Discriminant Analysis for Classification of Stressed Syllables in Arabic

A. Chentir, M. Guerti, and D. J. Hirst

Abstract— This paper proposes an approach for a Classification by Discriminant Analysis of stressed syllables in Standard Arabic. In this study, we exploited the acoustic parameters of fundamental frequency and energy by means of a classification by a discriminant analysis to detect stressed syllables of Standard Arabic words with the structure [CVCVCV] read by four native speakers (two male and two female). We obtained a global percentage of detection equal to 83% of the stressed syllables. These initial results need to be tested on larger corpora but our results suggest this could be a useful addition to existing algorithms, in the goal of improving systems of automatic synthesis and recognition in Standard Arabic.

***Index Terms*— Classification by Discriminant Analysis, lexical stress, Standard Arabic, fundamental frequency, energy, stressed syllable.**

I. Introduction

The quality of a Text to Speech Synthesis depends on the naturalness, on the intelligibility of the speech generated and the specific characteristics to the produced voice. These characteristics depend on the techniques and the methods of synthesis, but also on the care taken to linguistic and prosodic modelling. Several works underlines the fact that linguistic structures maintain the close links with the prosodic achievements. Concerning the Arab language, the models are based on the syllabic structure of the words, the stressing, the concept of markers intonates and very little on syntactic information [1].

Prosody plays an important role in the field of the identification of languages. It is also essential to the understanding and to the naturalness of speech and thus indispensable for speech synthesis. So, from the acoustic point of view, prosody refers to the phenomena linked to the variation in the time of the parameters of pitch, intensity and duration.

The perception of pitch is essentially related to fundamental frequency (F_0) which, at the physiological level of the production of the speech, corresponds to the frequency of vibration of the vocal cords. Intensity (I) is essentially connected to the energy of the sound while the acoustic duration (D) corresponds to its time of emission. These three parameters harmonize in uneven proportions to give to every language its particular prosodic characteristics.

There have been a number of studies concerning Arabic prosody and the importance of lexical stress in that language [2] – [6]. Bohas [5] showed that lexical stress plays a distinctive role. Rajouani [2] confirmed that the detection of the primary accent seems sufficient for the study of the Arabic intonation and found from his experiments, the following result: The hierarchy (F0, I, D) for the Arabic language.

In order to reinforce the existing systems of synthesis and recognition of Standard Arabic (SA), we made us in this study of a classification by discriminant analysis based on the acoustic parameters of fundamental frequency and energy to detect stressed syllables in Standard Arabic (SA) of type [CV]. Our choice was limited to three-syllable Arabic words. After manually segmenting and transcribing the corpus, we applied our algorithm based on discriminant analysis. A total percentage of detection equal to 83% of the stressed syllables was obtained, which shows the efficiency of such an approach which could reinforce existing methods based exclusively on fundamental frequency.

II. Lexical Stress in Arabic Standard

Arabic is a member of the Semitic language family. It is marked by a limited vocalic system and a rich consonantal system. There are typically three basic vowels a, I, u, which are attested in both their short and long forms. Semitic languages are also marked by a rich inventory of guttural consonants, which includes the laryngeals, the pharyngeals and the uvular fricatives [7]. The term Standard Arabic (SA) is used when referring to the literary language in general.

Although the classical Arab grammarians do not mention word stress, it is generally accepted that SA had it. For Ghalib [8], stress exists in Arabic but has no linguistic function and its importance is much less, as compared with English or German where it contributes to the meaning and grammatical function of some words of the lexicon. In Arabic, a shift of stress from one syllable to another changes neither the meaning of the word nor its grammatical function, even if such a movement can deform its correct pronunciation.

A. Syllable Structure in Standard Arabic

The allowed syllables in Arabic language are presented in Table 1, where V indicates a (long or short) vowel while C indicates a consonant. Arabic utterances can only start with a consonant.

A. Chentir is with the *University Saad Dahlab of Blida, Algeria, (e-mail: chentiramina@yahoo.fr*).
M. Guerti is with the *National Polytechnic College, Algeria, (e-mail: mhaniag@yahoo.fr*).
D. J. Hirst is with the *LPL, CNRS & University of Provence, France, (e-mail: daniel.hirst@lpl-aix.fr*)

Table 1: Classification of the syllables in Arabic [6]

Syllable	Open	Closed
Short	[CV]	
Long	[CVV]	[CVC], [CVVC], [CVCC], [CVVCC]

All Arabic syllables must contain at least one vowel. Also Arabic vowels cannot be initials and can occur either between two consonants or final in a word. Arabic syllables can be classified as short or long. The CV type is a short one while all others are long. Syllables can also be classified as open or closed. An open syllable ends with a vowel, while a closed syllable ends with a consonant. For Arabic, a vowel always forms a syllable nucleus, and there are as many syllables in a word as vowels in it [9].

