ISO/IEC JTC1/SC2/WG2 N 3475

Date: **2008-04-24**

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: **Disposition of comments**

Title: **Disposition of comments on SC2 N 3943 (FPDAM text for Amendment 5 to ISO/IEC 10646:2003)**

Source: **Michel Suignard (project editor)**
Project: JTC1 02.10646.00.05
Status: For review by WG2
Date: 2008-04-24
Distribution: WG2
Reference: SC2 N3982, N4004, WG2 N3389, N3413
Medium: Paper, PDF file

Comments were received from the following P-Members: China, Ireland, Japan, Korea (ROK), UK, and USA, and from the following O-Member: Sri Lanka. The following document is the disposition of those comments. The disposition is organized per country.

Note – The full content of the ballot comments have been included in this document to facilitate the reading. The dispositions are inserted between these comments and are marked in **Underlined Bold Serif text,** *with explanatory text in italicized serif***.**

As a result of these dispositions, Ireland, Japan, Korea (ROK), United Kingdom, and U.S.A. have changed their vote to Yes, meaning that all negative votes have been accommodated.

China, Positive with comments

China votes YES to SC2n3982 with the following comments on script TAI THAM:

T1. Names of digits

China requires to change character names of U+1A80 through U+1A89. The names should be changed to TAI THAM HORA DIGIT ZERO … NINE, as agreed on ad hoc in Chiang Mai early this year. (WG2n3379 Item 2)

Accepted

(Note that the WG2 document number reference above has been changed from 3389 to 3379 as clearly this was a typo.)

This comment is also implicitly supported by the comment T.3 from Ireland, T.2 from UK, and the comment T.8a) from USA. It is also understood and that was confirmed during the ballot resolution meeting that document WG2 N3379, especially the chart pages with representative glyphs and accompanying character names determines the consensus view of all interested parties.

T2. Fonts

China requires to change fonts to Khuen/Lue style, as agreed on ad hoc in Chiang Mai early this year. (WG2n3379 Item 4)

Accepted

It is the understanding of the editor that the Khuen/Lue style has been used for all iterations of the various amendments proposal including the Tai Tham script. So there is no need to change the font style. The ad-hoc simply re-iterates and confirms the current preference. Note the specific errata in comment T.4.

T3. Moving some characters

China requires to move U+1A56 to U+1A50, move U+1A50 through U+1A55 to U+1A51 through U+1A56, as agreed on ad hoc in Chiang Mai early this year. (WG2n3379 Item 6)

Accepted

Note that the Item 6 in N3379 describes a position move, but not the final position because consequent items in the same document reshuffle the code positions. The final positions of these characters in the chart section of that document are 1A4C-1A52.

T4. Representative Glyph for U+1A31

China requires to change the Representative Glyph for U+1A31 to Khuen/Lue style, as agreed on ad hoc in Chiang Mai early this year. (WG2n3379 Item 7)

Accepted

Note the character in question is TAI THAM LETTER HIGH RATHA and is now proposed at code position 1A2E.

T5. Remove Language Specific Letters

China requires to remove some language specific letters, as agreed on ad hoc in Chiang Mai early this year. (WG2n3379 Item 8)

Accepted

That item 8 proposes the removal of four letters (1A28, 1A2B, 1A2D, and 1A4F) and the renaming of the former 1A29 (now moved to 1A28 because of the removal of 1A28) to TAI THAM LETTER HIGH CHA

For details, read WG2n3379 please

Note that WG2 N3379 has the following additional requests:

Item 9: Addition of a new TAI THAM CONSONANT SIGN MA at position 1A5C (after reshuffle)

Item 10a: Removal of 1A5C TAI THAM SIGN KHUEN MAI KANG LAI and close the gap, change the glyph of the remaining TAI THAM SIGN MAI KANG LAI (before at 1A5B, now 1A58 after reshuffle) to the shape of the Northern Thai font style similar character described at position 1A5A in page 14 of WG2 N3207R (document N3379 is relatively imprecise there).

Item 10b: addition of a new TAI THAM SIGN KHUEN-LUE KARAN at 1A7C with the glyph shape of the removed 1A5B TAI THAM SIGN MAI KANG LAI (mentioned incorrectly as being located at 1A58 in FPDAM5 by N3379).

