International Journal on Advanced Science Engineering Information Technology

Vol.8 (2018) No. 4-2
ISSN: 2088-5334

# Automatic Detection of Shadda in Modern Standard Arabic Continuous Speech

Ammar Al-Sabri[#], Afzan Adam[#], Fadhilah Rosdi[*]

[1]*Center for Artificial Intelligence Technology*

[2] *Centre for Software Technology & Management*

*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi,43600, Selangor, Malaysia*
*E-mail: ammaralhabib@gmail.com; afzan@ukm.edu.my; fadhilah.rosdi@ukm.edu.my*

***Abstract*— The presence of diacritics *Shadda* in Arabic continuous speech may lead to the reduction of the accuracy of automatic Word Boundary Detection (WBD), which caused one word will be wrongly detected as two words. Therefore, this will affect the accuracy of Automatic Speech Recognition (ASR), if it is based on WBD. *Shadda* is one of the essential characteristics of the Arabic language which represents a consonant doubling. In this paper, a proposed method of automatic detection of *Shadda* in Modern Standard Arabic (MSA) continuous speech was introduced to improve the accuracy of WBD in MSA continuous speech. The prosodic features namely Short Time Energy (STE), Fundamental Frequency and Intensity were investigated for its ability as *Shadda* pattern detection in continuous MSA speech. We have analyzed the proposed features by implementing a separated algorithm for each feature to detect *Shadda* pattern automatically. In addition, a new proposed method which is a combination of STE and Intensity were introduced. The dataset in this work is a collection of 1-hour TV broadcast news from Aljazeera Arabic TV channel for 2018 - broadcasters. We found that the Shadda pattern is very similar to unvoiced regions of speech, and this represents a big challenge for the improvement of WDB using Shadda. Results showed that the detection of Shadda using Short Time Energy and Intensity outperforms the Fundamental frequency with 55% of accuracy. Intensity achieved 71.5% in accuracy. In addition, a combination between Intensity & STE features was performed and achieved good results with 67.15% in accuracy. The number of false positive too has been reduced compared to Intensity alone.**

***Keywords*— shadda; gemination; word boundary; modern standard Arabic; short time energy; fundamental frequency; intensity.**

## I. INTRODUCTION

Speech processing for the Arabic language is a field that worth study, because of the complexity and characteristics of the Arabic language, which must be taken in account when designing the recognizer of Arabic Speech Recognition (ASR) [1]. One particular and important field in speech processing is identifying the WBD in continuous speech [2]. WBD has been investigated by many researchers for several decades, due to the impact of this field in many speech applications and the nature of challenges of the problem, and it is still an active scientific field of research [3].

In the Arabic language, same as in other languages, the process of building an efficient ASR system is affected by identifying the word boundary process, because the accuracy of ASR recognizers depends on efficient detection of word boundaries [4]. Hence, to build an efficient ASR system for Arabic, a researcher must have enough knowledge of technical details as well as enough experience in the Arabic language, which is a barrier for many researchers [1]. For example, the Arabic language has complexity in morphology as compared to other languages such as English [5]. Hence, enough knowledge of Arabic language must be addressed by the researchers.

The Arabic language is an official language for 22 countries around the world [6]. The Arabic language is a Semitic language and it has three main forms: Classical Arabic (CA), MSA and dialectal Arabic [7]. CA is the language used in the resource of Muslims religious, such as Hadith and Quran and in ancient Arabic manuscripts such as poetry. MSA is an official version used by government and agencies [7]. Dialectal Arabic includes all forms of currently spoken Arabic in daily life [8]. Dialectal Arabic or sometimes called Colloquial Arabic, is widely used in social media and many of its words are derived from MSA [9]. The MSA language has 34 phonemes, six of which are basic vowels, three long and three short vowels, and 28 are consonants which are the Arabic Alphabet [10].

MSA has many characteristics and phonetic features that can distinguish it from another language. Some of these characteristics are the presence of particular consonants such

as pharyngeal, glottal and emphatic consonants [11]. In addition, one of the essential characteristics of the Arabic language is *Shadda*, which represents a consonant doubling and stressing [12]. *Shadda* is not represented by letters, but by diacritics. A diacritic is a short stroke appended above or below the consonant [11]. The meaning of a word with or without diacritics *Shadda* can be definitely different and leads to ambiguity. For example, the Arabic word /**درس**/ (DaRaSa) without *Shadda* 'he studied' differs than word /**درّس**/ (Da**R:R**aSa) with *Shadda* 'he taught' [13].

The presence of diacritics *Shadda* (double/geminate the consonant length) in Arabic speech, causes a special issue which may lead to errors in the process of WBD. *Shadda* always occurs on syllable boundaries within a word, the first (hypothetical) consonant belonging to the leading syllable and the second (hypothetical) consonant belonging to the following syllable [14]. This gemination is not realized as a doubling of a consonant only, but by increasing the duration of the pronunciation of the consonant, and this realization differs depending on the type of consonant as in [14].

Hence, with the presence of *Shadda*, a word may be detected as two words. This may add challenges in the process of WBD. Therefore, a good detection for S*hadda* pattern may lead to a good WBD.

Davis & Ragheb in [15], showed that Shadda might come in the middle of the word or at the end of the word. Furthermore, it is difficult to be detected [16].

In this research, we focus on the detection of the patterns of *Shadda* in MSA continuous speech to improve WBD. We showed to what extent the detection of *Shadda* patterns will improve WBD. In addition, other patterns which are very similar to the patterns of *Shadda* have been shown.

This paper is structured as follows: Section II presents material and methods which describes related work, features selection and description of the features used in this research, as well as the experimental setup and automatic detection for Shadda. Section III presents the results and discussion, followed by the conclusion in Section IV.

## II. Material and Methods

The presence of *Shadda* may affect the process of conducting automatic word boundary detection, in which one word may be wrongly detected as two words. Previous researchers have addressed the problem of *Shadda* from the point of the relationship between *Shadda* and feature of duration, F0, energy and intensity. However, none of the previous research address the detection of *Shadda* using these features. In this research, Shadda detection has been addressed and implemented as well. In this section, related works for *Shadda* and feature selection are presented. In addition, we introduced the experimental setup which presented the extraction of feature and the detection of Shadda.

### *A. Related Work*

Previous researchers have addressed the problem of WBD in MSA in their research, as well as the problem of *Shadda*. Some of them tried to improve WBD using different techniques, but most of them used a WBD as an input parameter in their work. In addition, only a few researches pointed to the problem of *Shadda* in Arabic, but, none of them introduced a separated paper related to *Shadda* patterns detection.