B. Word Stress Pattern in StandardArabic

Arabic is a language with word stress. This means that one of the syllables in content word is perceived as prominent and receives main stress.

The most commonly used rules are those established by Al-Ani [6] who speaks about of the presence of three degrees of stress:

- A first degree or Primary Stress (PS)
- A second degree or Secondary Stress (SS)
- A third degree or Weak Stress (WS)

The position and the distribution of the stress depend on the number and the types of syllables contained in the word. The rules which govern its place are defined as follows [6]:

- If all syllables of the word are of type [CV] then it is the first syllable which carries the PS, the other syllables receive a weak stress.
  Example: دَخَلَ [daxala]
- If there is a single long syllable, then this last receives the PS.
  Example: كَافَحَ [kaafaða]
- If there are two or more long syllables, then it is the last of these (but not counting the final syllable of the word) which receives the PS. The long syllable closest to the beginning of the word receives a SS; other syllables receive an weak stress.
  Example: حَيَوَانَاتٍ [ðajawaanaatin]

III. MATERIALS AND METHODS

We used 4 native Arabic-speakers (2 male and 2 female), each pronounced 5 Arabic words (Table 2) with the following three-syllable structure [S_1 S_2 S_3]: [C_1V C_2V C_3V] where [C_1], [C_2] and [C_3], corresponded to 3 different Arabic consonants and [V] to a vowel. These words were pronounced inside 5 carrier sentences. This made a total of 20 sentences with [C_1V] always corresponding to a syllable with primary stress.

The recording was made in an anechoic recording chamber in the Laboratoire Parole et Langage (LPL) in Aix-en-Provence. The Praat computer program [11] was then used to analyse and manipulate the speech data.

Table 2: Example of used Arabic words

Words in Arabic	كَتَبَ	عَبَثَ	بَرَزَ	خَبَزَ	حَزَنَ
IPA	kataba	ʔabaθa	baraza	xabaza	Hazana

A. Methods of Classification

Methods of classification are very useful tools because they make it possible to group objects according to their resemblance. They place some objects in the same group and separate them from the others by placing them in different groups.

Three big families can be distinguished (independently of the syntactic methods) [11]:

- search for similar forms by dynamic comparison
- probability where Hidden Markov Models (HMM) and Bayesian networks are by far the most commonly used in automatic speech recognition
- surfaces of decision and discriminant functions of forms

In all these methods, the choice of the distance or metric between vector forms is important. The Euclidian distance is often used:

$$dE(x,y)=\sqrt{(x-y)'(x-y)} \quad (1)$$

But the Mahalanobis distance [12] where C is the covariance matrix of the vector forms x and y is also interesting because it allows the taking into account of the correlation between the parameters of the forms:

$$dM(x,y)=\sqrt{(x-y)'C^{-1}(x-y)} \quad (2)$$

B. Discriminant analysis

Discriminant analysis is a statistical method which aims at describing, explaining and predicting the membership in predefined groups (classes, modalities of the variable to be predicted …) of a set of observations (individuals, examples…) from a set of predictive variables (descriptors, exogenous variables...) [12].

This analysis has two main purposes:

- description: among the known groups, what are the main differences which can be determined by means of the measured variables?
- Classification: can we determine the group of membership of a new observation only from the measured variables?

In other words, discriminant analysis aims at classifying an observation in the group for which the conditional probability of its belonging to this group according to the observed values is maximal.

There are several manners to evaluate the quality of a Discriminant Analysis (DA). Some appeal to probability hypotheses, while others don't. The percentage of well classified samples is the most used statistic and also the most revealing while being the simplest.

The idea is the following: we have a procedure of classification, then why not to apply it to the observations of which we know the real group and to check if we make a correct classification from the obtained matrix of confusion.

Generally, the matrix of confusion is a picture of dimensions g x g (where g is the number of groups), where the row represent the real memberships and the columns the assignment by the model. We can track down the number of erroneous and correct assignments there. The percentage of correct assignments with regard to the total number of individuals is a global indicator. Table 3, present an explanatory example of the confusion matrix.

From the matrix of confusion (also called classification table) above, we have 160/200 = 80% of samples which are correctly well. This is a strong percentage if we consider that a classification made completely at random would give on average 50% of correct classification. Furthermore, we note that the samples of the group 1 are classified correctly for 83% while those of group 2 are classified correctly for 78%. Group 1 is thus slightly more homogeneous than group 2.

Table 3: Example of confusion matrix

Groups AD		1	2
Real	1	50	10
Groups	2	30	110

The way of obtaining an evaluation which is considered more realistic consists in putting aside a certain proportion of the initial observations of every group, and apply the classification functions to the other observations then classifying the put aside observations. Another variant consists in putting aside an observation at the same moment and repeat the analysis and the classification n times.

