SIMPLIFIED SUB-NEURAL NETWORKS FOR PHONETIC FEATURES RECOGNITION OF ARABIC

This paper deals with a new indicatives features recognition system for Arabic witch uses a set of a simplified version of sub-neural networks (SNN). We have assigned classification sub-tasks to those sub-networks in order to globally identify the macro-classes and the Arabic phonetic aspects.

This approach leads to many advantages by the fact that the training phase become a straightforward task and consequently the SNN system can be incorporated easily in the information retrieval packages wich use human voice as media. In addition to that, the distributed structure seems to be appropriate to characterize the Arabic language in the field of phoneme recognition.

For the analysis of speech the perceptual linear predictive (PLP) technique is used. This technique use concepts from psychophysics of hearing to derive an estimate of the auditory spectrum. The performance of the system has been tested in experiments using 20 phonetically balanced phrases and 40 stimuli of CVCV non words uttered by 6 (3 male and 3 female) Algerian native speakers. Our interest goes to the particularities of Arabic such as geminate and emphatic consonants and vowel duration. The results show that SNN achieved well in pure identification and they proved that the automatic approach we carried out, constitutes a very promising and cheap way to integrate an Arabic voice recognition system in the modern packages dedicated to automatic translation and/or dictation.

KEYWORDS :

Neural networks - Arabic language - speech recognition - perceptual analysis.

1. Introduction

Les caractéristiques essentielles des réseaux neuromimétiques sont en général, leur capacité d'apprentissage à partir d'exemples, leur adaptabilité, leur robustesse aux données bruitées ou manquantes et en reconnaissance de la parole, leur puissance de discrimination pour diviser l'espace de paramètres acoustiques en classes phonétiques [11]. De nombreuses implémentations de ces réseaux ont été proposées dans la littérature [3]. La structure la plus répandue est celle du réseau multi-couches (Multi-layer Perception, MLP) ce type de réseau est capable d'apprendre et de généraliser sur les relations complexes et non linéaires reliant l'espace des vecteurs acoustiques et les classes phonétiques que l'on désire reconnaître.

Dans cet article, il est question de la reconnaissance automatique par des sous-réseaux neuronaux à multi-couches de macro-classes phonétiques de l'arabe. Nous focaliserons nos expérimentations sur les classes spécifiques à la langue arabe. Ces macro-classes sont les voyelles longues et brèves, les fricatives emphatiques ainsi que les fricatives géminées.

2- Reconnaissance de macro-classes par les réseaux de neurones

Toute utilisation des réseaux neuronaux implique nécessairement un ajustement des paramètres concernant la stratégie d'apprentissage ainsi que l'architecture du réseau. A défaut de méthode automatique de recherche de ces paramètres (tels que le nombre cachés, le nombre d'itérations, les constantes d'apprentissage etc...), nous proposons une méthode heuristique basée sur un aspect modulaire des réseaux de neurones auxquelles nous attribuons une sous-tâche de l'identification globale des macro-classes et une optimisation indépendante de ces modules. Celle-ci est réalisée au moyen d'une validation croisée qui permettra d'ajuster tous les paramètres des réseaux en observant leur taux de réussite pour différentes valeurs du paramètre à optimiser. Cette tâche est très aisée du fait de la simplicité du sous-réseau à entraîner. Le codage des séquences d'entrées

ainsi que la gestion de l'aspect dynamique de la parole sont à effectuer.

2.1. Le conditionnement

La structure du MLP comporte plusieurs couches, ce qui permet de séparer des formes non linéairement séparables. Cependant, il a été démontré que toute fonction continue de {-1, +1}n dans R, pouvait être approchée uniformément par un réseau possédant une seule couche cachée d'unités sigmoïdes.

Concernant la stratégie d'apprentissage, nous avons opté dans nos expériences pour l'utilisation d'un gradient quasi-stochastique. En effet, nous avons choisi de modifier les poids et biais du réseau après un cycle d'apprentissage afin que l'ordre de présentation des exemples à l'entrée n'influe pas sur l'apprentissage. Par ailleurs, cette approche nous affranchit de l'utilisation de facteurs tels que le moment d'inertie et la décroissance exponentielle. Ceux-ci étant surtout utiles pour les gradients déterministes afin d'éviter de tomber dans un minimum local.