In respect with the problem of *Shadda*, Hachour et al in [17] tried to resolve the problem of *Shadda* in standard Arabic from the side of speech synthesis. They depend on the comparison between the curve energy of the VC2V, (where V represents a vowel, C2 represents the gemination of a consonant). Ferrat & Guerti in [18] presented a study that used the acoustic feature (energy and durations) and articulatory features to analyze the pattern of Shadda in MSA. They showed how the energy is decreased during the pronunciation of *Shadda*. In addition, they showed the relation between the *Shadda* and the vowels following *Shadda*.

With regard to the improvements of WBD, the process of WBD is based on the segmentation of speech into small chunks by first. The speech segmentation is a process to divide continuous signals of speech waves into segmented waves that carry meaning such as words or phonemes [19]. There are other methods and techniques that were used to improve the WBD, different from *Shadda*. Biadsy et al [20] introduced Arabic pronunciation dictionary for phone and word recognition, depending on some linguistic pronunciations rules. Study showed the importance of using these rules in the improvements of the word recognition accuracy in MSA. The study improved the absolute accuracy of phone recognition by 3.77%–7.29%. Al-Irhaim & Saeed in [21] used Wavelet Neural Network to perform Arabic word recognition using amplitude to detect the end of the word. They suggested the beginning of the word depending on the amplitude crossing over a pre-defined threshold value. They showed that it is still at the voiced region while the amplitude of these signals still over the threshold value, and then the end of word detected if the amplitude became below the threshold for a predefined time. They achieved 77% in recognition accuracy.

AlDahri & Alotaibi [22] used Voice Onset Time (VOT) to perform classifying and recognizing two MSA stops, namely /t/ and /d/. The stop /d/ is voiced sound but /t/ is unvoiced. They based on the energy of the signals, and fundamental frequency to detect the start of stop release, closure, and voicing. They concluded that the VOT is always positive regardless of the stop voicing. In addition, the voiced consonant /d/ in the VOT is less than half of its value in unvoiced one. the VOT value of /t/ stop in MSA Arabic is higher than their values in other Arabic dialects and languages. Diehl et al [23] introduced a word-boundary context modelling for MSA. They improved the Cambridge Arabic Large Vocabulary Continuous Speech Recognition (LVCSR) Speech-to-Text (STT) system. They used word-boundary context information to mark the phonetic units of a word in the dictionary. Three indicators word-initial, I_, word-medial, M_, and word-final, F have been used. Also, they used full covariance Gaussian modelling in the Minimum Phone Error (MPE). They showed the importance of the presence of words indicators in the pronunciation, and how it varies according to the word location. In addition, they showed how these indicators could be used to detect word boundaries. They concluded that these indicators provide indirect information about short vowels and

nunation which may exist at the end of a noun or an adjective and this indicator can help to detect the end of some words.

Khalid [24] used zero crossing rate and the signal energy to detect word's start and end points as a part of their work. Elkourd & El Kourd in [25] used zero-crossing rate (ZCR) and STE to detect word boundaries, in order to produce a system for recognizing MSA Isolated Word.

### *B. Features Selection and Description*

This section describes the feature selection. We will show how these features were selected in order to detect the patterns of *Shadda*, as well as a brief description of these features.

*1) Features Selection*: As described in [12]&[14] that the Shadda is characterized by consonant doubling and stressing, hence, the duration of the pronunciation of the consonant is increased as in [18].

In [18], Ferrat & Guerti reported a relationship between the *Shadda* side and energy & duration side. Lass In [26], showed that the higher values of F0 give a robust cue for stress. In addition, Fry in [27] reported that the contour of F0 and duration gives distinctive cues for stress.

Hence, these features duration, energy and F0 can help to detect *Shadda*. Also, Mugair in [28], reported the relationship between gemination and intensity. The study showed that the gemination can include some emphatic functions such as intensity.

Therefore, in this research, the acoustic features STE, fundamental frequency (F0/pitch) and intensity has been selected to detect *Shadda* pattern in MSA.

2) *Features Description:* After we selected the feature above, these features are described below:

- *Short Time Energy (STE):* in [29], STE is calculated by the summation of the squared amplitude of the signal x(n) as shown below :

$$f(x) = \frac{1}{N}\sum_{n=0}^{N-1} x^2(n) \qquad (1)$$

The value of the energy of the voiced regions is always greater than the value of energy of silence region, and the value of energy of unvoiced regions is less than values of voiced regions but often greater than for silence. For the voiced region, the STE was observed permanently to be more than a dynamically calculated threshold value in this technique [30].

- *Fundamental Frequency (pitch):* As in [31], the measurement of the fundamental frequency (F0) and its harmonics, it is a framework that is important in the analysis of intonation pattern. It is referred to acoustic measurements (fundamental frequency, length, and/or amplitude). In addition, Sharma & Rajpoot in [30] defined F0/ pitch as the lowest frequency component, or partial, by which a strong relationship exists between this component and the other partials. The fundamental frequency was calculated for specific time step frames. It is the type of pitch that is growing up when someone speaks and then goes down when she/he stops. Hence, the fundamental frequency is permanently zero for the unvoiced and silence sound. The frequency values vary depending on the gender and the age of the speaker. It takes a range between 50-200 Hz for male, 150 to 450 Hz for female and 200-600 Hz for children [25].

- *Intensity:* It is measured by calculating the average amplitude of speech signals [14]. It depends on the energy of speech signals, which depends on the loudness of the voice of the speaker. If the speaker speaks loudly, the intensity goes high and vice versa [32]. Sometimes it is named as loudness which is calculated by narrow band approximation from the signal intensity as in [29].

### *C. Experimental Setup*

This section discusses the collection of datasets as well as the extraction of features and the study of such features. In addition, this section introduced the algorithms used to detect *Shadda* patterns automatically. In this research, the features STE, F0, Intensity and the combination between STE & Intensity were used to detect *Shadda* patterns in Modern Standard Arabic continuous speech.

*1) Data Set: The data*set in this research is a collection of 1 hour of TV broadcast news that was collected from Aljazeera Arabic news channel. These records are spoken by 9 adults: 7 males and 2 females. The records were carried out in a soundproof studio, they are clean and contain no soundtrack or echo. These records saved as wave format, 1 channel (mono) and a sample rate of 16000Hz. Recording files was splatted depending on the gender of the speaker (male/female) and was analysed separately.

Actual word boundaries and *Shadda* were marked manually with Praat software and have been saved as "TextGrid" file format to be readable from any programming language. Fig.1 shows a screenshot of Shadda segmentation from the dataset, labelled manually using Praat software. Those data have been done in pilot test.

