SPOTTING ARABIC PHONETIC FEATURES USING MODULAR CONNECTIONIST ARCHITECTURES AND A RULE-BASED SYSTEM

Sid-Ahmed SELOUANI
L.C.P Lab. Institute of electronics,
BP 32 EL Alia
U.S.T.H.B, Algiers, Algeria
e-mail:parole@ist.cerist.dz

Jean CAELEN
CLIPS/IMAG Lab.
BP 53, 38400, cedex9
U.J.F, Grenoble, France
e-mail:jean.caelen@imag.fr

Abstract

This paper reports the results of experiments in complex Arabic phonetic features identification using a rule-based system (SARPH) and modular connectionist architectures. The first technique we present, operates in the field of analytic approaches and intends to implement a relevant system for automatic segmentation and labeling through the use of finite state networks (FSN). For this task, an original ear model is used to calculate indicative features according to the phonetic and phonological matrix of standard Arabic we have established in earlier studies. The second method deals with a set of a simplified version of sub-neural-networks (SNN). A binary sub-task is assigned to these networks with the objective to recognize features as subtle as emphasis, gemination and semantically pertinent lengthening of vowels. This is proposed to be done at two different levels: at the gross level, by identifying the macro-classes, at the finer level, by detecting pertinent temporal distortion and emphasis. Serial and parallel architectures of SNN are investigated. A comparison between the two identification strategies is carried out using stimuli uttered by Algerian native speakers. The results show that SNN achieved well in rough identification while in the some cases of phonologic duration the rule-based system performs better.

1. Introduction

It is known that the standard Arabic is distinctive from Indo-European language because of its consonantal nature [1][20]. From an articulatory point of view, it is also characterized by the realization of some sounds in the rear part of the vocal tract. Another root of complexity is the pertinence of the sound duration for both vowels and consonants. These particularities added to the strong inflection of the language, constitutes a supplementary root of failure in the automatic speech recognition (ASR) systems dedicated to Arabic [8][9].

In this paper, we are concerned with the automatic recognition of phonetic macro-classes and complex features by multi-layer sub-neural-networks (SNN) and rule-based system. We will focus our experimentation on long and short vowels discrimination and the emphasis as well as the gemination detection.

The aim of the connectionist approach is to improve performances of Arabic ASR systems by integrating a mixture of neural experts individually specialized in the detection of a given feature. We are inspired by the hierarchized structure of experts introduced by Jakobs and Jordan [14][15] in order to solve non-linear regression problems. Moreover, it is worth to note that this structure has revealed more effective in the case of the spontaneous telephone speech recognition [6]. We compare the results obtained by this purely automatic approach to the one which uses the phonetic knowledge expressed by rules of the SARPH system (Arabic Recognition System using Phones)[23].

2. Phonetic knowledge

Arabic sounds can be divided into macro-classes such as: stop consonants, voiceless fricatives, voiced fricatives, nasal consonants, glide consonants, vowels, and semi vowels. The presence of emphatic and geminate consonants and the relevance of vowel duration are the main characteristics of the Arabic phonetic system.

2.1 The vowels

The vocalic system contains two phonological quantities for each tone. For each brief vowel /a/, /i/, /u/, there is respectively the associated long vowel /a:/, /i:/, /u:/. In Arabic, this temporal opposition is fundamental. For example, the two words /3amal/ “camel” and /3ama:l/ “beauty”, only have the length of the final vowel as a difference. In the traditional Arabic grammar, a long vowel is perceived as two brief vowels. El Ghazeli introduced the tension and laxity notion [11] to characterize the vocalic system. Jakobson in his *Preliminaries* [16] concluded that a tense vowel is longer than the lax vowel. In our case, we showed in [3] that the tense/lax indicative feature is very relevant for the long/brief vowel discrimination. The Caelen's model of the indicative features [4][5] permits us to quantify and encode the tense/lax feature in order to allow an automatic recognition of vowels.

2.2 Emphatic consonants

There are four emphatic consonants in Arabic: /t̲/, /s̲/, /d̲/, /δ̲/. They are articulated in the front zone of the oral cavity. The tongue root is carried against the pharynx. Acoustically they increase the first formant transition and decrease the second formant transition with contiguous vowels. By using the quantified sharp/flat feature, the discrimination between emphatic and non-emphatic consonant can be achieved [21][22].

