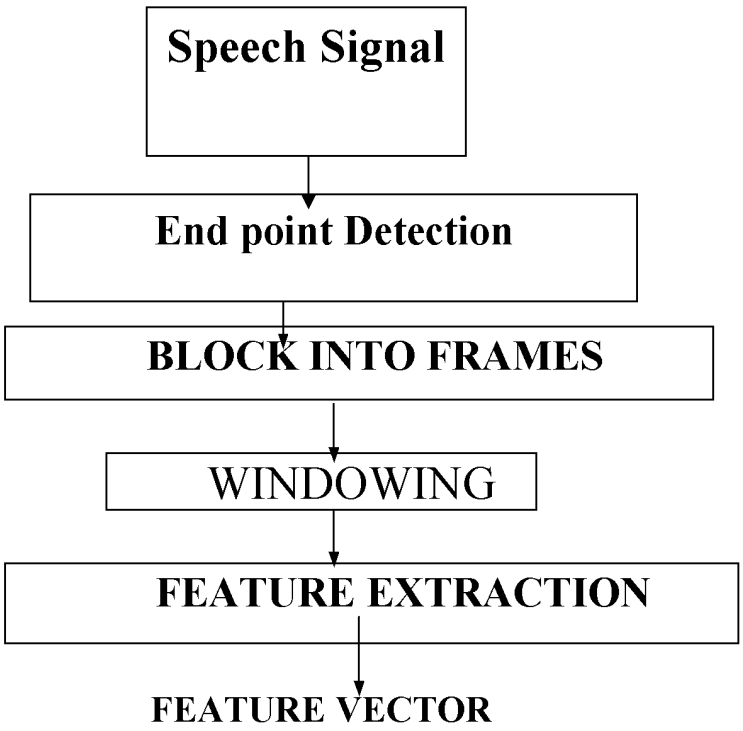
إدراك الحديث بالعربية باستخدام الحاسب الآلي

أ. د. محمد محمد حامد ، م. داليا وفيق
كلية الهندسة – بور سعيد - مصر معهد التكنولوجيا بالعاشر من رمضان - مصر

١- ملخص

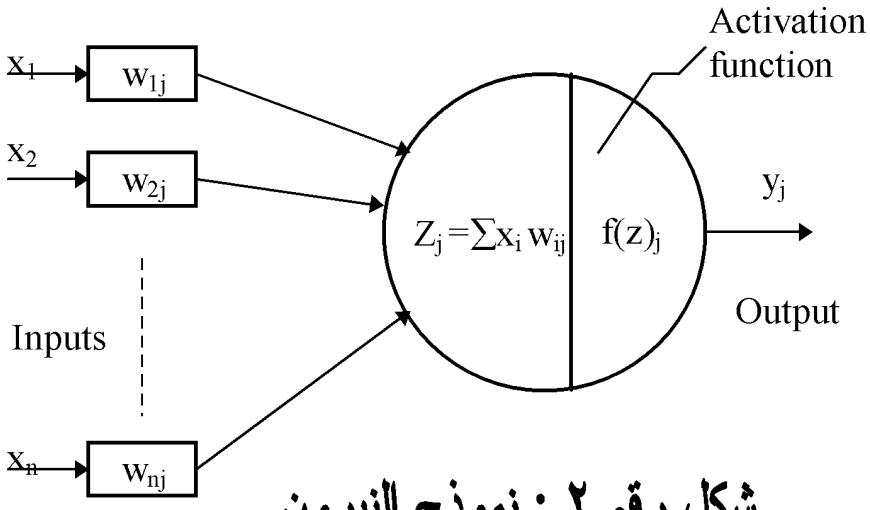
يقدم البحث منظومة جديدة للتعرف علي الحديث باللغة العربية باستخدام الحاسبات الآلية وذلك اعتمادا علي الشبكات العصبية الممثلة رياضيا علي الحاسب والتي تتكون من ثلاث طبقات متتالية متكاملة وبها نظم التحكم ومنظومات فرعية. تم الاستناد إلي أسلوب التجزيء لتقليل عدد الخواص المطلوبة عند التوزيع للكلمات تحت الاختيار مما سوف يقلل بدرجة كبيرة الزمن المطلوب لعمليات التشغيل علي الحاسب والمجهود المبذول لاستخلاص الخواص.