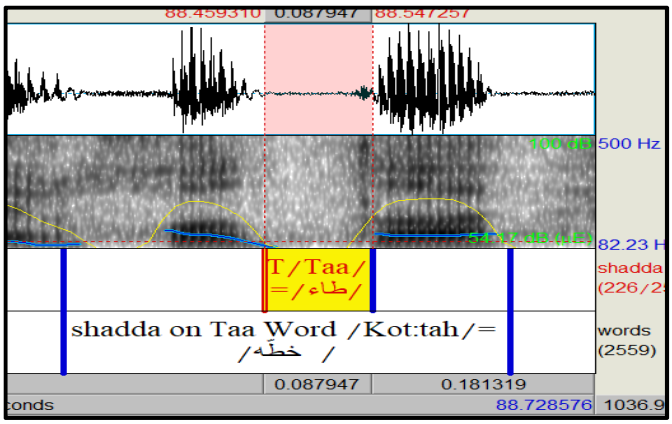


Fig.1 Screenshot of Shadda segmentation in Praat.

*2) Features Extraction:* The proposed flow diagram is shown in Fig.2 and it shows the features extraction process. The figure shows the flow of how each feature was extracted.

In respect to STE feature, this feature was extracted after performing pre-processing, framing and pre-emphasize. Then it was extracted using the equation (1) using MATLAB

software, then normalization was performed for STE frames' values, then a matrix of STE feature was obtained.

However, the F0 and Intensity features were extracted directly using Praat software after pre-processing of signals only. Praat software performed the framing process by its own, based on the input parameters from the user.

For F0, we used the default settings of option "To Pitch(ac)…" with only change of time step to be 0.02 seconds instead of "Auto". Then, a matrix of F0 feature was obtained.

In Intensity, we used the default setting of "to intensity" option with only changes of time step to be 0.02 seconds. Then, a matrix of Intensity feature was obtained.

Referring to STE, a brief description for pre-processing, framing, pre-emphasize and normalization steps was presented below:

- *Pre-processing:* It refers to everything done to the signals. In the pre-processing step, a down-mixing to a single channel (mono) was made for speech file was performed. Then, re-sampling for speech file to be 16000 Hz.

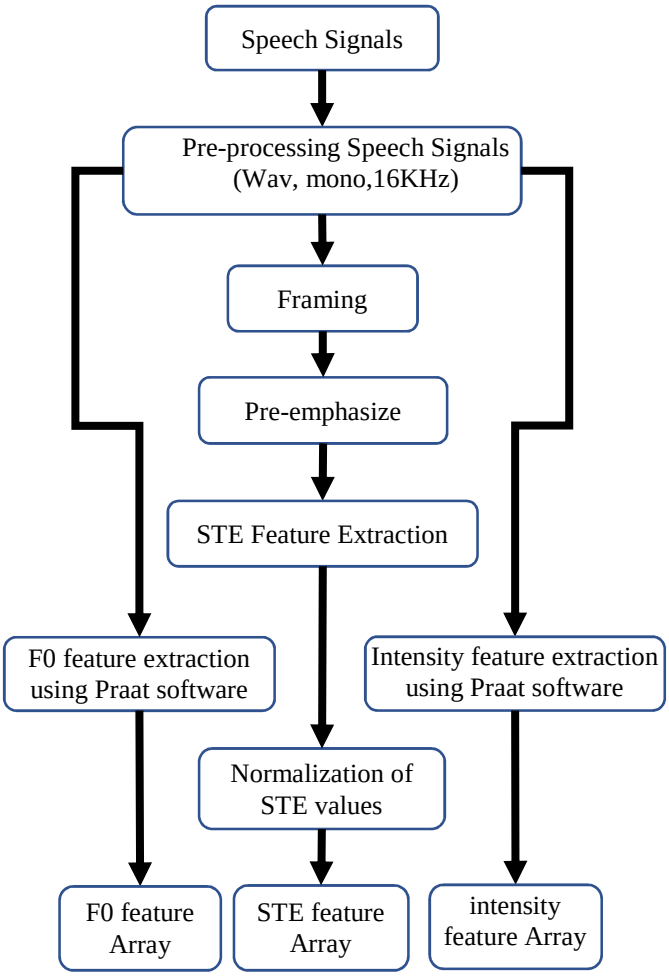


Fig.2 Flow diagram on features extraction process.

- *Framing:* As reported by [33], speech is not a stationary process, since the shape and size of this vocal tract goes on changing as human speak. Since it takes approximately 20-30ms for the vocal tract to change its shape and size, speech segment of 20-30ms duration can be considered to be stationary. Therefore, in this study, a frame duration of 20ms for each feature was taken.
- *Pre-emphasis:* It is the process of filtering the signal to attenuate frequency bands which carry important information. For speech processing, usually, it is a high-pass filter applied to a signal x(n) in order to emphasize information on formants [34]. In this work, a high-pass filter for each frame was applied.
- *Normalization*: Finally, after the STE frames' values obtained, the values of frames were normalized in order to make the values be comparable regardless of differences in magnitude. The normalization process was done by dividing the values of frames by maximum of absolute value of frames so that speech will be in the range from [0,1] file, as described in equation (2).

$$x(n) = \frac{x(n)}{\max(|x(n)|)} \qquad (2)$$

*3) Features Analysis:* After the extraction of features, and before going to automatic detection, these features were arranged alongside together in 3 columns, then the values of the features were traced, and studied separately and manually, in order to elicit the form of Shadda from each feature. Firstly, the values were traced and observed in general to observe the form of frames' values in each feature. Then the values of frames that contains Shadda were traced and observed manually for each feature. The observations were recorded as follow:

- *The observations of frames' values in general were recorded as follow:*

With regard to STE, it was observed that the values for speech frames are between 0.01 and 0.10 for unvoiced regions, and less than or equal 0.01 (STE <= 0.01) for silence regions and it is greater than 0.10 for voiced regions. Fig.3 shows the plotting of speech signals with STE.

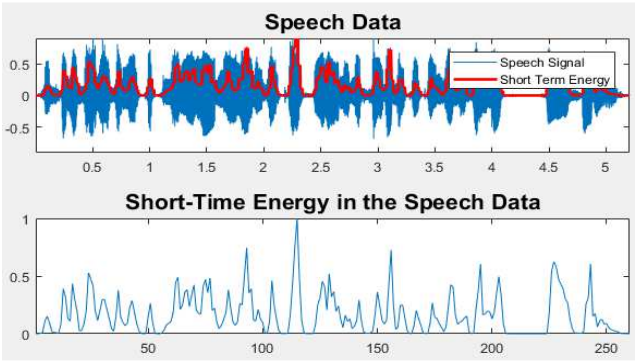


Fig.3 Plotting of speech data with short time energy.

In F0, it was observed that the values of frames are always zero for the unvoiced and silence regions. And it takes values greater than 90 for voiced regions. Fig.4 shows a screenshot of F0 contour in speech signals.

