Adoption of Neural Networks to Classified the Gender of the Speaker

Khalil I. Al-Saif **Mason Kh. Al-Nuaimi**

College of Computer Science and Mathematics

University of Mosul

Received on: 24/8/2010 **Accepted on: 10/11/2010**

ABSTRACT

In this research the neural network was adopted to classified the gender of the spoken, by creating the two dimension matrix from the parameters of the spoken speech signal which normal was snigle dimension array.

The porpose algorithm in this research divided in two stage :-

In the first stage the seven moment were calculated for a set of spoken signal of 50 persons , to be followed creating database depend on the seven moments .This database will be used to find the threshould value for both genders (male/female) which will be trained by neural network to classify any input tothe network.

In the second stage , speech of any spoken will be selected and the same feature will be extracted , as in the first stage , to be used as input to the neural network which was traind previously for gender recognition.

Back propagation neural network was achieved for recognition. The result of the applied algorithem on 10 spoken passed on 8 of them and 2 of them was failed **.**

Keywords: neural network, Back propagation network.

اعتماد الشبكات العصبية لتمييز جنس المتكلم

خليل إبراهيم السيف **ميسون خضر ألنعيمي**

كلية عوم الحاسبات والرياضيات

جامعة الموصل

تاريخ استلام البحث:٢٠١٠/٨/٢٤ **تاريخ قبول البحث:٢٠١٠/١١/10**

الملخص

تم في هذا البحث اعتماد الشبكات العصبية كوسيلة تصنيف لتحديد جنس المتكلم حيث يتم بناء مصفوفة ثنائية من معاملات إشارة صوت المتكلم (التي عادة تكون على شكل مصفوفة أحادية) حيث تضمنت الخوارزمية المعتمدة في البحث على مرحلتين :

في ألمرحله الأولى حساب العزوم السبعة للمصفوفة الثنائية الحاوية على عناصر إشارة الكلام لمجموعة تزيد على 50 متكلم والتي يتم بعدها بناء قاعدة بيانات يستخلص منها حد عتبة لفئتي الذكور وإناث , معتمدة على العزوم السبعة لإشارة الكلام لكل متكلم لأجل تدريب شبكة عصبية على إمكانية تحديد حد العتبة لأي إدخال.

أما المرحلة الثانية والتي يتم فيها اختبار صوت متكلم نستخلص منه نفس الصفات المعتمدة في بناء قاعدة البيانات ليتم بعدها إدخالها على الشبكة العصبية المدربة مسبقا لتحديد جنس المتكلم .

تم اعتماد شبكة الانتشار العكسي لأجل التمييز حيث أعطت النتائج المطبقة على 10 متكلمين تمييز كامل لثمانية منهم وأخفقت الخوارزمية على عينتين فقط.

الكلمات المفتاحية: الشبكة العصبية، الشبكة ذات الانتشار العكسي.

1- المقدمة

إن تحليل العلاقة بين السمات الصوتية، السمعية للكلام وجنس المتكلم لها تطبيقات كثيرة. حيث إن إشارة الكلام تحوي الكثير من المعلومات فضلا عن الرسالة التي يتم نقلها على لسان المتكلم ، فهي تعطي للمستمع فكرة عن جنس المتكلم، بالإضافة إلى حالته العاطفية. إن اختلاف نمط الكلام بين الإناث والذكور يأتي من الاختلافات في الخصائص اللغوية وغير اللغوية(paralinquistic) لإشارة الكلام. فمثلا، في حالة الإناث فأن الخصائص اللغوية تتمثل في وجود الترددات العالية، على العكس من ذلك بالنسبة للذكور.

كما أن التغييرات الفسلجية تؤثر على صوت المتكلم ، حيث تمر الحنجرة بعملية نضوج التي تبدأ في وقت مبكر من العمر وتستمر بالتغييرات خلال الحياة .حيث إن أكثر التغييرات إثارة تحدث خلال سن البلوغ ، ومع مرور الوقت فأن غضاريف الحنجرة تبدأ بالتكلس وتصبح متصلبة أكثر بقليل.كما أن المفاصل التي تسمح لحركة الطيات الصوتية ثلاثية الأبعاد تصبح أصلب وحجم عضلة الطيات الصوتية يقل.