شكل رقم ١: خطوات التشغيل للتعرف علي الكلمات العربية

٢- مقدمة

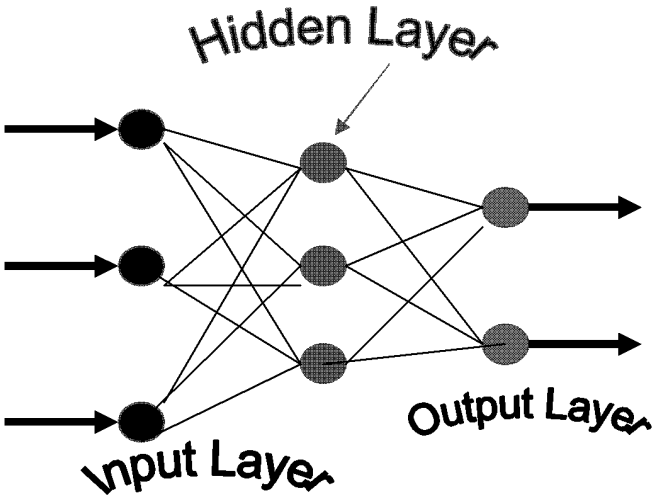
مع التطور العلمي المذهل في تطبيقات الحاسبات نحتاج إلي تطوير نظم التعامل مع اللغة العربية والتعرف عليها مباشرة دون تدخل العنصر البشري فكان علينا الاستعانة بنظم التعرف علي الحديث آليا Automatic



شكل رقم ٢ : نموذج النيرون

speech recognition (ASR) إضافة إلي أن تلك النظم مثل ASR system والمتميز بالقدرة علي التعامل مع اللغات والاتصال بين الإنسان والآلة.

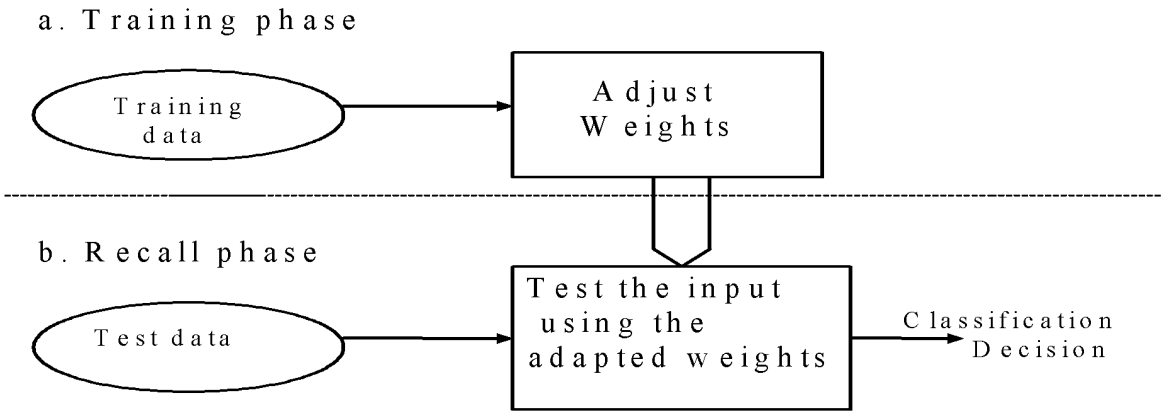
كما تهتم الدراسات الصوتية وتطبيقاتها الصناعية ببيانات وتعليمات أولية للحصول علي معدلات التمييز المناسبة للأداء الفعلي. تعتبر هذه العملية خطوة أساسية إلي الأمام نحو الاستغناء عن لوحة المفاتيح بالحاسب لتصل الكلمات مكتوبة مباشرة من خلال أجهزة الاستقبال الصوتية المعنية بهذا الموضوع [١].



