慢地走), go slowly)
        1. A monosyllabic word can also be written together with *de* (的/地/得): *wǒ de shū* / *wǒde shū* (我的书; 我的書, my book)
    5. Modal auxiliaries at the end of a sentence: *Nǐ zhīdào ma?* (你知道吗？; 你知道嗎？, Do you know?), *Kuài qù ba!* (快去吧！,

Go quickly!)

6. Exclamations and interjections: *À! Zhēn měi!* (啊！真美！), Oh, that's so beautiful!)
7. Onomatopoeia: *mó dāo huòhuò* (磨刀霍霍, honing a knife), *hōnglōng yī shēng* (轰隆一声; 轟隆一聲, rumbling)

10. **Capitalization**
    1. The first letter of the first word in a sentence is capitalized: *Chūntiān lái le.* (春天来了。; 春天來了。, Spring has arrived.)
    2. The first letter of each line in a poem is capitalized.
    3. The first letter of a proper noun is capitalized: *Beǐjīng* (北京, Beijing), *Guójì Shūdiàn* (国际书店; 國際書店, International Bookstore), *Guójiā Yǔyán Wénzì Gōngzuò Wěiyuánhuì* (国家语言文字工作委员会; 國家語言文字工作委員會, National Language Commission)
        1. On some occasions, proper nouns can be written in all caps: *BĚIJĪNG*, *GUÓJÌ SHŪDIÀN*, *GUÓJIĀ YŬYÁN WÉNZÌ GŌNGZUÒ WĚIYUÁNHUÌ*
    4. If a proper noun is written together with a common noun to make a proper noun, it is capitalized. If not, it is not capitalized: *Fójiào* (佛教, Buddhism),*Tángcháo* (唐朝, Tang dynasty), *jīngjù* (京剧; 京劇, Beijing opera), *chuānxiōng* (川芎, Szechuan lovage)

11. **Initialisms**
    1. Single words are abbreviated by taking the first letter of each character of the word: ***Beǐjīng*** (北京, Beijing) → *BJ*
    2. A group of words are abbreviated by taking the first letter of each word in the group: ***guójiā biāozhǔn*** (国家标准; 國家標準, Guobiao standard) → *GB*
    3. Initials can also be indicated using full stops: *Beǐjīng* → *B.J.*, *guójiā biāozhǔn* → *G.B.*
    4. When abbreviating names, the surname is written fully (first letter capitalized or in all caps), but only the first letter of each character in the given name is taken, with full stops after each initial: *Lǐ Huá* (李华; 李華) → *Lǐ H.* or *LǏ H.*, *Zhūgě Kǒngmíng* (诸葛孔明; 諸葛孔明) → *Zhūgě K. M.* or *ZHŪGĚ K. M.*

12. **Line Wrapping**
    1. Words can only be split by the character:
       guāngmíng (光明, bright) → *guāng-*
       míng, not *gu-*
       āngmíng
    2. Initials cannot be split:
       Wáng J. G. (王建国; 王建國) → *Wáng*
       J. G., not *Wáng J.-*
       G.
    3. Apostrophes disappear in line wrapping:
       Xī'ān (西安, Xi'an) → *Xī-*
       ān, not *Xī'-*

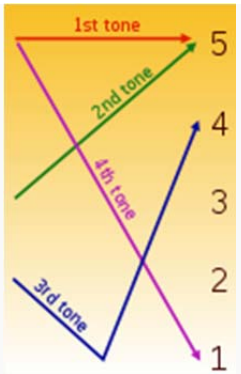
ān

4. When the original word has a hyphen, the hyphen is added at the beginning of the new line: *chēshuǐ-mǎlóng* (车水马龙; 車水馬龍, heavy traffic: "carriage, water, horse, dragon") → *chēshuǐ--mǎlóng*

13. **Hyphenation**: In addition to the situations mentioned above, there are four situations where hyphens are used.
    1. Coordinate and disjunctive compound words, where the two elements are conjoined or opposed, but retain their individual meaning: *gōng-jiàn* (弓箭, bow and arrow), *kuài-màn* (快慢, speed: "fast-slow"), *shíqī-bā suì* (十七八岁; 十七八歲, 17–18 years old), *dǎ-mà* (打骂; 打罵, beat and scold), *Yīng-Hàn*(英汉; 英漢, English-Chinese [dictionary]), *Jīng-Jīn* (京津, Beijing-Tianjin), *lù-hǎi-kōngjūn* (陆海空军; 陸海空軍, army-navy-airforce).
    2. Abbreviated compounds (略语; 略語; *luèyǔ*): *gōnggòng guānxì* (公共关系; 公共關係, public relations) → *gōng-guān* (公关; 公關, PR), *chángtú diànhuà*(长途电话; 長途電話, long-distance calling) → *cháng-huà* (长话; 長話, LDC).
    Exceptions are made when the abbreviated term has become established as a word in its own right, as in *chūzhōng* (初中) for *chūjí zhōngxué* (初级中学; 初級中學, junior high school). Abbreviations of proper-name compounds, however, should always be hyphenated: *Běijīng Dàxué* (北京大学; 北京大學, Peking University) → *Běi-Dà* (北大, PKU).
    3. Four-syllable idioms: *fēngpíng-làngjìng* (风平浪静; 風平浪靜), calm and tranquil: "wind calm, waves down"), *huījīn-rútǔ* (挥金如土; 揮金如土, spend money like water: "throw gold like dirt"), *zhǐ-bǐ-mò-yàn* (纸笔墨砚; 紙筆墨硯, paper-brush-ink-inkstone [four coordinate words]). (The AA-BB reduplication above is an instance of this.)[34]
        1. Other idioms are separated according to the words that make up the idiom: *bēi hēiguō* (背黑锅; 背黑鍋, to be made a scapegoat: "to carry a black pot"), *zhǐ xǔ zhōuguān fànghuǒ, bù xǔ bǎixìng diǎndēng* (只许州官放火，不许百姓点灯; 只許州官放火， 不許百姓點燈, Gods may do what cattle may not: "only the official is allowed to light the fire; the commoners are not allowed to light a lamp")

14. **Punctuation**
    1. The Chinese full stop (。) is changed to a western full stop (.).
    2. The hyphen is a half-width hyphen (-).
    3. Ellipsis can be changed from 6 dots (……) to 3 dots (…).
    4. The enumeration comma (、) is changed to a normal comma (,).
    5. All other punctuation marks are the same as the ones used in normal texts.

Tones



Relative pitch changes of the four tones

The pinyin system also uses diacritics to mark the four tones of Mandarin. The diacritic is placed over the letter that represents the syllable nucleus, unless that letter is missing (see below). Many books printed in China use a mix of fonts, with vowels and tone marks rendered in a different font from the surrounding text, tending to give such pinyin texts a typographically ungainly appearance. This style, most likely rooted in early technical limitations, has led many to believe that pinyin's rules call for this practice and also for the use of a Latin alpha (*ɑ*) rather than the standard style of the letter (*a*) found in most fonts. The official rules of *Hanyu Pinyin,* however, specify no such practice.