2.3 Geminated Consonants

The gemination arises by the intensification of the articulation and the sustained (prolongation) plosive closure. In the classical approach, the gemination is simply considered as the doubling of the consonant. Obviously, if the temporal difference between the geminate consonant and the simple consonant exists, Bonnot [2] advises us "to consider other pertinent indicative features". We have proposed in [21][22] to compute the tense/lax feature in order to detect the gemination.

3. SARPH system overview

The process conducted by SARPH can be divided into the following steps:

- A set of pre-processors operates on the signal in order to extract the distinctive features through the use of an ear model [4]. A segmental function is provided by the derivation of the sum of features' cues. A procedure based on delta coding and variable thresholds ensures the homogeneous phone detection.

For each detected phone a basic process is carried out. It consists in the determination of the segment duration and energy, formants, the degree of voicing and friction, fundamental frequency (F0), indicative features and their derivatives. These features are tense/lax, grave/acute, compact/diffuse, continuous/discontinuous, flat/sharp, and mellow/strident. The indicative features and their derivatives are encoded in order to facilitate the formalization of the knowledge.

- An identification task is carried out by a finite state network (FSN) which follows the signal segmentation. Each macro-class corresponds to a phonetic network that represents the knowledge of the phonetic macro-structure.

A phonetic FSN noted Rj is defined by a 5-tuplet:

Rj = { j, S(j), T, soj, sej }

With j: network name, S(j): phonetic node set, T: transition set, soj: initial node, sej: final node.

S(j) represents all possible realizations of acoustic phases for a given macro-class. It represents all the possible ways of SARPH labeling. Each realization is defined as the following set:

S(j) = {sk, sl, pi, Ci, Ai }

With sk and sl: transition endings, pi: transition score when the transition is over, Ci: constraint set which must be verified before running the transition, Ai: action set to be executed in case of transition success. The actions are procedures (evaluate a predicate or parameter) or call other rules.

The constraints Ci are divided into three categories :

-Conditions of realization of a given phone;
-Constraints from the previous context;
-Conditions generated by the network controller during the network exploration.

For example, the fricative network has seven nodes :

-Beginning: network entry;
-Vocalic friction: beginning of friction after a vowel;
-Closure onset: closure before friction;
-Closure offset: closure after friction;
-Voiceless friction: friction without voiced sound;
-Voiced friction: friction with voiced sound;
-End: network exit.

Figure 1 shows an example for voiceless friction detection.

- The labeling process of SARPH is composed of two parts. The first one deals with the localization of the macro-classes with the help of the FSN. The context constraints are supposed realized. The second one performs the accurate recognition of emphasis and gemination for consonants and the temporal opposition for vowels. For consonant networks, the average of flat/sharp indicative feature is computed over all the phones which constitutes the detected consonant. One of 5-level non linear code values is assigned to the feature average. In this way, if the code '+' or '++' is assigned, it means that the fricative is flat and hence it will be labeled as emphatic. The tense/lax feature average is also calculated and encoded. If it is '+' or '++' then the consonant is geminate.

The same mechanism of the fine recognition is carried out for the vowels. In addition to the macro-class detection, a complete classification of the six Arabic vowels is performed. The long/brief opposition is detected by the means of the encoded tense/lax feature.

The following part is an example of transition rule. It must be performed in order to reach either emphatic or geminate “voiceless friction” node.

Voiceless friction finite state network with emphasis and gemination detection

```
If   (cue (friction)>‘+’)
      And (F0=0)
And Noise + Threshold<energy
And Energy<Noise+3/4*signal/Noise
      And ((previous state = “voiceless friction”)
              Or (previous state = “closure onset”)
              Or (previous state = “vocalic friction”)
              Or (previous context =“pause”
                   And intensity slope > slope1
                   And previous state = beginning “))
Then actual state = “ voiceless friction”
     And procedural action
     (Beginning boundary  = actual phone)
     And action rule
     (Voiceless friction or closure offset or network exit)
If ( actual state =“ network exit”)
Then
   (Compute and code sharp/flat feature And if (cue (flat)≥‘+’ then  “emphasis”)
And (compute and code tense/lax feature And if (cue(tense)≥‘+’ then  “gemination”)
```

Figure1. SARPH transition rule for voiceless friction node.