2.2. l'initialisation

Les valeurs initiales des poids et biais sont déterminantes pour le temps et la qualité de l'apprentissage. Il faut se situer entre les deux cas extrêmes : allouer aux poids des valeurs initiales trop faibles a pour conséquence un apprentissage très long, en revanche des valeurs trop élevées entraînant une saturation des cellules sans qu'il y ait apprentissage effectif. Dans notre cas, une procédure particulière d'initialisation des poids et biais est utilisée : celle de NGUYEN-WINDROW [10]. Celle-ci consiste à initialiser tout d'abord à des valeurs comprises entre $-\mu$ et $+\mu$, les poids des unités notés w'ij, avec (i=1, ...,n) et (j=1,...,p), n étant le nombre d'unités d'entrées, p le nombre d'unités cachées. Pour déterminer les poids et biais retenus pour l'initialisation de l'apprentissage, nous définissons un facteur d'échelle noté β, avec $\beta=0.7p\ 1/n$, puis nous calculons le facteur de normalisation de la couche cachée, $||w'j||$. Les poids retenus pour l'initialisation de

l'apprentissage sont finalement données par l'expression suivante :

$$wij = \frac{\beta w'ij}{|w'j|}$$

Les biais sont initialisés à des valeurs aléatoires comprises entre $-\beta$ et $+\beta$.

Cette initialisation requiert l'utilisation de la fonction d'activation bipolaire. Comme le montre le tableau1 pour le problème du XOR, une réduction non négligeable du temps d'apprentissage est observée dans ce cas.

	Initialisation Aléatoire	Initialisation NGUYEN-WIDROW
XOR binaire	2891	1935
XOR bipolaire (-1, +1)	387	224
XOR bipolaire (-0.8, +0.8)	264	127

Tableau 1. Temps d'apprentissage (en nombre de cycles) pour différentes configurations de l'initialisation des réseaux

2.3. La normalisation des entrées

La gestion de la dynamique temporelle par les réseaux du type MLP reste leur grand point faible. Ceux-ci ne sont pas capables de gérer les distorions temporelles non apprises. Dans le cas de la parole, chaque segment contient un nombre variable de trames. Dans le cas d'une classification statique où l'architecture des réseaux est figée, cette difficulté doit être levée. Le système proposée ici, s'en affranchit après avoir dans un premier temps, segmenté les données sur les phases sables de chaque phonème, puis en divisant chaque segment en trois intervalles sur lesquelles on effectue un moyennage des vecteurs acoustiques. Par conséquent, le nombre de paramètres présentés à l'entrée est toujours fixe quelque soit la longueur du segment. Le nombre d'unités d'entrées sera toujours égal à 3 fois la taille du vecteur acoustique de la trame. Cette procédure est illustrée par le schéma donné en figure 1. une procédure particulière gère les cas extrêmes où

un nombre inférieur à 3 trames par segment est rencontré[1].

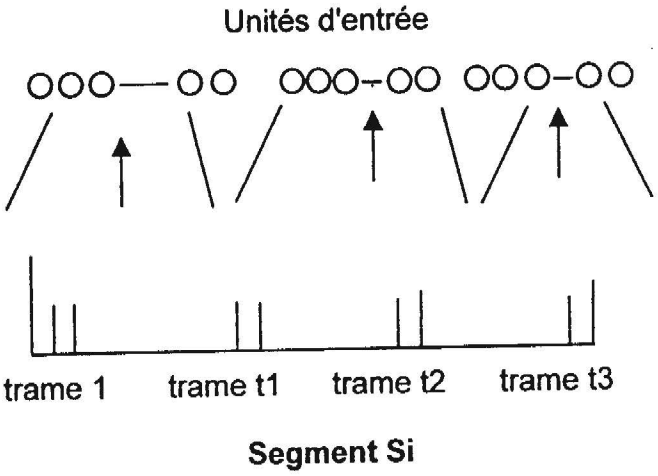

Figure 1. Schéma de la normalisation temporelle des données.

Si m est le nombre de trames par segment et P la dimension du vecteur acoustique, alors :

$$n1 = n3 \equiv m/3$$

et

$$n2 = m/3 + (m = 3)$$

n1, n2 et n3 étant respectivement le nombre de trames sur le premier, deuxième et troisième intervalle sur lequel s'effectue la moyenne des vecteurs de paramètres. Les entrées *Ej* présentées aux sous-réseaux sont données par l'expression suivante:

Pour K= {0,...,P-1}

$$Ej = \begin{cases} \frac{1}{n_1} \sum_{i=0}^{n1-1} C_{ijk} & \text{pour } j = \{0,1,\ldots,P-1\} \\ \frac{1}{n_2} \sum_{n1}^{n1+n2-1} C_{ijk} & \text{pour } j = \{P,\ldots,2P-1\} \\ \frac{1}{n_3} \sum_{n1+n2}^{n1-1} C_{ijk} & \text{pour } j = \{2P,\ldots,3P-1\} \end{cases}$$

Cijk : k ième composante du vecteur de la trame i, participant au calcul de l'entrée j.

