

Title: New proposal on the Hebrew vowel *HOLAM*
Source: Peter Kirk, Avi Shmidman, John Cowan, Ted Hopp, Trevor Peterson, Kirk Lowery, Elaine Keown, Stuart Robertson
Status: Individual Contribution
Action: For consideration by the UTC
Date: 2004-07-29

This proposal replaces the proposal made to the June 2004 UTC meeting as document L2/04-193 (also available as <http://qaya.org/academic/hebrew/Holam.pdf>). In this much shorter proposal there is no longer a set of options for consideration, but a single recommendation to the UTC. A separate background document L2/04-306 (also available as <http://qaya.org/academic/hebrew/Holam-background.pdf>) gives more details of the issues and options discussed during preparation of this proposal.

Background

The Hebrew point *HOLAM* combines in two different ways with the Hebrew letter *VAV*. In the first combination, known as *Holam Male*, the *VAV* is not pronounced as a consonant, and *HOLAM* and *VAV* together serve as the vowel associated with the preceding consonant. In the second combination, known as *Vav Haluma*, the *HOLAM* is the vowel of a consonantal *VAV*. In more exact typography *Holam Male* is distinguished from *Vav Haluma*: *Holam Male* is written with the *HOLAM* dot above the right side or above the centre of *VAV*; and *Vav Haluma* is written with *HOLAM* above the top left of *VAV*. The distinction is clear and significant in some texts, dating from the 10th century CE to the present day. In modern printing, the distinction is often made in biblical and liturgical texts, in poetry, and in educational materials; indeed in general where it is important to indicate the exact pronunciation of words which may not be familiar to readers. Normally the only graphical difference is in the relative positions of the *VAV* and *HOLAM* glyphs; occasionally small differences in one or other of the glyphs are also seen. See the samples in the figures below. But in common typography *Holam Male* and *Vav Haluma* are not distinguished, and usually both rendered with the *HOLAM* dot above the centre of *VAV*. *Holam Male* is very common in pointed Hebrew texts; *Vav Haluma* is much less common.

Note carefully that this is *not* a proposal to encode a *phonetic* distinction which is not made graphically. Rather, it is a proposal to encode a *graphical* distinction with a 1000 year history. This graphical distinction is made in a significant minority of modern texts, and it must be made when the phonetic distinction needs to be indicated unambiguously.

Unicode does not currently specify how to distinguish between *Holam Male*, *Vav Haluma*, and the undifferentiated combination. Several different ways have been used in existing texts, or recommended for use with Unicode Hebrew fonts. To avoid proliferation of *ad hoc* solutions, it is proposed here that the UTC indicate its approval of the specific representations proposed here.

For further details, see the separate background document.

Proposal

There has been an extensive debate, including at the June 2004 UTC meeting, about how best to distinguish between *Holam Male* and *Vav Haluma* in Unicode. A large number of options have been put forward and evaluated; see the separate background document for a list of these proposals and an evaluation of each of them. A consensus has now been reached among a group of users of both

biblical and modern Hebrew that the representations proposed here are the most likely to be generally acceptable. This group of users hereby requests the UTC to indicate its agreement that these representations are acceptable and should be recommended for general use to Hebrew users and to font designers; also to specify these representations in the text of the next version of *The Unicode Standard*. UTC agreement is required because the proposed representation involves the use of ZWNJ (i.e. U+200C ZERO WIDTH NON-JOINER).

The proposal is that *Vav Haluma* should be represented as <VAV, ZWNJ, HOLAM>, whenever there is a potential need to distinguish it from *Holam Male*. *Holam Male* should continue to be represented, as in the majority of existing texts, as <VAV, HOLAM>, and this same sequence may be used for a combination of VAV with HOLAM when a representation which does not distinguish between *Holam Male* and *Vav Haluma* is intended.