كذلك تتضمن العوامل المتعلقة بالمتكلم الخصائص الطبيعية (تشريحية و فسلجية) مثل الجنس، الوزن، الصحة، التدخين، الحالة الفسلجية. فمثلا في حالة الجنس حيث إن معظم الخصائص الصوتية تختلف بين الجنسين البعض بسبب اختلافات تشريحية مثل حجم الحنجرة الأكبر والمنطقة الصوتية (vocal tract) الأطول قليلا كمعدل في حالة الذكور.حيث ان حجم المنطقة الصوتية في حالة الإناث هي 15% أقصر من الذكور.

2- العزوم السبعة

تعتبر طريقة العزوم من أوائل الطرق التي اعتمدت في تمييز الحروف البصرية ولازالت تعتمد من قبل برامجيات تمييز الحروف البصرية اللاتينية خاصة المجهزة منها مع أجهزة الماسح الضوئي والتي تعتمد مبدأ الـ (off- Line) ويعزى سبب استخدامها إلى ثباتها ضد التدوير خاصة عندما يقوم جهاز الماسح الضوئي بعملية مسح الصورة (والتي تحتوي على الكتابة) مائلة عن الأفق بالنسبة لكاميرا جهاز الماسح ، طريقة العزوم الثابتة تعتبر العزوم السبعة هي عوامل ثابتة لأي صورة لا تتغير ولا تتأثر عند الدوران و التزحيف وتغير المقياس (Rotating / translating and scaling).

تم في هذا البحث اعتماد العزوم السبعة لمصفوفة إشارة الكلام التمييز باستخدام شبكة الانتشار العكسي العصبية. لقد أصبحت العزوم وسيلة لتصنيف الأشياء خلال الثلاثين سنة المنصرمة(Flusser,2000). وقد تمت الاستفادة من العزوم ووظائف العزوم كنمط في عدد من التطبيقات لتحقيق تمييز ثابت للصورة ذات البعدين.

منذ العام 1962 تم الاعتماد على إدخال ثوابت عزمية بالاستناد على طرق الثوابت الجبرية وباستخدام ربط غير خطي من العزوم المنتظمة يشار لها على إنها عزوم هندسية ، وقد تم اشتقاق مجموعة من العزوم الثابتة التي لها خاصية مرغوب في كونها ثابتة تحت التزحيف للصورة والتقييس والدوران. وان هذا النوع من العزوم يلائم بشكل أفضل مختلف أساليب التمييز بسبب خاصيتها الثابتة ضد التزحيف والتقييس والدوران.

وتعرف العزوم المركزية المقيسة بـ η_{pq} كما يلي [Gonzales and Wintz,1987]:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \qquad \ldots(1)$$

p, q = 0,1,2,...

حيث

Y,X يمثلان الإحداثيات السيني والصادي للوحدة الصورية على التوالي .

p,q يمثلان مرتبة العزم المنتظم .

f(x,y) تمثل قيمة دالة شدة إضاءة الصورة [وتكون قيمتها صفر أو واحد في الصور الثنائية].

x و y يمثلان المراكز القياسية للصورة باتجاه محور x و y .

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy \qquad \ldots(2)$$

حيث

$$\bar{x} = \frac{m_{10}}{m_{00}} \ , \ \bar{y} = \frac{m_{01}}{m_{00}}$$

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \qquad \ldots(3)$$

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad \ldots(4)$$

حيث

$$\gamma = 1/2(p+q) + 1 \qquad \ldots(5)$$

p + q = 2,3,...

يمكن اشتقاق سبعة عزوم غير متغيرة invariant moment من العزم الثاني والثالث تعطى بما يلي:

$$\phi_1 = \eta_{20} + \eta_{02} \qquad \ldots(6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\,\eta_{11}^2 \qquad \ldots(7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \qquad \ldots(8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \qquad \ldots(9)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2], + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \qquad \ldots(10)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2], + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \qquad \ldots(11)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2], + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \qquad \ldots(12)$$

لقد برهن(Hu 1962) إن هذه العزوم غير متغيرة بالنسبة للتزحيف translation والدوران Rotation وتغير المقياس Scaling [البدراني 2001].