شكل رقم ٣ : شبكة عصبية ثلاثية الطبقات

٣- عمليات التشغيل للحديث بالعربية

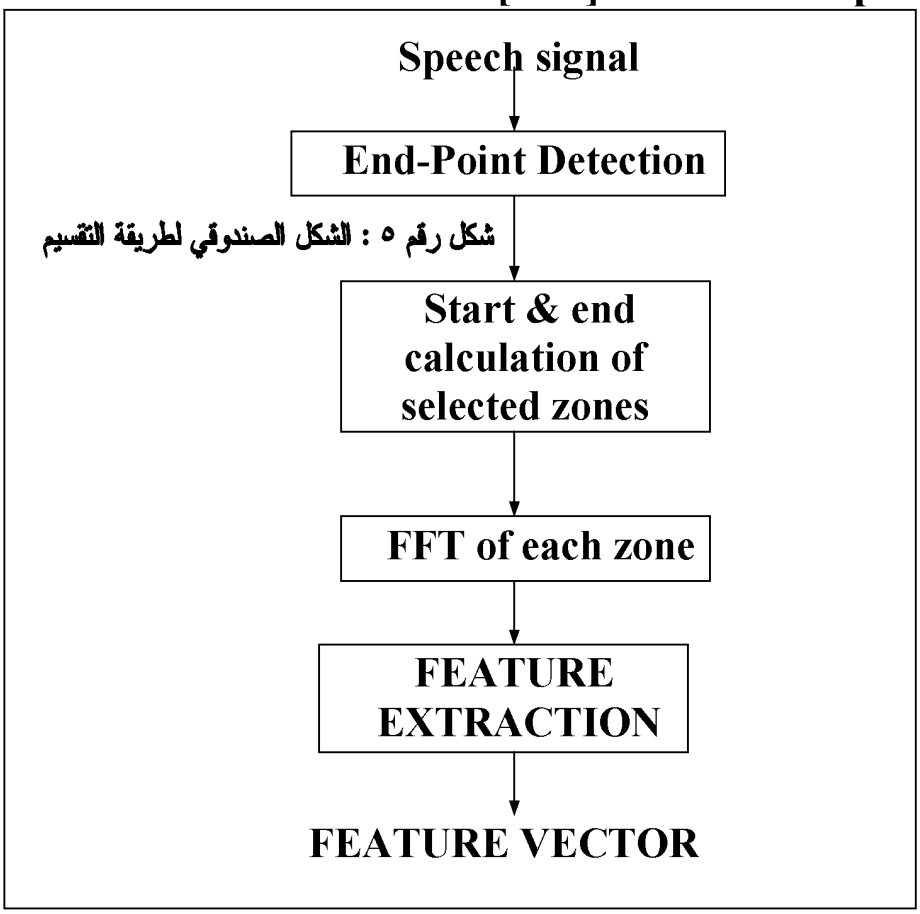
تتكون الإشارة الصوتية من عدد هائل من البيانات منها الصحيح بجانب تلك الشوائب والتي تحتاج إلي أسلوب تقني مخفض ومنقي لها لاستخراج البيانات الأساسية الصحيحة فقط ويستخدم في ذلك نظم علمية لتشغيل البيانات واستخراج الخواص مثل preprocessing or feature extraction (شكل رقم ١). باستخدام نافذة التنعيم يمكن وضع أوزان للإطارات الصوتية لتحديد نهايتها $W(n)$ ، والنظم ثلاثية النافذة هي الأكثر شيوعا مثل النافذة المستطيلة $W_R(n)$ أو نافذة هاونينج $W_N(n)$، أو نافذة هامينج $W_H(n)$ والتي يعبر عنها رياضيا بالمعادلة [٢]:



شكل رقم ٤

$$W_H(n) = 0.54 - 0.46\ COS\ (2\prod n/(N_{fs}-1)) \quad (1)$$

تتنوع طرق التحليل والتشغيل علي الحاسب علي إلي إما تحليل معاملات Parametric Analysis (PA) أو تحليل غير المعاملات Non-Parametric Analysis (NPA)، والأخير ينقسم إلي نوعين فالأول وهو الممكن في النطاق الزمني كطريقة الطاقة قصيرة الزمن أو معدل العبور الصفري Zero-crossing rate (ZCR) أو أسلوب تعديل الدالة الآلي Auto correlation function (ACF) ، أما الثاني فهو في نطاق الذبذبة frequency domain مثل طريقة فوريير Fast Fourier transform (FFT) والشعاع قصير الزمن short-time spectrum [٣] .