Item 11: Move 1A7E TAI THAM LETTER GREAT SA after TAI THAM LETTER LAE, meaning that it moves to 1A54.

Item 12: Rename TAI THAM SIGN LOLA HOY KONGNOT to TAI THAM CONSONANT SIGN LA TANG LAI now located at 1A57 after reshuffle.

Not explicitly mentioned in N3379 is the move to Consonants from Independent vowels for TAI THAM LETTER A and TAI THAM LETTER LOW HA which is reflected in the name list. However, after consultation with experts, this was confirmed as being correct. .

Ireland, Negative

Ireland **disapproves** the draft with the technical and editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments

T1. Page 9, Row 0D8, Sinhala.
Following on from discussions with members of the Sri Lankan standards body regarding a need for further consultation and study of the Sinhala digits, Ireland requests that the characters under ballot in columns 0D8E and 0D8F be removed from FPDAM 5 and placed on a subsequent ballot.
Accepted
Also supported by comment T.7 from USA and letter from the Sri Lanka Standards Institution. (Note that column 0D8E and 0D8F are understood as being 0DEx and 0DFx respectively, based on context)

T2. Page 14, Row 110: Hangul Jamo.
Ireland requests that the character name at U+11FD HANGUL JONGSEONG KIYEOK-KHIEUK be corrected to HANGUL JONGSEONG KIYEOK KHIEUKH since the final -H is missing
Accepted
Also requested by comment T.5 from USA. See also comment T.5..

T.3 Page 16, Row 1A2: Tai Tham.
With reference to ISO/IEC JTC1/SC2/WG2 N3379 “Tai Tham Ad-hoc Meeting Report”, Ireland requests that the character set and code table in FPDAM 5 be replaced by that in N3379
Accepted
See disposition from Chinese comments for further discussion and details.

T.4 Page 16, Row 1A2:
Tai Tham. With reference to ISO/IEC JTC1/SC2/WG2 N3384 “Tai Tham Subjoined Variants”, Ireland requests that U+1A5D TAI THAM CONSONANT SIGN BA and U+1A5E TAI THAM CONSONANT SIGN SA be added to FPDAM 5.
Accepted
Also requested by comment T.2 from UK and T.8b) from USA.

T.5 Page 14, Row 110: Hangul Jamo Extended-A.
Ireland requests that the character names at U+A96E HANGUL CHOSEONG RIEUL-KHIEUK and U+A973 HANGUL CHOSEONG PIEUP-KHIEUK be corrected to HANGUL CHOSEONG RIEUL-KHIEUKH and HANGUL CHOSEONG PIEUP-KHIEUKH respectively since the final -H is missing.
Accepted
Also requested by comment T.5 from USA.

T.6 Page 40, Row 10B0: Avestan.
Ireland reiterates its support for the retention of U+10B38 AVESTAN SEPARATION POINT, a character distinct from FULL STOP which is also used in Geldner’s edition of the Avesta. We have reviewed N3336 but do not agree that workarounds like lowering U+2E37 are the correct way to

encode the character in question. The complaint about the baseline drawing in N3197 is not well-made; those lines were indicative and were constructed by shifting out-of-text graphic elements (two horizontal hairlines) around in Quark XPress. The AVESTAN SEPARATION POINT does sit on the baseline; there is no real "discrepancy between the two documents regarding the location" of the character. We understand that the U.S. National Body's position is that there are too many dots in the standard; we maintain our view that script-specific punctuation is preferable, in general, to over-unification. Ireland stands by its comments to PDAM 5 on this point, which is in accord with the comments of Professor Jost Gippert, one of the specialists who worked on the encoding of Avestan. We would not agree to remove this character from FPDAM 5.

Withdrawn

See US comment T.6.

Editorial comments

E1. Page 12, Row 110: Hangul Jamo.

Ireland notes discrepancies between the font used in Unicode 5.0 and the font used in FPDAM 5. For instance, U+110C HANGUL CHOSEONG CIEUC and U+119E HANGUL JUNGSEONG ARAEA differ between them. Ireland requests that WG2 work with the Korean delegates, the Unicode Liaison, and other interested National Bodies in determining the preferred shapes for this and similar characters

Accepted

The differences are indeed visible. After further study, it is clear that Unicode 4.0 chart and ISO/IEC 10646:2003 share the same glyph set for the Hangul Jamos. Unicode 5.0 has a new glyph set and Amendment 5 is a variant of the Unicode 5.0 set. This makes the amendment 5 charts significantly different from the original 10646:2003 and introduces many more inconsistencies with other Jamo representations in the standard (Hangul Compatibility Jamos in 3130-316F, Jamo symbols in the 32xx block, Halfwidth Hangul variants in FFxx). This will be fixed in the next phase of this amendment.

