Center of Excellence for Urdu Informatics
National Language Authority
H-8/4 Islamabad

Dear Mark Davis!

I hope that you will be fine. I received your letter regarding the nuqta characters in which you stated the reason for keeping these out of the standard. I would like to address my concerns regarding this.

First of all, take the matter of the actual status of nuqta characters. Though they are supposed to act like combining marks but in their very essence these are characters that are needed most of the times right from publishing of primers to the archaic scripts. The nuqta characters added for Quran have a different status and a text processing client should be able to distinguish between the Quranic version of nuqta characters and the ordinary nuqta characters. The nuqta characters in the Quranic text have altogether different meanings than the proposed nuqta characters.

Secondly, I would like to emphasize that standards should not be developed either in isolation or inclined to protect a fewer aspects. The Arabic script is mostly used in countries and regions which can be filed under the third world. The declining costs of microprocessors have made it possible for the people of poor countries to have a desktop PC. Moreover, the falling rates of internet connectivity have also made it possible for these people to browse and surf the internet. I would not stuff my discussion with explaining the digital divide, but I would like to make a point that it's until recently that people have started to use the ARABIC SCRIPT on computers (Not Arabic language as it dates back earlier). And that era is from late nineties to date.

No matter how much a standard is well thought, potential problems cannot be anticipated until the standard comes in thorough and widespread usage. And now is the time when Arabic Code Block is being used for other than Arabic as well and this is where it needs some extensions.

Stated all the above, if the UTC decides not to include the nuqta characters into the standard then it simply leads to the closure of the Arabic script extensibility. I do not question about the health and accuracy of the UTC decision but its impacts. Moreover, when it's done for the Latin Script then why not for the Arabic Script? The diacritics for the Latin were not there from day 0.

The main obstacle in adoption of such Ideas is that Urdu is mistakenly considered the subset of Arabic. Though Urdu and other sister languages utilize the basic Arabic characters but they are grounds up different from the Arabic. Second thing is that sometimes problems are seen through the same window as the problems of Latin script are seen. We are facing even tough time in Nastaleeq while fitting it into the mold of OpenType. Same is true when we head to the semantics from the rendering.

The challenge for our nation (I am talking about Pakistan) is that there are almost 40+ regional languages written in Arabic script each having a different set of characters. This diversity simply cripples us to provide a single data entry interface in the form of a unified keyboard until and unless we have nuqta characters registered in the standard. And although the UTC might not be concerned with, it's indeed a matter of our national integrity.

I've explained the historical importance of the nuqta characters many times before but this time, let me show you evidence in this regard. Quran was initially in the form of base characters only without any diacritics even nuqta characters were not there. Later on, as the rest of the nations accessed Quran, they had difficulty in reciting it because of the fact that a single base characters may had different sounds based on the context (which only a native speaker knew based on his/her vocabulary) and secondly, when done with the sound of the base character, it may have different *Harkat* (movement which are now known to be *Airaab*) that too depending on the part of speech of that particular word. For the non native speakers to facilitate in reciting the Quran, Nuqta characters were introduced to show the exact sound and Harkat of a particular character. As these two elements were not part of the original Quranic text, they were written in a different color (most of the times red) to denote their special and separate status. Later on as the calligraphy evolved, they started it to write with a single colored ink: black, thus yielding the impact that nuqta characters are essential and un-separate able part of the script. The images of such scripts are attached at the end of this document for your reference. Moreover, dots are also used at the end of the Ayah and sit on the Ayah symbol already encoded in the Unicode Standard.

Now another problem that comes up is when we need to digitize these scriptures as it is. That turns out to be difficult because all the text processing clients treat the nuqta characters simply as part of the base character and thus no text operation is possible on such an entity which in its essence is independent but is unfortunately tied to the base character. And it's the place where in order to make it possible to process the nuqta characters separately; we need their separate presence in the Unicode Standard.

Regarding the stability, please reference my previous letter in which I already mentioned that every transition has associated costs. However, a comparison between the cost and benefit of the transition must be conducted and a decision must be made with optimal rationality. Current proposed transition is not that much "breaking" as its being envisioned. Normalization forms are to be involved for manipulating these compatibility equalances.

The spirit and philosophy of base/ghost characters have already been incorporated in the Arabic Code block in the form of dot less characters (U+066E and U+066F). The idea of a composite letter into its constituents is also not new for the Arabic Block (and for the Unicode Standard as a whole). For reference please refer to Table 1.

You may argue that those dots less characters are for the sake of archaic script. Very much right, if even so, Is a computer user able to type in the first piece of text at the end of this letter in which the color of dots is different from the base characters? Definitely not! Addition of Nuqta characters will enable us to type in such texts easily and save our historical scripts in their original form electronically. Moreover, this addition will stop the further proliferation of the Arabic Code block which otherwise will need extensive additions in the standard as the languages of Arabic script start their journey to the cyber age.

