

# The Unicode® Standard

## Version 11.0 – Core Specification

To learn about the latest version of the Unicode Standard, see <http://www.unicode.org/versions/latest/>.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

© 2018 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <http://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 11.0.

Includes index.

ISBN 978-1-936213-19-1 (<http://www.unicode.org/versions/Unicode11.0.0/>)

1. Unicode (Computer character set) I. Unicode Consortium.

QA268.U545 2018

ISBN 978-1-936213-19-1

Published in Mountain View, CA

June 2018

## Chapter 21

# *Notational Systems*

This chapter discusses various notational systems:

*Braille patterns*

*Western and Byzantine musical symbols*

*Ancient Greek musical notation*

*Duployan*

*Sutton SignWriting*

Braille consists of a related set of notational systems, using raised dots embossed on paper or other mediums to provide a tactile writing system for the blind. The patterns of dots are associated with the letters or syllables of other writing systems, but the particular rules of association vary from language to language. The Unicode Standard encodes a complete set of symbols for the shapes of Braille patterns; however the association of letters to the patterns is left to other standards. Text should normally be represented using the regular Unicode characters of the script. Only when the intent is to convey a particular binding of text to a Braille pattern sequence should it be represented using the symbols for the Braille patterns.

Musical notation—particularly Western musical notation—is different from ordinary text in the way it is laid out, especially the representation of pitch and duration in Western musical notation. However, ordinary text commonly refers to the basic graphical elements that are used in musical notation, so such symbols are encoded in the Unicode Standard. Additional sets of symbols are encoded to support historical systems of musical notation.

Duployan is an uncased, alphabetic stenographic writing system invented by Emile Duployé, and published in 1860. It was one of the two most commonly used French shorthands. The Duployan shorthands are used as a secondary shorthand for writing French, English, German, Spanish, and Romanian. An adaptation and augmentation of Duployan was used as an alternate primary script for several First Nations' languages in interior British Columbia, Canada.

Sutton SignWriting is a notational system developed in 1974 by Valerie Sutton and used for the transcription of many sign languages. It is a featural writing system, in which visually iconic basic symbols are arranged in two-dimensional layout to form snapshots of the individual signs of a sign language, which are roughly equivalent to words. The Unicode Standard encodes the basic symbols as atomic characters or combining character sequences.

## 21.1 Braille

### **Braille Patterns: U+2800–U+28FF**

Braille is a writing system used by blind people worldwide. It uses a system of six or eight raised dots, arranged in two vertical rows of three or four dots, respectively. Eight-dot systems build on six-dot systems by adding two extra dots above or below the core matrix. Six-dot Braille allows 64 possible combinations, and eight-dot Braille allows 256 possible patterns of dot combinations. There is no fixed correspondence between a dot pattern and a character or symbol of any given script. Dot pattern assignments are dependent on context and user community. A single pattern can represent an abbreviation or a frequently occurring short word. For a number of contexts and user communities, the series of ISO technical reports starting with ISO/TR 11548-1 provide standardized correspondence tables as well as invocation sequences to indicate a context switch.

The Unicode Standard encodes a single complete set of 256 eight-dot patterns. This set includes the 64 dot patterns needed for six-dot Braille.

The character names for Braille patterns are based on the assignments of the dots of the Braille pattern to digits 1 to 8 as follows:

1	●●	4
2	●●	5
3	●●	6
7	●●	8

The designation of dots 1 to 6 corresponds to that of six-dot Braille. The additional dots 7 and 8 are added beneath. The character name for a Braille pattern consists of `BRaille PATtern DOTS-12345678`, where only those digits corresponding to dots in the pattern are included. The name for the empty pattern is `BRaille PATtern BLANK`.

