Subject: Proposal to add “Deprecated” property to halfwidth hangul jamo characters

L2/17-118

Date: 2017-05-05 (originally submitted on 2017-04-28)

To: UTC

From: Jaemin Chung

Pages: 5 (13 including appendices)

I propose to add the “Deprecated” property to the following 52 characters.

U+	Char	Name
U+FFA0	ᅟ	HALFWIDTH HANGUL FILLER
U+FFA1	ᄀ	HALFWIDTH HANGUL LETTER KIYEOK
U+FFA2	ᄁ	HALFWIDTH HANGUL LETTER SSANGKIYEOK
U+FFA3	ᆪ	HALFWIDTH HANGUL LETTER KIYEOK-SIOS
U+FFA4	ᄂ	HALFWIDTH HANGUL LETTER NIEUN
U+FFA5	ᆬ	HALFWIDTH HANGUL LETTER NIEUN-CIEUC
U+FFA6	ᆭ	HALFWIDTH HANGUL LETTER NIEUN-HIEUH
U+FFA7	ᄃ	HALFWIDTH HANGUL LETTER TIKEUT
U+FFA8	ᄄ	HALFWIDTH HANGUL LETTER SSANGTIKEUT
U+FFA9	ᄅ	HALFWIDTH HANGUL LETTER RIEUL
U+FFAA	ᄚ	HALFWIDTH HANGUL LETTER RIEUL-KIYEOK
U+FFAB	ᄆ	HALFWIDTH HANGUL LETTER RIEUL-MIEUM
U+FFAC	ᄇ	HALFWIDTH HANGUL LETTER RIEUL-PIEUP
U+FFAD	ᄡ	HALFWIDTH HANGUL LETTER RIEUL-SIOS
U+FFAE	ᄐ	HALFWIDTH HANGUL LETTER RIEUL-THIEUTH
U+FFAF	ᄑ	HALFWIDTH HANGUL LETTER RIEUL-PHIEUPH
U+FFB0	ᄒ	HALFWIDTH HANGUL LETTER RIEUL-HIEUH
U+FFB1	ᄆ	HALFWIDTH HANGUL LETTER MIEUM
U+FFB2	ᄇ	HALFWIDTH HANGUL LETTER PIEUP
U+FFB3	ᄈ	HALFWIDTH HANGUL LETTER SSANGPIEUP
U+FFB4	ᄡ	HALFWIDTH HANGUL LETTER PIEUP-SIOS
U+FFB5	ᄉ	HALFWIDTH HANGUL LETTER SIOS
U+FFB6	ᄊ	HALFWIDTH HANGUL LETTER SSANGSIOS
U+FFB7	ᄋ	HALFWIDTH HANGUL LETTER IEUNG

U+	Char	Name
U+FFB8	ᄌ	HALFWIDTH HANGUL LETTER CIEUC
U+FFB9	ᄍ	HALFWIDTH HANGUL LETTER SSANGCIEUC
U+FFBA	ᄎ	HALFWIDTH HANGUL LETTER CHIEUCH
U+FFBB	ᄏ	HALFWIDTH HANGUL LETTER KHIEUKH
U+FFBC	ᄐ	HALFWIDTH HANGUL LETTER THIEUTH
U+FFBD	ᄑ	HALFWIDTH HANGUL LETTER PHIEUPH
U+FFBE	ᄒ	HALFWIDTH HANGUL LETTER HIEUH
U+FFC2	ᅡ	HALFWIDTH HANGUL LETTER A
U+FFC3	ᅢ	HALFWIDTH HANGUL LETTER AE
U+FFC4	ᅣ	HALFWIDTH HANGUL LETTER YA
U+FFC5	ᅤ	HALFWIDTH HANGUL LETTER YAE
U+FFC6	ᅥ	HALFWIDTH HANGUL LETTER EO
U+FFC7	ᅦ	HALFWIDTH HANGUL LETTER E
U+FFCA	ᅧ	HALFWIDTH HANGUL LETTER YEO
U+FFCB	ᅨ	HALFWIDTH HANGUL LETTER YE
U+FFCC	ᅩ	HALFWIDTH HANGUL LETTER O
U+FFCD	ᅪ	HALFWIDTH HANGUL LETTER WA
U+FFCE	ᅫ	HALFWIDTH HANGUL LETTER WAE
U+FFCF	ᅬ	HALFWIDTH HANGUL LETTER OE
U+FFD2	ᅭ	HALFWIDTH HANGUL LETTER YO
U+FFD3	ᅮ	HALFWIDTH HANGUL LETTER U
U+FFD4	ᅯ	HALFWIDTH HANGUL LETTER WEO
U+FFD5	ᅰ	HALFWIDTH HANGUL LETTER WE
U+FFD6	ᅱ	HALFWIDTH HANGUL LETTER WI
U+FFD7	ᅲ	HALFWIDTH HANGUL LETTER YU
U+FFDA	ᅳ	HALFWIDTH HANGUL LETTER EU
U+FFDB	ᅴ	HALFWIDTH HANGUL LETTER YI
U+FFDC	ᅵ	HALFWIDTH HANGUL LETTER I