4. Macro-class recognition by neural networks

The connectionist hierarchical structure we propose consists of a simplified neural network set to which a classification sub-tasks have been assigned in order to globally identify macro-classes and Arabic phonetic features. The training of each “specialized unit” is operated on all the learning corpus. The optimization of the parameters, such as the number of the network cells, the learning constant, the number and the quality of the inputs, is done individually on each sub-network. It is by the means of a cross validation [17][25] that the adjusting of all these parameters is realized. The principle is to notice their success rate for different values of the parameter to optimize. This task is very easy thanks to the simplicity of the sub-network to train [24][27].

4.1 Initialization

The initial values of the weights and biases are determinant for the quality and timing of the learning. In our case, the Nguyen-Widrow [19] procedure is used. Let n being the input units number and q the hidden units number The Nguyen-Widrow procedure consists firstly in initializing the hidden units weights, noted w’ij, at values comprised between -µ and +µ (typically µ=0.5 and j=0,…,q; input units indexed by i). Afterwards, in order to determine the retained weights and biases for the learning phase initialization, we define a level factor noted β with β=0,7q$^{1/n}$. Then, we calculate the normalization factor of the hidden layer, ||w’j ||. The retained weights for the learning initialization are finally expressed by:

$$w_{ij} = \frac{\beta w'_{ij}}{\left\| w'_{j} \right\|}$$

The biases are initialized by random values comprised between –β and +β. This initialization requires the use of a bipolar activation function with –0.8 and +0.8 as targets. We have noticed an improvement of the learning speed of 20 times with more reliability in contrast with the purely random initialization.

4.2 Input normalization

The dynamic temporal management by the multi-layer perceptron (MLP)-type networks remain their main weak points [12][26]. The networks are not capable to manage the temporal distortions which have not been learned. In the speech case, each segment consists of a variable number of frames. In the case of a static classification where the network architecture is stiffened, this difficulty must be discarded [7][28]. The system proposed here avoids this difficulty: it divides each segment into three intervals (onset-stabilization-end) on which we calculate the mean of the acoustic vectors. The first and the last interval include the contextual information (right and left). When the division result is not an integer, the middle interval is extended by the number of remaining frames (the stable phases are favored). Consequently, the number of parameters presented at the input is always fixed whatever the length of the segment. It will be always equal to three times of the acoustic vector size.

If m is the number of frames by a segment and p the acoustic vector size, then :

$$n_1 = n_3 = m/3$$

and

$$n_2 = m/3 + (m \equiv 3)$$

n_1, n_2, n_3, being respectively the number of frames on the first, the second and the third interval on which is operated the average of the parameters vectors. The inputs E_j presented to the sub-network are given by the following expression:

For k={0,...,p-1},

$$E_j = \begin{cases} \frac{1}{n_1}\sum_{i=0}^{n1-1} c_{ijk} & for \quad j = \{0,..., \ p-1\} \\ \frac{1}{n_2}\sum_{n1}^{n1+n2-1} c_{ijk} & for \quad j = \{p,..., 2p-1\} \\ \frac{1}{n_3}\sum_{n1+n2}^{m-1} c_{ijk} & for \quad j = \{2p,..., 3p-1\} \end{cases}$$

C_{ijk}: K uple component of the vector of the frame i, participating to the input j calculation. The number of input units will be 3p.

4.3 Acoustic analysis

Different acoustical analyses have been tested. The purpose is to determine the one which gives the best compromise between the learning speed and the generalization capability. For this purpose, a cross validation corpus has been established. It consists of 414 vowels, 246 fricatives, 214 plosives, 106 nasals and 101 liquids. The sub-network validation has been operated by using the linear predictive coding cepstral coefficients (LPCC), the perceptual linear predictive coefficients (PLP) [13], the energy (En) and the zero crossing rate (ZCR) as well as their first derivatives (of En and ZCR). The validation is operated on the vowel classification sub-network. The assigned task is the classification of the short vowels of the Arabic language. The PLP coefficients combined with the energy, ZCR and their derivatives are those which give the best result as it is shown in Table 1.