Le nombre d'unités d'entrés sera de 3P.

2.4. Analyse acoustique et type d'entrées

Différentes analyses acoustiques ont été testées. Le but étant de déterminer celle qui donnerait le meilleur compromis entre la rapidité d'apprentissage et la capacité de généralisation. Pour ce faire, un corpus de validation croisée a été établi. Un apprentissage des sous-réseaux a été effectué en

[1] Ce cas ne peut être évité en écoutant la durée de trame

utilisant les coefficients LPCC (coefficients cepstraux de prédiction) [4, l'énergie et le taux de passage par zéro ainsi que leurs dérivées premières. L'idée de base de l'analyse PLP est d'effectuer un pré-traitement sur le signal vocal dans le but d'en extraire une représentation reproduisant les caractéristiques de la réponse du système auditif humain. Une analyse par prédiction linéaire classique (LPC) est ensuite appliquée sur le signal résultant.

Les coefficients PLP combinés avec les dérivés de l'énergie et le TPZ sont ceux qui donnent les meilleurs résultats comme le montre le tableau 2. la validation est effectuée sur le sous-réseau de classification des voyelles. Pour ce faire, un corpus de validation croisée a été établi. Il est constitué de 414 voyelles, 246 fricatives, 214 plosives, 106 nasales et 101 liquides. La validation des sous-réseaux a été effectuée en utilisant des coefficients LPCC (coefficients cepstraux de prédiction), les coefficients PLP (perceptuels), l'énergie (En), et le taux de passage par zéro (TPZ) ainsi que leurs dérivées premières (dEn et dTPZ). Les mêmes conditions d'expériences (nombre d'itérations, constantes d'apprentissage, ...) ont été utilisées lors des différentes expérimentations. Les résultats de la validation montrent une nette supériorité de la modélisation auditive par rapport à la modélisation classique. Un gain en temps d'apprentissage (le nombre d'unités d'entrée est réduit) est également observé. La taille du réseau ayant également diminué le nombre de poids et biais à mémoriser diminuera sensiblement.

Types d'entrées	Unités d'entrées	Taux
36 LPPCC	36	89 %
15PLP	15	87 %
36LPCC + 3TPZ + 3EN	42	90 %
15 PLP + 3TPZ + 3EN	21	91 %
36LPC +3TPZ +3EN++3dEN+3dTPZ	48	91 %
15PLP+ +3TPZ +3EN++3Den+3dTPZ	27	92 %

Tableau 2 : Performance des sous-réseaux avec différents types d'analyses. Evaluation sur 384 segments de voyelles brèves

2.5 Architecture et constante d'apprentissage

Le même matériau (corpus d'apprentissage et de validation) est utilisé pour déterminer le nombre optimal d'unités cachées. Nous remarquons (courbe en figure 2) que

les performances de réseau après avoir augmenté, se dégradent à partir d'un certain seuil du nombre d'unités cachées par exemple, cette valeur est approximativement de 38 pour le réseau des fricatives. Cette expérience est effectuée en utilisant l'analyseur le plus performant à savoir celui utilisant les coefficients PLP, le TPZ, l'énergie et leurs dérivées.

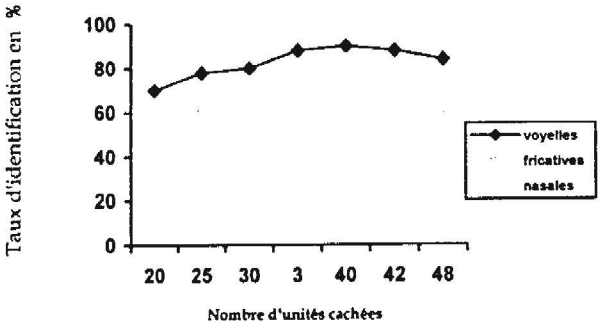

figure 2 : Validation du nombre optimal d'unités cachées

Un autre problème est le choix de la constante d'apprentissage du gradient notée généralement η . Une validation des sous-réseaux a été effectuée pour différentes valeurs de η avec le même nombre d'itérations. Le temps d'apprentissage dépend fortement du choix de cette constante. Pour les conditions d'initialisation énoncées précédemment, les résultats donnés en figure 3, montrent qu'une valeur de 0.4 pour η est un bon compromis entre le temps d'apprentissage et la généralisation.