The 256 Braille patterns are arranged in the same sequence as in ISO/TR 11548-1, which is based on an octal number generated from the pattern arrangement. Octal numbers are associated with each dot of a Braille pattern in the following way:

1	●●	10
2	●●	20
4	●●	40
100	●●	200

The octal number is obtained by adding the values corresponding to the dots present in the pattern. Octal numbers smaller than 100 are expanded to three digits by inserting leading zeroes. For example, the dots of `BRaille PATtern DOTS-1247` are assigned to the octal values of  $1_8$ ,  $2_8$ ,  $10_8$ , and  $100_8$ . The octal number representing the sum of these values is  $113_8$ .

The assignment of meanings to Braille patterns is outside the scope of this standard.

**Example.** According to ISO/TR 11548-2, the character LATIN CAPITAL LETTER F can be represented in eight-dot Braille by the combination of the dots 1, 2, 4, and 7 (BRAILLE PATTERN DOTS-1247). A full circle corresponds to a tangible (set) dot, and empty circles serve as position indicators for dots not set within the dot matrix:

1	●	●	4
2	●	○	5
3	○	○	6
7	●	○	8

**Usage Model.** The eight-dot Braille patterns in the Unicode Standard are intended to be used with either style of eight-dot Braille system, whether the additional two dots are considered to be in the top row or in the bottom row. These two systems are never intermixed in the same context, so their distinction is a matter of convention. The intent of encoding the 256 Braille patterns in the Unicode Standard is to allow input and output devices to be implemented that can interchange Braille data without having to go through a context-dependent conversion from semantic values to patterns, or vice versa. In this manner, final-form documents can be exchanged and faithfully rendered. At the same time, processing of textual data that require semantic support is intended to take place using the regular character assignments in the Unicode Standard.

**Imaging.** When output on a Braille device, dots shown as black are intended to be rendered as tangible. Dots shown in the standard as open circles are blank (not rendered as tangible). The Unicode Standard does not specify any physical dimension of Braille characters.

In the absence of a higher-level protocol, Braille patterns are output from left to right. When used to render final form (tangible) documents, Braille patterns are normally not intermixed with any other Unicode characters except control codes.

**Script.** Unlike other sets of symbols, the Braille Patterns are given their own, unique value of the Script property in the Unicode Standard. This follows both from the behavior of Braille in forming a consistent writing system on its own terms, as well as from the independent bibliographic status of books and other documents printed in Braille. For more information on the Script property, see Unicode Standard Annex #24, “Unicode Script Property.”

## 21.2 Western Musical Symbols

### *Musical Symbols: U+1D100–U+1D1FF*

The musical symbols encoded in the Musical Symbols block are intended to cover basic Western musical notation and its antecedents: mensural notation and plainsong (or Gregorian) notation, as well as closely related systems, such as Kievan notation. The most comprehensive coded language in regular use for representing sound is the common musical notation (CMN) of the Western world. Western musical notation is a system of symbols that is relatively, but not completely, self-consistent and relatively stable but still, like music itself, evolving. This open-ended system has survived over time partly because of its flexibility and extensibility. In the Unicode Standard, musical symbols have been drawn primarily from CMN. Commonly recognized additions to the CMN repertoire, such as quarter-tone accidentals, cluster noteheads, and shape-note noteheads, have also been included.

Graphical score elements are not included in the Musical Symbols block. These pictographs are usually created for a specific repertoire or sometimes even a single piece. Characters that have some specialized meaning in music but that are found in other character blocks are not included. They include numbers for time signatures and figured basses, letters for section labels and Roman numeral harmonic analysis, and so on.

Musical symbols are used worldwide in a more or less standard manner by a very large group of users. The symbols frequently occur in running text and may be treated as simple spacing characters with no special properties, with a few exceptions. Musical symbols are used in contexts such as theoretical works, pedagogical texts, terminological dictionaries, bibliographic databases, thematic catalogs, and databases of musical data. The musical symbol characters are also intended to be used within higher-level protocols, such as music description languages and file formats for the representation of musical data and musical scores.

Because of the complexities of layout and of pitch representation in general, the encoding of musical pitch is intentionally outside the scope of the Unicode Standard. The Musical Symbols block provides a common set of elements for interchange and processing. Encoding of pitch, and layout of the resulting musical structure, involves specifications not only for the vertical relationship between multiple notes simultaneously, but also in multiple staves, between instrumental parts, and so forth. These musical features are expected to be handled entirely in higher-level protocols making use of the graphical elements provided. Lack of pitch encoding is not a shortcoming, but rather is a necessary feature of the encoding.