Acoustical analysis	Number of inputs	rate
36 LPCC	36	89 %
15 PLP	15	87 %
36 LPCC+3ZCR+3 En	42	90 %
15 PLP+3ZCR+3En	21	91 %
36 LPCC+3ZCR+3En+3dEn+3dZCR	48	91 %
15 PLP+3ZCR+3En+3dEn+3dZCR	27	92 %

Table1. Performance of the vowel sub-network with different kinds of acoustical analysis.

The validation results show a distinct superiority of the auditory modeling opposed to the classical modeling. A gain in the learning time is also noticed. Since the size of the network has been also diminished, the number of weights and biases to store will decrease strongly.

4.4 Architecture and disposition of experts

The same material (learning and validation corpus) is used to determine the optimal number of the hidden units. We notice that the performances are decreasing from a certain threshold linked to the number of hidden units. For example, this value is approximately 38 for the fricative network. At the input both LPCC coefficients, ZCR, energy and their derivatives are used. Table 2 gives the success rate in validation for each sub-network with a proper optimal architecture. The inputs are the PLP coefficients, the ZCR, the energy and their first derivatives. These results permit *a posteriori* to organize into hierarchy the sub-networks according to their performances.

Sub-network	Architecture	failure	Success	Rate
Vowel/consonant	27-15-2	12	634	98 %
Vowel /a/,/u/,/i/	27-25-3	17	397	96 %
Fricatives	27-18-2	20	226	92 %
Plosives	27-20-2	39	175	82 %
Nasals	27-20-2	22	84	79 %
Liquids	21-15-2	23	78	77 %

Table 2. Macro-classes identification rate obtained in the cross validation.

During the learning, a flow of data segmented in macro-classes is presented at the network input. Since the learning is supervised, the data base is also labeled in macro-classes.

4.4.1 Serial structure

This structure is made up of serial disposition of MLP expert networks. Two types of classification of unknown sequences are accomplished [10]. A rough classification whose the objective is the macro-classes detection such as vowels, fricatives, plosives, nasals and liquids. A second classification, which is finer, makes an attempt to discriminate between long and brief vowels and detect the geminate aspect on all the macro-classes and the emphatic aspect on both plosives and fricatives. The progress in depth in the structure (cf. Figure 2) is conditioned by the opening of the logical gates (and/or). According to the activation of the two outputs for a given expert network, the process starts the identification of the emphatic feature or the geminate feature if the macro-class is detected, otherwise an activation of the contiguous network is operated. A failure of the overall system is accounted if the last network is reached without any discrimination.

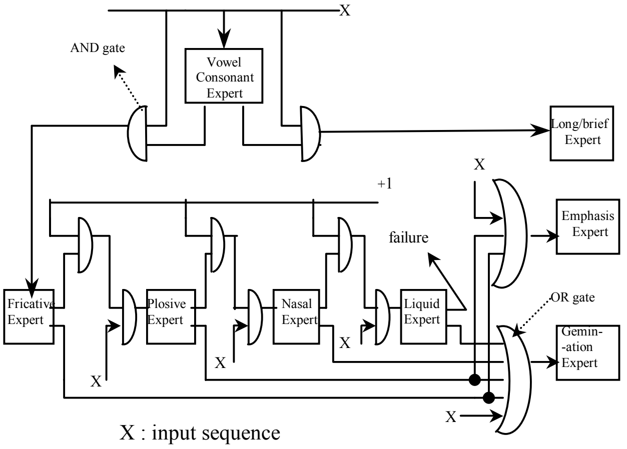

Figure 2. Serial connectionist structure for the identification of Arabic phonetic macro-classes and features.

The hierarchized structure (pipe-line) of this system seems inadequate since the network located deeply (the far right in the figure 2) are penalized. An architecture which puts the networks in the same level of competence seems less constraining. However, we can note that the experts of the highest levels (the far left in fig 2) justify their position in the test system by their high success rate during the cross validation. So, the penalization of the following levels is minimized.

