ARABIC LANGUAGE PROCESSING FOR RECOGNITION HOLY QURAN TAJWEED RULES USING FUZZY LOGIC AND PRODUCTION RULES

Dr. Kairoddin Omar*, UKM, Malaysia, email: ko@ftsm.ukm.my
Dr. Nazliyah Omar*, UKM, Malaysia, email: no@ftsm.ukm.my
Mr. Majdi Salameh, UKM, Malaysia, majdisalameh@yahoo.com, majdi_salameh@aabu.edu.jo.
Faculty of information Science and technology, University of Kapangsaam. Malaysia, Bangi, 43600, Selangor, Malaysia

Abstract
Arabic language is one of the natural languages which represents good domain for automatic processing, QURAN TAJWEED rules recognition is also one the interesting subjects in the past few years there were many attempts for processing Arabic language and just few which focused on Tajweed rules as result of the (Arabic Language Processing) [1].

This paper discuss new system that uses new techniques for computerize TAJWEED rules, the system uses QURAN text as an array of characters and vowels, which each have unique ASCII code representing it, then the array elements is analyzed to classify and recognize all TAJWEED rules that are in the selected text, all TAJWEED rules are stored in table, then the rules will be stored in a set with specific number of elements, after that each rule will be represent by a vector that contain ordered elements (characters, vowels and spaces), then the text will be scanned for TAJWEED rules, and these rules will be converted into vectors with specific format, by calculating the distance between the new rule vector and all TAJWEED rules vectors - the rule with the lowest distance-the correct Tajweed rule is recognized.

Keywords: Fuzzy Logic, production rules, AI, Tajweed Rules, Holy Quran, Arabic Language Processing, Language Processing.

Background: There were many attempts for creating software that deal with QURAN TAJWEED rules one of them is “computerize holy QURAN TAJWEED rules” [1] the research applies program that computerize TAJWEED rules using production rules (if-then) and BNF rules for TAJWEED rules, however this processing and the software “العاصم” have some errors in recognizing some of TAJWEED rules – not 100% correct, which cussed system error, however the suggested system will correct those errors by using fuzzy computations to find the best TAJWEED rule –the lowest distance—with threshold for accepting these TAJWEED rules, also the system uses extra features for reducing the amount of data to be searched by searching for specific conditions which will ignore the text and discard it from the searching process, which will fast up and enhance the system.

Tajweed rules: Tajweed rules they are the rules for pronouncing during reading QURAN words, the way of reading QURAN must be according to special rules specified with characters and vowels, unlike traditional arabic text which is not. all Muslims must know how to read the holy QURAN using tajweed rules, however there are many arabic and non arabic Muslims who don’t know these rules, the system help them to know, learn
and read the QURAN according to these rules.

the system also have a features that define the rules, explain them, color the characters that represent TAJWEED rules in the text with different color for each TAJWEED rule, to be easy recognized and classified.

The system cover most of QURAN TAJWEED rules, such as:

1- silent noon with sokon (سكون) and tanween rules (السكون والتونين) includes:
   a- ethhar halqi (اظهار حلفي)
   b- edgham (with ghonnah and without ghonnah)
   c- eqlab (القلب)
   d- ekhafa hakiki (أخفاء حقيق)

2- silent meem (with sokon) (ميم سكون) includes:
   a- ekhafa shafawi (أخفاء شفوي)
   b- edgham shafawi (الادغام شفوي)
   c- ethhar shafawi (إظهار شفوي)

3- meem and noon with shadah (ميم وسكون مداد) includes:
   a- ghonnah (غنة)
   b- qalqalah (قلة)
   c- b.1- grate qalqalah (قلة عملي)
   d- b.2- medium qalqalah (قلة وسطي)
   e- b.3- small qalqalah (قلة صغير)

4- ra’a rules (الحرام) includes:
   a- tafkheem (تخفيف)
   b- tarjeeq (entric)

**Quran Tajweed Rule System:**

the system is divided into many steps, at the beginning the text for the QURAN “ayah (اية)” is chosen either by written it directly in rich text box or by import it from the database where all QURAN ayah (اية) is already stored, ordered by the surah (سورة) it belongs to, and number that present the order of that ayah (اية) in the selected surah (سورة), the ayah is stored in one diminution array including the characters, the vowels, and even the spaces inside the ayah with the same order as they were in the text, then the text will be scanned searching for a match with any Tajweed rule by using production rules with fuzzy logic the best suitable rule that match the text is recognized as the TAJWEED rule for that text, and for all rules the text colored, defined and explained according to specified tajweed rule that been recognized.