E2. Page 24, Row A96: Hangul Jamo Extended-A.

The same comment applies here as in E1.

Accepted

E3. Page 24, Row A96: Hangul Jamo Extended-B.

The same comment applies here as in E1.

Accepted

E4. Page 48, Row 1300: Egyptian Hieroglyphs.

Ireland notes that the font is missing in the annotation for U+1309D EGYPTIAN HIEROGLYPH D036, and U+1313F EGYPTIAN HIEROGLYPH G001. There is also an extra comma in the note at U+133ED EGYPTIAN HIEROGLYPH Z004.

Accepted

As a result of this disposition, Ireland changes its vote to Yes.

Japan, Negative

Japan disapproves SC2 N3982 (ISO/IE10646 FPDAM 5) with the following comments.
Japan will change its vote if those comments are addressed appropriately.

Technical comments

JPT1-ARIB CJK Ideographs:
Japan supports the consensus on the IRG #29 meeting regarding so-called ARIB ideographs. Take a character, J_ARIB-757D, currently on FA6D, out of the compatibility ideographs and move it into the unified ideographs. See IRG document N 1347R for details. The amendment text and related files should be updated appropriately. (Also see JPE1.)
Accepted
There is a consensus to move J_ARIB-757D from FA6D into a location attributed to unified ideographs. For further details and processing of the remaining 5 proposed CJK characters see disposition of comment T.1 from UK.

JPT2: Reference to Unicode 5.2
The draft amendment text contains several references to the Unicode Standard 5.2, including the collection 309. Japan is not sure it will be published on time. Japan wants to know the plans and schedules of Unicode 5.2 publishing from the Unicode Consortium. Japan wants to remove those references from the amendment texts unless a satisfactory plan is provided.
Accepted in principle
It has been common practice for many amendments to do advance planning concerning synchronization with Unicode Standard version. In fact amendments 1, 2, and 4 have such reference to Unicode Standard versions. To allow parallel processing and to keep the two standard synchronized it is necessary to forecast future releases of Unicode. As Unicode planning firms up, the related text in the amendments is updated. For example, for a while it was thought that Unicode 5.1 would coincide with Amendment 3. At some point it became clear that Unicode 5.1 would instead synchronize with Amendment 4, therefore the references were moved from Amendment 3 to Amendment 4.

In the context of Amendment 5, a similar event occurred and at this point, it is clear that Unicode 5.2 will be synchronized with Amendment 6 instead; therefore references to Unicode 5.2 will be moved from Amendment 5 to Amendment 6.

JPT3: New Chart format
On the right column of page 3, there is a new text for 34.2. The third item reads "Normative character alias (one preceded by "※") which is a formal, unique, and stable alternate name for a character." The current sentence seems redundant for a normative clause. Considering these items are lead by a sentence "The following information items are normative" and a NOTE follows it, just a simple list item text such as "Character alias preceded by ※" is enough.
Accepted in principle
The text will be changed to remove all qualifiers related to normative or formal. The bullet point will read:

- Character alias (one preceded by "※") which is a unique and stable alternate name for a character.

Editorial comments

JPE1: On the left column of page 5, regarding the part of the code chart for new CJK UNIFIED IDEOGRAPHS (9FC4 and 9FC5), the current chat uses low quality images of the new characters. Use TrueType font supplied by Japan to typeset them. (Also see JPT1.)
Accepted

JPE2: On the left column of page 3, there is a new text for 27.3, "However, CJK Unified Ideographs Extension C uses a different format..." The world "However" seems inappropriate here and should be removed.
Accepted

Other comment:

JPO1: On the right column of page 2 to the next page, there is an "Editor's note". It is unsure what it means. Japan wants to see clearer explanation and proposed changes.
Accepted in principle
The editor's note covers two parts:

1. *First section, covering non controversial changes implied by the WG2 resolutions mentioned above (M51.9, 51.10, and 51.11) including the transition to the full numeric reference to the KangXi dictionary. For example, it was found that virtual references could not be used as normative references. Because these changes were not explicitly called for by the resolution, the editor felt it was preferable to mention them in the amendment itself.*
2. *Second section, covering controversial issues concerning KangXi referencing. This should be fixed before publication. Note the US comment T.4 in that effect.*

The following are all characters for which there is a difference between the Super CJK Database Version 14 (and Unihan as in Unicode 5.0) and the information that was provided to the editor for publication in Amendment5.