**Glyphs.** The glyphs for musical symbols shown in the code charts, are representative of typical cases; however, note in particular that the stem direction is not specified by the Unicode Standard and can be determined only in context. For a font that is intended to provide musical symbols in running text, either stem direction is acceptable. In some con-

texts—particularly for applications in early music—note heads, stems, flags, and other associated symbols may need to be rendered in different colors—for example, red.

**Symbols in Other Blocks.** U+266D MUSIC FLAT SIGN, U+266E MUSIC NATURAL SIGN, and U+266F MUSIC SHARP SIGN—three characters that occur frequently in musical notation—are encoded in the Miscellaneous Symbols block (U+2600.U+267F). However, four characters also encoded in that block are to be interpreted merely as dingbats or miscellaneous symbols, not as representing actual musical notes:

U+2669 QUARTER NOTE

U+266A EIGHTH NOTE

U+266B BEAMED EIGHTH NOTES

U+266C BEAMED SIXTEENTH NOTES

**Processing.** Most musical symbols can be thought of as simple spacing characters when used inline within texts and examples, even though they behave in a more complex manner in full musical layout. Some characters are meant only to be combined with others to produce combined character sequences, representing musical notes and their particular articulations. Musical symbols can be input, processed, and displayed in a manner similar to mathematical symbols. When embedded in text, most of the symbols are simple spacing characters with no special properties. A few characters have format control functions, as described later in this section.

**Input Methods.** Musical symbols can be entered via standard alphanumeric keyboard, via piano keyboard or other device, or by a graphical method. Keyboard input of the musical symbols may make use of techniques similar to those used for Chinese, Japanese, and Korean. In addition, input methods utilizing pointing devices or piano keyboards could be developed similar to those in existing musical layout systems. For example, within a graphical user interface, the user could choose symbols from a palette-style menu.

**Directionality.** When combined with right-to-left texts—in Hebrew or Arabic, for example—the musical notation is usually written from left to right in the normal manner. The words are divided into syllables and placed under or above the notes in the same fashion as for Latin and other left-to-right scripts. The individual words or syllables corresponding to each note, however, are written in the dominant direction of the script.

The opposite approach is also known: in some traditions, the musical notation is actually written from right to left. In that case, some of the symbols, such as clef signs, are mirrored; other symbols, such as notes, flags, and accidentals, are *not* mirrored. All responsibility for such details of bidirectional layout lies with higher-level protocols and is not reflected in any character properties. *Figure 21-1* exemplifies this principle with two musical passages. The first example shows Turkish lyrics in Arabic script with ordinary left-to-right musical notation; the second shows right-to-left musical notation. Note the partial mirroring.

**Format Characters.** Extensive ligature-like beams are used frequently in musical notation between groups of notes having short values. The practice is widespread and very predictable, so it is therefore amenable to algorithmic handling. The format characters U+1D173

**Figure 21-1.** Examples of Specialized Music Layout

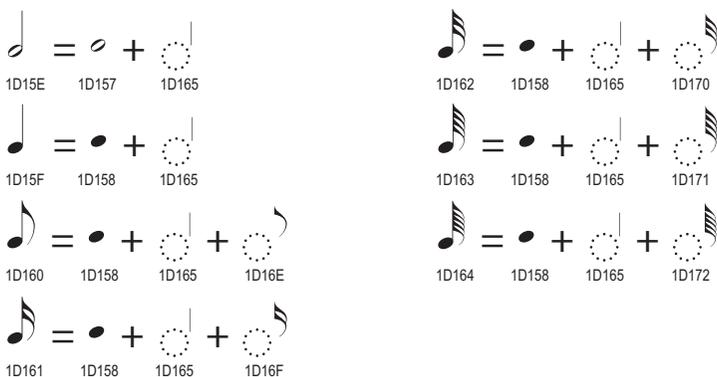
MUSICAL SYMBOL BEGIN BEAM and U+1D174 MUSICAL SYMBOL END BEAM can be used to indicate the extents of beam groupings. In some exceptional cases, beams are left unclosed on one end. This status can be indicated with a U+1D159 MUSICAL SYMBOL NULL NOTE-HEAD character if no stem is to appear at the end of the beam.