1. The first tone (Flat or High Level Tone) is represented by a macron (¯) added to the pinyin vowel:

ā (ɑ̄) ē ī ō ū ǖ Ā Ē Ī Ō Ū Ǖ

2. The second tone (Rising or High-Rising Tone) is denoted by an acute accent (´):

á (ɑ́) é í ó ú ǘ Á É Í Ó Ú Ǘ

3. The third tone (Falling-Rising or Low Tone) is marked by a caron/háček (ˇ). It is not the rounded breve (˘), though a breve is sometimes substituted due to font limitations.

ǎ (ɑ̌) ě ǐ ǒ ǔ ǚ Ǎ Ě Ǐ Ǒ Ǔ Ǚ

4. The fourth tone (Falling or High-Falling Tone) is represented by a grave accent (`):

à (ȁ) è ì ò ù ǜ À È Ì Ò Ù Ǜ

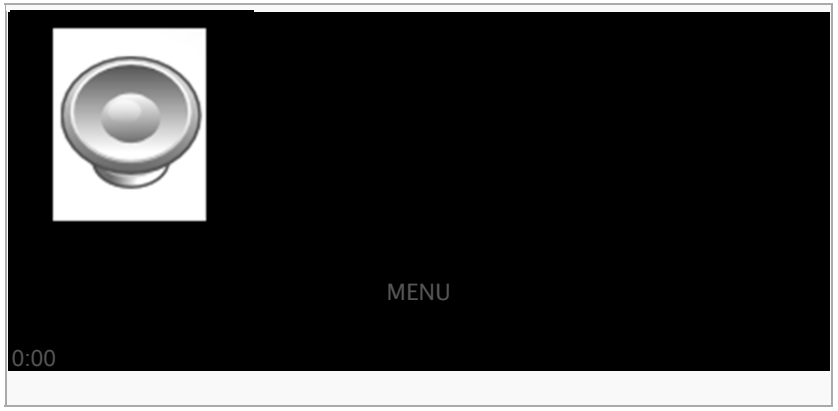
5. The fifth tone (Neutral Tone) is represented by a normal vowel without any accent mark:

a (ɑ) e i o u ü A E I O U Ü

(In some cases, this is also written with a dot before the syllable; for example, ·ma.)[citation needed]

These tone marks normally are only used in Mandarin textbooks or in foreign learning texts, but they are essential for correct pronunciation of Mandarin syllables, as exemplified by the following classic example of five characters whose pronunciations differ only in their tones:

The four main tones of Standard Mandarin, pronounced with the syllable *ma*.



Traditional characters:

媽 麻 馬 罵 嗎

Simplified characters:

妈 麻 吗 骂 吗

The words are "mother", "hemp", "horse", "scold" and a question particle, respectively.

Numerals in place of tone marks

Before the advent of computers, many typewriter fonts did not contain vowels with macron or caron diacritics. Tones were thus represented by placing a tone number at the end of individual syllables. For example, *tóng* is written *tong2.* The number used for each tone is as the order listed above, except the neutral tone, which is either not numbered, or given the number 0 or 5, e.g. *ma5* for 吗/嗎, an interrogative marker.

Tone	Tone Mark	Number added to end of syllable in place of tone mark	Example using tone mark	Example using number	IPA

First	macron (¯)	1	mā	ma1	mɑ˥
Second	acute accent (´)	2	má	ma2	mɑ˧˥
Third	caron (ˇ)	3	mǎ	ma3	mɑ˨˩˦
Fourth	grave accent (`)	4	mà	ma4	mɑ˥˩
"Neutral"	No mark or dot before syllable (·)	no number 5 0	ma ·ma	ma ma5 ma0	mɑ

Rules for placing the tone mark

Briefly, the tone mark should always be placed by the order—*a, o, e, i, u, ü*, with the only exception being *iu,* where the tone mark is placed on the *u* instead. Pinyin tone marks appear primarily above the nucleus of the syllable, for example as in *kuài,* where *k* is the initial, *u* the medial, *a* the nucleus, and *i* the coda. The exception is syllabic nasals like /m/, where the nucleus of the syllable is a consonant, the diacritic will be carried by a written dummy vowel.

When the nucleus is /ə/ (written *e* or *o*), and there is both a medial and a coda, the nucleus may be dropped from writing. In this case, when the coda is a consonant*n* or *ng,* the only vowel left is the medial *i, u,* or *ü,* and so this takes the diacritic. However, when the coda is a vowel, it is the coda rather than the medial which takes the diacritic in the absence of a written nucleus. This occurs with syllables ending in *-ui* (from *wei*: (wèi → -uì) and in *-iu* (from *you: yòu → -iù.*) That is, in the absence of a written nucleus the finals have priority for receiving the tone marker, as long as they are vowels: if not, the medial takes the diacritic.

An algorithm to find the correct vowel letter (when there is more than one) is as follows:[35]

1. If there is an *a* or an *e,* it will take the tone mark.
2. If there is an *ou,* then the *o* takes the tone mark.
3. Otherwise, the second vowel takes the tone mark.

Worded differently,

1. If there is an *a, e,* or *o,* it will take the tone mark; in the case of *ao,* the mark goes on the *a*.

2. Otherwise, the vowels are *-iu* or *-ui,* in which case the second vowel takes the tone mark.

If the tone is written over an *i*, the tittle above the *i* is omitted, as in *yī.*

Phonological intuition

The placement of the tone marker, when more than one of the written letters *a, e, i, o,* and *u* appears, can also be inferred from the nature of the vowel sound in the medial and final. The rule is that the tone marker goes on the spelled vowel that is not a (near-)semi-vowel. The exception is that, for triphthongs that are spelled with only two vowel letters, both of which are the semi-vowels, the tone marker goes on the second spelled vowel.

Specifically, if the spelling of a diphthong begins with *i* (as in *ia*) or *u* (as in *ua*), which here serves as a near-semi-vowel, this letter does not take the tone marker. Likewise, if the spelling of a diphthong ends with *o* or *u* representing a near-semi-vowel (as in *ao* or *ou*), this letter does not receive a tone marker. In a triphthongspelled with three of *a, e, i, o,* and *u* (with *i* or *u* replaced by *y* or *w* at the start of a syllable), the first and third letters coincide with near-semi-vowels and hence do not receive the tone marker (as in *iao* or *uai* or *iou*). But if no letter is written to represent a triphthong's middle (non-semi-vowel) sound (as in *ui* or *iu*), then the tone marker goes on the final (second) vowel letter.

Using tone colors

In addition to tone number and mark, tone color has been suggested as a visual aid for learning. Although there are no formal standards, there are a number of different color schemes in use.