Code position	New source Info	SuperCJK 14 / UniHan	Effect on FPDAM5	Consensus
03A6E 擧	G_KX0463.211	G_KX 0463.030	Virtual, removed	No, transient was 1062.441
22183 嚮	G_KX0338.240	G_KX 0338.231	0338.240	Yes
221CF 融	G_KX0342.180	G_KX 0343.031	0342.180	Yes
22A8F 扣	G_KX0420.160	G_KX 0420.170	0420.160	No, see KX source for 6287
2304A 敝	G_KX0477.061	G_KX 0477.020	Virtual, removed	No
23057 斁	G_KX0477.141	G_KX 0477.140	Virtual, removed	No
2305C 數	G_KX0477.181	G_KX 0477.160	Virtual, removed	No
23063 斠	G_KX0477.201	G_KX 0477.200	Virtual, removed	No
24799 狥	G_KX0712.071	G_KX 0711.110	Virtual, removed	Yes
2605F 緐	G_KX0929.190	G_KX 0929.290	0929.190	Yes
26B20 艹	G_KX0354.030	G_KX 1022.060	0354.030	No

For the 7 references where there is no consensus, the original SuperCJK14 values will be restored, resulting in the Editor's note in FPDAM5 being replaced with the following editor's instruction in FDAM5:

> *In addition to the changes mentioned above, the CJK G_KX source references for 3C08, 3DD7, and 21CED are replaced by G_HX because the KX entries for these characters are virtual and therefore cannot be used as source reference.*

Compared to FPDAM5, this results in the replacement of entries for:

- *3ACE (KX source added back),*
- *2304A, 23057, 2305C, 23063, and 24799 (KX source added back for these 5 characters),*
- *26B20 (KX source value changed).*

As a result of this disposition, Japan changes its vote to Yes.

Korea (ROK), Negative

XX_ DISAPPROVAL OF THE DRAFT FOR REASONS ON THE ATTACHED

_XX_ Acceptance of these reasons and appropriate changes in the text will change our vote to approval

Technical comments.

T1. p. 1, right column

T1.1 Change as follows:

1) (syllable-initial or initial consonant)

--> (syllable-initial character or initial consonant)

2) (syllable-peak or medial vowel)

--> (syllable-peak character or medial vowel)

3) (syllable-final or final consonant)

--> (syllable-final character or final consonant)

T1.2 Rationale:

1) Since "syllable-initial (-final)" implies consonantal,

- "syllable-initial character" seems better than "syllable-initial consonant".

2) Likewise, since "syllable-peak" implies vowel, "syllable-peak character" seems better than "syllable-peak vowel".

Accepted in principle

The suggested change would in fact undo a change introduced by amendment 3 which removed the term 'character' form these 3 annotations, recognizing the fact that Choseong, Jungseong, and Jongseong can be made of multiple characters each. Amendment 5 introduce a new model where new Jamo characters are being added allowing old hangul syllable to be represented by simple LVT sequences, but this does not deprecate the former model, they just co-exist. Moving the equivalence declaration inside the Choseong, Jungseong, and Jongseong description allowed the two models to be described without controversy.

An alternative solution agreeable to experts was found:
syllable-initial characters or initial consonants
syllable-peak characters or medial vowels
syllable-final characters or final consonants

T2. p. 12,

T2.1 Replace the glyph for U110D with a correct one.

T2.2 Rationale:

- The glyph for U110D is the same as that for U114F.
- The glyph for U114F is fine.
- However, the glyph for U110D is incorrect and therefore should be changed.
- We will provide the corrected font.

Accepted in principle

See Irish comment E.1. It is interesting to note that the Unicode 5.0 charts have the same issue, while Unicode 4.0 charts are aligned with ISO/IEC 10646:2003.

T3. Change six (6) phrases (three alternatives are given below).