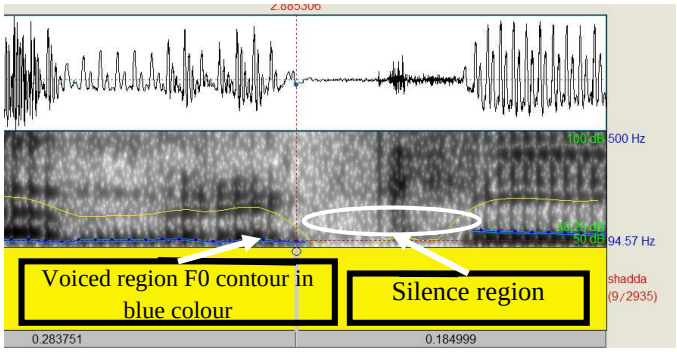


Fig.4 Screenshot of F0 contour in speech signals.

In respect with Intensity, the values for speech frames are between 45db and 60db for unvoiced regions, and less than 45db for silence regions and it is greater than 60db for voiced regions. Fig.5 shows screenshot of Intensity curve in speech signals.

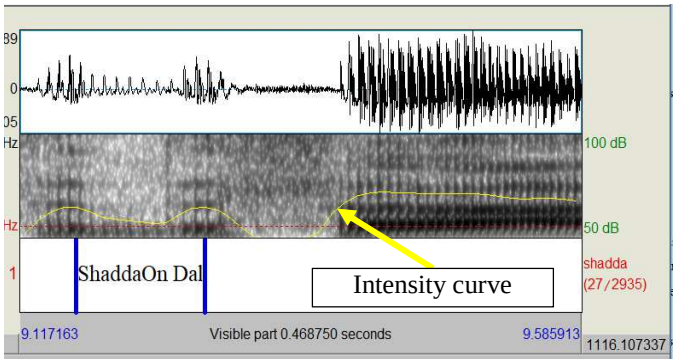


Fig.5 Screenshot of Intensity curve in speech signals.

The observations of frames values that contains Shadda (Patterns of Shadda) are described and discussed in "Results and Discussion" and recorded in tables (from Table I to Table X).

*4) Automatic Shadda Pattern Detection:* This step represents the implementation of this work. As described above, before moving to this step, the patterns of Shadda has been studied, traced manually in all 3 features obtained above. The previous step was performed, in order to study the form of Shadda pattern in each feature for all 3 features as well as to study the other patterns which are very similar to the patterns of Shadda.

In this step, the algorithms were implemented to perform automatic Shadda boundaries detection in the dataset. Finally, the results obtained were compared with the actual boundaries which were marked manually before, as described in section II.C.1.

Fig.6 shows the flow diagram for Shadda detection process using the features obtained from Fig.2.

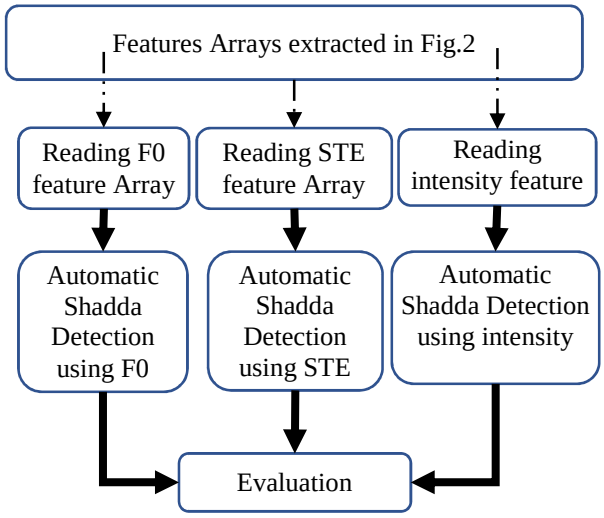


Fig.6 Flow diagram on Shadda detection process.

In this section, four experiments have been implemented using extracted features, in order to detect Shadda patterns automatically. These features were STE, F0, Intensity and a new proposed method which is a combination of (STE + Intensity). The evaluation of the results was conducted using equation (3).

The algorithms for such features are described below:

- *Algorithm of automatic Shadda Patterns Detection Using STE feature:* The pseudo code for detecting Shadda patterns using STE feature is shown below:

**Algorithm 1**

```
While i less than length of STE Array
  while frame value in silence region (i.e <=0.12)
    increment i;
    goto first;
  end
while i less than length of STE
  iF (STE (i) less than 0.05 and STE (i) not equal zero)
     keep the position of this frame i;
    let the count Of Frames Less Than 0.05=1;
     increment i;
   while i less than length of STE and STE(i) still less
   than 0.05
        increment i;
        increment the count Of Frames Less Than 0.05;
    end
    if (the count Of Frames Less Than 0.05 between 4 and
        add these frames to the list of Shadda patterns;
    else
      increment i;
    end
 else if (STE (i) less than 0.12 and STE (i) not equal zero)
     keep the position of this frame i;
    let the count Of Frames Less Than 0.12=1;
    increment i;
    while i less than length of STE and STE(i) still less
      than 0.12
        increment i;
        increment the count Of Frames Less Than 0.12;
    end
    if (the count Of Frames Less Than 0.12 between 4 and
        add these frames to the list of Shadda patterns;
    else
      increment i;
    end
  else
      increment i;
  endiF
```

```
  endwhile
endWhile
```

Depending on the observations of Shadda patterns in STE Feature in section III.A, the algorithm above was implemented to detect Shadda automatically using STE feature. The results of the algorithm are described in the results section in Table XI.

- *Algorithm of automatic Shadda Patterns Detection Using F0 feature:* The pseudo code for detecting Shadda patterns using F0 feature is shown below:

**Algorithm 2**

```
 While i < length of F0 Array -2
 iF F0(i) is between 90 and 110
keep the position of this frame i;
        let the countOfFrames =1;
         increment i;
while i less than length of F0 and F0(i) still less than or equal
100
          increment i;
          increment countOfFrames;
endwhile
   if (the countOfFrames between 5 and 8)
          add these frames to the list of Shadda patterns;
       else
        increment i;
       endif
  else
        increment i;
  endiF
endWhile
```

Depending on the observations of Shadda patterns in F0 Feature in section III.B, the algorithm above was implemented to detect Shadda automatically using F0 feature. The results of the algorithm are described in the results section in Table XI.

- *Algorithm of automatic Shadda Patterns Detection Using Intensity feature:* The pseudo code for detecting Shadda patterns using Intensity feature is shown below:

**Algorithm 3**

```
While i < length of intensity Array -2
iF intensity(i) <67 and intensity(i) >59 and
  intensity(i+1)>40 and intensity(i+1) < intensity(i)
  && intensity(i+2)< intensity(i)
keep the position of this frame i;
let the countOfFrames =3;
      increment i three times;// i=i+3;
        while i less than length of intensity and intensity
         (i) still less than 63
increment i;
increment countOfFrames;
        endwhile
        if (the countOfFrames between 3 and 10)
            add these frames to the list of Shadda patterns;
        else
          increment i;
        endif
          else
   increment i;
      endiF
endWhile
```

Depending on the observations of Shadda patterns in Intensity Feature in section III.C, the algorithm above was implemented to detect Shadda automatically using Intensity feature. The results of the algorithm are described in the results section in Table XI.