Les différentes formes à apprendre sont présentées de façon alternées. L'approche que nous avons utilisée a pour avantage de faciliter l'apprentissage car la tâche de discrimination binaire ne nécessite pas un grand nombre de cycles et chaque sous-réseau est 'initié' indépendamment des autres. Lors de la phase de reconnaissance, il n'est exigé de lui qu'une spécialisation dans le repérage d'une seule (et une seule) classe phonétique parmi les autres.

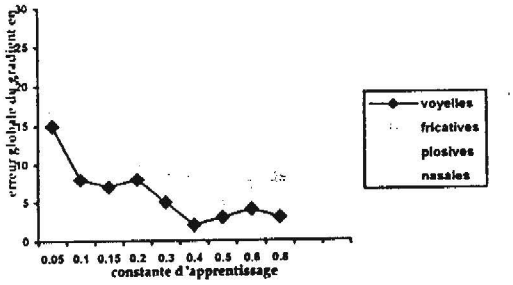

figure 3 Erreur globale en fonction de la constante d'apprentissage

3. Description générale du système

La structure du système que nous proposons (schématisée en figure 4) est constituée d'un ensemble de réseaux de neurones simplifiés auxquels nous avons attribué des sous-tâches de classification en vue de l'identification globale des macro-classes et de traits phonétiques de l'arabe. Chaque sous-réseau se spécialise dans la discrimination d'une classe (et une seule) par rapport aux autres. L'entraînement de chaque unité spécialisée s'effectue sur tout le corpus d'apprentissage.

Lors de l'apprentissage, un flot de données segmentées en macro-classes est présenté à l'entrée des réseaux. L'apprentissage étant supervisé, la base de données est également étiquetée en macro-classes voyelles, fricatives, plosives, nasales et liquides notées respectivement:V, S. Q. N et L.

3.1 L'identification

Deux types de classification de séquences inconnues sont effectuées. L'une grossière, ayant pour objectif la détection des macro-classes (V, S, Q, N, L), la seconde plus fine, tente de déceler le trait d'emphase et de gémination sur les plosives et fricatives. Un affinement de la détection des voyelles est également opéré par un réseau spécialisé.

Après avoir subi une normalisation temporelle les vecteurs acoustiques sont tout d'abord injectés dans le réseau Voyelle (noté 1 dans la figure 4) chargé de la discrimination grossière entre voyelle et consonne. Ils progressent ensuite dans les réseaux successifs S, Q, N et L (de gauche à droite et notés respectivement 2, 3, 4, 5 dans la figure 4). Selon l'activation des deux sorties d'un réseau donné, le processus s'arrête si la macro-classe est décelée sinon une activation du réseau adjacent est actionnée. Un échec du système global est comptabilisé si le dernier réseau est atteint sans qu'il y ait discrimination. Lorsque la macro-classe Q ou S est détectée les réseaux emphase (noté 2') et gémination (noté 3') sont activés.

