Identification of Nasalization (*Ghunnah*) in Classical Arabic Dialect Using ANN

Ali Meftah, Yousef Alotaibi

College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia.

{ameftah, yaalotaibi}@ksu.edu.sa

***Abstract* - Almost all languages contain nasalized vowels and/or consonants and there is much research in the field of classification, detection, and recognition of phonemic nasalization with different features. In Arabic language /m/ and /n/ are the nasal phonemes, but the rule of the recitation of The Holy Quran (THQ) requires the conversion or mixing of non-nasalized phonemes that come after /n/ or /m/ to produce the Ghunnah (nasalization). Our aim in this study is to classify phonemes in terms of whether they come after /n/ or /m/ or not, to produce a robust system used for classifying nasalized and non-nasalized phonemes. It is important to clarify our difficult goal which is the classify nasalized phonemes without dealing with inherent nasal phonemes themselves. We applied a multilayer perceptron classifier by using the first three formants. Our system accuracy was in the range 71.5 to 85.4% according to the size of the used training and testing data subsets.**

Keywords – Ghunnah, THQ, Arabic, formants, MLP

I. INTRODUCTION

More than 99% of languages contain nasalized vowels or consonants Nasalization, in very simple terms, is the nasal coloring of other sounds. Nasalization occurs when the *velum* (a flap of tissue connected to the posterior end of the hard palate) drops to allow coupling between the oral and nasal cavities. When this happens, the oral cavity is still the major source of output but the sound gets a distinctively nasal characteristic. When nasal consonants are produced, air flows through the nasal tract and is radiated from the nostrils. The closed oral cavity and the sinuses of the nose from shunting cavities to the main path substantially influence the resulting radiated sound [1], [2].

For the automatic detection and classification of vowel nasalization, Pruthi and Espy-Wilson [2] used F1 amplitude reduction, F1 bandwidth increase, spectral flattening, and extra pole-zero pairs. The amplitude of the first and second harmonic, the frequency of the first formant, the nasal peak before and after the first formant, and spectral intensity measures was used in [3]. Also, Pruthi and Espy-Wilson [4] used the **e**nergy ratio, low spectral peak measure, formant density, and an onset/offset measure for automatic classification of nasals and semivowels.

A. Arabic Language

There are three forms of the Arabic language: Classical Arabic or qur'anic language, Modern Standard Arabic (MSA), and the other form is the dialectical Arabic which can be classified into different forms depending on factors such as social factors and geographical linguistics. MSA has six vowels and 28 consonants where, /i/, /u/, and /a/ are the short vowels and /ii/, /uu/, and /aa/ are the three long counterpart vowels. Besides, MSA contains two distinctive classes of phonemes; pharyngeal and emphatic, which are found only in Semitic languages such as Arabic and Hebrew. Regarding syllabic structures, MSA contains the following six possible syllabic structures CV, CVV, CVC, CVCC, CVVC, and CVVCC, where C indicates a consonant and V indicates a vowel type (long or short) [5].

B. The Holy Quran (THQ) Ghunnah

Tajweed of The Holy Quran (THQ) is the knowledge and application of the rules of recitation [4]. One of these rules is the Ghunnah. The Ghunnah is a sound that is emitted from the nasal passage without any function of the tongue. It is an unconditional nasalized sound fixed on the /n/ and /m/; it is an inherent sound in the /m/ and /n/ whether they have a vowel or not. There are four levels of the Ghunnah in THQ; most complete Ghunnah, complete Ghunnah, incomplete Ghunnah, and most incomplete *Ghunnah*. In this work we will focus only on the first one of these levels; the most complete Ghunnah, which is the longest Ghunnah, and it appears in the following cases: double /m/, double /n/, or when noon sakinnah /n/ (noon without any succeeding vowel) or tanween (extra noon pronounced but not written) is followed by one letter of the *Idghaam Shafawi(* /j/, /n/, /m/, /w/) [6].

In this work, a Multi-Layer Perceptron (MLP) was used to classify nasalized phonemes from their non-nasalized counterparts in THQ recitations by using the first three formants only.

II. DATABASE

To build an excellent database that serves our stated purpose, we faced many problems such as choosing appropriate words from THQ, selecting high-quality samples that do not need pre-processing, and some well-known reciters do not have high-quality recordings. We recommend reviewing the College of Education, Department of Quranic Studies, to obtain samples without problems of pronunciation, background noise, or echoes.

We searched for the verses in THQ that contain the Ghunnah and determined the specific words that begin with the letters of the Idgham rule (/j/, /n/, /m/, /w/) and came after /m/, /n/ (without any vowels), or *Tanween*. Also, we aimed to select pair (nasalized and non-nazalized) with the same co-articulation effect. Then, we searched for words that contain the same letters of the Idgham rule but without Ghunnah (not preceded by /m/, /n/ (without any vowels), or tanween) as shown in Table I and Table II.

