Journal of Computer Science 4 (8): 668-673, 2008
ISSN 1549-3636

Classification by Discriminant Analysis of Energy in View of the Detection of Accented Syllables in Standard Arabic

[1,2]Chentir, A. [2]M. Guerti and [3]D.J. Hirst
[1]Electronic Deptment, Saâd Dahlab University, Blida, Algeria
[2] Electronic Deptment, National Polytechnic College (ENP), Algiers,
[3]LPL, CNRS and University of Provence, France

Abstract: Problem Statement: Current algorithms for the recognition and synthesis of Arabic prosody concentrate on identifying the primary stressed syllable of accented words on the basis of fundamental frequency. Generally, the three acoustic parameters used in prosody are: Fundamental frequency, duration and energy. **Approach:** In this study, we exploited the acoustic parameter of energy by means of a classification by a discriminant analysis to detect the primary accented syllables of Standard Arabic words with the structure [CVCVCV] read by four native speakers (two male and two female). **Results:** We obtained a percentage of detection equal to 78% of the accented syllables. **Conclusion:** These preliminary 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.

Key words: Classification by discriminant analysis, lexical accent, standard arabic, energy, accented syllable

INTRODUCTION

Arabic is a Semitic language. Standard Arabic counts 34 phonemes: 6 vowels and 28 consonants. The variety of Arabic to which we shall refer is called Unified Modern Arabic or the Standard Arabic. It is the language which is taught in the schools, and written and spoken in the official contexts.

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. 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 which, at the physiological level of the production of the speech, corresponds to the frequency of vibration of the vocal cords. Intensity is essentially connected to the energy of the sound while the acoustic duration corresponds to its time of emission [1]. These three parameters harmonize in uneven proportions to give to every language its particular prosodic characteristics.

Accentuation consists in giving prominence to one syllable in relation to those that surround it and that are thereby qualified as unaccented. For languages with fixed lexical stress, the syllable to accent is well defined. For example, in Finnish the accent is always carried on the last syllable of the word and the principle factors influencing it, are fundamental frequency (F_0) and duration (D) [2, 3]. For languages with free stress, like English, the accented syllable doesn't have a fixed position. Its place is variable and lexically specified. F_0 and intensity (I), seem to be the most highly correlated parameters [2].

In recent years, there have been a number of studies concerning Arabic prosody and the importance of lexical stress in that language [4-8]. Bohas[7] showed that lexical stress plays a distinctive role. Rajouani[4] 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 parameter of energy to detect the primary accent in SA in syllables 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 percentage of detection equal to 78% of the accented syllables was obtained, which shows the efficiency of such an approach which could

Corresponding Author: Amina Chentir, Electronic Department, Saâd Dahlab University, B.P. 270, Blida, ALGERIA.

reinforce existing methods based exclusively on fundamental frequency.

The Arabic language and lexical stress: As soon as we begin to investigate lexical stress in Arabic, we are confronted with various different observations reported in the studies of some researchers [4, 7-12].

The Arabic grammarians didn't give much consideration to the study of stress, for several reasons:

- The variety and the influence of different Arabic dialects didn't allow to a standardized account which would apply to all varieties.
- The role of lexical stress is not evident at first sight, to such point that some linguists denied its existence [4, 12]. Their argumentation was that the position of stress, if it existed, on any syllable of the word did not modify its sense.

For Ghalib[11], 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.

Different types of syllables in Standard Arabic: Any isolated word in Arabic receives an accent which will be carried on the stressed syllable. We are going to try to describe in a very simplified way the various types of syllables and the place of the stress in the word [9-11].

The structure of the syllable in Arabic is based on the properties of the phonemic system of the language. The nucleus is always the most dominant element of the syllable; it consists of a short vowel [V] or a long vowel [VV]. A syllable always begins with a consonant [C] and ends either by a silence [#], either one or two consonants (the case where the two final consonants are identical and where the final appendices of inflection [a], [u], [i], [an], [un] and [in] are omitted, this type is named pausal form).

Table 1: Classification of the syllables in Arabic

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

From this description, Al-Ani [10] describes the existence of 6 types of syllables: [CV], [CVV], [CVC], [CVCC], [CVVC] and [CVVCC] described as Short/ Long, Open/Closed (Table 1).

Place of stress in Arabic words: All authors admit that lexical stress is predictable in Arabic in the sense that it is absolutely a function of the syllabic structure of the word. There is, however, disagreement as to the rules governing the place of the stress [4].

The most commonly used rules are those established by Al-Ani [8] 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 [8]:

- 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]

MATERIALS AND METHODS

Corpus: 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.

Table 2: Example of used Arabic words

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

The recording was made in an anechoic recording chamber in the Laboratoire Parole et Langage (LPL) in Aix-en-Provence. The Praat computer program [17] was then used to analyse and manipulate the speech data.

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 [13].

Three big families can be distinguished (independently of the syntactic methods) [13]:

- 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 [14] 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:

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

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...) [14].