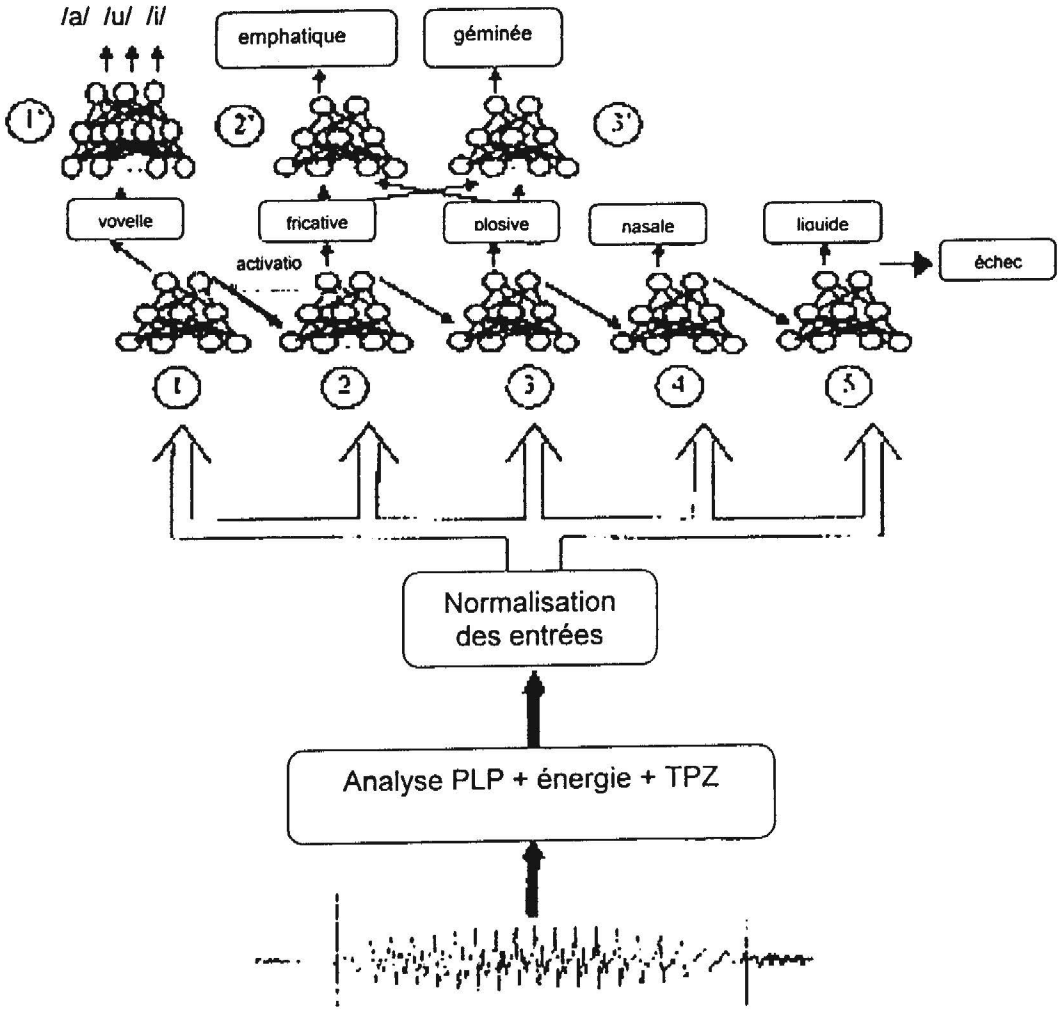

Figure 4. structure du système neuronal de détection des macro-classes arabes.

3.3 Expériences de classification

La méthode de validation croisée, utile pour affiner une architecture de réseau et déterminer les conditions optimales de fonctionnement, nous permet également d'hiérarchiser les sous-réseaux selon leur compétence à déceler la macro-classe considérée. En effet, au vu des résultats d'identification obtenus par chaque sous-réseau sur une partie du corpus d'apprentissage (cf tableau 3), il apparaît que certaines tâches sont plus aisément remplies que d'autres. Dans le système de test, les différents sous-réseaux sont sollicités par ordre décroissant de leur compétence.

Sous-réseau	Architecture	Succès	Echec	Taux
Voyelle-consonne	27-35-2	634	12	98%
Voyelles	27-35-3	397	17	92%
Fricatives	27-38-2	226	20	92%
Plosives	27-40-2	175	39	82%
Nasales	27-40-2	84	22	79%
Liquides	27-35-2	78	23	77%

Tableau 3.Taux moyen d'identification des macro-classes en validation croisée.

Dans le système hiérarchisé proposé (cf figure 4), une simple détection des voyelles est sollicitée dans un premier temps. Cette classification est rendue plus aisée par le fait que la langue arabe est une langue essentiellement consonantique et ne contient donc que peu de voyelles (3 voyelles brèves :/a/, /u/, /i/ et leurs correspondantes longues : /aa/, /uu/ et /ii/). Un deuxième argument plaide en faveur de cette discrimination et le fait que la langue arabe ne contient pas de voyelles nasales. La confusion voyelle / consonne due à la nasalité n'a pas lieu d'être.