- *Algorithm of automatic Shadda Patterns Detection Using a combination of (STE + Intensity) Features:* The pseudo code for detecting Shadda patterns using STE + Intensity feature is shown below:

**Algorithm 4**

```
   while i < length of intensity Array -2
   if intensity(i) <67 and intensity(i) >59 and intensity(i+1)>40 and
intensity(i+1)< intensity(i)  && intensity(i+2)< intensity(i) &&
STE(i+1)<=STE(i)
   keep the position of this frame i;
          let the countOfFrames =3;
           increment i three times;// i=i+3;
     while i less than length of intensity and intensity (i)
       still less than 63
   increment i;
   increment countOfFrames;
          endwhile
          if (the countOfFrames between 3 and 10)
              add these frames to the list of Shadda patterns;
          else
            increment i;
          endif
           else
      increment i;
        endiF
endwhile
```

The algorithm 4 is very similar to algorithm 3, the only change is the new condition which was added in line 4, and it is highlighted by **bold font**. The condition will not be met until verify that the value of the next frame in STE is less than or equal the value of the current frame in STE. The results of this algorithm are shown in Table XI.

## III. Results and Discussion

This section describes the findings and results for Shadda patterns, then followed by the results obtained from algorithms.

In section I, it had been described that the realization of Shadda differs depending on the type of consonant paired with Shadda. That is what was observed after we traced the values of frames for all 3 features in Shadda regions. The findings and results for patterns of Shadda are recorded as follow:

### *A. Patterns of Shadda in STE Feature*

With regard to STE feature, it was discovered that the patterns of *Shadda* differ depending on the type of consonant paired with *Shadda*. These patterns can be classified into three categories: *Shadda* paired with voiced consonants, *Shadda* paired with unvoiced consonants and *Shadda* paired with nasal consonants, as shown in tables I, II and III. The cells filled with color represent the actual frames of the patterns of *Shadda*, while other cells represent the values which precede or follow the patterns of *Shadda*.

Table I shows the patterns of *Shadda* paired with voiced consonants in STE feature. It was observed that the energy of frames of S*hadda* paired with voiced consonants falling down under 0.12 with a duration of 80-120ms in normal speech (4-6 sequenced frames of 20ms), and then raising up again above 0.12. The energy of the first frame in falling down frames does not equal to zero.

TABLE I
PATTERNS OF SHADDA IN VOICED CONSONANTS IN STE FEATURE

| | ***Shadda* with voiced consonants in STE** | | | |
|---|---|---|---|---|
| **Words** | **/عدّه/= /Eid:dah/** | **/صدّق/= /Sad:daq** | **/مرّات/= /Mar:rat/** | **/موظّفون/= /Mowad:d-afoon** |
| **Conson-ants** | **/د/=/Dal/** | **/د/=/Dal/** | **/ر/=/Raa'/** | **/ظ/=/Dhaa'/** |
| **Frames Values** | 0.24 | 0.57 | 0.16 | 0.40 |
| | 0.13 | 0.22 | 0.14 | 0.12 |
| | **0.01** | **0.01** | **0.10** | **0.01** |
| | **0.01** | **0.01** | **0.10** | **0.01** |
| | **0.00** | **0.01** | **0.06** | **0.02** |
| | **0.00** | **0.01** | **0.05** | **0.03** |
| | **0.00** | **0.01** | **0.16** | 0.15 |
| | 0.18 | 0.17 | 0.34 | 0.19 |

Table II shows the *Shadda* patterns paired with unvoiced consonants in STE feature. It was observed that the energy of frames of *Shadda* paired with unvoiced consonants falling down under 0.05 with a duration of 100-180ms in normal speech (5-8 sequenced frames of 20ms), and then raising again above 0.05. In some cases, it was observed that it takes 4-9 sequenced frames. The energy of the first frame in falling down frames doesn't equal to zero in most cases.

TABLE II
PATTERNS OF SHADDA IN UNVOICED CONSONANTS IN STE FEATURE

| | ***Shadda* with unvoiced consonants in STE** | | | |
|---|---|---|---|---|
| **Words** | **/شفّاف/= /Shaf:faf/** | **/السُّجون/= /As:sojon/** | **/ينصُّ/= /Yanos:so/** | **/التّعذيب/= /At:tazeb/** |
| **Conson-ants** | **/ف/=/Faa'/** | **/س/=/Seen/** | **/ص/=/Sad/** | **/ت/=/Taa'/** |
| **Frames Values** | 0.19 | 0.07 | 0.10 | 0.10 |
| | 0.13 | **0.02** | **0.01** | **0.03** |
| | **0.02** | **0.01** | **0.02** | **0.00** |
| | **0.00** | **0.01** | **0.03** | **0.00** |
| | **0.00** | **0.01** | **0.04** | **0.00** |
| | **0.00** | **0.01** | **0.04** | **0.00** |
| | **0.00** | **0.01** | **0.02** | **0.00** |
| | **0.00** | **0.04** | **0.03** | 0.20 |
| | 0.15 | 0.22 | 0.05 | 0.21 |

Table III shows the Shadda patterns paired with nasals consonants in STE feature. It was observed that the energy of frames of Shadda paired with nasals is not stationary. In most of cases, it is going up and down. It is very similar to the pattern of "voiced consonant without Shadda".

TABLE III
PATTERNS OF SHADDA IN NASALS CONSONANTS IN STE FEATURE

| | ***Shadda* with nasals consonants in STE** | | | |
|---|---|---|---|---|
| **Words** | **/إنّ/= /iin:na/** | **/محمّد/= /Moham:m-ed/** | **/تحمّل/= /Tataham:mal/** | **/مهمّة/= /Maham:m-ata/** |
| **Conson-ants** | **/ن/=/Noon/** | **/م/=/Meem/** | **/م/=/Meem/** | **/م/=/Meem/** |
| **Frames Values** | 0.25 | 0.18 | 0.23 | 0.09 |
| | 0.41 | **0.09** | **0.07** | 0.07 |
| | **0.31** | **0.13** | **0.15** | **0.05** |
| | **0.04** | **0.12** | **0.24** | **0.03** |
| | **0.44** | **0.12** | **0.18** | **0.03** |
| | **0.26** | **0.07** | **0.06** | **0.03** |
| | **0.27** | **0.12** | **0.01** | **0.07** |
| | 0.41 | **0.12** | **0.10** | 0.07 |
| | 0.42 | 0.28 | 0.13 | 0.02 |

Table IV shows the other patterns that are very similar to the patterns of *Shadda* in STE feature. These patterns are the patterns of unvoiced consonant in speech. It was observed that the Patterns of unvoiced consonants (consonants such as: such as /ق/, /ك/, /ت/, /س/, /ش/, /ص/, /ف/, /ح/, /هـ/), are very similar to the patterns of *Shadda*. Therefore, these patterns also might be detected as S*hadda*.