3- الشبكات العصبية:

طورت الشبكات العصبية الاصطناعية (Artificial neural networks ANN) لتحقيق أداء يشبه أداء الإنسان في حل بعض المشاكل عن طريق محاولة محاكاة فعاليات الدماغ إذ تتميز الخلايا العصبية بسرعتها في معالجة البيانات وقدرتها على التعلم والتعامل مع أنماط بيانات خاطئة مما جعلها مناسبة لكثير من المجالات التطبيقية مثل تنفيذ بعض المسائل المعقدة منها تمييز الأنماط، وتمييز الصوت والصورة...الخ [Rao, V.B. and Rao,H.V. ,1993]، وتقسم معالجة البيانات والمعلومات في الشبكات العصبية إلى طورين أساسيين [Abod L.K., 1998]:-

4- طور تدريب الشبكات العصبية الاصطناعية Artificial Neural Networks Training

التعلم طور في الشبكات العصبية عندما تنتج بيانات جديدة في الشبكة نتيجة تغير الوزن (weight). الشبكة تدرب على التطبيق المعطى أي على مجموعة الادخالات لإنتاج الإخراج المطلوب والتدريب ينجز بشكل متسلسل على متجه الإدخال عندما تتغير الأوزان للشبكة وفقاً لقوانين محددة [Lippmann R.P., 1987].
خلال التدريب تقترب أوزان الشبكة تدريجياً من القيم المثالية ويعمل الإدخال على إظهار الإخراج المطلوب ، ويكون التدريب في الشبكات العصبية الاصطناعية بإشراف أو بدون إشراف [Rao, V.B. and Rao,H.V. ,1993].

5- التدريب بدون أشراف (Unsupervised Training)

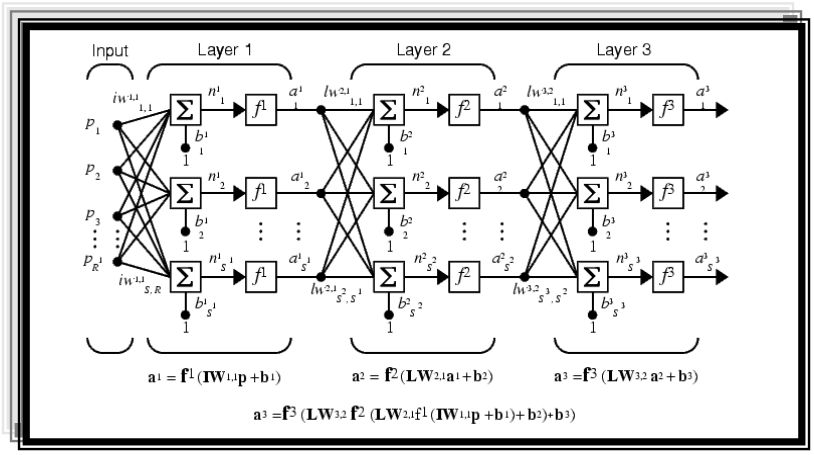
يمثل هذا التدريب الطور الأول في الشبكات العصبية ومعناه أن الشبكة تمتلك بعض المعلومات خلال التدريب أي أنها تمتلك الادخالات والأوزان فقط وليس لها معرفة عن ناتج الإخراج (بمعنى آخر لا يوجد إخراج منتخب) لأجل المقارنة [Chitra,S.P., 1993]. اكتشف هذا التدريب من قبل العالم كوهين وآخرين تتألف مجموعة التدريب من متجه الإدخال وخوارزمية التدريب لتغيير أوزان الشبكة لإنتاج متجه الإخراج الثابت ، إذ يطبق الإدخال لإيجاد الإخراج المحدد[Rao, V.B. and Rao,H.V., 1993] [Lippmann R.P., 1987].