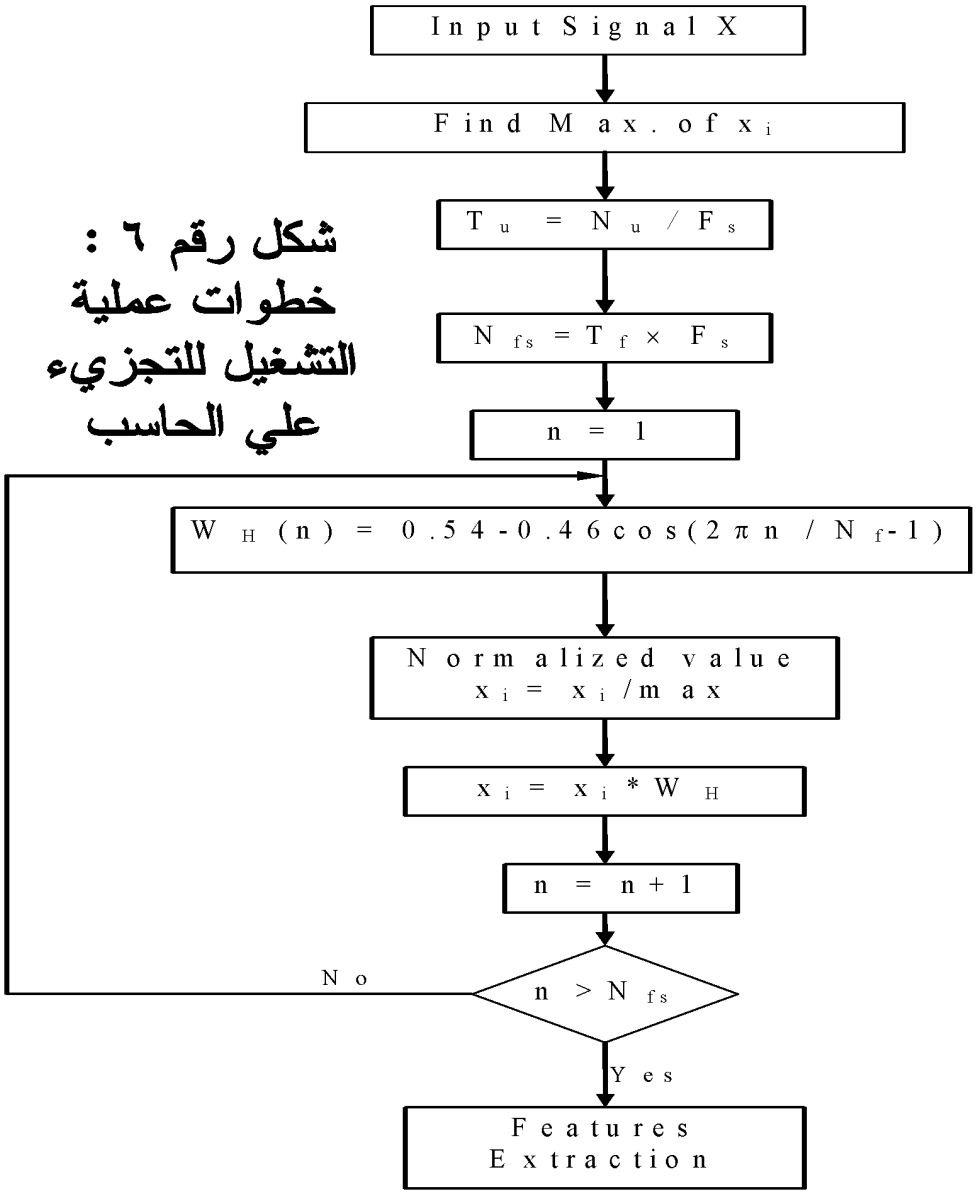


شكل رقم ٥ : الشكل الصندوقي لطريقة التقسيم

تعتبر طريقة التشفير التوقعي الخطية linear predictive coding (LPC) علي تحويل الإشارة الأصلية للحديث إلي مجموعة عينات خطية وبتقليل مربع الفارق squared difference بين الحديث الأصلي والعينات هذه نحصل علي معاملات التوقع predictor coefficients لتسهيل مهمة التشغيل علي الحاسب[٤] ويظهر تقسيم نظم التعرف علي الحديث عموما طبقا لعدد من المعاملات وتنحصر نظم التعرف علي الحديث بالعربية في: (شكل الكلمة وهي متباينة من صغيرة فمتوسطة ثم طويلة – نوع الحديث ويتبع إما الكلمة أو الجملة - طبيعة المتحدث وخواص الصوت) ففي التعرف علي الإطارات يتم أولا استخراج الخواص ثم التعرف والتقسيم لها [٥ ، ٦]، كما أنه هناك ثلاث نماذج للتقسيم (التجزيئ) بين

الإطارات وهي التقريب الرياضي (الإحصائية والتوصيل) أو قاعدة المعلومات أو الشبكات العصبية.

الشبكات العصبية الصناعية والممثلة علي الحاسب الآلي Artificial neural network (ANN) تعتمد علي التكوين الأصلي لها في الطبيعة كما يمثله الشكل رقم ٢، ٣ [٧] ، وتتنوع نظم التدريب بالشبكات العصبية بين ثلاث هم التعلم تحت الإشراف supervised أو learning التدريب بدونه unsupervised وأخيرا التدريب الذاتي self supervised learning ، وفي الطرق الإشرافية يتم إنشاء الجزء المناسب تلقائيا تبعا لمجموعة المدخلات وتعدل مع الأخطاء الخارجة عن الشبكة تكراريا وحتى نصل إلي مستوي الدقة المطلوبة الطرق غير الإشرافية لا تحتاج إلي تأثيرات خارجية أثناء التشغيل حيث يتم تقسيم المدخلات لها [٨ ، ٩].



