3.3 CJK Phonetics and Symbols

The CJK Phonetics and Symbols area of the Unicode standard includes the encoding of punctuation marks and symbols used in the CJK (Chinese, Japanese, Korean) phonetic alphabets.

Figure 3-11. CJK Phonetics and Symbols
CJK Symbols and Punctuation  U+3000 → U+303F

This block encodes punctuation marks and symbols in use with Han ideographs and Asian phonetic alphabets.

Most of these characters are found in East Asian standards.

Encoding Structure. The Unicode block for CJK Punctuation and Symbols is divided into the following ranges:

- U+3000 → U+3006: Ideographic space and punctuation
- U+3007: Ideographic zero
- U+3008 → U+3011: CJK quotation marks and brackets
- U+3012 → U+3013: CJK symbols
- U+3014 → U+301F: More CJK symbols and brackets
- U+3020: Postal mark face
- U+3021 → U+3029: Hangzhou numerals
- U+302A → U+302F: Tone marks
- U+3030 → U+3036: Other CJK symbols
- U+3037 → U+303E: Currently unassigned
- U+303F: Ideographic half-fill character
Hiragana $\text{U+3040} \rightarrow \text{U+309F}$

Hiragana is the cursive syllabary used to write Japanese words phonetically and also to write sentence particles and inflectional endings. Hiragana is commonly used as well to indicate the pronunciation of Japanese words. Hiragana syllables are phonetically equivalent to corresponding Katakana syllables.

**Standards.** The Unicode Hiragana block is based on the JIS X 0208-1990 standard, extended by the non-standard syllable U+3094 vu, which is included in some Japanese corporate standards.

**Non-spacing Marks.** Hiragana and the related script Katakana use the two non-spacing characters encoded in this block to generate voiced and semi-voiced syllables from the base syllables. All normal composites using these marks are already encoded as characters, and use of these composite forms is the predominant JIS usage. In the Unicode design, these non-spacing marks follow the base character. As most implementations and the JIS standard treat these as spacing characters, the Unicode standard also contains two corresponding spacing marks at U+309B and U+309C.

**Punctuation-like Characters.** These are the Hiragana specific iteration and voiced iteration marks.

**Encoding Structure.** The Unicode block for the Hiragana script is divided into the following ranges:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+3040</td>
<td>Mapping of the JIS X 0208 standard</td>
</tr>
<tr>
<td>U+3094</td>
<td>Variant form</td>
</tr>
<tr>
<td>U+3095</td>
<td>Currently unassigned</td>
</tr>
<tr>
<td>U+3099</td>
<td>Non-spacing diacritical marks</td>
</tr>
<tr>
<td>U+309B</td>
<td>Spacing diacritical marks</td>
</tr>
<tr>
<td>U+309D</td>
<td>Punctuation-like characters</td>
</tr>
<tr>
<td>U+309F</td>
<td>Currently unassigned</td>
</tr>
</tbody>
</table>
Katakana  U+30A0 → U+30FF

Katakana is the syllabary used to write Japanese words of Western origin. Katakana is commonly used as well to write Japanese words in order to create visual emphasis. Katakana syllables are phonetically equivalent to corresponding Hiragana syllables.

Standards. The Unicode Katakana block is based on the JIS X 0208-1990 standard.

Punctuation-like Characters. These are the Katakana conjunctive dot, the Hiragana/Katakana prolonged-syllable mark, the specific iteration and the voiced iteration marks.

Encoding Structure. The Unicode block for the Katakana script is divided into the following ranges:

<table>
<thead>
<tr>
<th>Code Point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+30A0 → U+30F6</td>
<td>Mapping of the JIS X 0208 standard</td>
</tr>
<tr>
<td>U+30F7 → U+30FA</td>
<td>Currently unassigned</td>
</tr>
<tr>
<td>U+30FB → U+30FE</td>
<td>Punctuation-like characters</td>
</tr>
<tr>
<td>U+30FF</td>
<td>Currently unassigned</td>
</tr>
</tbody>
</table>
Bopomofo  U+3100 → U+312F

Bopomofo are a set of letters used to annotate or teach the phonetics of Chinese, primarily the standard Mandarin language. They are used in dictionaries and teaching materials, not in the actual writing of Chinese text. Proper Chinese names for this alphabet would be Zhuyin-Zimu ("phonetic alphabet") or Zhuyin-Fuhao ("phonetic symbols"), but the informal term "Bopomofo" (analogous to "ABCs") provides a more serviceable English name. The Bopomofo were developed as part of a populist literacy campaign following the 1911 revolution; thus they are acceptable to all branches of modern Chinese culture, although in the People's Republic of China their function has been largely taken over by the Pinyin romanization.

Standards. The standard Mandarin set of Bopomofo are included in the People's Republic of China standard GB 2312-80 and in the Republic of China standard CNS 11643-86.