6- الشبكات العصبية وتمييز الأنماط :

تستخدم الشبكات العصبية الاصطناعية في تمييز الأنماط ، ان تمييز الأنماط هي العملية التي تتم على البيانات للحصول على معلومات مرتبة لتصنيف هذه البيانات. تمييز الأنماط يستخدم لعدة تطبيقات مثل تمييز الأرقام وتمييز الحروف المكتوبة يدويا والمطبوعة وتمييز الأصوات ومجالات لا مجال لحصرها.
الشبكات العصبية بنوعيها التي تعمل باشراف والتي تعمل بدون إشراف طبقة واحدة أو من عدة طبقات استطاعت وبنجاح تطبيق مختلف التصانيف ومسائل التمييز[AL_Yaseen Sh,2000].

7- الشبكة العصبية ذات الانتشار العكسي *Back Propagation Neural Network* BPNN:

كل شبكة عصبية قد تم تدريبها باستخدام خوارزمية الانتشار العكسي تدعى بالشبكة العصبية ذات الانتشار العكسي (الشكل 1). وتعتبر الشبكة العصبية ذات الانتشار العكسي نموذج مألوف جداً في الشبكات العصبية. وهي الأكثر استخداماً في معمارية الشبكات العصبية، والسبب الرئيسي لهذا هو : أنها سهلة التنفيذ، إذ أن الشبكات المستخدمة لخوارزمية الانتشار العكسي تتعلم المخططات المعقدة المتعددة الأبعاد بسهولة أكثر من استخدامها للخوارزميات الأخرى. والعديد من التطبيقات يمكن أن تُصاغ باستخدام شبكة الانتشار العكسي حيث يكون لديها نموذج لأغلب الشبكات العصبية المتعددة الطبقات. ومن ثم يمكننا القول إن أكثر طريقة تدريب شائعة الاستخدام في الشبكات العصبية المتعددة الطبقات هي خوارزمية الانتشار العكسي، وقد تم اختبارها بنجاح لأنواع مختلفة من المهام. وبالرغم من أن الشبكة لا تضم ارتباطات للتغذية العكسية، لكن الأخطاء تنشر عكسياً للخلف خلال عملية التدريب. والأخطاء في طبقة الإخراج تحدد قياسات الأخطاء في الطبقة المخفية، والتي تستخدم كانحياز لتعديل أوزان الربط بين طبقة الإدخال والطبقة المخفية. كل معالجة تكرار تضم تعديل مجموعتين من الأوزان على الأقل بين زوج من الطبقات وتدوير الإخراجات، وتستمر هذه المعالجة إلى أن تقع الأخطاء تحت مستوى تفاوت مسموح به tolerance [Rao,V.B. and Rao, H.V,1993] [Wassermann, P.D,1989].



الشكل (1) يوضح شبكة عصبية متعددة الطبقات

8- معمارية الشبكات العصبية ذات الانتشار العكسي BPNN :

تعتبر الشبكة العصبية ذات الانتشار العكسي هي نظام شبكة بتغذية أمامية وبصورة تدرجية، وتتألف من طبقتين أو أكثر من الطبقات لوحدات المعالجة المترابطة كلياً مع بعضها (طبقة الإدخال ليست من ضمنها). الخلايا العصبية لا ترتبط مع خلايا عصبية أخرى في نفس الطبقة (شبكة ذات ارتباط طبقي). بكلمات أخرى، كل خلية عصبية في طبقة الإدخال سوف ترسل إخراجها إلى كل خلية في الطبقة الوسطى، وكل خلية في الطبقة الوسطى سوف ترسل إخراجها إلى كل خلية في طبقة الإخراج [Wassermann, P.D,1989]. [Chitra,S.P,1993].

عدد الخلايا العصبية في الطبقة المخفية ربما يكون مختلف طبقاً إلى تعقيد المسألة، وحجم معلومات الإدخال. على أي حال، لعدد معطى من الإدخالات، إذا كانت الطبقة المخفية كبيرة جداً، ربما لا يكون من الممكن تطوير النمط

القابل للاستخدام . من جهة أخرى، الطبقة الوسطى التي تكون صغيرة جداً سوف يكون تأثيرها قوياً في زيادة عدد التكرارات iterations المطلوبة لتدريب الشبكة [Chitra,S.P,1993].