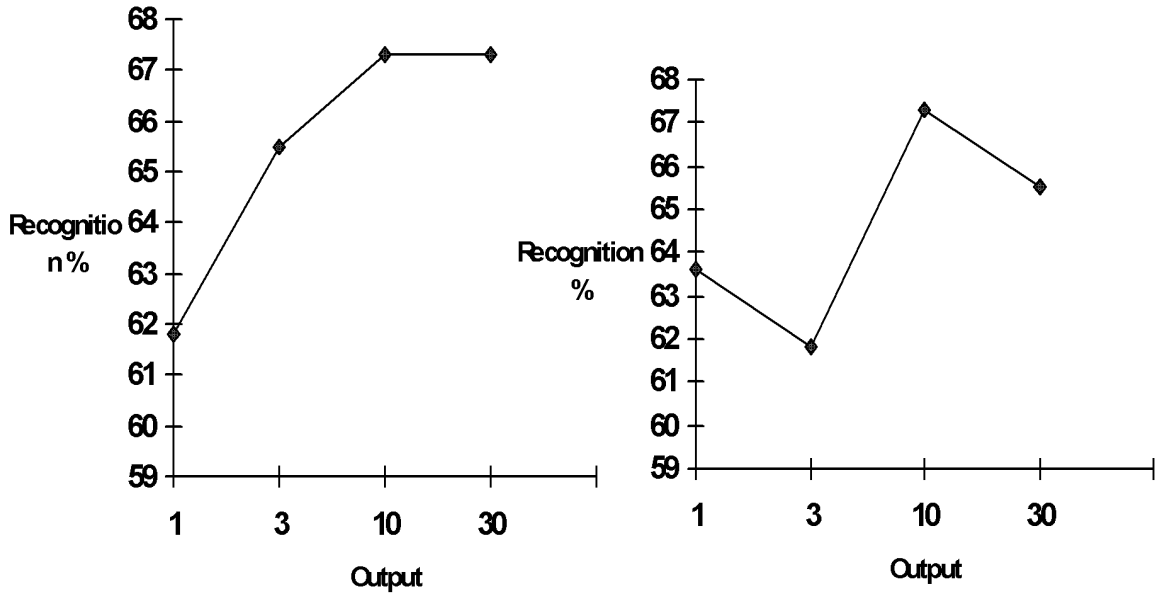
شكل رقم ٦ : خطوات عملية التشغيل للتجزيء علي الحاسب

قانون التعرف علي الحديث بناء علي ضبط الشبكة العصبية NN وللدقة المحددة بمجالي العمل (شكل رقم ٤) ، فالأول والمعروف باسم مجال التدريب *training phase* حيث يتم تدريب الشبكة العصبية علي مدخلات محددة تحت الاختبار وبتحديد الأوزان حتى نصل إلي الدقة المنشودة.

أما الثاني فهو مجال التنفيذ *recall phase* حيث يتم التحليل والتقسيم للكلمات ومن ثم الجمل [١٠].

توضع أرقاما صغيرة من أجل الفصل بين الكلمة والتالية لها thresholdبطريقة عشوائية ثم متجه الإدخال $X_p = [x_1 , x_2 , \ldots\ldots\ldots\ldots x_n]$ والخروج $Y_p =$ $[y_1 , y_2 , \ldots\ldots\ldots y_k]$لعدد من عقد الدخول input nodes n وتبسيطها مع عقد الخروج output nodes k للمعامل p في كلا من X_p and Y_p لحساب الخروج الفعلي الناتج $O_p = [o_1 , o_2 , \ldots\ldots o_k]$ وتحديثها حيث تتبع المعادلة:

$$w_{ij}(n+1) = w(n) + \eta\, \delta_{pj}\, x'_{pj} \qquad (2)$$



شكل رقم ٧ : معدل التعرف علي الحديث العربي بدالة التفعيل لظل الزاوية خطيا
(a) Hidden layer with 90 nodes (b) Hidden layer with 75 nodes

حيث $w_{ij}(n)$ تمثل الوزن من الطبقة الخفية أو الداخلة i في الوقت n ، و x'_{pj} الخروج بينما η هي الخروج الحقيقي و δ_{pj} تمثل الخطأ للشكل p علي العقدة j وبهذا نحصل عليه بدلالة الخروج المطلوب y_{pj} والفعلي o_{pj} :