4.4.2 Parallel structure

This configuration is established without any fixed condition concerning the relative individual performances of each expert. Experts have the same activation potential when a sequence to identify is presented at the input. This property allows a flexible use and permits to avoid the failure case. During the recognition phase it is required from each expert only a specialization in the scanning of one (and only one) phonetic class. A simple logic performs a final identification (decision) after processing activated outputs provided by the all binary discriminators.

This logic permits, contrarily to the serial configuration, to manage the cases in which several experts are simultaneously activating. This procedure is summarized as follows:

Considering S_{j1} and S_{j2}.

S_{j1} being the output 1 of the expert j. It takes state +1 if the feature (assigned to that expert j) is detected. In that case, we say that the expert is activated.

S_{j2} being the output 2 of the expert j. It takes state +1 if the aimed phonetic feature is absent.

The recognized class is the one whose the expert is activated. In the case in which the number of activated experts (noted K) is superior to 1, then the recognized class is given by :

$$\text{Argmax}_j\,(S_{j,1}-S_{j,2}) \qquad \text{with} \qquad j=1,\dots,K$$

Afterwards, the long/brief expert is activated if (and only if) the vowel output is retained. Otherwise, in the consonant case, whatever the activated expert, the "gemination expert" of the superior level is solicited. Oppositely, the emphatic expert is not solicited only if a plosive or a fricative has been detected by the inferior levels. This is carried out according to the Arabic grammar properties.

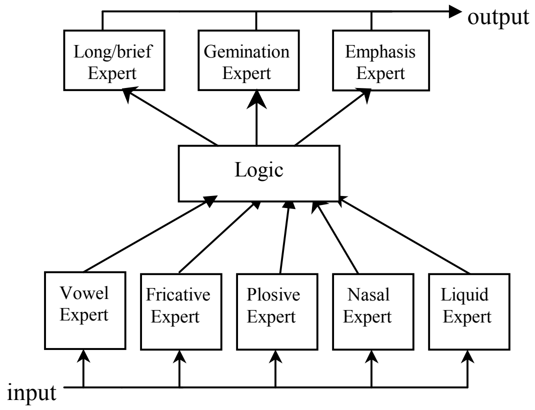

Figure 3. Parallel connectionist structure for the identification of the Arabic phonetic macro-classes and features.

5. Results and comments

The test corpus has been pronounced by six Algerian native speakers (3 men and 3 women). These speakers have participated to the learning and the cross validation. The stimuli are composed of 40 VCV utterances and 30 phrases, where the phoneme appearance frequencies are respected [18].

The test concerns :

- 14 fricatives : /f/, /s/, /s̲/, /z/, /h/, /ħ /, /ʃ/, /θ/, /χ/, /δ/, /δ̲/, /γ/, /ε/, /3/;
- 8 plosives : /t/, /t̲/, /k/, /b/, /d/, /d̲/, /q/, /?/;
- 2 liquids : /l/, /r/;
- 2 nasals : /m/, /n/;
- 3 short vowels : /a/, /u/, /i/;
- 3 long vowels : /aa/, /uu/, /ii/.

As a whole, the test has concerned 1082 vowels, 762 fricatives, 592 plosives, 304 nasals and 308 liquids. The semi-vowels are assimilated to their corresponding vowels. An additional sequence of 219 VCV utterances whose consonant is a geminate fricative has been tested.

The number of emphatic consonant (fricatives and plosives) tested is 184.

We must recall here that SARPH has permitted to validate experimentally (by the identification) some important phonetic aspects of the Arabic language. However, we have to note that the basis of its knowledge requires very often-empirical thresholds, which depend on the experimentation conditions (microphone, signal on noise ratio, utterance speed, etc.). This requires a rigorous management of an important number of parameters.

Either serial or parallel architectures realize mediocre scores in the particularly case of glottal and velar plosives /?/ and /q/. The rear fricatives (/h/, / ħ /, /γ/, /ε/) also cause problems. Their shortness and their sensibility to the utterance speed (co-articulation effects) make them merged into the vocalic context. We can conclude that VCV (Vowel-Consonant-Vowel) utterances are unfavorable material for the learning of this type of sounds by SNN (the omission percentage is very high).