TABLE I. NASALIZED SPEECH SEGMENTS EXAMPLE

Nasalized Speech Segments			
Surah No.	**Surah**	**Ayah No.**	**Words**
2	Al-Baqara	165	مَن يَتَّخِذُ
2	Al-Baqara	256	فَمَنْ يَكْفُرْ
7	Al-A'raaf	12	مِن نَّارٍ
33	Al-Ahzaab	7	وَمِن نُّوحٍ
14	Ibrahim	16	مِن مَّاء
15	Al-Hijr	54	أَن مَّسَّنِيَ
13	Ar-Ra'd	34	مِن وَاقٍ
18	Al-Kahf	26	مِن وَلِيٍّ

TABLE II. NON-NASALIZED SPEECH SEGMENTS EXAMPLE

Non-Nasalized Speech Segments			
Surah No.	**Surah**	**Ayah No.**	**Words**
3	Al-Imran	28	لَّا يَتَّخِذِ الْمُؤْمِنُونَ
2	Al-Baqara	271	وَيُكَفِّرُ عَنكُم
2	Al-Baqara	266	فِيهِ نَارٌ
3	Al-Imran	33	آدَمَ وَنُوحًا
2	Al-Baqara	22	مِنَ السَّمَاءِ مَاءً
7	Al-A'raaf	188	وَمَا مَسَّنِيَ السُّوءُ
13	Ar-Ra'd	37	وَلِيٍّ وَلَا وَاقٍ
2	Al-Baqara	257	اللَّهُ وَلِيُّ

We collected data until we reached 500 speech segments read by 5 different reciters (100 speech segments for each one). All samples of speech signals are taken from THQ recitations of well-known reciters and specific parts of THQ scripts. The first reciter is Abdullah Basfar, the second reciter is Mohammed Ayyoub, the third reciter is Ali Al-Hudhaify, the fourth reciter is Mshari Al-Afasi, and the fifth reciters is Ali Jaber. The total data equal 500 speech segments; 250 nasalized and 250 non-nasalized.

III. EXPERIMENTAL WORK

A. Artificial Neural Networks (ANN) Design

The ANN architecture plays a very important role in implementing an effective speech and phoneme recognition system. The structure cannot be measured by mathematical equations and calculations. The ANN ability to learn with examples makes it very flexible and powerful. The training process is a very important part in designing the ANN that enables it to do its job in the testing process, because of the ANN's fast response, it is useful for real-time systems or applications.

We used different architectures depending on data amount, but the system still has two possible outputs which are 1 to 0 for nasalized and -1 to 0 for non-nasalized. We investigated and experimented for the best MLP architecture and we found the best structure by practical experiment. For the final results, we took an average of 10 runs.

B. Preparation of Data and Feature Extraction

The ANN performance will certainly be affected by the amount of data and it will increase in accuracy as the amount of data increases. We used the following different sizes of datasets: 80, 120, 160, 200, 240, 280, 320, 360, 400, and 500. Each set consists of two halves, one of the nasalized speech segments and the other non-nasalized speech segments. Each set consists of two subsets: the training subset and the testing subset. The ratio between the training subset and the testing subset is 3:1. In this work, we used the first three formants by using PRAAT software [7]. we extracted 50 frames for each formant as shown in Table III.

TABLE III. INPUT CONFIGURATION

Feature configuration	Length (Rows)
F1	50
F2	50
F3	50
F2-F1	50
F3-F2	50
F3-F1	50
Total	300

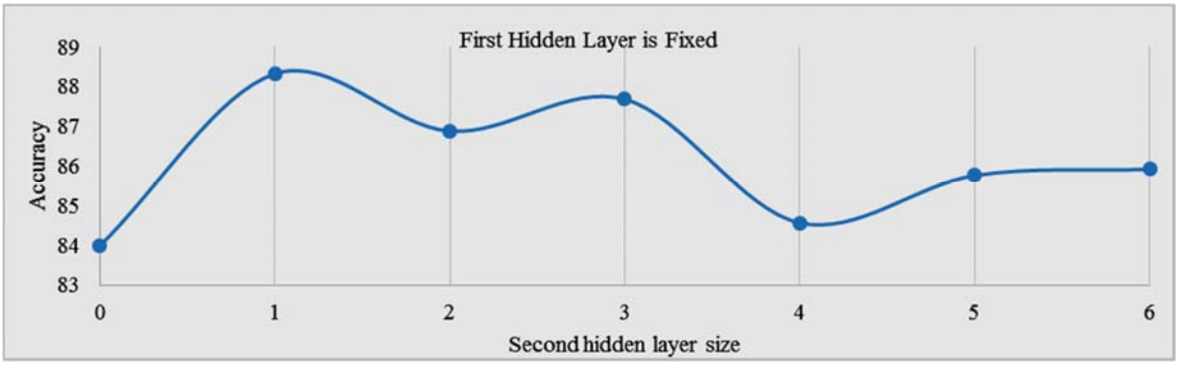

Fig. 1. First hidden layer is 3 neural nodes.