T3.1 six (6) relevant phrases

p. 13, left column, "Initial consonants"

p. 14, left column, "Medial vowels"

p. 14, right column, "Final consonants"

p. 24, right column, "Initial consonants"

p. 28, left column, "Medial vowels"

p. 28, left column, "Final consonants"

T3.2 Rationale:

- Our national position has been to use either "Choseong (Jungseong, Jongseong)" or "syllable-initial (-peak, -final) character", as used in ISO/IEC 10646 since the 1993 edition.

- We suggest that one of the following alternatives be used:

1) alternative 1: delete those six (6) phrases.

We have not had those phrases in ISO/IEC 10646 since the 1993 edition.

2) if we want to have some phrases in those 6 places, change as follows:

2-1) alternative 2A:

p. 13, left column, "Initial consonants"-> "Syllable-initial Characters"

p. 14, left column, "Medial vowels" -> "Syllable-peak Characters"

p. 14, right column, "Final consonants" -> "Syllable-final Characters"

p. 24, right column, "Initial consonants"-> "Syllable-initial Characters"

p. 28, left column, "Medial vowels" -> "Syllable-peak Characters"

p. 28, left column, "Final consonants" -> "Syllable-final Characters"

2-2) alternative 2B:

p. 13, left column, "Initial consonants"

-> "Syllable-initial Characters or Initial consonants"

p. 14, left column, "Medial vowels"

-> "Syllable-peak Characters or Medial vowels"

p. 14, right column, "Final consonants"

-> "Syllable-final Characters or Final consonants"

p. 24, right column, "Initial consonants"

-> "Syllable-initial Characters or Initial consonants"

p. 28, left column, "Medial vowels"

-> "Syllable-peak Characters or Medial vowels"

p. 28, left column, "Final consonants"

-> "Syllable-final Characters or Final consonants"

Not accepted

Because the aliasing is introduced in an earlier clause, it seems unnecessary to repeat the equivalence notation at every occurrence.

T4. Add an annotation for U+3180

- According to Resolution M50.34 (Hangul Jamo additions), WG2 accepted to add five annotations proposed in document WG2 N3172 (2006-09-27). However, one of those five annotations is missing in the FPDAM5.

- Therefore we request that one annotation be added as shown below:

U+3180 HANGUL LETTER SSANGIEUNG

=> HANGUL LETTER SSANGIEUNG (ssangyesieung)

- By the way, the relevant page (page 227 in ISO/IEC 10646:2003) is not included in FPDAM5.

Acceptance in principle

All these annotations were added by text located on top of page 5 of the amendment. So nothing is missing from the amendment and because the chart page including 3180 did not have any character addition or significant glyph change it was not included in the amendment, only instructions. It just happens that the other 4 annotations were located in a page where new characters were added.

However the text adding these annotations does not reflect the new chart format, so it will be amended as follows:

and new annotations at the following code positions:

11EC HANGUL JONGSEONG IEUNG-KIYEOK
= yesieung-kiyeok

11ED HANGUL JONGSEONG IEUNG-SSANGKIYEOK
= yesieung-ssangkiyeok

11EE HANGUL JONGSEONG SSANGIEUNG
= ssangyesieung

11EF HANGUL JONGSEONG IEUNG-KHIEUKH
= yesieung-khieukh

Add the following annotation to the character 3180.

3180 HANGUL LETTER SSANGIEUNG
= ssangyesienung

T5. p. 34, left column

- For UF9B8, the correct mapping seems U96B7, NOT U96B8.
- We suggest that
  1) delete 96B8 entry and
  2) keep 96B7 entry only.

Not accepted

The entry '≡ 96B8 隸 describes the fact that F9B8 is transformed into 96B8 by normalization which is a normative part of the standard (see clause 3). It is also related to the CJKC_SR.txt entry for that character: 0F9B8;096B8;;;;K0-6766;;
Removing the entry would just remove a useful piece of information which shows a de facto equivalence. The additional entry mentioning 96B7 simply shows that the original mapping is less than optimal but it cannot be changed without destabilizing normalization for that character.