We have noticed the total failure of all systems in the identification of the emphatic aspect for the consonant /d̲/. The explanation does not reside in the difficulty inherent to the consonant acoustical proprieties, but rather in the capability of the speaker to pronounce correctly. In fact, in a VCV context, it is very difficult to keep the emphatic character of /d̲/ and more often, it is its opposite by this feature (/d/) which is achieved[1].

Considering the obtained results (cf. Table 3), it seems clear that the brut serial and parallel configuration with respectively 16 % and 13 % of the mean error rate, are more performing than the SARPH system (18 %).

Class System	Vow.	Plos.	Fri.	Nas.	Liq.	Emp.	Gem.
SARPH	4.0	18.5	13.3	22.0	28.9	19.6	22.8
Serial SNN	2.0	16.1	11.9	16.4	19.9	16.8	32.7
Parallel SNN	2.0	13.7	10.2	14.3	13.8	13.9	26.9

Table 3. Error rate (in %) of the serial and parallel connectionist architectures and SARPH system. (Vow:Vowel, Plos:Plosives, Fri:Fricatives, Nas:Nasals, Liq:Liquids, Emp:Emphatic, Gem: Geminate).

In the serial architecture, the nasality is correctly detected in 84 % of cases but we have remarked that the cases of bad detection are generally due to the previous levels.

In the case of vowels, plosives, fricatives and liquids, the difference between scores is always in favor of the sub-networks. Parallel connectionist structure remain more reliable than serial structure with a difference of +3%, +2%, +2% and +6% respectively for plosive, fricative, nasal and liquid cases.

The correct detection rate of emphatic consonant is 81% for SARPH, 83% for serial structure and 86% for parallel structure.

SARPH detect gemination with a successful rate of 77%. At the opposite, serial and parallel neural systems have turned out to be a little failing with a rate of respectively 68% and 73%. We think it is due to the fact that the duration parameter which characterizes this feature is not integrated by this type of systems.

In the SARPH system, the long/short vowel distinction is realized with 78% success cases. A complete vowel detection is performed by using an algorithm of formants tracking (on the LPC spectrum). For SNNs, less than 68% of the success rate has been reached for serial connectionist configuration and 72 % is attained by the parallel structure.

6. Conclusion

We have presented the identification results of Arabic macro-classes by two systems having completely different strategies. The first one is based on phonetic rules and the second on a mixture of neural experts. These latter are composed of sub-neural-networks to which some binary discrimination tasks (∈ to the macro-class or ∉ to the macro-class) have been assigned. Two types of architecture are presented: serial structure of experts and parallel disposition of them.

Our objective is to test on Arabic language, the ability of the automatic systems operating a “blind” classification for detecting aspects as subtle as gemination, emphasis and relevant extension of vowels. These systems have been confronted to those operating an “intelligent classification” monitored by a human expert (SARPH knowledge).

In regard of obtained results, we can conclude that in the detection of complex phonetic features such as the phonological duration (long vowels and the gemination), the rule-based system remains more performing. At the opposite, when a rough discrimination is solicited (macro-class discrimination), neural networks are more adapted. In this latter case, we have to note that parallel architecture of SNN is the most reliable system.

[1] this defect is mainly due to the characteristic of the Algiers regional accent.

The connectionist mixture of experts we proposed is advantageous by the fact that it eases the learning because the binary discrimination does not need a large number of cycles (it is approximately 400). The generalization to the identification of other features as speaker gender, voiced-unvoiced marker, etc, may constitute a simple and powerful way to improve ASR systems.

References

[1] S.H. El-Ani, "Arabic phonology : an acoustical and physiological investigation", Mouton ed., the Hague, 1970.

[2] J.F. Bonnot, "Etude expérimentale de certains aspects de la gémination et de l'emphase en arabe", travaux de l'institut phonétique de Strasbourg, N°11, 1979, pp. 109-118.

[3] B. Boudraa, and S.A. Selouani**,** "Matrices phonétiques et matrices phonologiques arabes", proceedings XXèmes JEP, Tregastel, France, june 1994, pp. 345-350.

[4] J. Caelen, "Un modèle d'oreille, analyse de la parole continue, reconnaissance phonémique", thèse de doctorat ès Sciences, Toulouse, 1979.