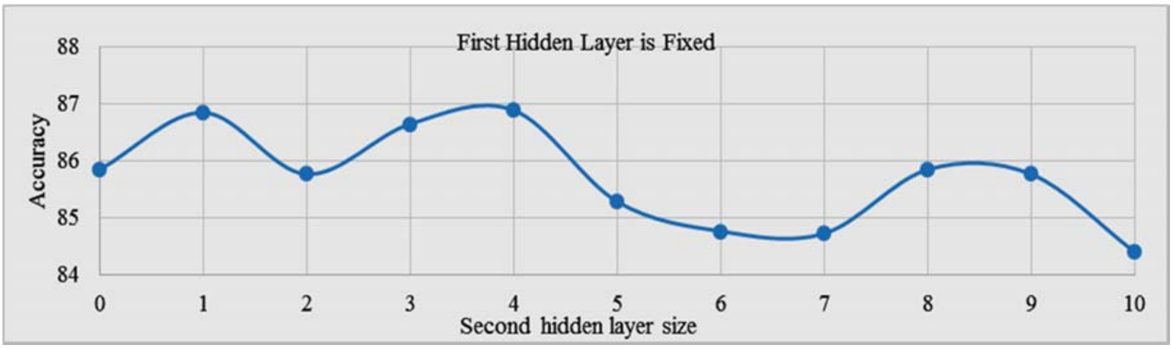

Fig. 2. First hidden layer is 5 neural nodes.

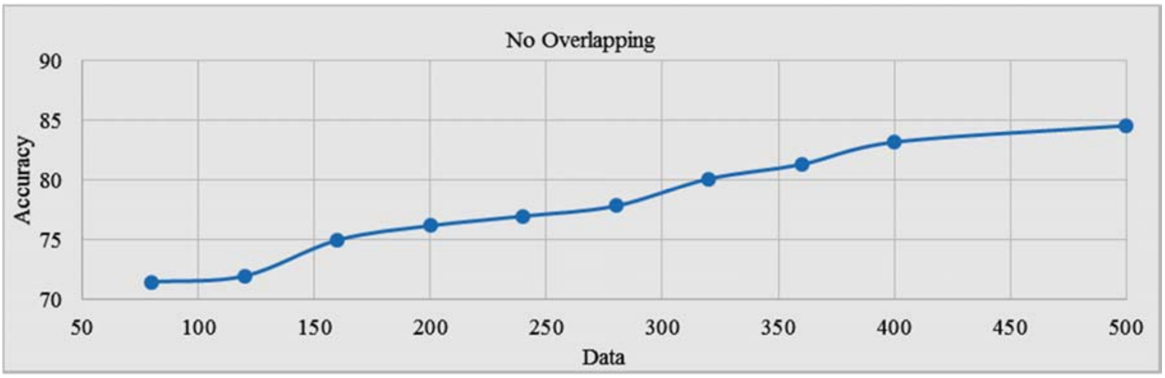

Fig. 3. Without reciters overlapping.

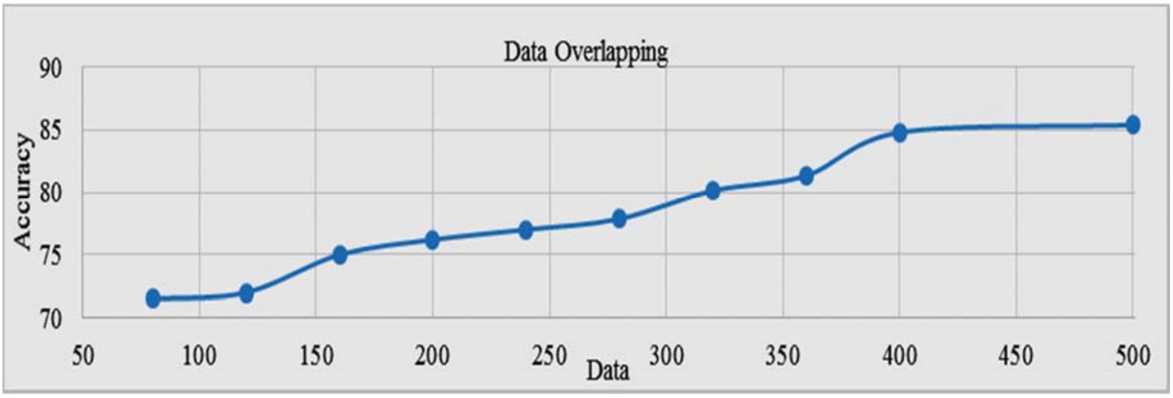

Fig. 4. With reciters overlapping.