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.

We have available a sample of n observations distributed in K groups of workforce n_k. Let us note Y the variable to be predicted, it takes its values in $\{y_1,\ldots y_K\}$. We have J predictive variables $X = (X_1,\ldots, X_J)$ and μ_k the centres of gravity of the clouds of conditional points, with W_k their matrix of variance-covariance.

The objective is to produce a rule of assignment F: $X \Rightarrow \{y_1,\ldots,y_K\}$ which allows us to predict, for a given observation w , its associated value of Y from the values taken by X.

The essence of discriminant analysis, then, comes down to proposing an estimate of the quantity:

$$P(X/Y) = y_k \quad (3)$$

With P (X/Y): Function of density of the X conditional to the class y_k.

Evaluation of the quality of the discriminant analysis: 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.

Used approach: In our approach, we followed the following stages:

- Stage 1: Segmentation and phonetic transcription of the recorded words
- Stage 2: Extraction then calculation of the medium-term spectre for every vowel detected inside the used word
- Stage 3: Make a discriminant analysis to classify all the vowels in an orderly structure and create the appropriate configuration
- Stage 4: Generate the confusion matrix to verify the conformity of the predictive classification with reality
- Stage 5: 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
- Stage 6: Generate the corresponding matrix of confusion.

RESULTS

To be able to interpret the results, we applied the bootstrap method[15] 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 [16] (Fig. 1).

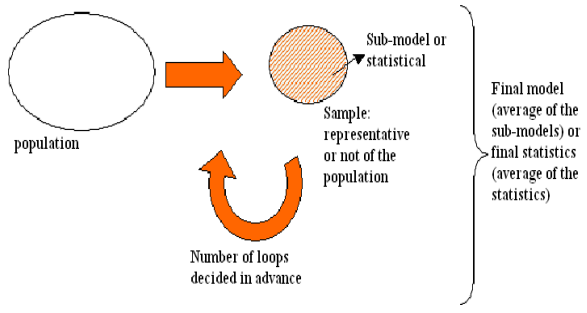

Fig. 1: Principle of the method of the Bootstrap

$$\begin{bmatrix} & S_1 & S_2 & S_3 \\ S_1 & & & \\ S_2 & & & \\ S_3 & & & \end{bmatrix}$$

Fig. 2: Look of the matrix of confusion

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.

For that purpose, we chose, during the learning phase, to always remove the same sentence pronounced by 2 different speakers, giving us 30 possible cases (5 x sentences x 6 possible combinations).

Table 4 presents an example of 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 accented syllable [S_1]. Table 4 presents the case 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 = 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 short-term spectrum, the number of used bands is limited to 3 each which a width equal to 500 Hz.
- Four speakers: 2 male (H_1 and H_2) and 2 female (F_1 and F_2)
- Twenty sentences in all (5 for every speaker): sentence 1 to sentence 5
- Eighteen sentences in the learning phase (18L) and 2 sentences in recognition (2R)
- The 2 sentences in recognition are the same for both speakers

We then calculated the matrix of total confusion for every sentence as well as the percentage of correct assignments (Table 4).

We noticed the following points:

- The correct classification obtained in the learning phase for the various vowels. We obtained a percentage of recognition superior to 50%
- The very good classification of the first vowel corresponding to the accented syllable with percentage of recognition superior to 70%

Table 4: Matrix of confusion in learning (L) and in recognition (R) and the percentages of affectation obtained according to every sentence

Removed sentence speakers X_1-X_2		Sentence 1		Sentence 2		Sentence 3		Sentence 4		Sentence 5
$X_1 - X_2$	108L	91 11 6 14 65 29 6 51 51 96.30% S1:84.26%	108L	83 20 5 21 62 25 7 48 53 61.11% S1:76.85%	108L	83 19 6 18 68 22 3 47 58 64.51% S1:76.85%	108L	84 18 6 18 65 25 7 48 53 62.35% S1:77.78%	108L	83 19 6 21 60 27 10 47 51 59.88% S1:76.85%
	12R	2 10 0 3 7 2 3 3 6 41.67% S1:16.67%	12R	9 0 3 0 8 4 0 6 6 63.89% S1 : 75%	12R	12 0 0 6 3 3 3 6 3 50% S1: 100%	12R	12 0 0 3 6 3 0 6 6 66.67% S1: 100%	12R	12 0 0 0 9 3 0 6 6 75% S1: 100%

- The bad classification of the sentence 1 during the test phase. We explain this weak result by the fact that this sentence was the first one pronounced by our 4 speakers
- The very good percentage of recognition of the vowel corresponding to the accented syllable (= 100%) for the last three sentences during the test phase

DISCUSSION

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.