- Dummitt's color scheme was one of the first to be used. It is tone 1 - red, tone 2 - orange, tone 3 - green, tone 4 - blue and neutral tone - black.[36]
- The Unimelb color scheme is tone 1 - blue, tone 2 - green, tone 3 - purple, tone 4 - red, neutral tone - grey
- The Hanping color scheme is tone 1 - blue, tone 2 - green, tone 3 - orange, tone 4 - red, neutral tone - grey.[37]
- The Pleco color scheme is tone 1 - red, tone 2 - green, tone 3 - blue, tone 4 - purple, neutral tone - grey
- The Thomas color scheme is tone 1 - green, tone 2 - blue, tone 3 - red, tone 4 - black, neutral tone - grey

Third tone exceptions

In spoken Chinese, the third tone is often pronounced as a "half third tone," in which the pitch does not rise. Additionally, when two third

tones appear consecutively, such as in 你好 (nǐhǎo, hello), the first syllable is pronounced with the second tone. In pinyin, words like "hello" are still written with two third tones (nǐhǎo).

The *ü* sound

An umlaut is placed over the letter *u* when it occurs after the initials *l* and *n* in order to represent the sound [y]. This is necessary in order to distinguish the front high rounded vowel in *lü* (e.g. 驴; 驢; "donkey") from the back high rounded vowel in *lu* (e.g. 炉; 爐; "oven"). Tonal markers are added on top of the umlaut, as in *lǘ*.

However, the *ü* is *not* used in the other contexts where it could represent a front high rounded vowel, namely after the letters *j*, *q*, *x* and *y*. For example, the sound of the word 鱼/魚 (fish) is transcribed in pinyin simply as *yú*, not as *yǘ*. This practice is opposed to Wade-Giles, which always uses *ü*, and *Tongyong Pinyin,* which always uses *yu*. Whereas Wade-Giles needs to use the umlaut to distinguish between *chü* (pinyin *ju*) and *chu* (pinyin *zhu*), this ambiguity cannot arise with pinyin, so the more convenient form *ju* is used instead of *jü*. Genuine ambiguities only happen with *nu*/*nü* and *lu*/*lü*, which are then distinguished by an umlaut.

Many fonts or output methods do not support an umlaut for *ü* or cannot place tone marks on top of *ü*. Likewise, using *ü* in input methods is difficult because it is not present as a simple key on many keyboard layouts. For these reasons *v* is sometimes used instead by convention. For example, it is common for cellphones to use*v* instead of *ü*. Additionally, some stores in China use *v* instead of *ü* in the transliteration of their names. The drawback is that there are no tone marks for the letter *v*.

This also presents a problem in transcribing names for use on passports, affecting people with names that consist of the sound *lü* or *nü*, particularly people with the surname 吕 (*Lǚ*), a fairly common surname, particularly compared to the surname 陆 (Lù), 鲁 (Lǔ), 卢 (Lú) and 路 (Lù). Previously, the practice varied among different passport issuing offices, with some transcribing as "LV" and "NV" while others used "LU" and "NU". On 10 July 2012, the Ministry of Public Securitystandardized the practice to use "LYU" and "NYU" in passports.[38][39]

Although *nüe* written as *nue*, and *lüe* written as *lue* are not ambiguous, *nue* or *lue* are not correct according to the rules; *nüe* and *lüe* should be used instead. However, some Chinese input methods (e.g. Microsoft Pinyin IME) support both *nve*/*lve* (typing *v* for *ü*) and *nue*/*lue*.

Pinyin in Taiwan

Taiwan (Republic of China) adopted *Tongyong Pinyin,* a modification of *Hanyu Pinyin,* as the official romanization system on the national

level between October 2002 and January 2009, when it switched to *Hanyu Pinyin*. *Tongyong Pinyin* ("official phonetic"), a variant of pinyin developed in Taiwan, was designed to romanize languages and dialects spoken on the island in addition to Mandarin Chinese. The Kuomintang (Chinese Nationalist Party) resisted its adoption, preferring the*Hanyu Pinyin* system used in Mainland China and in general use internationally. Romanization preferences quickly became associated with issues of national identity. Preferences split along party lines: the Kuomintang and its affiliated parties in the pan-blue coalition supported the use of Hanyu Pinyin while the Democratic Progressive Party and its affiliated parties in the pan-green coalition favored the use of Tongyong Pinyin.

Tongyong Pinyin was made the official system in an administrative order that allowed its adoption by local governments to be voluntary. A few localities with governments controlled by the Kuomintang (KMT), most notably Taipei, Hsinchu, and Kinmen County, overrode the order and converted to *Hanyu Pinyin* before the January 1, 2009 national-level switch,[4][5] though with a slightly different capitalization convention than mainland China. Most areas of Taiwan adopted Tongyong Pinyin, consistent with the national policy. Many street signs in Taiwan today still display *Tongyong Pinyin* but some, especially in northern Taiwan, display *Hanyu Pinyin.* It is still not unusual to see spellings on street signs and buildings derived from the older Wade-Giles, MPS2 and other systems.

The adoption of *Hanyu Pinyin* as the official romanization system in Taiwan does not preclude the official retention of earlier spellings. International familiarity has led to the retention of the spelling *Taipei* ("Taibei" in pinyin systems) and even to its continuation in the name of New Taipei, a municipality created in 2010. Personal names on Taiwanese passports honor the choices of Taiwanese citizens, who often prefer the Wade-Giles romanization of their personal names. Transition to Hanyu Pinyin in official use is also necessarily gradual. Universities and other government entities retain earlier spellings in long-established names, budget restraints preclude widespread replacement of signage and stationery in every area, and questions remain about the ability of the national government to enforce the standard island-wide.[40] Primary education in Taiwan continues to teach pronunciation using *zhuyin* (MPS or Mandarin Phonetic Symbols).

Comparison with other orthographies

Pinyin is now used by foreign students learning Chinese as a second language.

Pinyin assigns some Latin letters sound values which are quite different from that of most languages. This has drawn some criticism as it may lead to confusion when uninformed speakers apply either native or English assumed pronunciations to words. However this is not a specific problem of pinyin, since many languages that use the Latin alphabet natively assign different values to the same letters. A recent study on Chinese writing and literacy concluded, "By and large, pinyin represents the Chinese sounds better than the Wade-Giles system,

and does so with fewer extra marks."[41]

Pinyin is purely a representation of the sounds of Mandarin, therefore it lacks the semantic cues that Chinese characters can provide. It is also unsuitable for transcribing some Chinese spoken languages other than Mandarin.