آ	U+0622 ARABIC LETTER ALEF WITH MADDA ABOVE	U+0627 ARABIC LETTER ALEF | U+0653 ARABIC MADDAH ABOVE
أ	U+0623 ARABIC LETTER ALEF WITH HAMZA ABOVE	U+0627 ARABIC LETTER ALEF | U+0654 ARABIC HAMZA ABOVE
ؤ	U+0624 ARABIC LETTER WAW WITH HAMZA ABOVE	U+0648 ARABIC LETTER WAW | u+0654 ARABIC HAMZA ABOVE
إ	U+0625 ARABIC LETTER ALEF WITH HAMZA BELOW	U+0627 ARABIC LETTER ALEF | U+0655 ARABIC HAMZA BELOW
ئ	U+0626 ARABIC LETTER YEH WITH HAMZA ABOVE	U+064A ARABIC LETTER YEH | U+0654 ARABIC HAMZA ABOVE
اٴ	U+0675 ARABIC LETTER HIGH HAMZA ALEF	U+0627 ARABIC LETTER ALEF | U+0674 ARABIC LETTER HIGH HAMZA
وٴ	U+0676 ARABIC LETTER HIGH HAMZA WAW	U+0648 ARABIC LETTER WAW | U+0674 ARABIC LETTER HIGH HAMZA
ۇٴ	U+0677 ARABIC LETTER U WITH HAMZA ABOVE	U+06C7 ARABIC LETTER U | U+0674 ARABIC LETTER HIGH HAMZA
ىٴ	U+0678 ARABIC LETTER HIGH HAMZA YEH	U+064A ARABIC LETTER YEH | U+0674 ARABIC LETTER HIGH HAMZA
ۀ	U+06C0 ARABIC LETTER HEH WITH YEH ABVOE	U+06D5 ARABIC LETTER AE | U+0654 ARABIC HAMZA ABOVE

Table1: *Already Dual Mapped Characters in Arabic Block*

Hearing from us might be new thing but it's not new, the Unicode Standard has already adopted many things from our standardized ASCII Code Plate Version II from the range of 600-61F including many symbols and Honorifics. Those code plates also contained the nuqta characters but were not incorporated into the standard at that time.

Though it may sound a bit strange, but if our point of view is not properly taken by the consortium then definitely there is no point for us to be a regular member of the consortium.

Best regards,
Dr. Attash Durrani
Liaison Officer
Center of Excellence for Urdu Informatics
National Language Authority
Islamabad
Pakistan

وَصَّى بِهَا عَلَيْنَا عَنْ مَشَائِخِنَا وَمَحَبَّتِهِ

وَمُمْتَثِلًا عَلَى السُّنَّةِ وَالْجُمْلَةِ

وَالشَّوْقِ إِلَى لِقَائِكَ يَا غَالِبُ

الْجَلَالِ وَالْإِكْرَامِ

Page of maghribi script by al-Qandūsi. Reversed colour background between the letters on one side and the vowel and diacritical signs on the other.

The mujawhar *(pearly) variation of maghribi script.*

مقتدرہ قومی زبان پاکستان
اردو ضابطہ مختی (Code Plate) : Ver: 2.00

High Hex Digit

Low Hex Digit	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0			Sp	٠	@	إ	ڑ	م			للہ	O	[			ﹹ
1			!	ا	Hs	أ	ز	ل			جلڑ	ب	۱			
2			”	۲	نت اضا	آ	ژ	ن			بسم اللہ	ف	]			
3			#	۳	۱	ب	س	و			ﷺ	ۋ	Us			
4			Cr	۴	ؘ	پ	ش	ؤ			ص	ﹻ	{			
5			%	۵	ﹷ	ت	ص	ه			ع	ﹹ	؞			
6			&	۶	ﹶ	ٹ	ض	ۺ			رض	ﹺ	}			
7			؟	۷	ﹱ	ث	ط	؏			رح	ﹻ	Da			
8			(	۸	ﹻ	ج	ظ	ی			ﹽ	ﹸ	٥			
9			)	۹	ﹼ	چ	ع	ے			؏	ﹹ	٥			
a			*	؍	ﹾ	ح	غ	ھ			ﹽ	ﹸ	٠			
b			+	؛	ﹽ	خ	ف	X			صد	ﹹ	ﹽ			
c			،	<	د	ذ	ق	ﹼ			لم	ﹹ	ﹹ			
d			‐	=	ڈ	ڑ	ک	ﹷ			سنہ	ﹷ	ﹻ			
e			Dc	>	ب	ژ	گ	و			ﹽ	ﹹ	ﹹ			→
f			Dv	؟	ش	ر	ل				۷	ﻁ	م			

Figure 1: Urdu Zabta Takhti (Urdu Code Plate) Ver 2.00

Abbreviations

Sp: Space, Cr: currency, Dc: decimal dv: division,
HS: hard space, Us: underscore, Da: Dash.
----: code plate switching

Legend

Control area (not to be used)
Reserved area (for future use by the standards)
vendor area