**CONDITIONS FOR QURAN TAJWEED RULES:**

in this step there will be determination for each TAJWEED rule that is in the same word, in two different words or in two different sentence in QURAN text(ayah اية), preparing for tajweed rule recognition.
TAJWEED rules:
1- Allah : the word Allah (الله) حكم ( is in two rules
     a- tafkheem ( تفخيم) “Allah”: Allah (الله) ﺖ ﺖ
     * If the vowel on the letter before Allah (الله) is damah (ً) (ً)
     * If the vowel on the letter before Allah (الله) is fatha (فتحة) (‘)
     * If the vowel on the letters before Allah (الله) is sokon (سكون) ) and the vowel before sokon( سكون) is either fatha (فتحة) or damah (ً) (ً)

b- Tarqeeq ( ترقيق) “Allah”: I lam (الله)
     * If the vowel on the letter before Allah (الله) is kasrah (كسرة) ﺖ
     * If the vowel on the letter before Allah (الله) is sokon (سكون) and the Letter before sokon is kasrah (كسرة) ﺖ
     * If the vowel on the letter before Allah (الله) is tanween (التنوين)

1. tanween fath ( تنوين فتحة) ﺖ
2. tanween dam( تنوين ضمة) ﺖ
3. tanween kaser ( تنوين كسرة) ﺖ
4. shadah with tanween ( تنوين شدة) ﺖ

2- lam rules (silent lam)

     أحكام اللام الساكنة التي لا تحوي حركة

     a- lam “al shmosih” اللام الشمسية

     it must be (compound) (تدغم) with out ghonnah ( بدون غناء) with the character after it (لام) if it is one of 14 letters:

1 2 3 4 5 6 7
{tt, tha’a, sad, raaa, ta,, daa, tha}

3- al qalqalah

Al qalqalah is rule of Tajweed, which is done just for 5 arabic characters (characters) Which are

1 2 3 4 5 (qaf, tta, ba, jem, dal)

al qalqalah have 3 kinds

1- Small qalqalah: if the above (5 characters) one of them “is in the middle of the word and the vowel for it is sokon (سكون).

1- Middle qalqalah: if one of the 5 characters is in the end of the word and the vowel for it is sokon (سكون).

2- Grate qalqalah : if one of the 5 characters is in the end of the word – last character -and the vowel for it is shadah (شدة).

4- Noon (نون) with sokon (سكون) and al tanween

there are many rules for the noon(نون) and tanween(تنوين)
4- a – hiding “ekhafa”

the “non ﺛ” must be hid and composed with the followed character if
the non vowel is sokon and the character
after it is one of the following 15
characters:

1 2 3 4 5 6 7 8 9 10
{sad, thl, th, kaf, jem, shen, qf, sen, dal, tia,
ط، د، س، ق، ش، ح، ج، د، ص
11 12 13 14 15
zen, fa, taa dhad, tha}
ظ، ل، ث، ف

the non and the tanween may come in
three different places.
4-a-1 non char is in one word and one of
15 character is the same word- both in
the same word-
4-a-2 non is word and the character ( one
of the 15 characters)is in another word –
each one of the character in different
followed words .

4-a-3 tanween in one word and one of
(15 character) in another word.

4-b- eqlab

it is to convert the non and tanween
into mem before (ba ﺏ)

b-1 (non) ﺛ
if it is followed by (ba ﺏ) it must converted
into (mem ﻀ)

b-2- (tanween)
if it is (tanween) followed by (ba) it must be
converted into mem.