Rationale

1. Halfwidth hangul jamo characters are almost never used (or extremely rarely used) in Korean.
   1) For modern hangul, Korean input methods use Hangul Compatibility Jamo (U+3130–U+318F) and precomposed Hangul Syllables (U+AC00–U+D7AF). The characters in these two Unicode blocks are sufficient for representing modern hangul – halfwidth jamo characters are not needed.
      - All the halfwidth jamo characters have compatibility decomposition mappings to Hangul Compatibility Jamo characters.
   2) To begin with, most users do not even know halfwidth jamo characters exist (therefore do not even know how to enter them). Korean input methods do not provide a way to enter halfwidth jamo characters either.
   3) In addition, most Korean fonts do not cover halfwidth jamo characters.

   Note that these halfwidth jamo characters have been included in Unicode since version 1.0 (October 1991). After more than 25 years, there is no use case for these characters.

2. Their source is not well known (or unknown to most people), unlike halfwidth katakana characters (it is very well known that halfwidth katakana characters are from JIS X 0201).
   1) The halfwidth hangul jamo characters in Unicode seem to be from Annex 4 (titled "7-bit hangul jamo character code") of KS C 5601 / KS X 1001 (1987 and its subsequent versions). The arrangement of characters, including unused code points, is identical to what is in U+FFxx.
   2) Xerox Character Code Standard (XCCS) has an almost identical set of characters (including unused code points as well), but has two filler characters instead of one. It refers to a historical and obsolete standard KS C 5601-1982 (titled "8-bit Roman and Korean Character Code").

   And the fact that one needs to look for the source clearly shows that their source is not well known (or unknown to most people).

   See Appendix A for excerpts.

3. They are not covered by any legacy character encoding, unlike halfwidth katakana characters (which are covered by Shift_JIS).

4. Even if there happens to be data stored in a single-byte hangul jamo character encoding,
   1) the jamo characters in that character encoding do not necessarily need to be mapped to halfwidth jamo characters – they can simply be mapped to Hangul Compatibility Jamo.
      (That is, one can simply use ᄒ ᅡ ᄂ ᄀ ᅳ ᄅ instead of 하ᄂ그ᄅ when making a Unicode mapping table of that character encoding.)
   2) in order for the text to be properly displayed and processed, sequences of jamo stored in that character encoding need to be reconverted to precomposed Hangul Syllables anyway.
      (That is, whether the text is first converted to ᄒ ᅡ ᄂ ᄀ ᅳ ᄅ or to 하ᄂ그ᄅ, it needs to be reconverted to 한글 anyway.)

   See Appendix B for a character code conversion program (and this program does not use halfwidth jamo characters at all).

(Document continued on the next page)

5. KS C 5601 / KS X 1001 made an incompatible change in its 7-bit jamo character code after the 1992 version.

Right below the Table 1 (표 1) of Annex 4 (부속서 4) of KS C 5601 / KS X 1001, one can notice the following difference (see Appendix A for excerpts).

In KS C 5601-1989 and in KS C 5601-1992,

> 비고: 4/0 "채움"은 한글낱자 빈자리 표시를 위한 부호로서 필요시 사용할 수 있다.
> (Note: 4/0 "filler" is a code for an indication of an absence of hangul jamo. It may be used when needed.)