الجدول رقم ١ : الترقيم العددي للكلمات الممثلة للحروف الهجائية وأرقامها بالعربية

الكود	الكلمة	الكود	الكلمة
15 /	ضاد/	1 /29	ألف/ صفر
16 /	طه/	2 /30	به/ واحد
17 /	ظه/	3 /31	ته/ اتنين
18 /	عين/	4 /33	ثه/ أربعه
19 /	غين/	5 /34	جيم/خمسه
20 /	فه/	6 /35	حه/سته
21 /	قاف/	7 / 36	خه/سبعه
22 /	كاف/	8 / 37	دال/تمانيه
23 /	لام/	9 / 38	ذال/تسعه
24 /	ميم/	10 / 39	ره/عشره
25 /	نون/	11 /	زين /
26 /	هه/	12 /	سين/
27 /	واو/	13 /	شين/
28 /	يه/	14 /	صاد/

$$\delta_{pj} = o_{pj}\,(1- o_{pj})(y_{pj} - o_{pj}) \qquad (3)$$

أما عند العقدة j من الطبقة الخفية فيصبح الخطأ هو :

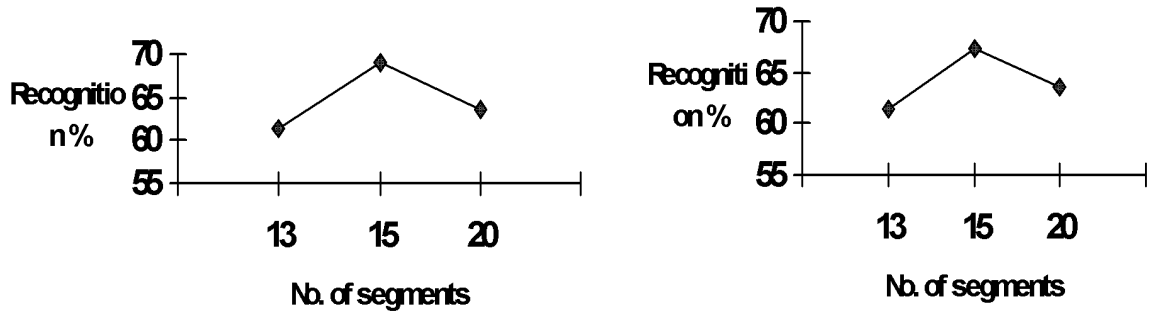
$$\delta_{pj} = x'_{pj}(1- x'_{pj})\ \sum \delta_{pj}\, w_{jk} \qquad (4)$$

أخيرا نستطيع أن نحصل علي الدقة المطلوبة في التعرف علي الكلمات بالعودة دائريا في حلقات متتابعة وممتالية.

٤- التجارب المعملية

جدول رقم ٢ : مجموعات الكلمات تبعا للطاقة المحسوبة

كود الكلمات	رقم	المدى
1,3	1	400 : 350 M
2,4,6,7	2	300 : 350 M
5,8,9,10,13,15	3	300 : 250 M
11,12,13,14,16,17,18,19,21	4	250 : 200 M
26,27,28,20,22,23,24,25	5	200 : 150
29:39	6	0 : 50 M