Chart of comparison with other romanizations

Vowels *a, e, o, i*

IPA	ä	ɔ	ɛ	ɯʌ	aɪ	eɪ	ɑʊ	ou	än	ən	ɑŋ	ɤŋ	ɑɻ	i	iɛ	iou	iɛn	in	iŋ
Pinyin	a	o	ê	e	ai	ei	ao	ou	an	en	ang	eng	er	yi	ye	you	yan	yin	ying
Tongyong Pinyin	a	o	e	e	ai	ei	ao	ou	an	en	ang	eng	er	yi	ye	you	yan	yin	ying
Wade–Giles	a	o	eh	o/ê	ai	ei	ao	ou	an	ên	ang	êng	êrh	i	yeh	yu	yen	yin	ying
Zhuyin	ㄚ	ㄛ	ㄝ	ㄜ	ㄞ	ㄟ	ㄠ	ㄡ	ㄢ	ㄣ	ㄤ	ㄥ	ㄦ	ㄧ	ㄧㄝ	ㄧㄡ	ㄧㄢ	ㄧㄣ	ㄧㄥ
example	阿	哦	呗	俄	艾	黑	凹	偶	安	恩	昂	冷	二	一	也	又	言	音	英

Vowels *u, y*

IPA	u	uɔ	ueɪ	uən	uɤŋ	ʊŋ	y	yœ	yɛn	yn	iʊŋ
Pinyin	wu	wo/o	wei	wen	weng	ong	yu	yue	yuan	yun	yong
Tongyong Pinyin	wu	wo/o	wei	wun	wong	ong	yu	yue	yuan	yun	yong
Wade–Giles	wu	wo/o	wei	wên	wêng	ung	yü	yüeh	yüan	yün	yung
Zhuyin	ㄨ	ㄨㄛ/ㄛ	ㄨㄟ	ㄨㄣ	ㄨㄥ		ㄩ	ㄩㄝ	ㄩㄢ	ㄩㄣ	ㄩㄥ
example	五	我	位	文	翁	中	玉	月	元	云	用

Non-sibilant consonants

IPA	puɔ	pʰuɔ	muɔ	fɤŋ	tioʊ	tueɪ	tuən	tʰɯʌ	ny	ly	kɯʌɻ	kʰɯʌ	xɯʌ
Pinyin	bo	po	mo	feng	diu	dui	dun	te	nü	lü	ger	ke	he
Tongyong Pinyin	bo	po	mo	fong	diou	duei	dun	te	nyu	lyu	ger	ke	he

Wade–Giles	po	p'o	mo	fêng	tiu	tui	tun	t'ê	nü	lü	kêrh	k'o	ho
Zhuyin	ㄅㄛ	ㄆㄛ	ㄇㄛ	ㄈㄥ	ㄉㄧㄡ	ㄉㄨㄟ	ㄉㄨㄣ	ㄊㄜ	ㄋㄩ	ㄌㄩ	ㄍㄜㄦ	ㄎㄜ	ㄏㄜ
example	玻	婆	末	封	丟	兌	頓	特	女	旅	歌儿	可	何

Sibilant consonants

IPA	tɕiɛn	tɕiʊŋ	tɕʰin	ɕyɛn	ʈʂɯ˅	ʈʂɨ	ʈʂʰɯ˅	ʈʂʰɨ	ʂɯ˅	ʂɨ	ʐɯ˅	ʐɨ	tsɯ˅	tsuɔ	tsɯ	tsʰɯ˅	tsʰɯ	sɯ˅	sɯ
Pinyin	jian	jiong	qin	xuan	zhe	zhi	che	chi	she	shi	re	ri	ze	zuo	zi	ce	ci	se	si
Tongyong Pinyin	jian	jyong	cin	syuan	jhe	jhih	che	chih	she	shih	re	rih	ze	zuo	zih	ce	cih	se	sih
Wade–Giles	chien	chiung	ch'in	hsüan	chê	chih	ch'ê	ch'ih	shê	shih	jê	jih	tsê	tso	tzu	ts'ê	tz'u	sê	szu
Zhuyin	ㄐㄧ	ㄐㄩ	ㄑㄧ	ㄒㄩ	ㄓㄜ	ㄓ	ㄔㄜ	ㄔ	ㄕ	ㄕ	ㄖ	ㄖ	ㄗㄜ	ㄗㄨ	ㄗ	ㄘㄜ	ㄘ	ㄙ	ㄙ

	ㄋ	ㄙ	ㄣ	ㄋ					ㄜ		ㄜ			ㄛ				ㄜ	
example	件	窘	秦	宣	哲	之	扯	赤	社	是	惹	日	仄	左	字	策	次	色	斯

Tones					
IPA	**mä˥˥**	**mä˧˥**	**mä˨˩˦**	**mä˥˩**	**mä**
Pinyin	mā	má	mǎ	mà	ma
Tongyong Pinyin	ma	má	mǎ	mà	mȧ
Wade–Giles	ma^{1}	ma^{2}	ma^{3}	ma^{4}	ma^{0}
Zhuyin	ㄇㄚ	ㄇㄚˊ	ㄇㄚˇ	ㄇㄚˋ	•ㄇㄚ
example (traditional/simplified)	媽/妈	麻/麻	馬/马	罵/骂	嗎/吗

Braille

Derivation of the 26 letters of the alphabet from the 10 numeric digits (black)

a/1	b/2	c/3	d/4	e/5	f/6	g/7	h/8	i/9	j/0
k	l	m	n	o	p	q	r	s	t
u	v	x	y	z					w

The next ten letters (the next "decade") are the same again, but with dots at both 3 and 6 (green dots). Here *w* was left out as not being part of the basic French alphabet; the French braille order is *u v x y z ç é à è ù* (⠥⠧⠭⠽⠵⠯⠿⠷⠮⠹).[8] The next ten, ending in *w*, are the same again, except that for this series position 6 (purple dot) is used without position 3. These are *â ê î ô û ë ï ü ö*

w (⠽). The *a–j* series lowered in dot space (⠂ ⠆ ⠒ ⠲ ⠢ ⠖ ⠶ ⠦ ⠔ ⠴) are used for punctuation. Letters *a* ⠁ and *c* ⠉, which only use dots in the top row, were lowered two places for the apostrophe and hyphen: ⠄ ⠤. (These are the decade diacritics, at left in the table below, of the second and third decade.) In addition, there are ten patterns that are based on the first two letters (⠁ ⠃) shifted to the right; these were assigned to non-French letters (*ì ä ò* ⠌ ⠜ ⠬) or serve non-letter functions: ⠈ (superscript; in English the accent mark), ⠰ (currency prefix), ⠠(capital, in English the decimal point), ⠼ (number sign), ⠸ (emphasis mark), ⠐ (symbol prefix).