L'architecture et la disposition de la macro-classe dans le système global dépendent des résultats de la validation croisée obtenu par chaque sous-réseau (cf tableau 3). La structure hiérarchisée (pipeline) de ce système peut sembler inadéquate dans la mesure où les réseaux situés en profondeur (les plus à droite) sont pénalisés. Une architecture parallèle mettant au même niveau de

compétence les réseaux paraît moins contraignante. Cependant, dans le cas particulier de la langue arabe, des argument plaident au contraire pour la structure que nous avons adoptée. Ces arguments sont les suivants :

- les fricatives représentent à elles seules 50 % du système consonantique arabe ;
- les plosives se réalisent dans des lieux d'articulation dispersés (uvulaire, glottale, vélaire alvéolaires, et bilabiale). Il n'existe pas de /p/ ni de /g/ en arabe ;
- il n'existe pas de voyelles nasales. Si une nasalité est rencontrée, elle ne peut être que la réalisation d'une consonne nasale ;
- les classes situées dans les bas niveaux (liquides et nasales) contiennent peu d'éléments (2 liquides, 2 nasales).

Ainsi, les tâches dévolues aux niveaux supérieurs sont caractérisées par 3 aspects importants qui justifient leur position dans le système de test (cf figure 4) et qui minimisent la pénalisation des niveaux inférieurs, ce sont :

- Les fréquences d'apparition des macro-classes lors de l'élocution (à eux seuls les 2 premiers niveaux :voyelles et fricatives traitent environ 90 % des cas, l'élocution)
- la simplicité de leur tâche de discrimination car les cas traités sont dispersés dans le lieu d'articulation (pour les fricatives et les plosives)
- leur taux de réussite lors de la validation croisée.

Lorsque une identification arrive au dernier réseau (celui des liquides) sans possibilité de classification, une ambiguïté est décrétée. Dans nos expériences, ce dernier cas est considéré comme un échec. Dans le cas de la discrimination voyelle brève / voyelle longue, nous pouvons préjuger des difficultés de notre système à la réaliser. En effet, et c'est là un problème majeur des réseaux de neurones: l'intégration de la composante temporelle des événements acoustiques. Des réseaux

spécialisés dans la détection des traits phonétiques d'emphase et de gémination sont adjoints dans la perspective de mesurer l'aptitude des réseaux neuromimétiques à déceler ce type de traits phonétiques (très subtils) spécifiques à la langue arabe. L'architecture de ces 2 sous-réseaux est analogue à celle des plosives (27-40-2).

4. Résultats et commentaires

L'originalité de la phonétique arabe se fonde, pour une grande partie sur la pertinence de la durée dans le système vocalique et sur la présence de consonnes emphatiques. Une autre caractéristique déterminante est la gémination. Celle-ci joue un rôle fondamental dans le développement morphologique nominal et verbal. Ces aspects particuliers focaliseront notre intérêt dans la comparaison des deux systèmes d'identification des macro-classes.

Le corpus de test a été prononcé l'apprentissage et à la validation croisée. Les stimuli sont constitués de 40 occurrences VCV et de 20 phrases (Arabe standard) où les fréquences d'apparition des phonèmes, sont respectées [7]. Le test concerne : les 14-fricatives :

ف،ث،س،ص،ز،ح،ه،ش،ث،خ،ذ،ظ،غ،ع،ج

soit respectivement en AP[2] : /f/, /s/, /s/,/z/,/h/, /$\overline{h}$/, /ʃ/, /θ/, /χ/, /δ/, /δ/, /γ/,/ε/,/3/

- les 8 plosives :

ت،ط،ك،ب،د،ض،ق،ء

- soit respectivement en API :

/t/,/$\underline{t}$/,/k/,/b/,/d/,/$\underline{d}$/,/q/,/?/ ;

- les 2 liquides : /l,ل/ ; /r,ر/ ;

- les2nasales: /n,ن/ ; /m,م/ ;

- les 3 voyelles brèves : /a/, /u/, /i/;

- les 3 voyelles longues :/aa/, /uu/, /ii/.

Au total, le test a porté sur 852 voyelles, 384 fricatives, 248 plosives, 164 nasales et 168 liquides.

Les semi-voyelles sont assimilées aux voyelles correspondantes.

Une suite supplémentaire de 108 occurrences VCV dont la consonne est une fricative géminée a été testée. Le choix de tenter la détection de la gémination indépendamment des autres consonnes est délibéré, dans la mesure où phonétiquement il n'existe pas de consonnes géminées (les utiliser dans

le corpus général déséquilibrerait phonétiquement celui-ci). Le nombre de consonnes emphatiques (fricatives et plosives) testé est de 84.