In KS X 1001:2004,

> 비고: 4/0 채움 문자의 용법은 부속서 3에 나오는 2바이트 조합형 부호계의 채움 문자 용법과 같다. 다시 말하여 한 글자마디는 세 바이트로 나타내며, 첫소리, 가운뎃소리, 끝소리 글자가 없을 때, 그 자리에 각각 채움 문자를 넣는다.
> (Note: The usage of 4/0 filler character is the same as the usage of the filler character in the 2-byte johab character code in Annex 3. In other words, a single hangul syllable is represented as three bytes, and the filler character is inserted when there is no leading consonant (*choseong*), vowel (*jungseong*), or trailing consonant (*jongseong*).)

The 1989 and 1992 versions do not mandate the use of the filler character at 4/0. However, the 2004 version mandates the use of the filler character whenever there is no jamo. This significant and incompatible change in the usage of the filler character shows that the 7-bit jamo character code in KS C 5601 / KS X 1001 is never used (or extremely rarely used) in practice.

Therefore, even if those halfwith jamo characters were indeed from Annex 4 of KS C 5601 / KS X 1001, they can be safely deprecated without any problems.

(Document continued on the next page)

Proposed changes

1. In the http://www.unicode.org/Public/UNIDATA/PropList.txt file, replace the following entries

```
0149          ; Deprecated # L&       LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
0673          ; Deprecated # Lo       ARABIC LETTER ALEF WITH WAVY HAMZA BELOW
0F77          ; Deprecated # Mn       TIBETAN VOWEL SIGN VOCALIC RR
0F79          ; Deprecated # Mn       TIBETAN VOWEL SIGN VOCALIC LL
17A3..17A4    ; Deprecated # Lo   [2] KHMER INDEPENDENT VOWEL QAQ..KHMER INDEPENDENT VOWEL QAA
206A..206F    ; Deprecated # Cf   [6] INHIBIT SYMMETRIC SWAPPING..NOMINAL DIGIT SHAPES
2329          ; Deprecated # Ps       LEFT-POINTING ANGLE BRACKET
232A          ; Deprecated # Pe       RIGHT-POINTING ANGLE BRACKET
E0001         ; Deprecated # Cf       LANGUAGE TAG

# Total code points: 15
```

with these:

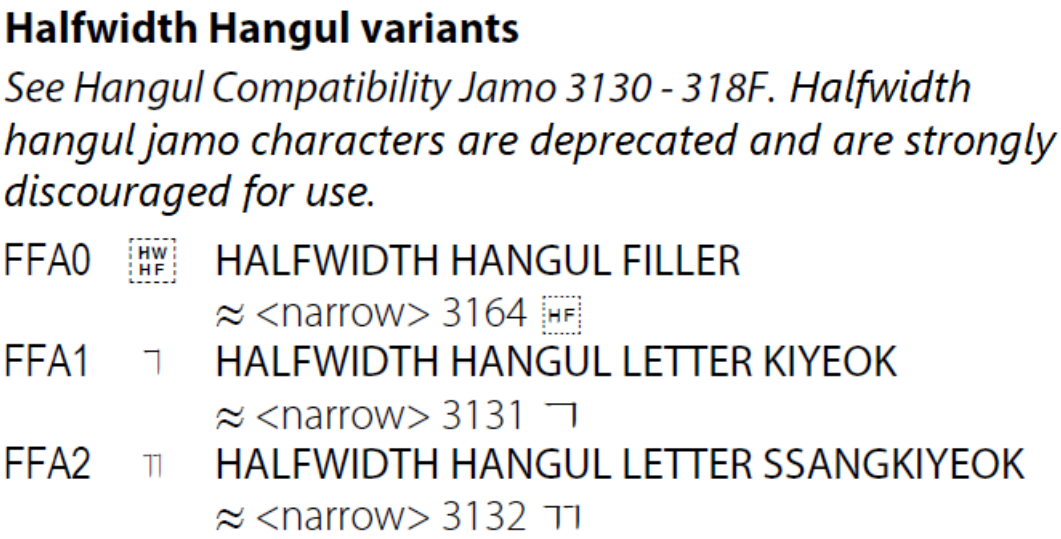
```
0149          ; Deprecated # L&       LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
0673          ; Deprecated # Lo       ARABIC LETTER ALEF WITH WAVY HAMZA BELOW
0F77          ; Deprecated # Mn       TIBETAN VOWEL SIGN VOCALIC RR
0F79          ; Deprecated # Mn       TIBETAN VOWEL SIGN VOCALIC LL
17A3..17A4    ; Deprecated # Lo   [2] KHMER INDEPENDENT VOWEL QAQ..KHMER INDEPENDENT VOWEL QAA
206A..206F    ; Deprecated # Cf   [6] INHIBIT SYMMETRIC SWAPPING..NOMINAL DIGIT SHAPES
2329          ; Deprecated # Ps       LEFT-POINTING ANGLE BRACKET
232A          ; Deprecated # Pe       RIGHT-POINTING ANGLE BRACKET
FFA0..FFBE    ; Deprecated # Lo  [31] HALFWIDTH HANGUL FILLER..HALFWIDTH HANGUL LETTER HIEUH
FFC2..FFC7    ; Deprecated # Lo   [6] HALFWIDTH HANGUL LETTER A..HALFWIDTH HANGUL LETTER E
FFCA..FFCF    ; Deprecated # Lo   [6] HALFWIDTH HANGUL LETTER YEO..HALFWIDTH HANGUL LETTER OE
FFD2..FFD7    ; Deprecated # Lo   [6] HALFWIDTH HANGUL LETTER YO..HALFWIDTH HANGUL LETTER YU
FFDA..FFDC    ; Deprecated # Lo   [3] HALFWIDTH HANGUL LETTER EU..HALFWIDTH HANGUL LETTER I
E0001         ; Deprecated # Cf       LANGUAGE TAG

# Total code points: 67
```