The 64 braille cells[b]

decade		numeric sequence										shift right	
1st	○ ○ ○ ○ ○ ○	● ○ ○ ○ ○ ○	● ○ ● ○ ○ ○	● ● ○ ○ ○ ○	● ● ○ ● ○ ○	● ○ ○ ● ○ ○	● ● ● ○ ○ ○	● ● ● ● ○ ○	● ○ ● ● ○ ○	○ ● ● ○ ○ ○	○ ● ● ● ○ ○	○ ● ○ ○ ○ ○	○ ● ○ ● ○ ○
2nd	○ ○ ○ ○ ● ○	● ○ ○ ○ ● ○	● ○ ● ○ ● ○	● ● ○ ○ ● ○	● ● ○ ● ● ○	● ○ ○ ● ● ○	● ● ● ○ ● ○	● ● ● ● ● ○	● ○ ● ● ● ○	○ ● ● ○ ● ○	○ ● ● ● ● ○	○ ● ○ ○ ● ○	○ ● ○ ● ● ○
3rd	○ ○ ○ ○ ● ●	● ○ ○ ○ ● ●	● ○ ● ○ ● ●	● ● ○ ○ ● ●	● ● ○ ● ● ●	● ○ ○ ● ● ●	● ● ● ○ ● ●	● ● ● ● ● ●	● ○ ● ● ● ●	○ ● ● ○ ● ●	○ ● ● ● ● ●	○ ● ○ ○ ● ●	○ ● ○ ● ● ●
4th	○ ○ ○ ○ ○ ●	● ○ ○ ○ ○ ●	● ○ ● ○ ○ ●	● ● ○ ○ ○ ●	● ● ○ ● ○ ●	● ○ ○ ● ○ ●	● ● ● ○ ○ ●	● ● ● ● ○ ●	● ○ ● ● ○ ●	○ ● ● ○ ○ ●	○ ● ● ● ○ ●	○ ● ○ ○ ○ ●	○ ● ○ ● ○ ●

5th	shift down		○○ ●○ ○○	○○ ●○ ●○	○○ ●● ○○	○○ ●● ○●	○○ ●○ ○●	○○ ●● ●○	○○ ●● ●●	○○ ●○ ●●	○○ ●○ ●○	○○ ○● ●○	○○ ○● ●●		○○ ○● ○○	○○ ○● ○●

Assignment

Historically, there have been three principles in assigning the values of a linear script (print) to braille: Using Louis Braille's original French letter values; reassigning the braille letters according to the sort order of the print alphabet being transcribed; and reassigning the letters to improve the efficiency of writing in braille.

Under international consensus, most braille alphabets follow the French sorting order for the 26 letters of the basic Latin alphabet, and there have been attempts at unifying the letters beyond these 26 (see international braille), though differences remain, for example in German Braille and the contractions of English Braille. This unification avoids the chaos of each nation reordering the braille code to match the sorting order of its print alphabet, as happened in Algerian Braille, where braille codes were numerically reassigned to match the order of the Arabic alphabet and bear little relation to the values used in other countries (compare modernArabic Braille, which uses the French sorting order), and as happened in an early American version of English Braille, where the letters *w, x, y, z* were reassigned to match English alphabetical order. A convention sometimes seen for letters beyond the basic 26 is to exploit the physical symmetry of braille patterns iconically, for example, by assigning a reversed *n* to *ñ* or an inverted *s* to *sh*. (See Hungarian Braille and Bharati Braille, which do this to some extent.)

A third principle was to assign braille codes according to frequency, with the simplest patterns (quickest ones to write) assigned to the most frequent letters of the alphabet. Such frequency-based alphabets were used in Germany and the United States in the 19th century (see American Braille), but none are attested in modern use. Finally, there are braille scripts which don't order the codes numerically at all, such as Japanese Braille and Korean Braille, which are based on more abstract principles of syllable composition.

Academic texts are sometimes written in a script of eight dots per cell rather than six, enabling them encode a greater number of symbols. (See Gardner–Salinas braille codes.) Luxembourgish Braille has adopted eight-dot cells for general use; for example, it adds a dot below each letter to derive its capital variant.

Form

Silver wedding bands with names*Henri(que)* and *Tita* written in braille

Braille was the first writing system with binary encoding.[5] The system as devised by Braille consists of two parts:[6]

1. Character encoding that mapped characters of the French alphabet to tuples of six bits (the dots),
2. The physical representation of those six-bit characters with raised dots in a braille cell.

Within an individual cell, the dot positions are arranged as two columns of three positions. A raised dot can appear in any of the six positions, producing sixty-four (2^6) possible patterns, including one in which there are no raised dots. For reference purposes, a pattern is commonly described by listing the positions where dots are raised, the positions being universally numbered, from top to bottom, as 1 to 3 on the left and 4 to 6 on the right. For example, dot pattern 1-3-4 describe a cell with three dots raised, at the top and bottom in the left column and at the top of the right column: that is, the letter ⠍ *m*. The lines of horizontal braille text are separated by a space, much like visible printed text, so that the dots of one line can be differentiated from the braille text above and below. Different assignments of braille codes (or code pages) are used to map the character sets of different printed scripts to the six-bit cells. Braille assignments have also been created for mathematical and musical notation. However, because the six-dot braille cell allows only 64 (2^6) patterns, including the space, the characters of a braille script commonly have multiple values, depending on their context. That is, character mapping between print and braille is not one-to-one. For example, the character ⠙ corresponds in print to both the letter *d* and the digit *4*.

In addition to simple encoding, many braille alphabets use contractions to reduce the size of braille texts and to increase reading speed. (See Contracted braille)

Writing braille

Braille typewriter

Braille may be produced by hand using a slate and stylus in which each dot is created from the back of the page, writing in mirror image, or it may be produced on a braille typewriter or Perkins Brailler, or an electronic Brailler or eBrailler. Because braille letters cannot be effectively erased and written over if an error is made, an error is overwritten with all six dots (⠿).*Interpoint* refers to braille printing that is offset, so that the paper can be embossed on both sides, with the dots on one side appearing between the divots that form the dots on the other (see the photo in the box at the top of this article for an example). Using a computer or other electronic device, braille may be produced with a braille embosser (printer) or arefreshable braille display (screen).

Braille has been extended to an 8-dot code, particularly for use with braille embossers and refreshable braille displays. In 8-dot braille the additional dots are added at the bottom of the cell, giving a matrix 4 dots high by 2 dots wide. The additional dots are given the numbers 7 (for the lower-left dot) and 8 (for the lower-right dot). Eight-dot braille has the advantages that the case of an individual letter is directly coded in the cell containing the letter and that all the printable ASCII characters can be represented in a single cell. All 256 (2^8) possible combinations of 8 dots are encoded by the Unicode standard. Braille with six dots is frequently stored as Braille ASCII.

Letters

The first 25 braille letters, up through the first half of the 3rd decade, transcribe *a–z* (skipping *w*). In English Braille, the rest of that decade is rounded out with the ligatures *and, for, of, the,* and *with*. Omitting dot 3 from these forms the 4th decade, the ligatures *ch, gh, sh, th, wh, ed, er, ou, ow* and the letter *w*.

⠉	⠓	⠹
ch	sh	th

(See English Braille.)

Formatting

Various formatting marks affect the values of the letters that follow them. They have no direct equivalent in print. The most important in English Braille are:

⠠	⠼
Capital follows	Number follows

That is, ⠠⠁ is read as capital 'A', and ⠼⠁ as the digit '1'.