TABLE IV
PATTERNS OF UNVOICED CONSONANTS IN STE FEATURE

| | **Unvoiced consonants pattern in STE** | | | |
|---|---|---|---|---|
| **Words** | **/بقوة/= /Bi:qowat/** | | **/للإطاقة/= /Lil:itahati/** | **/□تة/= /Sit:tate/** |
| **Conson-ants** | **/ق/=/Qaf/** | **/ة/=/Taa'/** | **/ح/=/Haa'/** | **/ـة/=/Taa'/** |
| **Frames Values** | 0.12 | 0.28 | 0.16 | 0.03 |
| | 0.00 | 0.01 | 0.04 | 0.01 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.01 | 0.03 | 0.00 | 0.00 |
| | 0.14 | 0.07 | 0.00 | 0.01 |
| | 0.18 | 0.06 | 0.08 | 0.05 |
| | 0.19 | 0.01 | 0.16 | 0.06 |

### *B. Patterns in Fundamental Frequency (pitch) Feature*

In F0 feature, it was observed that the patterns of *Shadda* can be classified into two categories: *Shadda* paired with voiced consonants and *Shadda* paired with unvoiced consonants, as shown in tables V, VI.

Table V shows the patterns of *Shadda* paired with voiced consonants in F0 feature. It was observed that the values of *Shadda* frames which paired with voiced consonants in F0 are not stationary and it does not take fixed forms. No rules could be elicited to help in the detection of this pattern automatically.

Table VI shows the patterns of *Shadda* paired with unvoiced consonants in F0 feature. It was observed that the values of *Shadda* frames which paired with unvoiced consonants in F0 are always zero, it is falling down under 110Hz for 5-8 frames, then raising up above 100Hz.

TABLE V
PATTERNS OF SHADDA IN VOICED CONSONANTS IN F0 FEATURE

| | ***Shadda* with voiced consonants in F0** | | | |
|---|---|---|---|---|
| **Words** | **/ضّ/= /Had:da/** | **/غضّ/= /Taghod:da/** | **/غزّة/= /Gaz:za/** | **/لعلّه/= /La'al:laho/** |
| **Conson -ants** | **/ض/=/Daad/** | | **/ز/=/Zaa/** | **/ل/=/Lam/** |
| **Frames Values** | 146.92 | 176.69 | 128.94 | 201.95 |
| | 136.70 | 172.75 | 129.40 | 210.29 |
| | 128.39 | 158.06 | 123.88 | 213.33 |
| | 119.63 | 150.00 | 118.98 | 211.62 |
| | 112.32 | 140.94 | 113.33 | 208.82 |
| | 107.84 | 131.53 | 112.83 | 205.88 |
| | 107.56 | 162.55 | 117.31 | 203.79 |
| | 124.81 | 158.64 | 126.61 | 204.19 |
| | 124.25 | 157.59 | 141.22 | 219.00 |

TABLE VI
PATTERNS OF SHADDA IN UNVOICED CONSONANTS IN F0 FEATURE

| | ***Shadda* with unvoiced consonants in F0** | | | |
|---|---|---|---|---|
| **Words** | **/شفّاف/= /Shaf:faf/** | **/السُّجون/= /As:sojon/** | **/الخاصّة/= /Alkhas:sa-h/** | **/موثّرون/= /Muth:thir-on/** |
| **Conson -ants** | **/ف/=/Faa'/** | **/س/=/Seen/** | **/ص/=/Sad/** | **/ث/=/Thaa'/** |
| **Frames Values** | 175.24 | 134.80 | 116.59 | 179.77 |
| | 162.84 | 0.00 | 0.00 | 155.33 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 186.15 |
| | 197.15 | 0.00 | 0.00 | 183.42 |
| | 190.42 | 147.87 | 253.89 | 179.77 |

Table VII, shows the other patterns that are very similar to the patterns of *Shadda* in F0 feature. These patterns are the patterns of the unvoiced consonant in speech. As in STE, it was observed that the Patterns of unvoiced consonants (consonants such as: such as /ق/, /ك/, /ت/, /س/, /ش/, /ص/, /ف/, /ح/, /هـ/), is very similar to the patterns of *Shadda* in F0. Therefore, these patterns also might be detected as S*hadda*.

TABLE VII
PATTERNS OF UNVOICED CONSONANTS IN F0 FEATURE

| | **Unvoiced consonants pattern in F0** | | | |
|---|---|---|---|---|
| **Words** | **/الحفاظ/= /Alhifadh/** | **/المرشد/= /Almorshed/** | **/نوات/= /Sanawat/** | |
| **Conson -ants** | **/ح/=/Haa'/** | **/ش/=/Sheen/** | **/س/=/Seen/** | **/ت/=/Taa'/** |
| **Frames Values** | 117.15 | 118.83 | 115.11 | 141.69 |
| | 0.00 | 0.00 | 111.67 | 138.65 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.00 | 0.00 | 124.73 | 0.00 |
| | 117.91 | 130.90 | 118.18 | 139.81 |

### *C. Patterns of Shadda in Intensity Feature*

In Intensity feature, it was observed that the patterns of *Shadda* can be classified into two categories: *Shadda* paired with voiced or unvoiced consonants and *Shadda* paired with nasal consonants, as shown in tables VIII, IX.

Table VIII shows the patterns of *Shadda* paired with voiced or unvoiced consonants in Intensity feature. It was observed that the values of frames start with a frame value between 67 and 59db and then falling down for two frames, keeping values under 63db for a duration of 4-8 frames and then raises up above 63db.

TABLE VIII
PATTERNS OF SHADDA IN VOICED OR UNVOICED CONSONANTS IN INTENSITY FEATURE

| | ***Shadda* with voiced consonants** | | | ***Shadda* with unvoiced consonants** | |
|---|---|---|---|---|---|
| **Words** | **/الأوربّيه/= /al'uwr-ub:biya/** | **/منظّمة/= /muna-dh:dha-ma/** | **/حضّ/= /tahod-:do/** | **/مكّن/= /tamak:k-an/** | **/الطّرف/= /at:taraf/** |
| **Conson -ants** | **/ب/=/Baa'/** | **/ظ/=/Dhaa'/** | **/ض/=/Dad/** | **/ك/=/Kaf/** | **/ط/=/Taa'/** |
| **Frames Values** | 65.80 | 72.65 | 73.27 | 72.41 | 75.58 |
| | 69.30 | 69.81 | 72.69 | 69.36 | 72.68 |
| | 66.42 | 62.57 | 64.44 | 57.90 | 60.70 |
| | 56.00 | 60.01 | 61.99 | 40.04 | 49.85 |
| | 44.31 | 59.68 | 61.18 | 38.78 | 49.06 |
| | 42.32 | 61.75 | 62.57 | 43.38 | 48.19 |
| | 43.91 | 66.51 | 72.19 | 58.63 | 49.69 |
| | 65.81 | 73.14 | 69.61 | 64.95 | 50.11 |
| | 71.88 | 74.44 | 70.15 | 72.79 | 68.21 |

Table IX shows the patterns of *Shadda* paired with nasals consonants in Intensity feature. It was observed that frames containing *Shadda* paired with nasals always start with a frame value between 70 and 65 and keeping values between 65 and 70 for a duration of 6-8 frames, and then raising up above 70 or falling down under 65db.