4-c- edgham

it is to compose the non with vowel sokon
and tanween with the character after it if the
character is one of the following
{ya, non, mem, waw, lam, ra}

the edgham (إدغام) is two kinds :

4-c-1 with ghonnah ( ﻛ) if the letter
(character) follows the non or tanween is one
of {ya, non, mem, waw }
4-c-2 with out ghonnah ( ﻰ): if the character
follows the non or tanween is either
{l, ra}

إظهار

it is to appear the non with sokon and
tanween. if the non or tanween came after
following 6 characters :
1 2 3 4 5 6
{alef, ha, ein, ha, gein, kha}

4-d- ethhar

it is to appear the non with sokon and
and the tanween may come in
three different places.
4-a-1 non char is in one word and one of
15 character is the same word- both in
the same word-
4-a-2 non is word and the character ( one
of the 15 characters)is in another word –
each one of the character in different
followed words .

4-a-3 tanween in one word and one of
(15 character) in another word.

4-b- eqlab

it is to convert the non and tanween
into mem before (ba ﺏ)

b-1 (non) ﺛ
if it is followed by (ba ﺏ) it must converted
into (mem ﻀ)

b-2- (tanween)
if it is (tanween) followed by (ba) it must be
converted into mem.

4-c- edgham

it is to compose the non with vowel sokon
and tanween with the character after it if the
character is one of the following
{ya, non, mem, waw, lam, ra}
11 12 13 14 15 16
17 18 19
non, hah, waw, ein, giev, ya, lam}
20 21 22 23 24 25 26

6- Hamza wasel

6- a - if it is at the beginning of the word, the hamza appeared and it must be specked
6- b - if the hamza is not at the beginning it is not specked but must appear.
6-c- if it is in the middle and if it is with “edgham”
6-d - convert the “hamza” with “mad”
if the hamza vowel is fatha after hamza for question

1- al-edgham “general”
It is to compose more than one character and pronounce anew character, if they
followed each other, there are many kinds of edgham :

7-a - general heterogeneous edgham (الاضغام العام المتجانس)
7-a-1- if the characters followed to others at any order
{ tta، ، ، ، dal } are
{ د، ت، ت، ط }

7-a -2 - if the following characters follow each other in any position (order)
{ tha، tha، tha، tha، tha، tha، tha } { ت، ت، ت، ت، ت، ت، ت }
7-a-3 - if the following characters follows each other in any order
(mem， ba) ( ب، م)

7- b -general edgham Simi similar

7-b-1 - if the following characters follows each other in any position (order)
(lam، ra) ( ل، ر)

7-b-2 - if the following characters follow each other in any order
(qaf، kaf)

**System stages:**

1- read and load the QURAN text
in this step there will be two options, first is
chosen the QURAN text or ayah(آية) from
database that store all QURAN surah(سورة),
and from specific surah one ayah or more will
be chosen, performed using lists and combo
boxes, or by display the QURAN text page
for the specific surah in al Othmani copy of
Quran, the text may be written in different
font type, font size , and font style, also the
text may be written with vowel and in multi
differences than it is in the holy QURAN,
also the text may be copied or imported from
external source.

2- QURAN text process and
analyst
in this step all characters, vowels an
spaces in the ayah will be stored in array
of characters, after a copy of that array
will be stored in one another array of
integer representing the ASCII, then the
index for all characters vowels and spaces
is stored, the number of words in the ayah
is counted, the position for each word –
the index of the first character in the word
– is located and stored, and the number of
characters in each word - excluding the vowels- is counted . all these stored data and information is going to be used later in next steps.