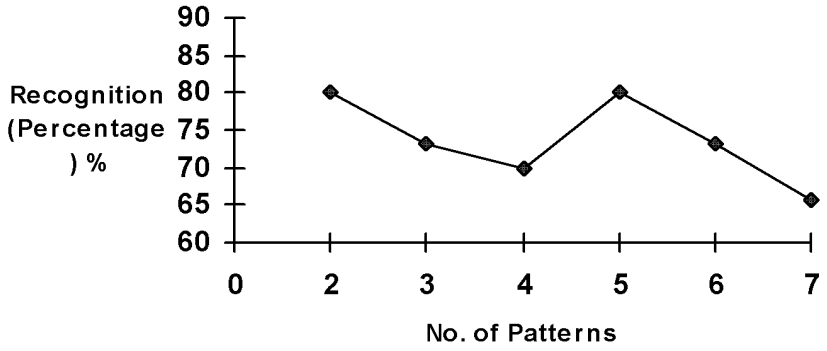


شكل رقم ٨ : معدل التعرف باستخدام:
(b) using LPC data (a) using LPC & RC data

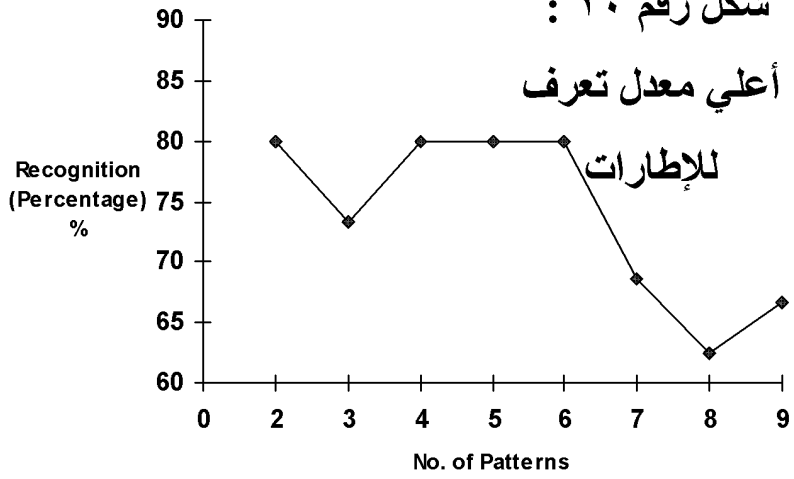
تمت التجارب علي عينات من الكلمات باستخدام كارت صوتي بالمواصفات (Sound Blaster 16 Advanced Signal Processing (SB 16 ASP) card and its utility programs.SB 16 ASP) والمحددة في المرجع رقم ١١
وبالطريقة المقترحة هنا معتمدا عل محول ١٦ بت بمعدل ١١ كيلو هيرتز ، حيث يتم تقسيم الكلمة إلي ثلاث مقاطع متساوية الزمن (٤٤ ميلي ثانية) ، ويقدم الجدول رقم ١ الترقيم التسلسلي لها لسهولة المتابعة والشرح . والخواص التي تستخرج من كل كلمة (مقطع / كلمة) تتمثل في العبور الصفري والنهاية العظمي والصغرى لقيمة الإشارة تبعا للرسم رقم ٥ . وأظهرت النتائج لكلمات الجدول رقم ١ أن التقسيم ممكن فتم التعرف علي العديد من الكلمات (مثل كود ٢ ، ٤) من العبور الصفري وقد تم

الفصل بين الكلمات الناطقة للحروف الكلامية عن تلك الخاصة بالأرقام من خلال الطاقة المحسوبة لكل منها كما أمكن تقسيم الكلمات في عدد من المجموعات (جدول رقم ٢)، وهذا أتاح الفرصة لوضع شبكتين للتعرف علي الكلمات أحدهما للحروف الهجائية والأخرى للأرقام وبكل منها ٧ وحدات خفية .
واستكملت التجارب علي الطاقة الكلية بالأسلوب الخطي فكانت الحاجة للشبكة LPC للتعرف علي بقية الكلمات (شكل رقم ٦) . يقدم الجدول رقم ٣ معدلات التعرف هذه إضافة للشكل رقم ٧ مع تكرار التجارب مع تغيير الأجزاء بين ١٣ و ٢٠ وكانت النتائج كما في شكل ٨ حيث تأكد علي أنسب التقطيع في الكلمة وهو ١٥ وبالتعرف المطلوب .

شكل رقم ٩ : تأثير عدد الإطارات علي معدل التعرف



شكل رقم ١٠ : أعلي معدل تعرف للإطارات

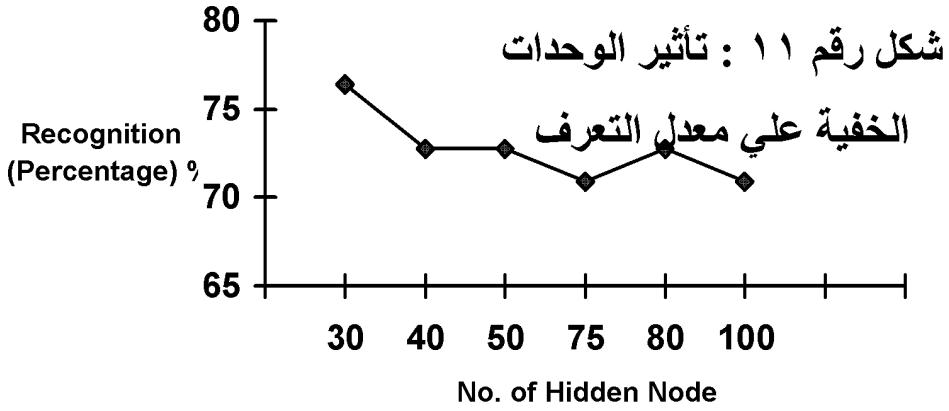


جدول رقم ٣ : معدل التعرف علي الشبكات المستخدمة (%)