Pour les plosives et particulièrement / ?/ et /q/ des scores médiocres ont été réalisés. Pour les fricatives ce sont les fricatives arrières (h/, /h /, /δ/, /ε/) qui posent le plus de problèmes. Leur brièveté et leur sensibilité au débit (effet de coarticulation) font qu'elles sont le plus souvent fondues dans le contexte vocalique. Les logatomes VCV constituent un matériau défavorable pour l'apprentissage de ce type de fricatives car le pourcentage d'omissions est très élevé. Il faut également, comme c'est le pour certaines plosives (glottales et vélaires), attacher la plus grande importance à la segmentation aux phases d'apprentissage et de test.

Les nasalité est détectée en moyenne dans 72 % des cas. Les cas de mauvaise détection sont dans beaucoup de cas dus aux niveaux précédents. Les scores moyens obtenus individuellement sur chacun des six locuteurs pour les différentes macro-classes sont donnés en figure 5. Il est également possible d'affiner la détection de macro-classe en ajoutant pour les consonnes, un sous-réseau spécialisé dans la détection du voisement.

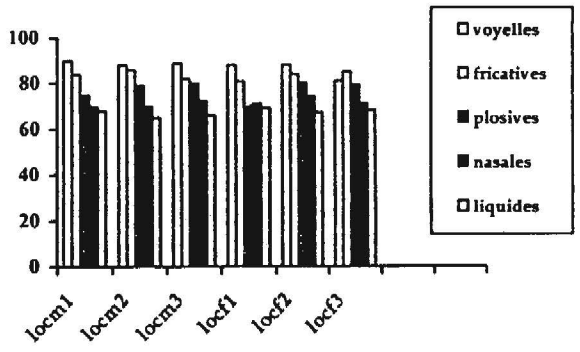

Figure 5. Résultats de classification par locuteur des sous-réseaux spécialisés.

4.1 Détection de l'emphase

L'emphase est un trait phonétique caractérisant 4 consonnes, 2 plosives : /t̲/,/d/ et 2 fricatives : /δ̲/, /s̲/).

Ces consonnes sont articulées dans la partie antérieure de la cavité buccale, la racine de la langue est reportée en arrière contre la parois pharyngale postérieure et un creusement de la langue est observé. Acoustiquement, elles se caractérisent par l'élévation de la transition de FI et

la baisse de la transition de F2 de la voyelle précédente et suivante.

Le taux de détection correcte est de 73% . Notons la défaillance totale du système dans l'identification de ce trait pour la consonne /d/. L'explication n'est pas dans une difficulté inhérente aux propriétés .acoustiques de la consonne, mais plutôt dans la capacité des locuteurs à la prononcer correctement. En effet. dans un contexte VCV, il est très difficile de garder le caractère emphatique de /d/ et le plus souvent c'est son opposé par ce trait, /d/ qui est réalisée[2] .

4.2 Détection de la gémination et de la durée phonologique

Concernant la durée phonologique, double problème se pose en reconnaissance automatique de l'arabe : Il faut déceler les phonèmes allongés tout en s'assurant que ce prolongement est pertinent c'est à dire en le distinguant des allongements dus au débit d'élocution, à un accent particulier du locuteur etc.... par exemple, les deux mots /jamal/ (chameau) et /jamaal/ (beauté) ne diffèrent que par l'allongement de la voyelle finale. On exige du système de reconnaissance de déceler les 2 voyelles sans altérer la propriété temporelle. Un alignement temporel, au contraire, pénaliserait cette détection.

Quant à la gémination, nous avons montré, confirmant les thèses de Bonnot [2], de Jackobson [5] et de Boudraa & Selouani [1] et contrairement à l'école traditionaliste des grammairiens arabes qui considère que le trait de gémination est un simple dédoublement de la consonne, que l'indice tendu/lâche peut déceler ce trait. Ceci a permis d'ailleurs, au système à base de connaissances (SARPM que nous avons développé en [8][9], d'opérer une discrimination géminée-simple d'une manière satisfaisante. Par contre, le système neuronal hiérarchisé s'est avéré quelque peu défaillant avec un taux de 56%. Nous pensons améliorer ce taux en intégrant un paramètre quantifiant l'indice tendu-lâche. Lic même problème s'est posé dans la détermination du trait de durée pour les voyelles. Cette tâche est réalisée au moyen d'un sous-réseau spécialisé qui est initié par le réseau supérieur des voyelles. Le taux obtenu avoisine les 58 %. De la même

[2] *déformation qui est d'ailleurs caractéristique de l'accent régional algérois.*

manière que pour le trait de Gémination, une amélioration de ce score est attendue en incluant dans les versions futures du système, des indices acoustiques à mêmes d'améliorer les performances des sous-réseaux.