Similarly, format characters have been provided for other connecting structures. The characters U+1D175 MUSICAL SYMBOL BEGIN TIE, U+1D176 MUSICAL SYMBOL END TIE, U+1D177 MUSICAL SYMBOL BEGIN SLUR, U+1D178 MUSICAL SYMBOL END SLUR, U+1D179 MUSICAL SYMBOL BEGIN PHRASE, and U+1D17A MUSICAL SYMBOL END PHRASE indicate the extent of these features. Like beaming, these features are easily handled in an algorithmic fashion.

These pairs of characters modify the layout and grouping of notes and phrases in full musical notation. When musical examples are written or rendered in plain text without special software, the start/end format characters may be rendered as brackets or left uninterpreted. To the extent possible, more sophisticated software that renders musical examples inline with natural-language text might interpret them in their actual format control capacity, rendering slurs, beams, and so forth, as appropriate.

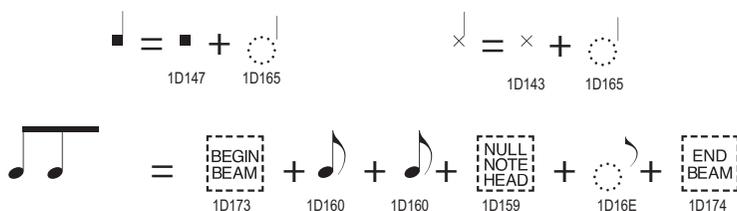
**Precomposed Note Characters.** For maximum flexibility, the character set includes both precomposed note values and primitives from which complete notes may be constructed. The precomposed versions are provided mainly for convenience. However, if any normalization form is applied, including NFC, the characters will be decomposed. For further information, see *Section 3.11, Normalization Forms*. The canonical equivalents for these characters are given in the Unicode Character Database and are illustrated in *Figure 21-2*.

**Figure 21-2.** Precomposed Note Characters



**Alternative Noteheads.** More complex notes built up from alternative noteheads, stems, flags, and articulation symbols are necessary for complete implementations and complex scores. Examples of their use include American shape-note and modern percussion notations, as shown in the first line of *Figure 21-3*.

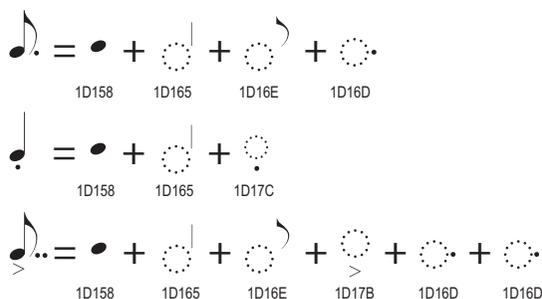
**Figure 21-3.** Alternative Noteheads



U+1D159 MUSICAL SYMBOL NULL NOTEHEAD is a special notehead that has no distinct visual appearance of its own. It can be used as an anchor for a combining flag in complicated musical scoring. For example, in a beamed sequence of notes, the beam might be extended beyond visible notes, as shown in the second line of *Figure 21-3*. Even though the *null notehead* has no visual appearance of its own, it is not a default ignorable code point; some indication of its presence, as for instance a dotted box glyph, should be shown if displayed outside of a context that supports full musical rendering.

**Augmentation Dots and Articulation Symbols.** Augmentation dots and articulation symbols may be appended to either the precomposed or built-up notes. In addition, augmentation dots and articulation symbols may be repeated as necessary to build a complete note symbol. Examples of the use of augmentation dots and articulation symbols are shown in *Figure 21-4*.