[5] J. Caelen, and H. Tattegrain "Le décodeur acoustico-phonétique dans le projet DIRA", proceedings XIIèmes JEP, Nancy,1988, pp 115-121.

[6] G.D. Cook, S.R. Waterhouse, and A.J Robinson, "Ensemble methods for connectionist acoustic modeling", ESCA, Eurospeech97, Rhodes, Greece, 1997, pp 1959-1962.

[7] L. Devilliers, "Reconnaissance de parole continue avec un système hybride neuronal et markovien", thèse de doctorat, Paris XI Orsay, (1992).

[8] M. Djoudi, D. Fohr, J.P. Haton, "Phonetic study for automatic recognition of Arabic", European Conference on speech and technology, 1989, pp. 268-271.

[9] O. Emam, "Speech recognition of Arabic" Technical notice on www-page of IBM Cairo scientific center, 1997.

[10] P. Gallinari., S. Thiria, and F. Badran, "On the relations between discriminant analysis and Multi-Layer Perceptrons". neural networks Vol 4, 1991, pp. 349-360.

[11] S. Ghazeli, "Du statut des voyelles en arabe", études arabes N°2-3, 1979, pp. 199-219.

[12] J.P. Haton, "Modèles neuronaux et hybrides en reconnaissance de la parole : état des recherches", fondements et perspectives en traitement automatique de la parole, éditions H. Méloni, 1995, pp. 139-154.

[13] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech", Journal Acouc. Soc. Am. N° 87 (4), 1990, pp.1738-1752.

[14] R.A. Jacobs "Methods for combining experts probability assessments", Neural computation, Volume 7(5), 1995, pp. 867-888.

[15] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton "Adaptative mixtures of local experts", Neural computation, Volume 3(1), 1991, pp. 79-87.

[16] R. Jakobson, G.M. Fant, and M. Halle, "Preliminaries to speech analysis: The distinctive features and their correlates", MIT press, Cambridge, 1963.

[17] A. Krogh, and J. Vadelsby, "Neural networks ensembles, cross validation, and active learning", Advances in Neural information processing Systems, Volume 7, MIT press, 1995.

[18] M. Mrayati., "Statistical studies of Arabic roots", Applied Arabic linguistics and signal and information processing, Hamshire publishing, 1987.

[19] D. Nguyen, B. Widrow, "Improving the learning speed of two-layer neural networks by choosing initial values of the adaptative weights", International Joint Conference on Neural Networks, San Diego, CA, Vol. III, 1990, pp.21-26.

[20] A. Roman. "Le système phonologique de l'arabe classique au VIII eme siecle d'apres el Kitab de Sibawayhi", Travaux de l'institut de phonetique, Aix en provence, vol. 2, 1975, pp. 203-232.

[21] S.A. Selouani, and J. Caelen, "Experiments on Arabic phone recognition using automatically derived indicative features", IVth ISSPA, Gold coast, Australia, 1996.

[22] S.A. Selouani, and J. Caelen., "Validation de traits phonétiques par un système de reconnaissance de l'arabe standard", Proceedings XXI JEP, Avignon France, june 1996, pp. 347-350.

[23] S.A. Selouani, and J. Caelen, "Experiment in automatic speech recognition of standard Arabic", Proceedings of KFUPM workshop on information and computer science, Dhahran Saudi Arabia, 1996, pp. 161-171.

[24] K. Takuya., and T. Shuji, "Simplified sub-neural-networks for accurate phoneme recognition', proceedings ICSLP, Yokohama Japan, 1994, pp. 1571-1574.

[25] M. Stone, "Cross validatory choice and assessment of statistical predictions", Journal of the royal statistical society series B, Volume 36, 1974, pp. 111-147.

[26] J. Tebelskis, "Speech recognition using neural Networks", PHD thesis, CMU, Pittsburgh, Pennsylvania, 1995.

[27] S. R. Waterhouse and G.D. Cook, "Ensembles for phoneme classification ", Advances in Neural information processing Systems, Volume 9, MIT press, 1996.

[28] R.L. Watrous, and L. Shastri, "Learning phonetic features using connexionist networks: an experiment in speech recognition", proceedings ICASSP Vol 4, San Diego California, June 1987, pp. 381-388.