2. In the code chart, add the following sentence under the subhead “Halfwidth Hangul variants”:

    Halfwidth hangul jamo characters are deprecated and are strongly discouraged for use.

The code chart should look something like this:



(End of document, without appendices)

Appendix A: Excerpts

(All of these scans are from Dr. Ken Lunde.)

1. Excerpt from KS C 5601-1989

부속서 4 7단위 한글낱자 부호

1. 제정 목적 현재 난립해 있는 여러종류의 7단위 한글낱자 부호들이 한 종류로 통일되도록 유도하기 위하여 7단위 한글낱자 부호의 권장안을 제시한다.

2. 적용 범위 이 권장안은 정보교환용 부호를 내부적으로 사용하지 못하는 시스템과 관련장비에서 사용하는 한글 부호의 표현형식을 나타내며, 정보교환용으로 사용하지 않는 것을 원칙으로 한다.

3. 부 호 계

(1) 7단위 한글낱자 부호는 한글 자모음자 1문자를 표현한다.

(2) 한글낱자 부호는 **부속서4 표1**에 따른다.

(3) 로마문자와 한글 병용시에는 **KS C 5636** 의 기능 문자 부호인 SI 및 SO를 사용하여 구분한다. 이 경우 SI에 연속되는 부호군은 로마문자용 부호를, SO에 연속되는 부호군은 **부속서4 표1의** 한글낱자 부호를 뜻하는 것으로 한다.

(4) 부호표상의 위치는 부호표 내에서의 "열 번호/행 번호"로서 표시한다.

부속서4 표1 7단위 한글 자모용 부호

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1	열 / 행	0	1	2	3	4	5	6	7
0	0	0	0	0					(채움)	ㄹㅎ		
0	0	0	1	1					ㄱ	ㅁ		
0	0	1	0	2					ㄲ	ㅂ	ㅏ	ㅛ
0	0	1	1	3					ㄱㅅ	ㅃ	ㅐ	ㅜ
0	1	0	0	4					ㄴ	ㅂㅅ	ㅑ	ㅝ
0	1	0	1	5					ㄴㅈ	ㅅ	ㅒ	ㅞ
0	1	1	0	6	기				ㄴㅎ	ㅆ	ㅓ	ㅟ
0	1	1	1	7	능				ㄷ	ㅇ	ㅔ	ㅠ
1	0	0	0	8	문				ㄸ	ㅈ		
1	0	0	1	9	자				ㄹ	ㅉ		
1	0	1	0	10					ㄹㄱ	ㅊ	ㅖ	ㅡ
1	0	1	1	11					ㄹㅁ	ㅋ	ㅚ	ㅢ
1	1	0	0	12					ㄹㅂ	ㅌ	ㅗ	ㅣ
1	1	0	1	13					ㄹㅅ	ㅍ	ㅘ	
1	1	1	0	14					ㄹㅌ	ㅎ	ㅙ	
1	1	1	1	15					ㄹㅍ		ㅟ	

비 고 4/0 "채움"은 한글낱자 빈자리 표시를 위한 부호로서 필요시 사용할 수 있다.