IV. USED APPROACH

In our approach, knowing that Arabic stress is influenced respectively by the fundamental frequency followed by the intensity, we choose to detect the stressed syllable with a Classification by Discriminant Analysis. The fact that every syllable in Arabic contains only one vowel, we extract every vowel present in our words and we then followed the following stages:

Stage1: Segmentation and phonetic transcription of the recorded words

Stage2: Extraction of the fundamental frequency, for each vowel, detected inside the used word.

Stage3: Extraction then calculation of the Long Term Average Spectrum (LTAS) for every vowel detected inside the used word

Stage4: Make a discriminant analysis to classify all the vowels in an orderly structure and create the appropriate configuration

Stage5: Generate the confusion matrix to verify the conformity of the predictive classification with reality

Stage6: Consider values for additional vowels not present in the training sample. We shall thus manage predict the values of new observations in the classification of the already existing groups

Stage7: Generate the corresponding matrix of confusion.

V. RESULTS

To be able to interpret the results, we applied the bootstrap method [13] to our corpus. Its purpose is to supply indications on statistics other one than its value (dispersal, distribution, reliable intervals) to know the precision of the realized estimations. This method is based on a technique of re-sampling, accompanied by a large number of iterations which result from the application of the Monte Carlo method [14] (Fig. 1).

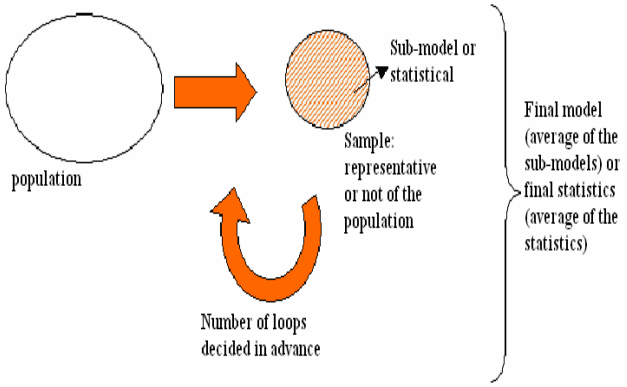

Fig. 1: Principle of the method of the Bootstrap

We proceeded to the learning of 18 of the 20 sentences of our corpus. Once carried out, we continued to the recognition of 2 sentences not included in the first phase of calculation.

To end on the efficiency of the used method, we appealed to the principle of the Bootstrap method (defined previously) and we then calculated the matrix of total confusion corresponding to the tested corpus.

For that purpose, we chose, during the learning phase, to always remove the same sentence pronounced by 2 different speakers, giving us 540 possible cases (5 x 18 sentences x C_2^4 possible combinations) and for the recognition phase, we have 60 possible cases (5 x 2 sentences x C_2^4 possible combinations).

Table 4 presents the various confusion matrices as well as the percentage of correct assignment with regard to the total number of vowels as well as that corresponding to the vowel [V] allocated to the 1st stressed syllable [S_1]. And where [V] is one other than the vowel [a].

The confusion matrix established in Table 4, is illustrated in Fig. 2, where rows represent the real membership and columns the assignments obtained by the calculated model. S_i=1…, 3, are the three syllables present in the used words.

The data and the symbols used throughout our calculation are:

- During the calculation of the Long-Term Average Spectrum, the number of used bands is limited to 6 each which a width equal to 500 Hz.
- The mean value of fundamental frequency is calculated (F_0).
- Four speakers: 2 male (H_1 and H_2) and 2 female (F_1 and F_2)
- 540 sentences in the Learning phase (540L) and 60 sentences in Recognition phase (60R)
- Sentences in recognition are the same for both speakers.

Table 4: Matrices of confusion in Learning and in Recognition phases and the percentages of affectation obtained according to every syllable

Removed sentences speakers X_1-X_2		sentences		
$X_1 - X_2$	540L	503	37	0
		34	473	33
		10	51	479
	Fraction Correct: 89.81 %			
	S1 : 93.51 %			
	S2 : 87.59 %			
	S3 : 88.70 %			
	60R	51	9	0
		5	49	6
		3	7	50
	Fraction Correct: 83.33 %			
	S1 : 85 %			
	S2 : 81.67 %			
	S3 : 83.33 %			

VI. DISCUSSION

Table 4 allows us to conclude as follows:

- The learning phase gives a good total percentage of recognition equal to 89.81%.
- The stressed syllable S_1 is classified with a good rate equal to 93.51% in learning phase.
- The global recognition phase (83.33 %) is superior to the threshold corresponding to a random classification (50%).

To show the efficiency of this method, we call for our previous work [15], where we have done only a classification of energy by a discriminant analysis to detect the primary stressed syllables (the number of used bands was limited to only 3 bands). Table 5 shows results obtained in this approach.