5. Conclusions et perspectives

Nous avons présenté les résultats d'identification des macro-classes de l'arabe par un système ayant une stratégie de reconnaissance basée sur une structure hiérarchisée de sous-réseaux de neurones état de fait est compensé par la simplicité auxquels on a alloué des tâches de discrimination binaire ($\in$ à la macro- classe ou $\notin$ à la macro-classe). Notre souci majeur dans la conception d'un tel Système étant la simplicité des réseaux la facilité de leur apprentissage et la souplesse d'utilisation. La base de données est certes très réduite mais elle est suffisante Pour le Propos de cet article.

Nos expérimentations ont posé dans le cas de la langue arabe le Problème l'aptitude des systèmes (tout) automatiques de classification (aveugle) à déceler des traits aussi subtils que la gémination, l'emphase et l'allongement pertinent des voyelles. Au regard des résultats obtenus, nous pouvons conclure que dans la détection de traits Phonétiques (fins) tels que la durée Phonologique (voyelles longues et gémination) l'approche doit être perfectionnée. Par contre, lorsqu'une discrimination grossière est sollicitée (discrimination des macro-classes), les réseaux connexionnistes bien adaptés. Un compromis consiste peut être, à utiliser les méthodes connexionnistes en injectant à l'entrée non pas des données brutes mais en incluant des connaissances à priori sur les formes à classifier. Nos réseaux de PMC s'y prêtent, grâce à leur structure très souple qui Permet d'ajouter des informations de nature différente à l'entrée. Les critiques que l'on peut porter sur notre système sont celles inhérentes aux méthodes connexionnistes en général. Celles-ci se heurtent au problème de passage d'un espace qui a subit des distorsions temporelles à un espace discret de symboles. Il s'agit d'un double problème de classification et de segmentation. Cette dernière

conditionne d'une manière certaine les performances du système. Ceci a été vérifié pour le cas des consonnes glottales et vélaires. Notons que les sous-réseaux mis en jeu dans notre système sont optimisés séparément et la solution globale est sous-optimale. Cet état de fait compensé par la simplicité de la tâche exigée des sous-réseaux. L'approche proposée affinée (en incluant des réseaux par genre de locuteur, un réseau de Boisement, architecture parallèle, etc ...) peut servir comme système d'appoint à d'autres techniques complémentaires, performantes dans la normalisation temporelle telles que les HMM.

Références

1. Boudraa B.,Selouani s.A, matrices phonétiques et matrices phonologiques arabes" XXèmes JEP Tregastel (1994).
2. Bonnot J.F 'étude expérimentale de certains aspects de la gémination et de l'emphase en arabe,, travaux de L'institut phonétique de Strasbourg N11, pp. 109-1 18, (1979).
3. Haton JP, 'Modèles neuronaux et hybrides en reconnaissance de la parole: état des recherches,, in fondements et perspectives en TAP, H. Méloni éd. PP 139-154, 1995.
4. Hermansky H., 'perceptual linear predictive (PLP) analysis of speech', JASA Journal, 87 (4), pp 1738-1752,1990.
5. Jakobson R.,Fant, G.M., Halle M., « preliminaires to speech analysis : The distinctive features and their correlates", MIT press, 1963.
6. Komori Y., Hatazaki K., Tanaka T. 'Phoneme recognition expertsystem using spectrogram reading knowledge and neural networks' technical report of ATR laboratories, Kyoto, Japan, 1990.
7. Mrayati M- 'Staistical studies of Arabic roots-, Applied Arabic liguinstics and signal information processing, Hamshire Publishing, (1987).
8. Selouani i S-A, Caelen J?, Experiments on Arabic phone recognition usiong automatically derived indicative features', IV th ISSPA, God coast, Australia August 1996.
9. Selouani S.A Caelen J., 'validation de traits, phonétiques par un système de reconnaissance de l'arabe standard', Proc des XXI JEP, Avignon, France, pp 347350, Juin 1996.
10. TakuYa K-, Shuji T.,'Simplified sub-neural-networks for accurate phoneme recognition, ICSLP,Yokohama, Japan 1571-1574,(1994).
11. Watrous R-L, Shastri L., 'learning Phonetic features using connexionist networks : an experiment in speech regonition', Vol 4, IEEE, San diégo, california, june 21-24 pp 381-388,(1987).