부속서 4 7단위 한글낱자 부호

1. 제정 목적 현재 난립해 있는 여러종류의 7단위 한글낱자 부호들이 한 종류로 통일되도록 유도하기 위하여 7단위 한글낱자 부호의 권장안을 제시한다.

2. 적용 범위 이 권장안은 정보교환용 부호를 내부적으로 사용하지 못하는 시스템과 관련장비에서 사용하는 한글 부호의 표현형식을 나타내며, 정보교환용으로 사용하지 않는 것을 원칙으로 한다.

3. 부 호 계

(1) 7단위 한글낱자 부호는 한글 자모음자 1문자를 표현한다.

(2) 한글낱자 부호는 **부속서4 표1**에 따른다.

(3) 로마문자와 한글 병용시에는 **KS C 5636**의 기능 문자 부호인 SI 및 SO를 사용하여 구분한다. 이 경우 SI에 연속되는 부호군은 로마문자용 부호를, SO에 연속되는 부호군은 **부속서4 표1**의 한글낱자 부호를 뜻하는 것으로 한다.

(4) 부호표상의 위치는 부호표 내에서의 "열 번호/행 번호"로서 표시한다.

부속서4 표1 7단위 한글 자모용 부호

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1	열 / 행	0	1	2	3	4	5	6	7
0	0	0	0	0					(채움)	ㄹㅎ		
0	0	0	1	1					ㄱ	ㅁ		
0	0	1	0	2					ㄱㄱ	ㅂ	ㅏ	ㅛ
0	0	1	1	3					ㄱㅅ	ㅂㅂ	ㅐ	ㅜ
0	1	0	0	4					ㄴ	ㅂㅅ	ㅑ	ㅝ
0	1	0	1	5					ㄴㅈ	ㅅ	ㅒ	ㅞ
0	1	1	0	6	기				ㄴㅎ	ㅆ	ㅓ	ㅟ
0	1	1	1	7	능				ㄷ	ㅇ	ㅔ	ㅠ
1	0	0	0	8	문				ㄸ	ㅈ		
1	0	0	1	9	자				ㄹ	ㅉ		
1	0	1	0	10					ㄹㄱ	ㅊ	ㅖ	ㅡ
1	0	1	1	11					ㄹㅁ	ㅋ	ㅒ	ㅢ
1	1	0	0	12					ㄹㅂ	ㅌ	ㅗ	ㅣ
1	1	0	1	13					ㄹㅅ	ㅍ	ㅘ	
1	1	1	0	14					ㄹㅌ	ㅎ	ㅙ	
1	1	1	1	15					ㄹㅍ		ㅚ	

비 고 4/0 "채움"은 한글낱자 빈자리 표시를 위한 부호로서 필요시 사용할 수 있다.

부속서 4 7비트 한글 낱자 부호계

1. 제정 목적 현재 난립해 있는 여러 종류의 7비트 한글 낱자 부호계가 한 종류로 통일되도록 유도하기 위하여, 7비트 한글 낱자 부호계의 권장안을 제시한다.

2. 적용 범위 이 권장안은 정보 교환용 부호계를 내부적으로 사용하지 못하는 시스템과 관련 장비에서 사용하는 한글 부호계의 표현 형식을 나타내며, 정보 교환용으로 사용하지 않는 것을 원칙으로 한다.

3. 부호계

a) 7비트 한글 낱자 부호값은 한글 낱자 1자를 표현한다.

b) 한글 낱자 부호계는 **부속서 4 표 1**에 따른다.

c) 로마 문자와 한글 병용시에는 **KS X 1003(KS C 5636)**의 기능 문자 부호인 SI 및 SO를 사용하여 구분한다. 이 경우 SI에 연속되는 부호값군은 로마 문자용 부호값을, SO에 연속되는 부호값군은 **부속서 4 표 1**의 한글 낱자 부호값을 뜻하는 것으로 한다.

d) 부호표상의 위치는 부호표 내에서의 "열 번호 / 행 번호"로 표시한다.