**Figure 21-4.** Augmentation Dots and Articulation Symbols



**Ornamentation.** Table 21-1 lists common eighteenth-century ornaments and the sequences of characters from which they can be generated.

**Table 21-1.** Examples of Ornamentation

	1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D1A0 STROKE-6 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A3 STROKE-9
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D1A2 STROKE-8 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19F STROKE-5
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D1A1 STROKE-7 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19F STROKE-5
	1D1A2 STROKE-8 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D1A0 STROKE-6 + 1D19D STROKE-3
	1D19B STROKE-1 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3
	1D19B STROKE-1 + 1D19C STROKE-2 + 1D19C STROKE-2 + 1D19D STROKE-3 + 1D19E STROKE-4
	1D19C STROKE-2 + 1D19D STROKE-3 + 1D19E STROKE-4

**Gregorian.** The *punctum*, or Gregorian *brevis*, a square shape, is unified with U+1D147 MUSICAL SYMBOL SQUARE NOTEHEAD BLACK. The Gregorian *semibrevis*, a diamond or lozenge shape, is unified with U+1D1BA MUSICAL SYMBOL SEMIBREVIS BLACK. Thus Gregorian notation, medieval notation, and modern notation either require separate fonts in practice or need font features to make subtle differentiations between shapes where required.

**Kievan.** Kievan musical notation is a form of linear musical notation found in religious chant books of the Russian Orthodox Church, among others. It is also referred to as East Slavic musical notation. The notation originated in the 1500s, and the first books using Kievan notation were published in 1772. The notation is still used today.

Unlike Western plainchant, Kievan is written on a five-line staff (encoded at U+1D11A) with uniquely shaped notes, and several distinct symbols, including its own C clef and flat signs. U+1D1DF MUSICAL SYMBOL KIEVAN END OF PIECE is analogous to the Western U+1D102 MUSICAL SYMBOL FINAL BARLINE.

Beaming is used in Kievan notation occasionally, and the existing musical format characters encoded between U+1D173 and U+1D17A may be used in implementations of beaming in higher-level protocols.

## 21.3 Byzantine Musical Symbols

### ***Byzantine Musical Symbols: U+1D000–U+1D0FF***

Byzantine musical notation first appeared in the seventh or eighth century CE, developing more fully by the tenth century. These musical symbols are chiefly used to write the religious music and hymns of the Christian Orthodox Church, although folk music manuscripts are also known. In 1881, the Orthodox Patriarchy Musical Committee redefined some of the signs and established the New Analytical Byzantine Musical Notation System, which is in use today. About 95% of the more than 7,000 musical manuscripts using this system are in Greek. Other manuscripts are in Russian, Bulgarian, Romanian, and Arabic.

**Processing.** Computer representation of Byzantine musical symbols is quite recent, although typographic publication of religious music books began in 1820. Two kinds of applications have been developed: applications to enable musicians to write the books they use, and applications that compare or convert this musical notation system to the standard Western system. (See *Section 21.2, Western Musical Symbols.*)

Byzantine musical symbols are divided into 15 classes according to function. Characters interact with one another in the horizontal and vertical dimension. There are three horizontal “stripes” in which various classes generally appear and rules as to how other characters interact within them. These rules, which are still being specified, are the responsibilities of higher-level protocols.

## 21.4 Ancient Greek Musical Notation

### *Ancient Greek Musical Notation: U+1D200–U+1D24F*

Ancient Greeks developed their own distinct system of musical notation, which is found in a large number of ancient texts ranging from a fragment of Euripides' *Orestes* to Christian hymns. It is also used in the modern publication of these texts as well as in modern studies of ancient music.

The system covers about three octaves, and symbols can be grouped by threes: one symbol corresponds to a “natural” note on a diatonic scale, and the two others to successive sharpenings of that first note. There is no distinction between enharmonic and chromatic scales. The system uses two series of symbols: one for vocal melody and one for instrumental melody.