TABLE IX
PATTERNS OF SHADDA IN NASALS CONSONANTS IN INTENSITY FEATURE

| | ***Shadda* with nasals consonants in Intensity** | |
|---|---|---|
| **Words** | **/إنّكم/= /iin:nakum/** | **/عمّان/= /Am:man/** |
| **Consonants** | **/ن/=/Noon/** | **/م/=/Meem/** |
| **Frames Values** | 71.42 | 69.12 |
| | 66.65 | 67.15 |
| | 66.65 | 65.27 |
| | 66.60 | 66.66 |
| | 66.27 | 67.66 |
| | 67.45 | 68.22 |
| | 68.63 | 68.77 |
| | 64.97 | 69.57 |
| | | 70.23 |

Table X shows the other patterns that are very similar to the patterns of *Shadda* in Intensity feature. These patterns are the patterns of unvoiced consonant in speech. As in STE and F0, it was observed that the Patterns of unvoiced consonants (consonants such as: such as /ق/, /ك/, /ت/, /س/,

/ش/, /ص/, /ف/, /ح/, /هـ/), is very similar to the patterns of *Shadda* in Intensity feature. Therefore, these patterns also might be detected as S*hadda* in this feature.

TABLE X
PATTERNS OF UNVOICED CONSONANTS IN INTENSITY FEATURE

| | Unvoiced consonants pattern in Intensity | | | |
|---|---|---|---|---|
| **Words** | /□تة=<br>**/Sit:tate/** | | /مستمرة=<br>**/Mostam-er:raton/** | /ونفى=<br>**/Wanafa/** |
| **Conson-ants** | /س/=**/Seen/** | /ـة/=**/Ta'a/** | /س/=**/Seen/** | /ف/=**/Faa'/** |
| **Frames Values** | 69.49 | 69.22 | 69.27 | 73.05 |
| | 68.73 | 66.52 | 69.04 | 69.15 |
| | 63.96 | 62.41 | 65.51 | 55.14 |
| | 55.01 | 53.87 | 57.53 | 50.48 |
| | 51.27 | 43.23 | 53.73 | 50.84 |
| | 51.38 | 41.98 | 49.18 | 54.14 |
| | 47.98 | 57.59 | 50.43 | 72.29 |
| | 62.78 | 65.25 | 50.09 | 74.62 |
| | 66.58 | 65.73 | 65.04 | |
| | 60.96 | | 71.32 | |

### *D. Results of Automatic Shadda Detection*

After the pattern of *Shadda* has been obtained in all 3 features, the algorithms were implemented to detect the boundaries of Shadda automatically in the dataset.

Previous researchers have addressed the problem of Shadda from the point of the relationship between Shadda and feature of duration, F0, energy and intensity. But, none of the previous research implemented the detection of Shadda using these features. In this research, Shadda detection has been addressed and implemented as well using such features. In addition, we introduced an implementation for a new method which is a combination of (STE + Intensity).

To do the implementation, the patterns which obtained from three features have been used to perform three algorithms separately. Then the results have been compared with the ground truth boundaries which were built before. The final results are shown in Table XI.

Then a combination of (F0 & STE) and (F0 & Intensity) was performed separately, in order to increase the accuracy of Shadda detection. But, the results were not good enough, due to the instability of F0 frames values in Shadda paired with voiced consonants, as described in Section III.B.

Finally, the combination between STE & Intensity was performed to find out to what extent it can enhance the results. This method shows approximate results to the results of Intensity. Results are described in Table XI.

The accuracy was calculated using the formula (3) below, where the acc means accuracy:

$$acc = \frac{correctly\ detected}{Actual\ Total} \qquad (3)$$

TABLE XI
FINAL RESULTS OF AUTOMATIC SHADDA BOUNDARIES DETECTION

| Feature | Actual Shadda patterns in dataset | correctly detected | False negative (Shadda but not detected) | False positive (not Shadda but detected as Shadda) |
|---|---|---|---|---|
| **STE** | 228 | 103 | 125 | 1331 |
| **F0** | 228 | 38 | 190 | 360 |
| **Intensity** | 228 | 163 | 65 | 1712 |
| **STE+ Intensity** | 228 | 152 | 76 | 1519 |

As described in table XI, STE performs 45% in the accuracy, but the number of false positive (not *Shadda* but detected as *Shadda*) is high. F0 shows a poor accuracy. Intensity performs better than STE in automatic detection of *Shadda*, it performed 71% in the accuracy, but the false positive is still high. The combination of STE & Intensity has managed to reduce the number of the false positive, and it performs 66% in the accuracy.

Further investigation into misclassified cases or false negative (*Shadda* but not detected), these patterns were traced manually. It was observed that it occurs in case of an unvoiced consonant precedes (follows) *Shadda* pattern, in this case, the algorithms include frames of unvoiced consonant and frames of *Shadda* pattern together. Hence, the total number of frames will be greater than 8 or 9 in the patterns of *Shadda* paired with the unvoiced consonant, and greater than 6 for the patterns of *Shadda* paired with the voiced consonant. In this case, the pattern of *Shadda* will not be detected and it will be excluded by the algorithms.

Similarly, false positive cases were traced manually as well, and it was observed that most of these patterns are patterns of unvoiced consonants which are very similar to *Shadda* pattern. The patterns of unvoiced consonants were described in Tables IV, VII and X. These patterns might occur more than one time per a word. If there are any methods that can distinguish between *Shadda* patterns and unvoiced consonants patterns, the outcomes will be much better.

## IV. CONCLUSIONS

In this paper, a proposed method of automatically detection of *Shadda* in Modern Standard Arabic (MSA) continuous speech was introduced in order to improve WBD in MSA continuous speech. Prosodic features namely STE, Fundamental Frequency, Intensity and a new proposed method which is a combination of (STE + Intensity) were implemented to detect Shadda patterns automatically. Results from our dataset showed that the detection of Shadda using STE achieved an accuracy of 45%, Intensity achieved 71% outperforms the Fundamental frequency with 55% of accuracy. The combination method of (Intensity & STE) achieved good results with 67.15% in accuracy. The number of false positive too has been reduced compared to Intensity alone. In addition, we found that the Shadda pattern is very similar to the patterns of unvoiced consonants, and this represents a big challenge for WBD improvements using Shadda. Therefore, relevant features to differentiate Shadda

patterns and unvoiced consonants patterns need to be further investigated.