부속서 4 표 1 7비트 한글 낱자 부호계

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1	열 / 행	0	1	2	3	4	5	6	7
0	0	0	0	0	기능 문자				(채움)	ㅀ		
0	0	0	1	1					ㄱ	ㅁ		
0	0	1	0	2					ㄲ	ㅂ	ㅏ	ㅛ
0	0	1	1	3					ㄳ	ㅃ	ㅐ	ㅜ
0	1	0	0	4					ㄴ	ㅄ	ㅑ	ㅝ
0	1	0	1	5					ㄵ	ㅅ	ㅒ	ㅞ
0	1	1	0	6					ㄶ	ㅆ	ㅓ	ㅟ
0	1	1	1	7					ㄷ	ㅇ	ㅔ	ㅠ
1	0	0	0	8					ㄸ	ㅈ		
1	0	0	1	9					ㄹ	ㅉ		
1	0	1	0	10					ㄺ	ㅊ	ㅕ	ㅡ
1	0	1	1	11					ㄻ	ㅋ	ㅖ	ㅢ
1	1	0	0	12					ㄼ	ㅌ	ㅗ	ㅣ
1	1	0	1	13					ㄽ	ㅍ	ㅘ	
1	1	1	0	14					ㄾ	ㅎ	ㅙ	
1	1	1	1	15					ㄿ		ㅚ	

비 고 4/0 채움 문자의 용법은 **부속서 3**에 나오는 2바이트 조합형 부호계의 채움 문자 용법과 같다. 다시 말하여 한 글자마다는 세 바이트로 나타내며, 첫소리, 가운뎃소리, 끝소리 글자가 없을 때, 그 자리에 각각 채움 문자를 넣는다.

Identifier Octal	Dec	Hex	Shape	Character description — Character set 342_8: Phonetic
333_8	219	DB	j	Voiced palatal fricative or approximant
334_8	220	DC	ʎ	Palatal lateral
335_8	221	DD	ŋ	Velar nasal
336_8	222	DE	k	Voiceless velar plosive
337_8	223	DF	g	Voiced velar plosive
340_8	224	E0	x	Voiceless velar fricative
341_8	225	E1	ɣ	Voiced velar fricative
342_8	226	E2	ɰ	Velar approximant
343_8	227	E3	g	Velar implosive
344_8	228	E4	ɴ	Uvular nasal
345_8	229	E5	ɳ	Japanese syllabic nasal (obsolete)
346_8	230	E6	q	Voiceless uvular plosive
347_8	231	E7	ɢ	Voiced uvular plosive
350_8	232	E8	χ	Voiceless uvular fricative
351_8	233	E9	ʁ	Voiced uvular fricative
352_8	234	EA	ʀ	Uvular trill, tap, or flap
353_8	235	EB	ħ	Voiceless pharyngeal fricative
354_8	236	EC	ʕ	Voiced pharyngeal fricative
355_8	237	ED	ʔ	Glottal plosive
356_8	238	EE	h	Voiceless glottal fricative
357_8	239	EF	ɦ	Voiced glottal fricative