T6. Change K5 entries in CJKU_SR.txt file

- We decided to change the K5 reference format from "K5Hddddd" (decimal) to "K5-hhhh" (hexadecimal) [see p. 2, right column of FPDAM5].
- The change is well reflected in the code table of FPDAM5.
- However, the change is NOT reflected in CJKU_SR.txt file of FPDAM5.
- Therefore, we request that the change be reflected in CJKU_SR.txt file.
- An example is shown below:

  2A710;;;;K5H00418;;;;;

  -->

  2A710;;;;K5-01A2;;;;;

Accepted in principle

Due to production errors, an incorrect CJKU_SR.txt file was initially distributed with the ballot documents. The file was later corrected and the member bodies were alerted. The corrected file contains the appropriate K5 format, for example:
2A70F;;;;K5-01A2;;;;;
(note that the characters is at a slightly different location which was always correct in the chart section of FPDAM5).
The editor regrets the confusion.

T7. Delete one line from CJKU_SR.txt file.

- According to M51.11.g, we decided to delete one Hanja suggested by Rep. of Korea: its reference is K5H00029 (or K5-001D in a new format).
- CJKU_SR.txt file still contains that character.
- We suggest that the following line be removed from CJKU_SR.txt file:

2B265;;;;K5H00029;;;;;

Accepted in principle
See previous comment. Due to reshuffle of CJK Ext C, it would have now been located between 2B22B and 2B22C and is clearly not there:
2B22B;;TD-4445;;;;;;;
2B22C;;TD-4448;;;;;;;

As a result of this disposition, Korea (ROK) changes its vote to Yes.

UK, Negative

The UK votes to DISAPPROVE the amendment, with the following technical comments. Satisfactory resolution of comments T.1 and T.2 will change our vote to APPROVAL.

Technical comments

T.1 ARIB Unified and Compatibility Ideographs
The UK has grave concerns over the encoding of the two unified ideographs 9FC4..9FC5 and the four compatibility ideographs FA6B..FA6E requested by Japan.

T1.a) In particular, the UK objects to the encoding of U+9FC4 (ARIB#47) as a unifiable variant of U+6881. We note that this character has the perhaps unique distinction of having been removed from Amd.5 with one hand (it was U+2ACAD in PDAM5) and added back to Amd.5 with the other (see Resolutions M51.10-11).

Not accepted
Note that the revised version of WG2 N3318R justifies the addition as being cognate with U+6881 but with different abstract shape. Concerning the removal from Extension C, it seems in fact a wise move to remove it from the main ext C block if it is being processed in another venue in the same amendment. Note that the character still seems to be in Extension D but should now be removed.
After discussion, the proposal from WG2 N3318R was accepted with 3 CJK Unified Ideographs located at 9FC4-9FC6 (former FA6D moved to 9FC6), and 3 CJK Compatibility Ideographs located at FA6B-FA6D (former FA6E moved up to FA6D).In addition, the character 9FC4 gets an additional source reference: TC-4A76.

T1.b) On the other hand U+FA6D (ARIB#93) is not cognate with U+7953 according to the Chinese source, and so it may be a candidate for encoding as a unified ideograph.

Accepted
See disposition of T1.a and Japanese comment JPT1.

T1.c) There is also the wider issue of whether any new compatibility ideographs should be accepted for encoding now that the Ideographic Variation Database (IVD) is ready to be incorporated into the standard. We do not believe that it is helpful to have two different mechanisms for encoding unifiable ideographs (either as compatibility ideographs or by means of IVS sequences), as it can only cause confusion amongst users as to which mechanism is the appropriate one to use. With regard to the six ARIB characters, we note that five of them could be represented by means of the Adobe-Japan1 IVS collection <http://www.unicode.org/ivd/data/2007-12-14/ >:
U+FA6B (ARIB#39) = <U+6075 U+E0101>
U+FA6C (ARIB#67) = <U+242EE U+E0101>
U+FA6D (ARIB#93) = <U+7953 U+E0101>
U+FA6E (ARIB#105) = <U+8218 U+E0101>
U+9FC4 (ARIB#47) = <U+6881 U+E0101>

We do not believe that new compatibility ideographs should be encoded when they could be represented as IVSes, which have the advantage over compatibility ideographs of not being lost during the process of normalization. However, it is not clear to us whether it would be appropriate to represent these ARIB characters by means of the corresponding Adobe-Japan1 IVSes or whether a new ARIB IVS collection should be defined instead. We hope that the issue of the quality control of the IVD, and the relationship between the IVD and ISO/IEC 10646 will be discussed at WG2 #52.