Punctuation

Basic punctuation marks in English Braille include:

○○ ●○ ○○	○○ ●○ ●○	○○ ○○ ●○	○○ ●● ○○	○○ ○○ ●●	○● ○○ ○●
Comma	Semicolon	Apostrophe	Colon	Hyphen	Decimal point
○○ ●● ○●	○○ ●● ●○	○○ ●○ ●●	○○ ○● ●●	○○ ●● ●●	○● ○○ ●○
Full stop (Period)	Exclamation point	Open quote, question mark	Close quote	Bracket (Parentheses)	Slash (Fraction)

⠦ is both the question mark and the opening quotation mark. Its reading depends on whether it occurs before a word or after.
⠶ is used for both opening and closing parentheses. Its placement relative to spaces and other characters determines its interpretation.
Punctuation varies from language to language. For example, French Braille uses ⠢ for its question mark and swaps the quotation marks and parentheses (to ⠶ and ⠦ ⠴); it uses the period (⠲) for the decimal point, as in print, and the decimal point (⠨) to mark capitalization.

Contractions

For a full list of abbreviations and contractions in English, see English Braille#Contractions.

Braille contractions are words and affixes that are shortened so that they take up fewer cells. In English Braille, for example, the word *afternoon* is written with just three letters, ⠖⠝⠋ ⟨afn⟩, much like stenoscript. There are also several abbreviation marks that create what are effectively logograms.[6] The most common of these is dot 5, which combines with the first letter of words. With the letter ⠍ *m*, the resulting word is ⠐⠍ *mother*. There are also ligatures ("contracted" letters), which are single letters in braille but correspond to more than one letter in print. The letter ⠯ *and*, for example, is used to write words with the sequence *a-n-d* in them, such as ⠓⠯ *hand*.

⠁⠋⠝	⠍⠕⠞	⠓⠁⠝
afternoon *(a-f-n)*	*mother* *(dot 5-m)*	*hand* *(h-and)*

Page dimensions

Most braille embossers support between 34 and 37 cells per line, and between 25 and 28 lines per page.

A manually operated Perkins braille typewriter supports a maximum of 42 cells per line (its margins are adjustable), and typical paper allows 25 lines per page.

A large interlining Stainsby has 36 cells per line and 18 lines per page.

An A4-sized Marburg braille frame, which allows interpoint braille (dots on both sides of the page, offset so they do not interfere with each other) has 30 cells per line and 27 lines per page.

Braille transcription

Braille on a box of tablets. The raised Braille reads 'P L A V I X'.

Braille book and the same book in standard inkprint

Although it is possible to transcribe print by simply substituting the equivalent braille character for its printed equivalent, in English such a character-by-character transcription (known as *uncontracted braille*) is only used by beginners.

Braille characters are much larger than their printed equivalents, and the standard 11" by 11.5" (28 cm × 30 cm) page has room for only 25 lines of 43 characters. To reduce space and increase reading speed, most braille alphabets and orthographies use ligatures, abbreviations, and contractions. Virtually all English Braille books are transcribed in this*contracted braille,* which adds an additional layer of complexity to English orthography: The Library of Congress's*Instruction Manual for Braille Transcribing*[19] runs to over 300 pages and braille transcribers must pass certification tests.

Fully contracted braille is known as *Grade 2 Braille*. There is an intermediate form between Computer Braille—one-for-one identity with print—and Grade 2, which is called Grade 1 Braille. In Grade 1 the capital-sign and Number sign are used, and most punctuation marks are shown using their Grade 2 values.

The system of contractions in English Braille begins with a set of 23 words which are contracted to single characters. Thus the word *but* is contracted to the single letter *b, can* to *c*, *do* to *d*, and so on. Even this simple rule creates issues requiring special cases; for example, *d* is, specifically, an abbreviation of the verb *do;* the noun *do* representing the note of the musical scale is a different word, and must be spelled out.

Portions of words may be contracted, and many rules govern this process. For example, the character with dots 2-3-5 (the letter "f" lowered in the braille cell) stands for "ff" when used in the middle of a word. At the beginning of a word, this same character stands for the word "to"; the character is written in braille with no space following it. (This contraction was removed in the Unified English Braille Code.) At the end of a word, the same character represents an exclamation point.

Some contractions are more similar than their print equivalents. For example, the contraction ⟨lr⟩, meaning 'letter', differs from ⟨ll⟩, meaning 'little', only in adding one dot to the second ⟨l⟩: ⠇⠇ *little*, ⠇⠗ *letter*. This causes greater confusion between the braille spellings of these words and can hinder the learning process of contracted braille.[20]

The contraction rules take into account the linguistic structure of the word; thus, contractions are generally not to be used when their use would alter the usual braille form of a base word to which a prefix or suffix has been added. Some portions of the transcription rules are not fully codified and rely on the judgment of the transcriber. Thus, when the contraction rules permit the same word in more than one way, preference is given to "the contraction that more nearly approximates correct pronunciation."

Grade 3 Braille[21] is a variety of non-standardized systems that include many additional shorthand-like contraction. They are not used for publication, but by individuals for their personal convenience.

Braille translation software

When people produce braille, this is called braille transcription. When computer software produces braille, this is called braille translation. Braille translation software exists to handle most of the common languages of the world, and many technical areas, such as math, music, and tactile graphics.

Braille-reading techniques

Since braille is one of the few writing systems where tactile perception is used, as opposed to visual perception, a braille reader must develop new skills. One skill important for braille readers is the ability to create smooth and even pressures when running one's fingers along the words. There are many different styles and techniques used for the understanding and development of braille, even though a study by B. F. Holland[22] suggests that there is no specific technique that is superior to any other.

Another study by Lowenfield & Abel[23] shows that braille could be read "the fastest and best... by students who read using the index fingers of both hands." Another important reading skill emphasized in this study is to finish reading the end of a line with the right hand and to find the beginning of the next line with the left hand simultaneously. One final conclusion drawn by both Lowenfield and Abel is that children have difficulty using both hands independently where the right hand is the dominant hand. But this hand preference does not correlate to other activities.

International uniformity

Main article: International uniformity of braille

Braille plate in *Duftrosengarten* inRapperswil, Switzerland

When braille was first adapted to languages other than French, many schemes were adopted, including mapping the native alphabet to the alphabetical order of French – e.g. in English W, which was not in the French alphabet at the time, is mapped to braille X, X to Y, Y to Z, and Z to the first French accented letter – or completely rearranging the alphabet such that common letters are represented by the simplest braille patterns. Consequently, mutual intelligibility was greatly hindered by this state of affairs. In 1878, the International Congress on Work for the Blind, held in Paris, proposed an international braille standard, where braille codes for different languages and scripts would be based, not on the order of a particular alphabet, but on phonetic correspondence and transliteration to Latin.[24]

This unified braille has been applied to the languages of India and Africa, Arabic, Vietnamese, Hebrew, Russian, and Armenian, as well as nearly all Latin-script languages. Greek, for example, *gamma* is written as Latin *g*, despite the fact that it has the alphabetic position of *c*; Hebrew *bet*, the second letter of the alphabet and cognate with the Latin letter *b*, is sometimes pronounced /b/ and sometimes /v/, and is written *b* or *v* accordingly; Russian *ts* is written as *c*, which is the usual letter for /ts/ in those Slavic languages that use the Latin alphabet; and Arabic *f* is written as *f*, despite being historically *p*, and occurring in that part of the Arabic alphabet (between historic *o* and *q*).