The symbols are based on Greek letters, comparable to the modern usage of the Latin letters A through G to refer to notes of the Western musical scale. However, rather than using a sharp and flat notation to indicate semitones, or casing and other diacritics to indicate distinct octaves, the Ancient Greek system extended the basic Greek alphabet by rotating and flipping letterforms in various ways and by adding a few more symbols not directly based on letters.

**Unification.** In the Unicode Standard, the vocal and instrumental systems are unified with each other and with the basic Greek alphabet, based on shape. *Table 21-2* gives the correspondence between modern notes, the numbering used by modern scholars, and the Unicode characters or sequences of characters to use to represent them.

**Table 21-2.** Representation of Ancient Greek Vocal and Instrumental Notation

Modern Note	Modern Number	Vocal Notation	Instrumental Notation
g <sup>♮</sup>	70	2127, 0374	1D23C, 0374
	69	0391, 0374	1D23B, 0374
	68	0392, 0374	1D23A, 0374
f <sup>♮</sup>	67	0393, 0374	039D, 0374
	66	0394, 0374	1D239, 0374
	65	0395, 0374	1D208, 0374
e <sup>♮</sup>	64	0396, 0374	1D238, 0374
	63	0397, 0374	1D237, 0374
	62	0398, 0374	1D20D, 0374
d <sup>♮</sup>	61	0399, 0374	1D236, 0374
	60	039A, 0374	1D235, 0374
	59	039B, 0374	1D234, 0374
c <sup>♮</sup>	58	039C, 0374	1D233, 0374
	57	039D, 0374	1D232, 0374
	56	039E, 0374	1D20E, 0374
b <sup>♮</sup>	55	039F, 0374	039A, 0374

**Table 21-2.** Representation of Ancient Greek Vocal and Instrumental Notation (Continued)

<b>Modern Note</b>	<b>Modern Number</b>	<b>Vocal Notation</b>	<b>Instrumental Notation</b>
	54	1D21C	1D241
	53	1D21B	1D240
a'	52	1D21A	1D23F
	51	1D219	1D23E
	50	1D218	1D23D
g'	49	2127	1D23C
	48	0391	1D23B
	47	0392	1D23A
f	46	0393	039D
	45	0394	1D239
	44	0395	1D208
e'	43	0396	1D238
	42	0397	1D237
	41	0398	1D20D
d'	40	0399	1D236
	39	039A	1D235
	38	039B	1D234
c'	37	039C	1D233
	36	039D	1D232
	35	039E	1D20E
b	34	039F	039A
	33	03A0	03FD
	32	03A1	1D231
a	31	03F9	03F9
	30	03A4	1D230
	29	03A5	1D22F
g	28	03A6	1D213
	27	03A7	1D22E
	26	03A8	1D22D
f	25	03A9	1D22C
	24	1D217	1D22B
	23	1D216	1D22A
e	22	1D215	0393
	21	1D214	1D205
	20	1D213	1D21C
d	19	1D212	1D229
	18	1D211	1D228
	17	1D210	1D227
c	16	1D20F	0395
	15	1D20E	1D211
	14	1D20D	1D226

**Table 21-2.** Representation of Ancient Greek Vocal and Instrumental Notation (Continued)

Modern Note	Modern Number	Vocal Notation	Instrumental Notation
B	13	1D20C	1D225
	12	1D20B	1D224
	11	1D20A	1D223
A	10	1D209	0397
	9	1D208	1D206
	8	1D207	1D222
G	7	1D206	1D221
	6	1D205	03A4
	5	1D204	1D220
F	4	1D203	1D21F
	3	1D202	1D202
	2	1D201	1D21E
E	1	1D200	1D21D

**Naming Conventions.** The character names are based on the standard names widely used by modern scholars. There is no standardized ancient system for naming these characters. Apparent gaps in the numbering sequence are due to the unification with standard letters and between vocal and instrumental notations.