Mandarin Tone Marks. Small modifier letters used to indicate the five Mandarin tones are part of the Bopomofo system, but in the Unicode standard they have been unified into the Modifier Letter range, as follows:

<table>
<thead>
<tr>
<th>Tone</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>U+02C9</td>
<td>MODIFIER LETTER MACRON</td>
</tr>
<tr>
<td>second</td>
<td>U+02CA</td>
<td>MODIFIER LETTER ACUTE</td>
</tr>
<tr>
<td>third</td>
<td>U+02C7</td>
<td>MODIFIER LETTER HACEK</td>
</tr>
<tr>
<td>fourth</td>
<td>U+2CB0</td>
<td>MODIFIER LETTER GRAVE</td>
</tr>
<tr>
<td>light</td>
<td>U+0209</td>
<td>SPACING DOT ABOVE</td>
</tr>
</tbody>
</table>

Standard Mandarin Bopomofo. The order of Bopomofo letters is standard worldwide. The code offset of the first letter U+3105 BOPOMOFO LETTER B from a multiple of 16 is included to match the offset in the ISO-registered standard GB 2312-80. The character U+3127 BOPOMOFO LETTER I is usually written as a horizontal stroke when the Bopomofo text is set vertically; in the Unicode standard this is considered to be a rendering variation and not a separate character code.

Non-Mandarin Letters. These are very rarely used, but are included for completeness. There are no standard Bopomofo letters for the phonetics of Cantonese or other dialects.

Encoding Structure. The Unicode block for Bopomofo is divided into the following ranges:

- elsewhere
- U+3105 → U+3129 Standard Mandarin Bopomofo
- U+312A → U+312C Dialect (non-Mandarin) letters
Hangul Elements  U+3130 → U+318F

Standards. The Unicode standard follows KS C 5601 for Hangul elements.

Encoding Structure. The Unicode block for Hangul Elements is divided into the following ranges:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+3130</td>
<td>Currently unassigned</td>
</tr>
<tr>
<td>U+3131 → U+3163</td>
<td>Modern Jamo elements</td>
</tr>
<tr>
<td>U+3164 → U+318E</td>
<td>Archaic Jamo elements</td>
</tr>
<tr>
<td>U+318F</td>
<td>Currently unassigned</td>
</tr>
</tbody>
</table>

NOTE: The Jamos encoded in the first version of the Unicode standard are non-combining and are provided for use with data from current Korean standards which contain non-combining Jamos.
CJK Miscellaneous  U+3190 → U+31FF

This block currently contains only a set of Kanbun marks used in Japanese texts to indicate the Japanese reading order of classical Chinese texts. They are not encoded in any current character encoding standards, but are widely used in literature.

*Encoding structure.* The Unicode block for CJK Miscellaneous is divided into the following ranges:

- U+3190 → U+319F  Kanbun marks
- U+31A0 → U+31FF  Currently unassigned
Enclosed CJK Letters and Ideographs  U+3200 → U+32FF

Standards. The CJK Enclosed block provides mapping for all the enclosed Hangul elements from Korean standard KS C 5601 as well as parenthesized ideographic characters from JIS 0208-1990 standard, CNS 11643, and several corporate registries.

Encoding Structure. The Unicode block for CJK parenthesized letters and ideographs is divided into the following ranges:

- U+3200 → U+320D Parenthesized Hangul elements
- U+320E → U+321C Parenthesized Hangul syllables
- U+321D → U+321F Currently unassigned
- U+3220 → U+3243 Parenthesized ideographs
- U+3244 → U+325F Currently unassigned
- U+3260 → U+326D Circled Hangul elements
- U+326E → U+327B Circled Hangul syllables
- U+327C → U+327E Currently unassigned
- U+327F Korean standard symbol
- U+3280 → U+3280 Circled ideographs
- U+3281 → U+32CF Currently unassigned
- U+32D0 → U+32FE Circled Katakana
- U+32FF Japanese Industrial Standard symbol
CJK Squared Words  U+3300 → U+337F

CJK squared Katakana words are Katakana-spelled words that fill a single character position if intermixed with ideographic Han (Kanji) characters. Likewise, squared Latin abbreviation symbols are designed to fill a single character position when mixed with Han characters.

Standards. Squared Katakana words are derived from various corporate registries.

Encoding Structure. The Unicode block for CJK Squared Words is divided into the following ranges:

U+3300 → U+3357 Squared symbolic Katakana words
U+3358 → U+337A Currently unassigned
U+337B → U+337E Japanese era names

The Japanese era names refer to the following dates:

U+337B Heisei era 1989/1/7 to present day
U+337C Showa era 1926/12/24 to 1989/1/6
U+337D Taishou era 1912/7/29 to 1926/12/23
U+337E Meiji era 1867 to 1912/7/28
CJK Squared Abbreviations  U+3380 → U+33FF

CJK squared abbreviations are encoded solely for compatibility with existing standards.

Standards. Squared Latin abbreviation symbols are derived from the KS C 5601 and CNS 11643 standards.

Encoding Structure. The Unicode block for CJK Squared Abbreviations is divided into the following ranges:

U+3380 → U+33DD     Squared Latin abbreviation symbols
U+33DE → U+33FF       Currently unassigned
Korean Hangul Syllables  U+3400 → U+3D2F

*Standards.* The Hangul Syllables are taken from KS C 5601 and are encoded in the same order as that standard. This means that the Unicode and KS code values can be computed from each other via a simple formula, rather than requiring table lookup.

*Encoding Structure.* The Unicode block for Hangul syllables is divided into the following ranges:

- U+3400 → U+3D2D  KS C 5601 Hangul syllables
- U+3D2E → U+3D2F  Currently unassigned
This page intentionally left blank. The material is undergoing review.
The revised version will appear in Volume 2.