Other braille conventions

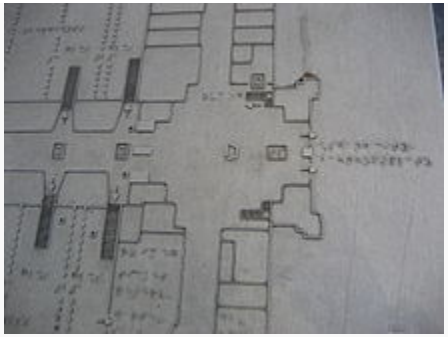
Other systems for assigning values to braille patterns are also followed, beside the simple mapping of the alphabetical order onto the original French order. Some braille alphabets start with unified braille, and then diverge significantly based on the phonology of the target languages, while others diverge even further.

In the various Chinese systems, traditional braille values are used for initial consonants and the simple vowels. In both Mandarin and Cantonese Braille, however, characters have different readings depending on whether they are placed in syllable-initial (onset) or syllable-final (rime) position. For instance, the cell for Latin *k*, ⠅, represents Cantonese *k* (*g* in Yale and other modern romanizations) when initial, but *aak* when final, while Latin *j*, ⠚, represents Cantonese initial *j* but final *oei*.

Novel systems of braille mapping include Korean, which adopts separate syllable-initial and syllable-final forms for its consonants, explicitly grouping braille cells into syllabic groups in the same way as hangul. Japanese, meanwhile, combines independent vowel dot patterns and modifier consonant dot patterns into a single braille cell – an abugida representation of each Japanese mora.

Uses

A bottle of Chapoutier wine, with braille on the label



An embossed map of a German train station, with braille text

The current series of Canadian banknotes has a tactile feature consisting of raised dots that indicate the denomination, allowing bills to be easily identified by visually impaired people. It does not use standard braille; rather, the feature uses a system developed in consultation with blind and visually impaired Canadians after research indicated that braille was not sufficiently robust and that not all potential users read braille. Mexican bank notes, Indian rupee notes, Israeli New Shekelnotes,[25] Russian Ruble and Swiss Franc notes also have special raised symbols to make them identifiable by the visually impaired.

In India there are instances where the parliament acts have been published in braille, such as *The Right to Information Act*.[26]

In the United States, the Americans with Disabilities Act of 1990[27] requires various building signage to be in braille.

In the United Kingdom, it is required that medicines have the name of the medicine in Braille on the labelling.[28]

Braille phone

In May 2014 a Braille phone was introduced by London-based manufacturer OwnFone. Constructed using 3D printingtechniques, the device has print raised text on the keypad to help those who cannot read Braille.[29]

In August 2014 an Australian MVNO KISA Phone launched the first Australian-designed Braille mobile phone. The phone was designed with input from organizations such as Vision Australia and Guide Dogs Victoria. The Braille text is large and Braille buttons can accommodate up to seven characters. Other versions of the KISA Phone are available, including high-contrast designs.[30]

Unicode

Main article: Unicode braille patterns

Braille was added to the Unicode Standard in September, 1999 with the release of version 3.0.

Most braille embossers and refreshable braille displays do not support Unicode, using instead 6-dot braille ASCII. Because of this, they are unable to display this article. Some embossers have proprietary control codes for 8-dot braille or for full graphics mode, where dots may be placed anywhere on the page without leaving any space between braille cells, so that continuous lines can be drawn in diagrams, but these are rarely used and are not standard.

The Unicode standard encodes 8-dot braille glyphs according to their binary appearance, rather than following their assigned numeric order. Dot 1 corresponds to the least significant bit of the low byte of the Unicode scalar value, and dot 8 to the high bit of that byte.

The Unicode block for braille is U+2800 ... U+28FF:

2800 Braille Patterns 28FF

280	281	282	283	284	285	286	287	288	289	28A	28B	28C	28D	28E	28F
	⠁	⠂	⠃	⠄	⠅	⠆	⠇	⠈	⠉	⠊	⠋	⠌	⠍	⠎	⠏
2800	2801	2802	2803	2804	2805	2806	2807	2808	2809	280A	280B	280C	280D	280E	280F
⠐	⠑	⠒	⠓	⠔	⠕	⠖	⠗	⠘	⠙	⠚	⠛	⠜	⠝	⠞	⠟
2810	2811	2812	2813	2814	2815	2816	2817	2818	2819	281A	281B	281C	281D	281E	281F
⠠	⠡	⠢	⠣	⠤	⠥	⠦	⠧	⠨	⠩	⠪	⠫	⠬	⠭	⠮	⠯
2820	2821	2822	2823	2824	2825	2826	2827	2828	2829	282A	282B	282C	282D	282E	282F
⠰	⠱	⠲	⠳	⠴	⠵	⠶	⠷	⠸	⠹	⠺	⠻	⠼	⠽	⠾	⠿
2830	2831	2832	2833	2834	2835	2836	2837	2838	2839	283A	283B	283C	283D	283E	283F

⠀ ⠁ ⠂ ⠃ ⠄ ⠅ ⠆ ⠇ ⠈ ⠉ ⠊ ⠋ ⠌ ⠍ ⠎ ⠏

2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 284A 284B 284C 284D 284E 284F

⠐ ⠑ ⠒ ⠓ ⠔ ⠕ ⠖ ⠗ ⠘ ⠙ ⠚ ⠛ ⠜ ⠝ ⠞ ⠟

2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 285A 285B 285C 285D 285E 285F

⠠ ⠡ ⠢ ⠣ ⠤ ⠥ ⠦ ⠧ ⠨ ⠩ ⠪ ⠫ ⠬ ⠭ ⠮ ⠯

2860 2861 2862 2863 2864 2865 2866 2867 2868 2869 286A 286B 286C 286D 286E 286F

⠰ ⠱ ⠲ ⠳ ⠴ ⠵ ⠶ ⠷ ⠸ ⠹ ⠺ ⠻ ⠼ ⠽ ⠾ ⠿

2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 287A 287B 287C 287D 287E 287F

⡀ ⡁ ⡂ ⡃ ⡄ ⡅ ⡆ ⡇ ⡈ ⡉ ⡊ ⡋ ⡌ ⡍ ⡎ ⡏

2880 2881 2882 2883 2884 2885 2886 2887 2888 2889 288A 288B 288C 288D 288E 288F

⡐ ⡑ ⡒ ⡓ ⡔ ⡕ ⡖ ⡗ ⡘ ⡙ ⡚ ⡛ ⡜ ⡝ ⡞ ⡟

2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 289A 289B 289C 289D 289E 289F