3- exclusive words discarding
this step will discard some of the words in the ayah (ая) because it is impossible to have rule that match any of Tajweed rules [exclusive rules], the main purpose of this step is to reduce the source text which will be processed to recognize TAJWEED rule and speed up the recognition process to enhance the performance of the system, the following exceptions represents the words which will be discard :
1- If the word consists of less than three characters, by checking the length of each word, which have been calculated from the previous step.
2- if the word were one of the following words
{Waw (و) , etha ,  kathalek (كذلك) , allaty (اللاتي) , alallaty (اللاتي) , allaawaat (اللاتي) ، ayoha(ايثا ) , thalek(ذلك) , telk (ذلك)}
   a. if the first and the last character in the word is noon (ن)  
3- often the words that begin with characters some characters don’t mach any TAJWEED rules , characters such as :
   { خ , ذ , ث , د , ن , س , ض , ط , ظ , و } 
4- if the word is ( Allah ) the rule is tafkheem lam
5- if the word was one of the following
   {ب , ت , ض , س , ضم , طم , تم , دم , الناقة , ن ، هؤلاء ، الضالین ، ق ، ص ، حم } 
   the rule is mad with 6 interval .
6- if the text is ( من راق ) then the rule is silent after noon(ن)
7- if the text is ( بل ران ) then the rule is silent after noon(ن).
8- if the word was one of the following
   { الدنيا , بنیان , صنوان , قوان } the rule is ethhar (اظهار).

4- tajweed rules recognition using production rules :

The recognition in this step done by using the production rules (IF – THEN), the production rules are used with the BNF (Backus Naur Form) [1].

5- TAJWEED rules Fuzzification
and recognition:
in this step fuzzy logic principles are used , with soft computing to find the best TAJWEED rule for specific text (recognition) if it is not found by the production rules , first let v be a set of all vowels used in TAJWEED rules , v={ shadah(شدة) , tanween fath(تونين فتح), tanween dam(تونين ضم), tanween kaser(تونين كسر), sokon(سكون), mad(مد) } and presented by ASCII code set, where each vowel is presented by its ASCII code, and let C be a set of all characters used in TAJWEED rules , C ={all Arabic alphabetical }, also presented by there ASCII code values . let each TAJWEED rule be presented by vector of ordered elements of both sets (v ,C) , let each rule be denoted with r with different sub number ,
   r1= (ethhar halqi )
   r2= edgham (with and without ghonna)
   r3= eqlab
   r4= ekhafa hakiki (real hiding)
   r5= ekhafa shafawi ( حقيقي شفوي)
   r6= edgham shafawi (にとって شفوي)
   r7= ethhar shafawi ( اظهار شفوي)
   r8= ghonna (و)
   r9= grate qalqalah (عظم)
   r10= medium qalqalah (efeller وسطي)
   r11= small qalqalah (عقله صغير)
   r12= tafkheem (تخفیم)
   r13= tarqeeq (ترقیق)
after that any checked text will be converted into a vector of characters and vowels using the one diminution array the text (ayah) with many conditions, these conditions are compounded of the number of parameters such as: word characters, the position of the word, the first elements value, the last element value, counter for spaces … etc.

6- TAJWEED rule recognition:

In this step the each text, paragraph, sentence, and word is scanned, and analyzed searching for specific vowel, spaces and characters in proper order to match with one of TAJWEED rules that been stored before, from previous step all rules are stored as vectors of elements (vowel, character, and space) in specific order, the vector from the QURAN text will be compared with all vowels using similarity or distance. The smallest distance value from all vectors is going to be the most suitable rule that match with the vector from QURAN.

\[
\text{Distance}(\text{input vector, } r_1) = | \text{input vector( element1)} - R_1(\text{element 1}) + | R \text{ input vector( element2)} - 1(\text{element 2}) | +... | R \text{ input vector( element n)} - R_1(\text{element n}) | \text{ For all } R \text{ 1 to m}
\]

7- Tests:

In this system ten surah (سورة) were tested, and the results of recognition for the selected surah (سورة) shown in table 1.

<table>
<thead>
<tr>
<th>TAJWEED rule</th>
<th>Rule recognition</th>
<th>Wrong recognition</th>
<th>Coloring result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real ekhafa</td>
<td>19</td>
<td>0</td>
<td>1 error</td>
</tr>
<tr>
<td>Real ethhar</td>
<td>15</td>
<td>0</td>
<td>No error</td>
</tr>
</tbody>
</table>

![Figure 1 (chart of table1)](image)

Also another test on all TAJWEED rules were done separately, and table two shows the results of recognition for all TAJWEED rules.