We obtained then the Table 5 which allows us to conclude as follows:

- The learning phase gives a good percentage of recognition equal to 62.35%. It is clear that it is the classification of both unaccented syllables (S_2 and S_3) that are at the origin of this reduction
- The accented syllable S_1 is classified with a good rate equal to 78.52%
- The global test phase is only slightly superior to the threshold corresponding to a random classification. However, we note the very good classification of the syllable S1 (78.33%)

Table 5: Matrices of confusion in learning and in recognition and the percentages of affectation obtained according to every sentence

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

CONCLUSION

In this study, we made use of the classification by discriminant analysis based on the acoustic parameter energy to detect the primary accent in 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 detection equal to 78% of the accented syllable was obtained.

This is only a first approach to the detection of accent in standard Arabic by a discriminant analysis on the prosodic parameter of energy. The 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 can be a supplementary parameter for the detection of accented syllables which could enrich the already existing methods of recognition that are based only on the parameter of fundamental frequency.

REFERENCES

1. Monaghan, A., 2002. Prosody in Synthetic Speech: Problems, Solutions and Challenges. In: Improvements in Speech Synthesis, Part II, Issues in Prosody, Keller, Bailly, Terken and Huckvale (Eds.). John Wiley and Sons, Ltd., London, pp: 408. ISBN: 9780471499855.
2. Livonen, A., 1998. Intonation in Finnish. In: Intonations Systems: A Survey of Twenty Languages, Hirst and D. Cristo, (Eds.). Cambridge University Press, Cambridge, pp: 314-330. ISBN-10: 052139550X.
3. Hirst, D.J., 1999. Intonation in British English. In: Intonations Systems: A Survey of Twenty Languages, Hirst and D. Cristo, (Eds.). Cambridge University Press, Cambridge, pp: 500. ISBN-10: 052139550X.
4. Zaki, A., A. Rajouani, Z. Luxey and M. Najim, 2002. Rules based model for automatic synthesis of F0 variation for declarative arabic sentences. Proceedings of the 1st International Conference on Speech Prosody, April 11-13, Aix en Provence, France, pp:719-722. http://aune.lpl.univ-aix.fr/sp2002/pdf/zaki-etal.pdf.

5. Elgendy, A.M. and L.C.W. Pols, 2001. Mechanical versus perceptual constraints as determinants of articulatory strategy. Proceedings of the Eurospeech Scandinavia 7th European Conference on Speech Communication and Technology, Sept. 3-7, Aalborg, Denmark, pp: 269-272. http://www.isca-speech.org/archive/eurospeech_2001/e01_0269.html.
6. Hanna, A.N. and N.A. Ghattas, 2005. Text-to-speech synthesis by diphones for modern standard arabic. An-Najah University J. Res. Natural Sci., 19: 159-166. http://www.najah.edu/nnu_portal/researches/294.pdf.
7. Bohas, G., J.P. Guillaume and D. Kouloughli , 2006. The Arabic Linguistic Tradition. Georgetown University Press, pp: 163. ISBN-10: 158901085X.
8. AL-Ani, S.H., 1970. Arabic Phonology: An Acoustical and Physiological Investigation. Walter De Gruyter Inc., pp: 104. ISBN-10: 9027907277.
9. McCarthy, J.J., 1979. On stress and syllabification. In Linguistic Inquiry, 10: 443-465. http://cat.inist.fr/?aModele=afficheN&cpsidt=12672707.
10. Zemirli, Z., S. Khabet and M. Mosteghanem, 2007. An effective model of stressing in an arabic text to speech system. IEEE/ACS International Conference on Computer Systems and Applications, May 13-16, IEEE Xplore, Jordan, pp: 700-707. DOI: 10.1109/AICCSA.2007.370708 .
11. Ghalib, G.B.M., 1984. An Experimental Study of Consonant Gemination in Iraqi Colloquial Arabic. Unpublished Ph.D. Thesis, University of Leeds (Department of Linguistics and Phonetics), UK. http://lib.leeds.ac.uk/record=b1455013~S4
12. Blau, J., 1972. Middle an Old Arabic Material for the history of Stress in Arabic. Bull. School Oriental African Studies, 35: 476-484. http://www.jstor.org/pss/612899.
13. Melody, Kiang Y., 2003. A comparative assessment of classification methods. Decision Support Sys., 35: 441-454. DOI: 10.1016/S0167-9236(02)00110-0.
14. McLachlan, G.J. 2004. Discriminant Analysis and Statistical Pattern Recognition. Wiley-Interscience, pp: 552. ISBN-10: 0471691151.
15. Efron, B. and R.J. Tibshirani, 1994. An introduction to the Bootstrap. Chapman and Hall/CRC, USA., pp: 456. ISBN-10: 0412042312.
16. Landau, D.P. and K. Binder, 2000. A Guide to Monte Carlo Simulations in Statistical Physics. Cambridge University Press, Cambridge, pp: 398. ISBN-13: 978-0521653664.
17. Boersma, P. and D. Weenink, 2008. Praat: Doing phonetics by computer (Version 5.0.32) [Computer program]. http://www.fon.hum.uva.nl/praat/.