Justification

This proposal is based on the fundamental nature of *Holam Male* and *Vav Haluma* as distinct renderings of the combination of the same pair of characters VAV and HOLAM. From a graphical viewpoint they differ primarily in that in the former the HOLAM dot is placed in a different position from its normal one relative to the base character, indicating a special close connection between VAV and HOLAM. Thus *Holam Male* and *Vav Haluma* are respectively more and less connected renderings of the same character pair VAV and HOLAM. Indeed, *Holam Male* is commonly understood, and is implemented in many existing fonts, as a ligature between VAV and HOLAM; this also reflects its logical and linguistic nature, because *Holam Male* represents a single sound, long O, whereas *Vav Haluma* represents a sequence of separate sounds, VO. Because *Holam Male* is much more common than *Vav Haluma*, this ligature is taken as the default. The function of ZWNJ in the proposed representation of *Vav Haluma*, in accordance with its description in section 15.2 of *The Unicode Standard (TUS)* version 4.0.1 (<http://www.unicode.org/versions/Unicode4.0.0/ch15.pdf>), is to inhibit this ligature formation or equivalently to select the less connected rendering of VAV with HOLAM, appropriate for *Vav Haluma*, in which the HOLAM dot is placed in its regular top left position relative to the base character.

This use of a sequence including ZWNJ is in accordance with the revised definitions in *TUS* version 4.0.1 (<http://www.unicode.org/versions/Unicode4.0.1/>), in that ZWNJ is used within a combining character sequence immediately after the base character. According to the approved minutes of the February 2004 UTC meeting (<http://www.unicode.org/consortium/utc-minutes/UTC-098-200402.html>) the UTC made a specific decision to allow such sequences:

[98-C33] **Consensus:** Allow U+200D ZERO WIDTH JOINER and U+200C ZERO WIDTH NON-JOINER in combining character sequences. The interpretation of a joiner or a nonjoiner between two combining marks is not yet defined.

There is a precedent for such a sequence in the <base character, ZWNJ, combining mark> sequence defined for Bengali Reph and Ya-phalaa in *TUS* version 4.0.1.

The proposed representation should cause no difficulties for rendering engines which support ZWNJ as recommended in *TUS* version 4.0.1 section 15.2. The following implementation suggestions are based on the Implementation Notes in that section: *Holam Male* may be implemented, according to existing common practice at least with OpenType fonts, as a ligature formed by default for <VAV, HOLAM>; the inserted ZWNJ in the proposed sequence for *Vav Haluma* may automatically inhibit formation of this ligature and allow the rendering engine to position the HOLAM dot relative to the VAV glyph by the same positioning rule as used with every other base character. It is recognised that

there are some short term practical difficulties with certain current rendering engines in rendering the proposed sequence for *Vav Haluma*, especially on the rare occasions (essentially only in the biblical text) in which an accent is also combined with this VAV and HOLAM. However, encoding decisions should be based on the principles decided by the UTC rather than on the peculiarities of current implementations.

The main reason for preferring this proposal to other suggestions, especially those involving encoding of new characters, is that it is least disruptive of existing data. There is a considerable body of existing pointed Hebrew data in which *Holam Male* is represented as <VAV, HOLAM> (including for example 6,290 web pages found by Google containing the common word <LAMED, VAV, HOLAM>). Changing the representation of this very common letter at this stage, or recommending continuing use of two alternative and incompatible representations, would result in massive data representation ambiguities for Hebrew data. The continuing existence of incompatible representations would create a significant data mapping problem at the interface between the domains of the two different representations of Hebrew texts. *Holam Male* would be represented in biblical, liturgical, poetic and educational texts by a Unicode sequence which would appear, in rendering, to be the existing widely used sequence <VAV, HOLAM>, but which would in fact not be treated as equivalent to this sequence. This would create a *de facto* situation where the same Hebrew data would be represented in Unicode in one way in biblical, liturgical, poetic and educational texts and in an incompatibly different way outside such texts.

In most of the current data *Vav Haluma*, when it occurs, is represented by the same sequence <VAV, HOLAM>, but it is very much less common than *Holam Male* (a little over 1% of the frequency of *Holam Male* in the Hebrew Bible). Therefore the disruption to existing data in changing its representation, although the same in principle as for *Holam Male*, is quantitatively much less serious.

Obviously, in order to distinguish *Holam Male* from *Vav Haluma* in plain text it is necessary to change the Unicode representation of one or the other, or of both. But the practical adverse consequences of a change of representation are considerably reduced if a new representation is chosen which automatically falls back to the existing representation when processed by processes (including rendering, collation and general character and text processing) which have not been specifically set up to recognise the distinction between *Holam Male* and *Vav Haluma*. Precisely this automatic fallback is the default if a representation is used which consists of the existing representation plus a default ignorable control character. Variation selectors as currently defined cannot be used with combining characters, and CGJ cannot support a graphical distinction. But ZWJ and ZWNJ, as defined in *TUS* version 4.0.1, are available for control of ligature formation in this context, and so are suitable for distinguishing *Holam Male* from *Vav Haluma*. Specifically, ZWNJ is appropriate for a marked representation of *Vav Haluma*, because this is graphically and logically a less connected rendering of VAV with HOLAM than *Holam Male*.