معدل التعرف (%)					
عدد الوحدات الخفية	O/P	T-P	L-P	L-L	T-T
90	0:3:30	67.3	65.5	-----	-----
	0:0.3:3	65.5	63.6	-----	-----
	0:1:10	67.3	63.6	-----	-----
	0:0.1:1	61.8	63.6	61.3	61.3
75	0:3:30	65.5	failed	-----	-----
	0:0.3:3	61.8	65.5	-----	-----
	0:1:10	67.3	63.6	-----	-----
	0:0.1:1	63.6	63.6	61.8	61.8

٥- التقسيم

من أجل زيادة معدل التعرف تقسم البيانات الداخلة إلي مجموعتين أحدهما سهلة التعرف بها طبقا لمجال التدريب السابق وباستخدام الشبكة ART1 network والثانية للكلمات الأخرى . الكلمات متشابه متجهات الدخول ستأتي معا عند عقدة الخروج ويبين الشكل رقم ٩ تأثير عدد الإطارات علي معدل التعرف مشيرا إلي ارتفاعا في التعرف وخصوصا عندما عدد الكلمات يكون قليلا ، ويعرض الشكل رقم ١٠ لبعض من الكلمات عند ثبوت عدد الإطارات محددا الأفضل عند الرقم من ٢ وحتى ٩ . اعتمادا علي المقسم بالانتشار الرجعي مع تغيير عدد الوحدات الخفية (٣٠-١٠٠) فتحدد أعلي معدل تعرف عند ٣٠ وحدة (الشكل رقم ١١).



شكل رقم ١١ : تأثير الوحدات الخفية علي معدل التعرف

٦- الخلاصة

مما سبق في هذا البحث نستطيع أن نخلص إلي:

١- يكفي ١٥ جزء لكل كلمة عربية من مجموعات الاختبار للحصول علي معدل التعرف الجيد

٢- خواص الطاقة الصوتية تعطي تميزا واضحا للتمييز بين الكلمات

٣- معاملات كل من LPC وكذلك RC تعطي نجاحا في التمييز بين الكلمات العربية

٤- يوصي البحث بالاستعانة بدالة تفعيل ظل الزاوية بخواص خطية في طبقتي الوسط (الخفية) والإخراج للشبكة التكاملية NN

٥- يمكن الاعتماد علي شبكات الانتشار الرجعي في تحديد الكلمات العربية

٦- معدل التعرف علي الكلمات العربية المنطوقة يعتمد علي عدد الكلمات المطلوب التعرف عليها والعلاقة فيما بينهم .

٧- المراجع

[1] D. Wafik: "Arabic speech Perception, M Sc Thesis, Faculty of Engineering, Port Said, Egypt, 1997.

[2] A. Khalil: "A New System for Continuous Speech Recognition with Application to Arabic Digits". A Thesis for Ph. D. - Cairo University - Cairo 1992.

[3] J. R. Ullman: "Pattern Recognition Techniques", Butterworths 1973.

[4] M. Hamed: "A quick neural network for computer vision of gray images", Circuits, systems and signal processing journal, vol. 16, no. 1, (41-58), 1997.

[5] Jeff Dalton, Atul Deshmane: "An approach to increasing machine intelligence", IEEE Potentials, Artificial Neural Networks, pp. 33-36, 1991.

[6] G. Bebis & M. Georgiopoulos: "Feed-forward neural networks", IEEE Potentials, Pp. 27-31 Oct / November, 1994.

[7] J. Hertz, A. Krogh & R. Polemar: "Introduction To The Theory Of Neural Computation', Addison Weslely Publisher, New York, 1991.

[8] J. Mathew Palakal And Michael J. Zoran: "A neural based learning system for speech processing", Expert Systems With Applications, Vol. 2, 59-71, 1991.

[9] M. M. Awad: "An Artificial Neural Network Based-System For Automatic Gas Recognition", M. Sc. Thesis, Suez Canal Univ., Port said, Egypt, 1997.

[10] A. JamesFreeman & David M. Skapura: "Neural Networks: Algorithms, Applications, and Programming Techniques", Addison Weslely Publishing Company, New York, 1991.

[11] *M. Hamed,* F. W. Zaki, R. Atta: Arabic speech recognition using neural networks, PSERJ, ISSN 1110- 6603, vol. 4-1, 2000 (56-79)

ABSTRACT

Since the Arabic speech has not processed by computers as English as enough, the more density is required to cover its use and applications. This is an important item due to the high developing speed in the field of science and implementation. However, the speech is a main tool in the order system in all fields. This paper presents a new generalized system for the Arabic speech recognition based on 3-layer integrated neural networks consisting of control network and several sub-networks. The segmentation technique is accounted to minimize the selected features. This will reduce the computational time required for feature extraction as well as the computational effort. Different methods, applied to English, are introduced in the work. High recognition rate is reached by sequential processing.