IV. RESULTS AND DISCUSSIONS

To find the best MLP architecture, we did many practical experiments to determine the best number of MLP hidden layers and the number of neurons. Figures 1 and 2 are an example of these experiments outcomes.

The performance worsens as the number of neural nodes in both hidden layers increase. The best performance was 88.32% with ANN layer of 3-1-1 (input-hidden-output) MLP architecture while the worst was 76.8% at layering of10-20-1 MLP architecture with the data set that contains 300 speech segments.

As we mention in the above, we used the following speech corpus sizes: 80, 120, 160, 200, 240, 280, 320, 360, 400, and finally 500 for both training and testing. Figure 3 and Figure 4 show the effect of the amount of data in two cases; with and without speakers' overlap between training and testing subsets. The difference between Figure 3 and

Figure 4 is the subsets of the testing data. In Figure 3, we used three reciters for training and one different reciter for testing. In Figure 4, we used 75 speech segments from each reciter for training and 25 speech segments from each reciter for testing. So, the overlapping here means: Does the reciter have speech segments in both training and testing subsets, or can his speech segments be found only in one subset?

As we can see in Figure 3 and Figure 4, we notice that the system accuracy is increasing steadily according to the amount of data, where is the accuracy range is 71.5 % to 84.58% in the case of no overlapping between the training and testing sets, and 71.5% to 85.4% in the overlapping case. It is clear that the system accuracy is better in the case of the reciters overlapping (85.4%); clearly, overlapping makes a difference.

In general, we found the system can recognize more than 94% of the data successfully, there are only 7 remaining speech segments which the system failed to recognize properly. To answer the question of why our system failed to recognize some speech segments, we consulted linguists for their opinion on these words. After interviewing the linguists, we found that there are two main reasons for language problems: one of which is the lack of nasalized speech segment or hypernasality, the other reason is a sudden change in tone of the sound.

V. CONCLUSION

As a result of this analysis, we found that formant frequencies play an important role in the process of speech sound classification and analysis. There is a difference in formant frequencies and shapes for the same word between nasalized and non-nasalized sounds, and by using these formants, the different sounds can be recognized by both humans and machines.

We designed an MLP classifier for nasalized Arabic speech recognition. We are targeting the Arabic nasalized phonemes not the originally nasal ones, namely, /m/ and /n/. This is going to give the problem some kind of default and challenge.

Although sounds differ from person to person, MLP can detect the nasalized parts from each sound. By increasing the amount of the data in training subset, the system accuracy of the ANN will increase, and the hidden layer will have a considerable impact if we have a large amount of data. The best way to train the MLP is to use many sounds for one word, which will train the ANN to be able to detect any nasalized or non-nasalized sounds with more accuracy for a larger data set. Also, ANN can detect whether the reciter has read the word correctly or incorrectly imposing or ignoring nasalization.

To build an effective system that can recognize the nasalized sounds with high efficiency, we must use more data and different speech features to reduce the errors and make the system more reliable. Also, we must improve the system and set appropriate ANN training parameters depending on the amount of data.

ACKNOWLEDGMENT

This project was funded by the National Plan for Science, Technology, and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (11-INF1968-02).

REFERENCES

[1] S. Najnin and C. Shahnaz, “Detection and classification of nasalized vowels in noise based on cepstra derived from differential product spectrum,” Circuits, Syst. Signal Process., vol. 36, no. 1, pp. 181–201, 2017.

[2] T. Pruthi and C. Y. Espy-Wilson, “Acoustic parameters for the automatic detection of vowel nasalization,” in Eighth Annual Conference of the International Speech Communication Association, 2007.

[3] I. Dutta and A. Pandey, “Acoustics of articulatory constraints: Vowel classification and nasalization,” in Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[4] T. Pruthi and C. Espy-Wilson, “Automatic classification of nasals and semivowels,” in ICPhS 2003-15th International Congress of Phonetic Sciences, 2003, pp. 3061–3064.

[5] Y. A. Alotaibi and S.-A. Selouani, “Evaluating the MSA West Point Speech Corpus,” Int. J. Comput. Process. Lang., vol. 22, no. 4, pp. 285–304, 2009.

[6] K. C. Czerepinski and A. D. A. R. Swayd, Tajweed Rules of the Qur'an. Dar Al-Khair Islamic Books Publisher Jeddah, 2006.

[7] D. Boersma, Paul & Weenink, “Praat: doing phonetics by computer.” 2014.