An additional argument against solutions involving new characters is that, from the abstract character perspective, *Holam Male* and *Vav Haluma* are made up of the same VAV and HOLAM characters, but in different combinations. It is important for all kinds of character processing that the fundamental identities of the Hebrew characters VAV and HOLAM not be confused by representing either of them with two different Unicode characters. Indeed, it would be a breach of the Unicode character/glyph model to encode a new HOLAM character for what is essentially a contextual glyph variant of a single abstract character.

The current proposers wish to minimise the extent of disruption of existing Hebrew data, as well as to represent the abstract characters of the Hebrew script properly according to the Unicode character/glyph model. For this reason they wish to indicate the following definite preferences:

- Solutions are preferred in which <VAV, HOLAM> continues to represent *Holam Male*, and a different, marked representation is used only for the graphically distinct *Vav Haluma*;
- Solutions are preferred in which the marked representation is distinguished from the unmarked by default ignorable control characters, and which do not require definition of any new combining characters.

The representations in the current proposal agree most closely of all of the options considered with these preferences as well as with the general definitions in *The Unicode Standard*. They are therefore recommended to the UTC for its approval.

Samples

<i>Codex Leningradensis</i> (1006-7)	Lisbon Bible (1492)	Rabbinic Bible (1524-5)
Ginsburg/BFBS edition (1908)	<i>Biblia Hebraica Stuttgartensia</i> (1976)	Stone edition of Tanach (1996)

Figure 1: *Holam Male* (marked in red) and *Vav Haluma* (marked in blue) distinguished in ancient and modern editions of the Hebrew Bible - these words are from Genesis 4:13. (If the colours are not visible: In each image, the third base character from the right, with the dot above its right side or its centre, is *Holam Male*; the third base character from the left, with the dot above its left side, is *Vav Haluma*.)



Figure 2: *Holam Male* (left, twice, red, from p.529) and *Vav Haluma* (right, blue, from p.528) contrasted in Keil & Delitzsch *Commentary on the Old Testament*, vol.1, reprint by Hendrickson, 1996 (Hebrew words quoted in English text).



Figure 3: *Holam Male* (right Hebrew word, red) and *Vav Haluma* (left word, blue) contrasted in *Langenscheidt's Pocket Hebrew Dictionary*, p.243.



Figure 4: *Holam Male* (red) written with a different glyph from a regular VAV (blue), from *Siddur Tikkun Meir Hashalem*, R. Greenfield, 1982.

<p>Yose ben Yose (5th century), from <i>sidrei avodah</i> for <i>yom hakipurim</i> ("etain <i>tehila</i>"), in Goldschmidt, <i>Mahzor L'yamim Nora'im</i>, Koren Publishing 1970, p464</p>	<p>R. Elazar Hakalir (poetry of the late 6th century), from <i>piyyut</i> for <i>Shavuot</i>, "<i>eret mateh</i>", in Shulamit Elizur, <i>Kedushtaot l'yom matan torah</i>, Meketzzei Nirdamim, 2000, p116</p>	<p><i>Midrash Tanchuma</i> (8th century), <i>Or haHayim</i>, v1, 1998, p185</p>
<p>Yannai (poet of the early 6th century), from <i>kedushta piyyut</i> "<i>ashrei mo'asei alrla</i>", in Zaulai, <i>Piyyute Yannai</i>, Shoken Publishing, 1938, p32</p>		

Figure 5: *Holam Male* (red) and *Vav Haluma* (blue) distinguished in modern editions of mediaeval Hebrew poetry and midrashic literature.

<p><i>Mahzor Yom Hakippurim</i>, Israel Ariel, ed., Makhon Hamikdash / Carta Publishing, 1995, p92</p>	<p><i>Siddur Tefila</i>, Koren Publishing, 1996, p60</p>	<p><i>Hagada Shel Pesach</i>, Torat Chaim series, Mosad Harav Kook, 1998, p142</p>

Figure 6: *Holam Male* (red) and *Vav Haluma* (blue) distinguished in modern editions of liturgical texts. Note the larger and higher HOLAM dots in *Vav Haluma* in the right hand two examples; other idiosyncratic distinctions are made especially in Koren Publishing editions of such liturgical texts.