If a symbol is used in both the vocal notation system and the instrumental notation system, its Unicode character name is based on the vocal notation system catalog number. Thus U+1D20D GREEK VOCAL NOTATION SYMBOL-14 has a glyph based on an inverted capital lambda. In the vocal notation system, it represents the first sharp of B; in the instrumental notation system, it represents the first sharp of d'. Because it is used in both systems, its name is based on its sequence in the vocal notation system, rather than its sequence in the instrumental notation system. The character names list in the Unicode Character Database is fully annotated with the functions of the symbols for each system.

**Font.** Scholars usually typeset musical characters in sans-serif fonts to distinguish them from standard letters, which are usually represented with a serified font. However, this is not required. The code charts use a font without serifs for reasons of clarity.

**Combining Marks.** The combining marks encoded in the range U+1D242..U+1D244 are placed over the vocal or instrumental notation symbols. They are used to indicate metrical qualities.

## 21.5 Duployan

### **Duployan:** U+1BC00–U+1BC9F

The Duployan shorthands are used to write French, English, German, Spanish, and Romanian. The original Duployan shorthand was invented by Emile Duployé, and published in 1860 as a stenographic shorthand for French. It was one of the two most commonly used French shorthands. There are three main English adaptations from the late 19th and early 20th centuries based on Duployan: Pernin, Sloan, and Perrault. None were as popular as the Gregg and Pitman shorthands.

An adaptation and augmentation of Duployan by Father Jean Marie Raphael LeJeune was used as an alternate primary script for several First Nations' languages in interior British Columbia, including Chinook Jargon, Okanagan, Lilloet, Shushwap, and North Thompson. Its original use and greatest surviving attestation is from the *Kamloops Wawa*, a Chinook Jargon newsletter of the Catholic diocese of Kamloops, British Columbia, published 1891–1923. Chinook Jargon was a trade language widely spoken from southeast Alaska to northern California, from the Pacific to the Rockies, and sporadically outside this area. The Chinook script uses the basic Duployan inventory, with the addition of several derived letterforms and compound letters.

**Structure.** Duployan is an uncased, alphabetic stenographic writing system. The model letterforms are generally based on circles and lines. It is a left-to-right script.

The basic inventory of consonant and vowel signs has been augmented over the years to provide more efficient shorthands and has been adapted to the phonologies of languages other than the original French. The Romanian Pernin, Perrault, and Sloan stenographic orthographies add a few letters or letterforms, ideographs, and several combined letters.

The core repertoire of Duployan contains several classes of letters, differentiated primarily by visual form and stroke direction, and nominally by phonetic value. Letter classes include the line consonants (P, T, F, K, and L-type), arc consonants (M, N, J, and S-type), circle vowels (A and O vowels), nasal vowels, and orienting vowels (U/EU, I/E). In addition, the Chinook writing contains spacing letters, compound consonants, and a logograph. The extended Duployan shorthand includes four other letter classes—the complex letters (multisyllabic symbols with consonant forms), and high, low, and connecting terminals for common word endings.

The extended Duployan shorthand includes four other letter classes—the complex letters (multisyllabic symbols with consonant forms), and high, low, and connecting terminals for common word endings. The repertoire also includes U+1BC9D DUPLOYAN THICK LETTER SELECTOR, which modifies a preceding Duployan character by causing it to be rendered bold.

***Shorthand Format Controls: U+1BCA0–U+1BCAF***

Many systems of shorthand use overlapping letters to indicate abbreviations and initialisms. (Initialisms are abbreviations that are pronounced one letter at a time, such as IBM or HTML.) Such non-default text flow may be controlled with the shorthand format controls. U+1BCA0 SHORTHAND FORMAT LETTER OVERLAP indicates a single letter overlap, with the text continuing to flow as if that overlapping character did not exist. U+1BCA1 SHORTHAND FORMAT CONTINUING OVERLAP indicates a continuing overlap where the text flow proceeds from the overlapping character. In Duployan, the overlapping behavior is limited to consonants, circle vowels, and orienting vowels overlapping consonants.

There are two other “step” format controls used with word endings and contractions in specific contexts. U+1BCA2 SHORTHAND FORMAT DOWN STEP indicates downstep, which means that a following character should be rendered below the previous character, with any subsequent joined characters proceeding relative to the lowered glyph. U+1BCA3 SHORTHAND FORMAT UP STEP indicates upstep, which causes the following word or stenographic full stop to be raised.