## ACKNOWLEDGEMENT

This research is supported by Young Researchers Incentive Grant GGPM-2017-020 from the Universiti Kebangsaan Malaysia. In addition, it is supported by SAMAASOFT company.

## REFERENCES

[1] A. Ali, Y. Zhang, P. Cardinal, N. Dahak, S. Vogel, and J. Glass, "A complete kaldi recipe for building arabic speech recognition systems," in Spoken Language Technology Workshop (SLT), 2014 IEEE, 2014, pp. 525-529.

[2] S. Kanneganti, "Design Of An Automatic Word Boundary Detection System Using The Counting Rule," Temple University Libraries, 2011.

[3] A. Tsiartas, P. K. Ghosh, P. Georgiou, and S. Narayanan, "Robust word boundary detection in spontaneous speech using acoustic and lexical cues," in Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, 2009, pp. 4785-4788.

[4] V. Naganoor, A. K. Jagadish, and K. Chemmangat, "Word boundary estimation for continuous speech using higher order statistical features," in Region 10 Conference (TENCON), 2016 IEEE, 2016, pp. 966-969.

[5] E. A. Mohammed and M. J. Ab Aziz, "Towards English to Arabic Machine Translation," International Journal of Advanced Research in Computer Science, vol. 2, 2011.

[6] B. Bataineh, S. N. H. S. Abdullah, and K. Omar, "Generating an Arabic Calligraphy Text Blocks for Global Texture Analysis," International Journal on Advanced Science, Engineering and Information Technology, vol. 1, pp. 150-155, 2011.

[7] R. E. Salah and L. Qadri binti Zakaria, "A Comparative Review of Machine Learning for Arabic Named Entity Recognition," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, pp. 511-518, 2017.

[8] S. Khoja, "APT: Arabic part-of-speech tagger," in Proceedings of the Student Workshop at NAACL, 2001, pp. 20-25.

[9] R. E. Salah and L. Q. binti Zakaria, "Arabic Rule-Based Named Entity Recognition Systems Progress and Challenges," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, pp. 815-821, 2017.

[10] Y. Alotaibi, S.-A. Selouani, and D. O'shaughnessy, "Experiments on automatic recognition of nonnative Arabic speech," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2008, p. 679831, 2008.

[11] K. Daqrouq, M. Alfaouri, A. Alkhateeb, E. Khalaf, and A. Morfeq, "Wavelet LPC with neural network for spoken arabic digits recognition system," British Journal of Applied Science & Technology, vol. 4, p. 1238, 2014.

[12] I. L. Learning and A. P. com, Learn Arabic - Level 1: Introduction to Arabic: Volume 1: Lessons 1-25, 2017.

[13] M. Alkhalifa and H. Rodríguez, "Automatically extending named entities coverage of Arabic WordNet using Wikipedia," International Journal on Information and Communication Technologies, vol. 3, pp. 20-36, 2010.

[14] N. Halabi, "Modern standard Arabic phonetics for speech synthesis," University of Southampton, 2016.

[15] S. Davis and M. Ragheb, "Geminate representation in Arabic," in Perspectives on Arabic Linguistics XXIV-XXV. vol. 1, ed: John Benjamins Publishing Company, 2014, pp. 3-19.

[16] H. Al-Haj, R. Hsiao, I. Lane, A. W. Black, and A. Waibel, "Pronunciation modeling for dialectal Arabic speech recognition," in Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on, 2009, pp. 525-528.

[17] O. Hachour, N. Mastorakis, and M. GUERTI, "The problem of “cheddah” In Standard Arabic Language," 2016.

[18] K. Ferrat and M. Guerti, "An Experimental Study of the Gemination in Arabic Language," Archives of Acoustics, vol. 42, pp. 571-578, 2017.

[19] A. Radman, N. Zainal, C. Umat, and B. A. Hamid, "Effective arabic speech segmentation strategy," Jurnal Teknologi, vol. 77, pp. 9-13, 2015.

[20] F. Biadsy, N. Habash, and J. Hirschberg, "Improving the Arabic pronunciation dictionary for phone and word recognition with linguistically-based pronunciation rules," in Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2009, pp. 397-405.

[21] Y. F. Al-Irhaim and E. G. Saeed, "Arabic word recognition using wavelet neural network," in Scientific Conference on Information Technology, 2010, pp. 29-30.

[22] S. S. AlDahri and Y. A. Alotaibi, "Phonetic investigation of MSA Arabic stops (/t, d/)," in Image and Signal Processing (CISP), 2010 3rd International Congress on, 2010, pp. 3524-3527.

[23] F. Diehl, M. J. F. Gales, X. Liu, M. Tomalin, and P. C. Woodland, "Word Boundary Modelling and Full Covariance Gaussians for Arabic Speech-to-Text Systems," in INTERSPEECH, 2011, pp. 777-780.

[24] S. Khalid, "Arabic Speech Recognition Using Hidden Markov Model," 2014.

[25] A. M. Elkourd and K. El Kourd, "Arabic Isolated Word Speaker Dependent Recognition System," ed: Islamic University, Gaza, Palestine Deanery of Higher Studies Faculty of Engineering Computer Engineering Department, 2014.

[26] N. Lass, Contemporary issues in experimental phonetics: Elsevier, 2012.

[27] D. B. Fry, "Experiments in the perception of stress," Language and speech, vol. 1, pp. 126-152, 1958.

[28] S. K. Mugair, "A Linguistic Study of Gemination of Arabic Languages," 2018.

[29] F. Eyben, Real-time speech and music classification by large audio feature space extraction: Springer, 2015.

[30] P. Sharma and A. K. Rajpoot, "Automatic identification of silence, unvoiced and voiced chunks in speech," Journal of Computer Science & Information Technology (CS & IT), vol. 3, pp. 87-96, 2013.

[31] J. O. Uguru, "Fundamental frequency as cue to intonation: Focus on Ika Igbo and English rising intonation," Proceedings of Meetings on Acoustics, vol. 19, p. 060231, 2013.

[32] L. Fu, X. Mao, and L. Chen, "Relative speech emotion recognition based artificial neural network," in Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on, 2008, pp. 140-144.

[33] P. R. Rao, Communication Systems: Tata McGraw-Hill Education, 2013.

[34] F. Eyben, "Acoustic Features and Modelling," in Real-time Speech and Music Classification by Large Audio Feature Space Extraction, ed: Springer, 2016, pp. 9-122.