Character Set $343_8 = 227_{10} = E3_{16}$: Korean Hangul

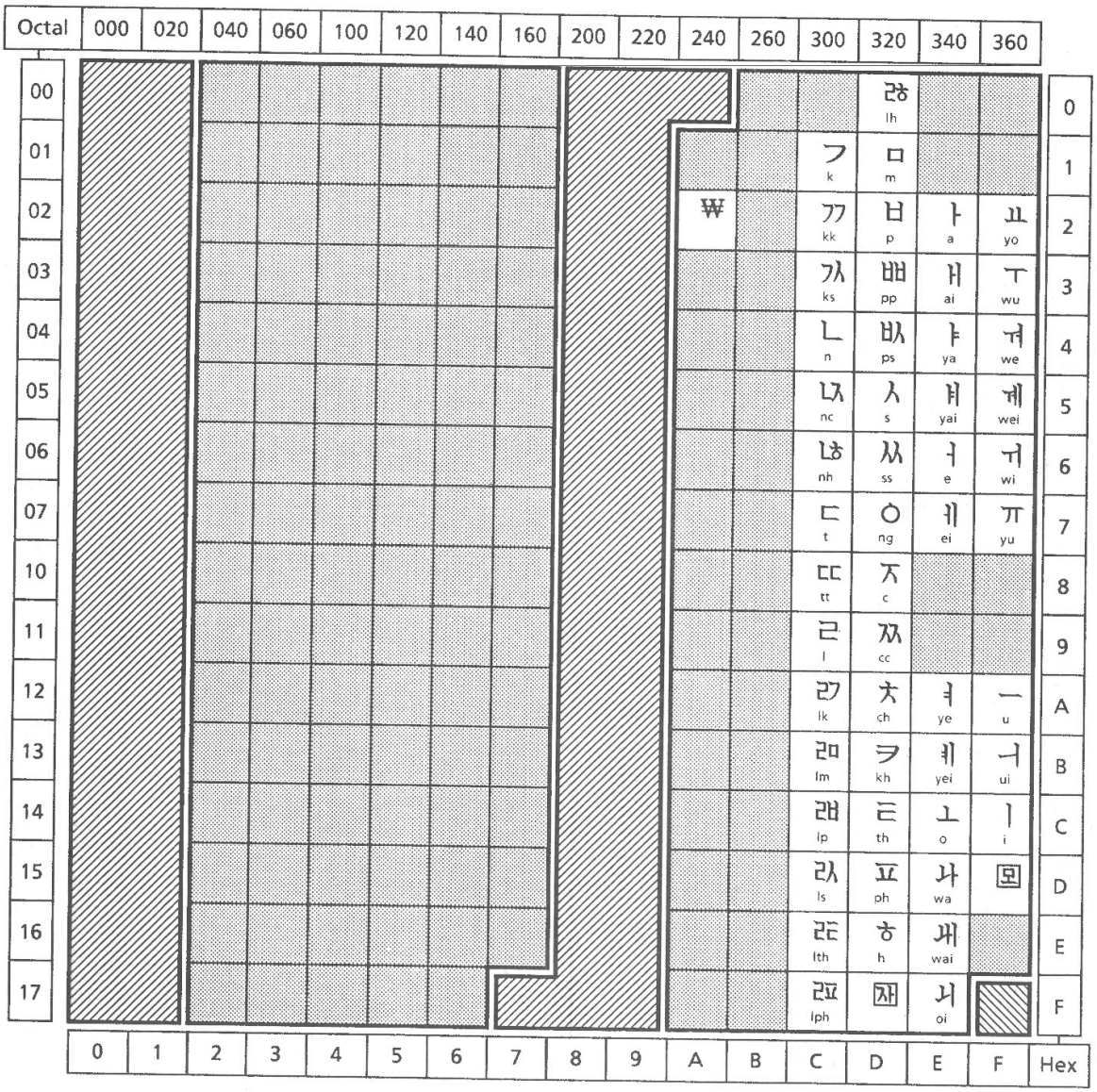
Character Set 343_8 contains characters defined in the Korean Standard KS C 5601 - 1982, 8-bit Roman and Korean Character Codes [27]. The characters selected from the Korean Hangul alphabet are retained in the same order and code positions as in the national standard.

The following are character codes (low-order byte) within Character Set 343_8 (see reference charts in appendix B):

Identifier Octal	Dec	Hex	Shape	Character description — Character set 343_8: Korean
242_8	162	A2	₩	Korean Wen (Won) sign
301_8	193	C1	ㄱ	Korean letter K
302_8	194	C2	ㄲ	Korean letter KK
303_8	195	C3	ㄳ	Korean letter KS
304_8	196	C4	ㄴ	Korean letter N
305_8	197	C5	ㄵ	Korean letter NC
306_8	198	C6	ㄶ	Korean letter NH
307_8	199	C7	ㄷ	Korean letter T
310_8	200	C8	ㄸ	Korean letter TT
311_8	201	C9	ㄹ	Korean letter L
312_8	202	CA	ㄺ	Korean letter LK
313_8	203	CB	ㄻ	Korean letter LM
314_8	204	CC	ㄼ	Korean letter LP
315_8	205	CD	ㄽ	Korean letter LS
316_8	206	CE	ㄾ	Korean letter LTH
317_8	207	CF	ㄿ	Korean letter LPH
320_8	208	D0	ㅀ	Korean letter LH
321_8	209	D1	ㅁ	Korean letter M
322_8	210	D2	ㅂ	Korean letter P
323_8	211	D3	ㅃ	Korean letter PP
324_8	212	D4	ㅄ	Korean letter PS
325_8	213	D5	ㅅ	Korean letter S
326_8	214	D6	ㅆ	Korean letter SS
327_8	215	D7	ㅇ	Korean letter NG
330_8	216	D8	ㅈ	Korean letter C
331_8	217	D9	ㅉ	Korean letter CC
332_8	218	DA	ㅊ	Korean letter CH
333_8	219	DB	ㅋ	Korean letter KH
334_8	220	DC	ㅌ	Korean letter TH
335_8	221	DD	ㅍ	Korean letter PH
336_8	222	DE	ㅎ	Korean letter H
337_8	223	DF	ㅐ	Null Korean Hangul Consonant
342_8	226	E2	ㅏ	Korean letter A
343_8	227	E3	ㅑ	Korean letter AI