Table 2 - Table of tajweed recognition results

<table>
<thead>
<tr>
<th>TAJWEED rule</th>
<th>Rule counts</th>
<th>Coloring result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethhar shafawi</td>
<td>25</td>
<td>25 correct</td>
</tr>
<tr>
<td>Ekhafa shafawi</td>
<td>2</td>
<td>2 correct</td>
</tr>
<tr>
<td>Edgham shafawi</td>
<td>1</td>
<td>1 correct</td>
</tr>
<tr>
<td>Edgham without ghonnah</td>
<td>20</td>
<td>18 correct</td>
</tr>
<tr>
<td>Edgham without ghonnah</td>
<td>7</td>
<td>4 correct</td>
</tr>
<tr>
<td>Eqlab</td>
<td>5</td>
<td>5 correct</td>
</tr>
<tr>
<td>Real ekhafa</td>
<td>16</td>
<td>15 correct</td>
</tr>
<tr>
<td>Real ethhar</td>
<td>16</td>
<td>15 correct</td>
</tr>
<tr>
<td>Ghonnah</td>
<td>46</td>
<td>46 correct</td>
</tr>
<tr>
<td>Tafkheem</td>
<td>36</td>
<td>36 correct</td>
</tr>
<tr>
<td>Tarqeeq</td>
<td>9</td>
<td>9 correct</td>
</tr>
<tr>
<td>Continues Mad</td>
<td>8</td>
<td>8 correct</td>
</tr>
<tr>
<td>Separated mad</td>
<td>11</td>
<td>11 correct</td>
</tr>
<tr>
<td>Grate qalqalah</td>
<td>2</td>
<td>2 correct</td>
</tr>
<tr>
<td>Medium qalqalah</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Small qalqalah</td>
<td>10</td>
<td>10 correct</td>
</tr>
</tbody>
</table>

![Figure 2 (chart of table 2)](image)
8- Results:

The results were very good for the first test as shown two errors of 227 recognized rule with (99.992 %) correct percentage for recognition.

The results of the second test were also very good with seven errors of 214 recognized rule (99.968 %) correct percentage for recognition.

9- Conclusion:

it is possible to use production rules or crisp – hard computing - to recognize holy QURAN TAJWEED rules, but there will be some errors not because of the production rules it self but because of the text it process, and because of some special vowels and characters in holy QURAN, this study used both crisp and soft computing operations by using production rules with fuzzy logic principles to recognize TAJWEED rules, and the results were ((99.98 % %), which are better than the previous studies which use production rules only, because there will be no missing for any of TAJWEED rule if it was checking and it will take the closest rule that may match the correct TAJWEED rule, with suitable threshold and correct condition.

10- Problems:

Some rules are recognized in correct because it takes the nearest rule that may match, but those rules in other studies are missing and not recognized, but still have to solve this problem.

the program used in this research coloring characters that catch any of the TAJWEED rules, some colors covers the colors which was before it, then the character will have the latest used TAJWEED rule color, this case appears in shared TAJWEED rules as been mentioned before in this paper and in some examples part of the characters representing one TAJWEED rule is colored differently.

11- Future work:

- Making speaking TAJWEED rules for holy QURAN.
- Enhancing this research to have 100% correctness percentage.

Special thanks: special thanks for the students who help in developing the program of Tajweed rules using production rules and the results of the program. (Ms. Noor and yasmin).

References:

[4] المثير في احكام التحديد اعداد د. أحمد شكري ، د. أحمد القضاة ، د. محمد سليمان ، د. محمد المجالي ، د. محمد القضاة، عبد الرحمن أبو غليون علي الجيسي, عمر حماد، مامو الشمالي، 2005
[5] arabic character recognition using structural approach by Mustafa ramadneh