9- الخوارزمية المعتمدة في البحث :

تم اقتراح خوارزمية لأجل اعتماد العزوم السبعة لمصفوفة ثنائية حاوية على عناصر إشارة الكلام بعد أجراء توافق لخزن إشارة الكلام (ذو البعد الواحد)داخل مصفوفة ثنائية مربعة.حيث يتم إدخال العزوم السبعة إلى شبكة عصبية وممكن تلخيصها بمرحلتين:

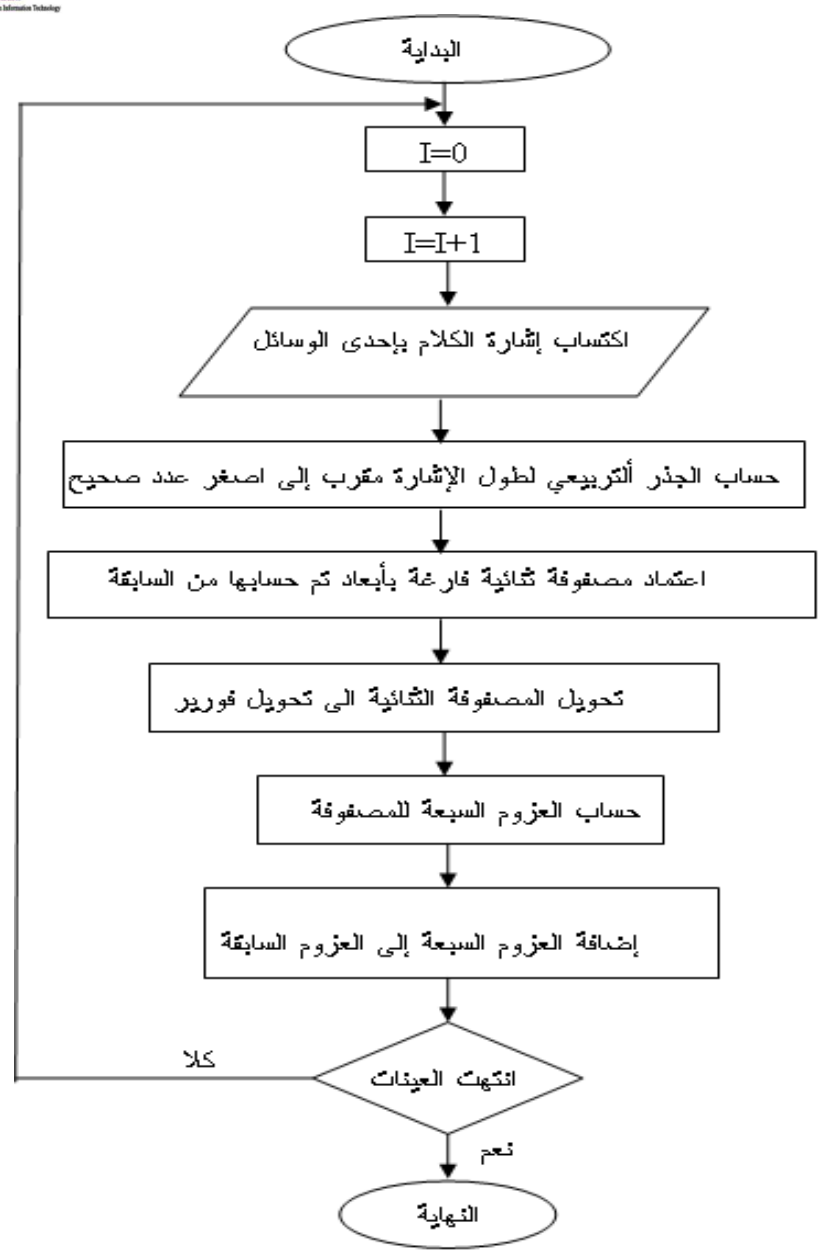
المرحلة الأولى:

يتم العمل على تدريب شبكة عصبية لأجل التمييز بين فئتي الذكور والإناث حيث يتم اعتماد الشبكة العصبية ذات الانتشار العكسيBPNN . من خلال تهيئة مدخلات الشبكة المتضمنة العزوم السبعة للمصفوفة الثنائية الحاوية على عناصر إشارة الكلام بصيغة التردد(fourier transform).