⡠ ⡡ ⡢ ⡣ ⡤ ⡥ ⡦ ⡧ ⡨ ⡩ ⡪ ⡫ ⡬ ⡭ ⡮ ⡯

28A0 28A1 28A2 28A3 28A4 28A5 28A6 28A7 28A8 28A9 28AA 28AB 28AC 28AD 28AE 28AF

⡰ ⡱ ⡲ ⡳ ⡴ ⡵ ⡶ ⡷ ⡸ ⡹ ⡺ ⡻ ⡼ ⡽ ⡾ ⡿

28B0 28B1 28B2 28B3 28B4 28B5 28B6 28B7 28B8 28B9 28BA 28BB 28BC 28BD 28BE 28BF

⢀ ⢁ ⢂ ⢃ ⢄ ⢅ ⢆ ⢇ ⢈ ⢉ ⢊ ⢋ ⢌ ⢍ ⢎ ⢏

28C0 28C1 28C2 28C3 28C4 28C5 28C6 28C7 28C8 28C9 28CA 28CB 28CC 28CD 28CE 28CF

⠐⠀ ⠐⠁ ⠐⠂ ⠐⠃ ⠐⠄ ⠐⠅ ⠐⠆ ⠐⠇ ⠐⠈ ⠐⠉ ⠐⠊ ⠐⠋ ⠐⠌ ⠐⠍ ⠐⠎ ⠐⠏

28D0 28D1 28D2 28D3 28D4 28D5 28D6 28D7 28D8 28D9 28DA 28DB 28DC 28DD 28DE 28DF

⠐⠐⠀ ⠐⠐⠁ ⠐⠐⠂ ⠐⠐⠃ ⠐⠐⠄ ⠐⠐⠅ ⠐⠐⠆ ⠐⠐⠇ ⠐⠐⠈ ⠐⠐⠉ ⠐⠐⠊ ⠐⠐⠋ ⠐⠐⠌ ⠐⠐⠍ ⠐⠐⠎ ⠐⠐⠏

28E0 28E1 28E2 28E3 28E4 28E5 28E6 28E7 28E8 28E9 28EA 28EB 28EC 28ED 28EE 28EF

⠐⠐⠐⠀ ⠐⠐⠐⠁ ⠐⠐⠐⠂ ⠐⠐⠐⠃ ⠐⠐⠐⠄ ⠐⠐⠐⠅ ⠐⠐⠐⠆ ⠐⠐⠐⠇ ⠐⠐⠐⠈ ⠐⠐⠐⠉ ⠐⠐⠐⠊ ⠐⠐⠐⠋ ⠐⠐⠐⠌ ⠐⠐⠐⠍ ⠐⠐⠐⠎ ⠐⠐⠐⠏

28F0 28F1 28F2 28F3 28F4 28F5 28F6 28F7 28F8 28F9 28FA 28FB 28FC 28FD 28FE 28FF

Appendix J: Text to Speech Tips

TIPS FOR TEXT TO SPEECH: ** Note: Acronyms and short form words should be avoided whenever possible. When in doubt, spell it out.

	CORRECT	INCORRECT
Age	• 42 to 45 years old • 42 years old	• 40-45 • 42 yr old • 42 yrs old
Height	• 5 feet 6 inches • 165 centimetres (or centimeters) • 1.2 meters	• 5 foot 6 • 5 ft 6 in • 165cm, • 165 cms (cms is not a unit of measure) • 1.2m
Speed	• Kilometres (kilometers) per hour	• km per hour

		• km/h • KmH • kms/hr • km/hour
Temperature	• -30 degrees Fahrenheit • +30 degrees Celsius	• -30° F • +30 degrees C
Date	• MM/DD/YYYY • YYYY-MM-DD ISO preferred Not ambiguous	• 02/12/2013 = February 12th, 2013 • 2013/2/12 = 2013, February 12th • Only recognizes M-D-Y format. Some countries use DD/MM/YYYY Which is ambiguous
Time	• 10:00 AM • 10:00 PM	• 10 o'clock • 2200 hours (avoid using the 24 hour clock, recipients may not understand this format.)
Weight	12 lbs (must have a space) 12 pounds 13 kg (plural form does not exist)	12lbs 12 pds 13kgs wt

Directions	• North • Northeast • East • Southeast • South • Southwest • West • Northwest	• N • NE • E • SE • S • SW • W • NW
Directions of Travel	• Northbound, heading north • Southbound, heading south • Westbound, heading west • Eastbound, heading east	• NB (northbound) • SB (southbound) • WB (westbound) • EB (eastbound)
License Plate	• New Jersey • A B C 1 2 3 (must have a space between each character)	• NJ • ABC123 • ABC 123 Non alpha numeric
	• and • percent • at	• & • % • @

	• #1(incl number = “number 1”) • _(will be pronounced “underscore”)	
Addresses	• Ensure proper punctuation and capitalization. North, not N • 14225 142nd Street • 100 Ave. to 118 Ave. (requires period with Ave.) • “Suite” needs to be spelled out in full • Remember that numbers are spoken out in the tens and hundreds. So 12445 is twelve thousand four hundred forty five. So use spaces between such numbers as number and street number. Not non-breaking space • New Jersey	• 506 2nd Street N • Be careful about dual use abbreviations. St. = “Saint” rather than “Street”. 506 2nd St. N becomes “Five Hundred and Six, second Saint N”. Type out the full text to insure proper pronunciation. • Spell out “Drive” and “Highway” in full • NJ
Telephone Numbers	780 980 8758 Do not use non-breaking space for phone numbers recommended. 9 1 1 (spaces need to be included in between each number) Currently Excel does not recognize international phone number format e.g +64(7)321-4567	780-980-8758 7809808758 780.980.8758 (780)9808758 911
Space in numbers	1 234 567.89 use non-breaking space is ISO recommended. 123.45 ISO is decimal point, not a comma which is a string delimiter.	1,234,567.89 123,45

	Currently Excel does not recognize a non-breaking space as a number format although it is ISO.	Non-breaking space is CTRL-SHIFT-SPACE in Word.
Acronyms	• Pronounce acronyms to right • Pronounce unit of measure • Pronounce unit of measure	• All uppercase English and IPA (International Phonetic Alphabet) except that SHALL, NOT, MAY, SHOULD and RECOMMENDED are not included. • Mix of English and Greek • digit(s) followed by space or no-break space then some lower case including e.g. kg, mm, mi, kyd, psi, Gb/s, gal (US or Imperial).

Failure to use proper punctuation may result in a word being mispronounced. ***WHEN IN DOUBT, SPELL IT OUT***

Apparently the IPAWS to NWEM interface has a limited character set. So instead of creating a NWEMtext parameter (similar to the CMAMtext parameter) or translating character sets for systems with text limitations, originators can't use the full character set.

The NOAA/NWEM character set is limited to letters, digits, a period (.), and ellipsis (...) [only three dots, not two, not four). No other punctuation is allowed, such as comma (,), semi-colon (;), colon (:), etc.

As long as IPAWS and NWEM have that character set limitation, it will be more challenging for text-to-speech systems.

A considerable assistance in finding source material was provided by Wikipedia and Google.