Table 5: Matrices of confusion in Learning and in Recognition phases and the percentages of affectation obtained according to every syllable

Removed sentences speakers X_1-X_2		sentences		
$X_1 - X_2$	540L	424	87	29
		92	320	128
		33	241	266
	Fraction Correct : 62.35%			
	S1 : 78.52%			
	S2 : 59.26%			
	S3 : 49.26%			
	60R	47	10	3
		12	33	15
		6	27	27
	Fraction Correct : 59.44%			
	S1 : 78.33%			
	S2 : 55%			
	S3 : 45%			

So, we can say that results obtained in Table 4 show the improvement which brings this new approach to detect stressed syllables. The learning phase is better than in the previous method. We noticed the good classification of S_2 and S_3 syllables while they were worst previously. The same goes for the recognition phase, where we notice the increasing of fraction correct (from 59.44 % to 83.33 %) and the good classification of all syllables.

VII. CONCLUSION

In this study, we made use of the classification by discriminant analysis based on the fundamental frequency and the energy to detect stress type for SA in syllables of type [CV]. Our choice limited itself to the three-syllabic Arabic words. After segmenting and transcribing manually the corpus, we applied our algorithm based on discriminant analysis. A percentage of fraction correct equal to 89.81% in learning phase and 83.33% in recognition phase were obtained. And a percentage of detection equal to 85 % of the primary stressed syllable was obtained. We noticed then the improvement of this method by comparison with our previous work.

It is clear that results obtained need to be tested on larger corpora of Arabic. But already, we can say that the classification by discriminant analysis of the criterion energy added to the formants could enrich the already existing methods of recognition.

REFERENCES

[1] Z. Zemirli and S. Khabet, “TAGGAR : Un analyseur morphosyntaxique destiné à la synthèse vocale de textes arabes voyellés”. JEP – TALN - RECITAL, Fes , Morocco, 19-22 April 2004. http://aune.lpl.univ-aix.fr/jep-taln04/proceed/actes/arabe.htm

[2] A. Zaki, A. Rajouani, Z. Luxey and M. Najim, “Rules Based Model for Automatic Synthesis of F0 Variation for Declarative Arabic Sentences”. In B. Bel & I. Marlien (eds) Proceedings of the First International Conference on Speech Prosody, April 11-13, 2002, Aix en Provence, France.

[3] A.M. Elgendy and L.C.W. Pols, “Mechanical versus perceptual constraints as determinants of articulatory strategy”. In Proceedings Eurospeech, Scandinavia 7th European Conference on Speech Communication and Technology, Vol.1, Aalborg, Denmark, September 3-7, 269-272, 2001.

[4] A.N. Hanna and N.A. Ghattas, “Text-to- Speech Synthesis by Diphones for Modern Standard Arabic”. An-Najah University Journal for Research, Natural Sciences (A), 2005.

[5] Bohas, G., J.-P. Guillaume and D. Kouloughli, 2006. The Arabic Linguistic Tradition. Georgetown University Press, 2006.

[6] S.H. AL-Ani, “Arabic Phonology. An Acoustical and Physiological Investigation”. Walter De Gruyter Inc., Mouton Ed., The Hague, 1970.

[7] J.C.E. Watson, “The Phonology and Morphology of Arabic”. Oxford University Press Inc., New York, 2002.

[8] G.B.M. Ghalib, G.B.M., “An Experimental Study of Consonant Gemination in Iraqi Colloquial Arabic”. Unpublished Ph.D. Thesis, University of Leeds (Department of Linguistics and Phonetics), UK, 1984.

[9] H. Satori M. Harti and N. Chenfour, “Arabic Speech Recognition System Based on CMU Sphinx”. International Symposium on Computational Intelligence and Intelligent Informatics, ISCIII’07, pp. 31-35, Agadir, Morocco, 28-30 March 2007.

[10] P. Boersma and D. Weenink, “Praat: doing phonetics by computer”, (Version 5.0.32) [Computer program], 2008. http://www.praat.org

[11] M.Y. Kiang, “A comparative assessment of classification methods. In Decision Support Systems”. Elsevier Science Publishers B. Amsterdam, The Netherlands. 35 (4), 441- 454, 2003.

[12] G.J. McLachlan, “Discriminant Analysis and Statistical Pattern Recognition”. Wiley-Interscience, 2004.

[13] B. Efron and R.J. Tibshirani, “An introduction to the Bootstrap”. Chapman and Hall/CRC, USA, 1994.

[14] D.P. Landau and K. Binder, “A guide to Monte Carlo simulations in statistical physics”. Cambridge University Press, Cambridge, 2000.

[15] A. Chentir, M. Guerti and D.J. Hirst, “Classification by Discriminant Analysis of Energy in View of the Detection of Accented Syllables in Standard Arabic”. Journal of Computer Science 4 (8), pp. 668-673, 2008.