Not accepted
Similar comment T.2 from USA. The consensus is that, long term, CJK compatibility Ideographs should not be added anymore for the reasons exposed by both the UK and the US NBs. However clarification should be first brought concerning the use of registered IVS contained in the IVD as a replacement mechanism and described in a new contribution. Because they are already about 1000 CJK Compatibility Ideographs, the addition of 3 similar characters is acceptable as a short term solution, given the relative size.

T1.d) In the light of these concerns, and given that there has not been unanimous agreement on the encoding of these six characters at either WG2 #51 or IRG #29, we request that 9FC4..9FC5 and FA6B..FA6E be removed from Amd.5.
Not accepted
In fact the WG2#51 was quasi unanimous in accepting inclusion these characters in the amendment ballot (with all countries present, including UK, in favor with only China abstaining) in its resolution M51.10. Concerning IRG#29, the resolution 29.9 was again accepted quasi-unanimously, with only US abstaining, on accepting these characters and asking Japan to revise its document for submission to WG2 (document WG2 N3318R, also IRG N1347R). The updated document has been now accepted as the basis for the inclusion of these 6 characters. See disposition of comment T1a) for further details.

T.2 Tai Tham
T.2.a) The UK supports the recommendations given in the Tai Tham Ad-hoc Meeting Report (N3379).
Accepted
See also comment T.3 from Ireland and T.8a from USA.

T.2.b) We also request the addition of the two subjoined characters proposed in N3384 to Amd.5.
Accepted
See also comment T.4 from Ireland and T.8b from USA.

As a result of this disposition, despite the partial non acceptance of comment T1, U.K. changes its vote to Yes.

USA: Negative:

The US National Body is voting *no* with technical and editorial comments on the following SC2 ballot SC2 N3982 […] Acceptance of technical comments T3-T8 will change our vote to *yes.*

Technical Comments:

T.1 U Source reference

The amendment currently adds the Unicode U source "UTC" with the reference "The Unicode Standard 5.1-2008." We recommend to replace this reference by "Unicode Technical Report #45, U-source Ideographs", as this new technical report provides more details about the origin of those characters.

Accepted in principle

Pending availability of the technical report

T.2 New collection numbers (page 3)

The end of the third paragraph of Clause 20.5 and the following note currently read:

> Variations sequences composed of a unified ideograph as the base character and one of VARIATION SELECTOR-17 to VARIATION SELECTOR-256 from the Supplementary Special-purpose Plane (SSP) are registered in the Ideographic Variation Database defined by Unicode Technical Standard #37.
>
> NOTE 2 - The Ideographic Variation Database is currently empty. When entries are registered, these variation sequences will be referenced by this standard.

Following the procedure defined by Unicode Technical Standard #37, a new version of the Ideographic Variation Database has been accepted on December 12, 2007. This version is identified as '2007-12-14', and contains 14,651 sequences, covering the repertoire of the Adobe Japan1 repertoire. We suggest to replace NOTE 2 by the following paragraph or similar text:

> This version of the standard incorporates by reference the variation sequences listed in version 2007-12-14 of the Ideographic Variation Database, as described at <http://www.unicode.org/ivd/data/2007-12-14>.

Accepted

See disposition of comment T1.c) from UK.

T.3 CJK Compatibility ideographs

While we entirely agree with the goal of establishing round-tripping between the ARIB character set and ISO/IEC 10646, we believe that there is now a better solution than encoding compability ideographs.

Consider the case of 恵 and 惠 which are distinct in the ARIB character set. In the model of ISO/IEC 10646, those two forms are unified as U+6075. To achieve round-tripping, the two forms must be mapped to different sequences of ISO/IEC 10646 characters.

The usual solution is to encode a compatibility ideograph, U+FA6B in this case, and to establish a canonical decomposition of that compatibility ideograph into the unified form, U+6075 in this case. The purpose of the decomposition is to account for the unification. Under that solution 恵 is mapped to the sequence <U+6075>, and 惠 is mapped to the sequence <U+FA6B>. However, this approach imposes a very severe constraint on implementations, as they can never normalize data; any normalization transforms <U+FA6B> into <U+6075> and prevents round-tripping. Essentially, the canonical decomposition defeats the purpose of the compatibility ideograph.