## 21.6 Sutton SignWriting

### *Sutton SignWriting: U+1D800–U+1DAAF*

Sutton SignWriting is a notational system developed in 1974 by Valerie Sutton and used for the transcription of many sign languages. It is designed to represent physical formations of sign language signs precisely, and is used in a number of publications. More information about the notational system and catalogs of signs can be found on the Sutton SignWriting websites <http://www.signwriting.org/> and <http://www.signbank.org/>.

**Structure.** Sutton SignWriting is a featural writing system, in which visually iconic basic symbols are arranged in two-dimensional layout to form snapshots of the individual signs of a sign language, which are roughly equivalent to words. The Unicode Standard encodes the basic symbols as atomic characters or combining character sequences. The spatial arrangement of the symbols is an essential part of the writing system, but constitutes a higher-level protocol beyond the scope of the Unicode Standard.

**Repertoire.** The repertoire of Sutton SignWriting is comprised of characters for hand-shapes, which are the configurations that the hands take in signing, as well as characters for contact, movement, head and face, body, and location. The repertoire also includes five punctuation marks and twenty characters that indicate fill and rotation.

The head and face characters are used in combining character sequences to represent facial expressions. The character sequences are formed with U+1D9FF SIGNWRITING HEAD as base, followed by nonspacing marks from the ranges U+1DA00..U+1DA36 and U+1DA3B..U+1DA6C. These nonspacing marks represent expressions or movements of the eyes, cheeks, mouth, and so on, and include such characters as U+1DA17 SIGNWRITING EYE BLINK SINGLE and U+1DA3E SIGNWRITING MOUTH SMILE.

**Modifiers.** The fill and rotation characters are nonspacing combining marks that modify a base character to create various realizations of the base character. For example, the handshape U+1D800 SIGNWRITING HAND-FIST INDEX can be modified by a fill character, a rotation character, or both to represent different positions of that handshape and to distinguish between the left and the right hand.

There are five fill modifiers, U+1DA9B SIGNWRITING FILL MODIFIER-2 through U+1DA9F SIGNWRITING FILL MODIFIER-6, and fifteen rotation modifiers, U+1DAA1 SIGNWRITING ROTATION MODIFIER-2 through U+1DAAF SIGNWRITING ROTATION MODIFIER-16. There are no explicit modifiers encoded for *fill-1* or *rotation-1*, as those values are considered inherent in the base character. When both a fill and a rotation modifier are used in a combining character sequence, the fill modifier precedes the rotation modifier in the sequence.

The effect of a fill modifier depends on the character sequence it appears in. For example, when applied to a handshape character such as U+1D800 SIGNWRITING HAND-FIST INDEX, a fill modifier selects one of six possible fills representing as many palm orientations. When applied to a tempo symbol such as U+1D9F7 SIGNWRITING DYNAMIC FAST, a fill modifier alters the shape of the base character. When used in a character sequence such as

<U+1D9FF SIGNWRITING HEAD, U+1DA16 SIGNWRITING EYES CLOSED, *fill*>, the fill modifier selects between one eye and both eyes closed.

The rotation modifiers turn a base character by 45 degree increments. In combination with handshape characters, the rotation modifiers also distinguish between the right and left hand characters. U+1DAA4 SIGNWRITING ROTATION MODIFIER-5 turns a base character by 180 degrees. For a handshape that distinguishes between right and left hand shapes, U+1DAAC SIGNWRITING ROTATION MODIFIER-13 turns the left hand shape 180 degrees.

**Punctuation.** Sutton SignWriting uses five script-specific punctuation marks. These include U+1DA8B SIGNWRITING PARENTHESIS, which represents an opening parenthesis. A closing parenthesis is represented with the sequence <U+1DA8B SIGNWRITING PARENTHESIS, U+1DAA4 SIGNWRITING ROTATION MODIFIER-5>.