Identifier Octal	Dec	Hex	Shape	Character description — Character set 343_8: Korean
344_8	228	E4	ㅑ	Korean letter YA
345_8	229	E5	ㅒ	Korean letter YAI
346_8	230	E6	ㅓ	Korean letter E
347_8	231	E7	ㅔ	Korean letter EI
352_8	234	EA	ㅕ	Korean letter YE
353_8	235	EB	ㅖ	Korean letter YEI
354_8	236	EC	ㅗ	Korean letter O
355_8	237	ED	ㅘ	Korean letter WA
356_8	238	EE	ㅙ	Korean letter WAI
357_8	239	EF	ㅚ	Korean letter OI
362_8	242	F2	ㅛ	Korean letter YO
363_8	243	F3	ㅜ	Korean letter WU
364_8	244	F4	ㅝ	Korean letter WE
365_8	245	F5	ㅞ	Korean letter WEI
366_8	246	F6	ㅟ	Korean letter WI
367_8	247	F7	ㅠ	Korean letter YU
372_8	250	FA	ㅡ	Korean letter U
373_8	251	FB	ㅢ	Korean letter UI
374_8	252	FC	ㅣ	Korean letter I
375_8	253	FD	ㅱ	Null Korean Hangul Vowel

XEROX
Character Set 343_8
Korean Hangul



Appendix B: KS 7-bit hangul jamo character code conversion program (for Windows)

Open the Attachments panel of this PDF document.

This program (under the GNU GPL) is originally made by June-Yub Lee in the early 1990s, and is available at the following URL:

http://ftp.kaist.ac.kr/hangul/incoming/hcode2.1-mailpatch3.tar.gz

I made the following changes before compiling for Windows.

- In `hcode.c`: replaced `#include <strings.h>` with `#include <string.h>` (line 17)
- In `mail.c`: replaced `strncasecmp` with `strnicmp` (7 occurrences; lines 175, 176, 189, 198, 213, 508, and 512)
- Made another version of `h3Bcode.h` (also attached in this document)

1. In the Attachments panel of this document, right-click on `hcode1.txt` or `hcode2.txt`, click on "Save Attachment…" and save it to any folder. After that, change the file extension from `txt` to `exe`.

Difference between `hcode1` and `hcode2`:

- `hcode1` maps a compound jamo to a single code point (e.g. ᅫ to `0x6E`, ᇹ to `0x50`).
- `hcode2` maps a compound jamo (except ㄲ, ㄸ, ㅃ, ㅆ, ㅉ, ㅐ, ㅒ, ㅔ, and ㅖ) to a sequence of two code points (e.g. ᅫ to `0x6C 0x63` (ㅗ + ㅐ), ᇹ to `0x49 0x5E` (ㄹ + ㅎ)). The original version is `hcode2`.

2. Open Command Prompt (`cmd`), and use the `cd` command to go to the folder where the conversion program is.

3. To convert KS 7-bit hangul jamo character code to EUC-KR or Johab (note that this is a very old program – it does not support Unicode encodings), use the following commands:

- To EUC-KR: `hcode# -nk (input_file_name) (output_file_name)`
- To Johab: `hcode# -nt (input_file_name) (output_file_name)`

(`#` is either `1` or `2`, depending on the version you are using)

Note: When converting to EUC-KR, this program uses 8-byte sequences (hangul filler (`0xA4D4`) + three jamo characters) for hangul syllables that are outside of the 2350 in KS X 1001 proper (e.g. [filler] ㅇ ㅒ ㅆ for 얬).

4. When opening the output file (or to convert it to Unicode), select (or use) code page 949 for EUC-KR, and select (or use) code page 1361 for Johab.

In addition, to convert EUC-KR 8-byte sequences to precomposed hangul syllables in Unicode (e.g. [filler] ㅇ ㅒ ㅆ → 얬), open the converted EUC-KR file with Firefox (https://www.mozilla.org/firefox/), press F10 to display the menu bar, go to View > Text Encoding and click Korean.

(End of document, with appendices)