With the advent of variation sequences, we have a better solution at our disposal. Indeed the variation sequence <U+6075, U+E0100> is targeting the form 恵 and the variation sequence <U+6075, U+E0101> is targeting the form 惠, so the ARIB characters can be mapped to those sequences and support round-tripping. Unlike the sequences of the usual solution, these variation sequences remain unchanged by normalization. This gives a much greater freedom to implementations.

The Ideographic Variation Database also contains sequences for the other three pairs of ARIB characters which are unified in ISO/IEC 10646.

In conclusion, we believe that the proposed characters U+FA6B..FA6E fail to effectively achieve the goal of round-tripping the ARIB character set, and that this goal can be achieved today using variation sequences already in the Ideographic Variation Database. We propose to not encode those four characters.
Withdrawn
See disposition of comment T.1c) from UK.

T4. Conflicting sources
The editor's note at the bottom of page 2 mentions that there is unresolved conflicting information concerning KangXi source references. We would like these conflicts to be resolved before further progression of the Amendment.
Accepted
See disposition of comment JPO1 from Japan.

T5. Names of Hangul jamos
The names of the three characters 11FD, A96E and A973 should have an additional "H" at their end. The correct names are:
11FD HANGUL JONGSEONG KIYEOK-KHIEUKH
A96E HANGUL CHOSEONG RIEUL-KHIEUKH
A973 HANGUL CHOSEONG PIEUP-KHIEUKH
(Apparently, the original version of the proposal WG2 N3168 had incorrect names, which in turn led to incorrect names in WG2 N3242, which is what was accepted by motion M50.34. As described in WG2 N3257, a revision of WG2 N3168 included the correct names as above.)
Accepted
See disposition of comments T2 and T5 from Ireland.

T6. Avestan separation point
The US NB remains opposed to the encoding of yet another middle dot punctuation at position 10B38 (AVESTAN SEPARATION POINT).
Accepted
See also disposition of comment T6 from Ireland.

T7. Archaic Sinhala numerals
The US NB has received information that indicates that more investigation is needed for the Sinhala archaic digits and numbers (0DE7-0DEF and 0DF5-0DFF). The US NB would like those characters to be moved to a future amendment.
Accepted
Also supported by comment T.1 from Ireland and letter from the Sri Lanka Standards Institution.

T8. Tai Tham
T8a) The US NB supports the recommendations of the Tai Tham ad-hoc meeting as documented in WG2 N3379, …
Accepted
Also supported by comment T.3 from Ireland and T8a) from UK.

T8b) …as well as the inclusion of the two additional characters requested in WG2 N3384.
Accepted
Also supported by comment T.4 from Ireland and T8b) from UK.

Technical Comments:

E1. Incorrect U Source header
On page 5, the header of the additional code chart fragment for the new characters 9FC4 and 9FC5 is "U Unicode". Those characters only have a J source, so the header should be "J".
Accepted

As a result of this disposition, the U.S. NB changes its vote to Yes.

Sri Lanka (O-member)

Sri Lanka Standards Institution (SLSI) is the national Standards body in Sri Lanka, and is also a member of ISO and IEC representing Sri Lanka. …

SLSI recently found that a paper has been submitted by Mr. Michael Everson for inclusion of archaic Sinhala numerals in the Sinhala Code range, and that JTC 1 SC2 has accepted and registered as ISO/IEC 10646 : 2003 IFPD Amd 5 which is being processed to issue an amendment to Sinhala Character Code range.

…, a preliminary investigation by ICTA and the University of Colombo School of Computing (UCSC) shows that while Mr. Everson addresses only ONE type of Sinhala Numerals, but there are in fact at least five types that had been used in our past of history over 2000 years.
…
In view of the above points, it is our considered view that there is currently no agreement on what constitutes Sinhala Numerals and how they should be represented in UCS. It is also clear that there is no pressing or urgent need to immediately encode Sinhala Numerals. We do not agree with the justification provided in Mr. Everson's proposal that contact been made to members of the user community, as the Institutions such as SLSI, ICTA and the Department Official Languages of Sri Lanka were not even made aware of the proposal, and any such urgent need has not arisen for this amendment.

This letter was the basis for postponing the encoding of the proposed Sinhala historical digits and number. See comment T1 from Ireland and T7 from USA.