المرحلة الثانية:

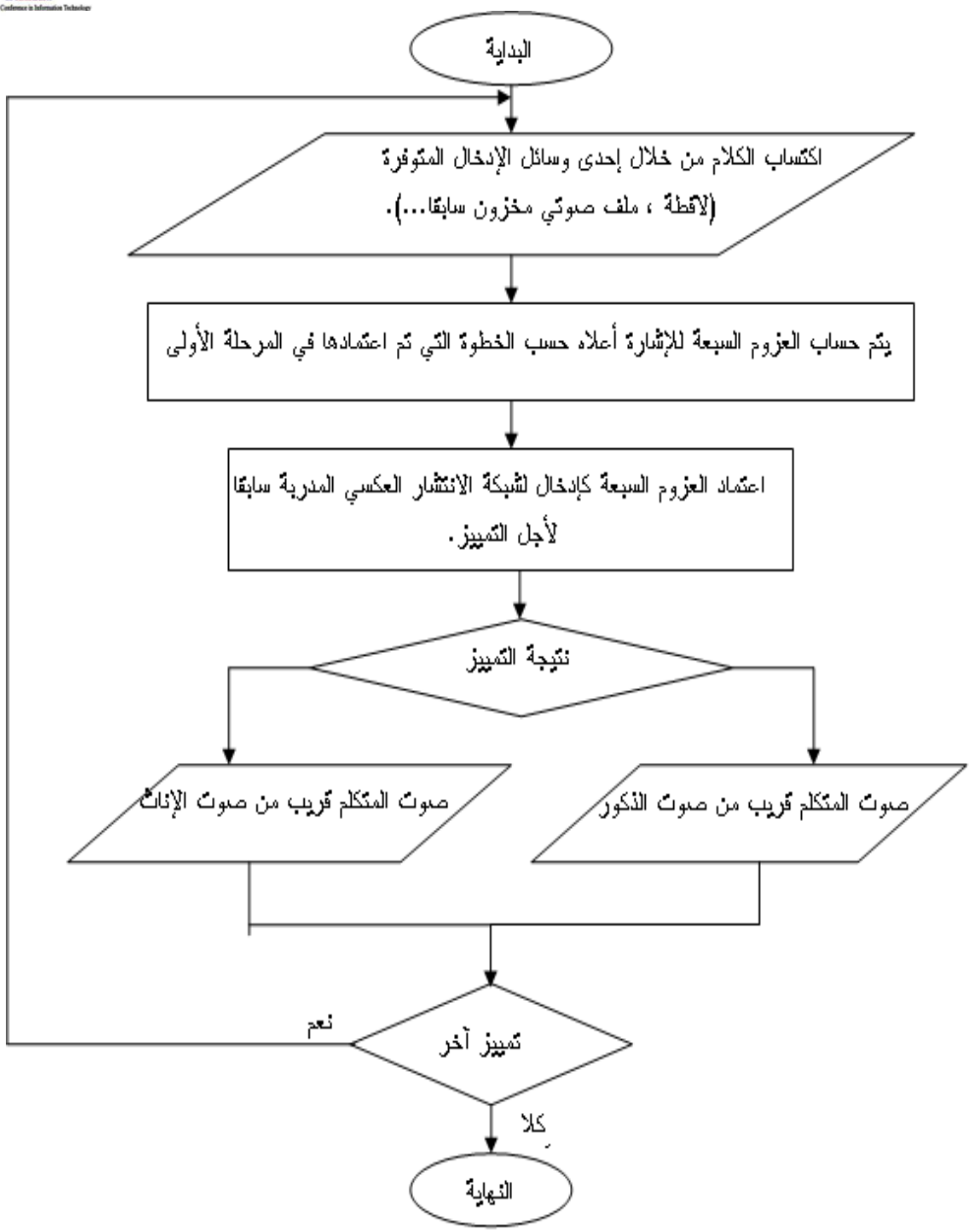
1. اكتساب الكلام من خلال إحدى وسائل الإدخال المتوفرة (لاقطة , ملف صوتي مخزون سابقا...).
2. حساب بعد المصفوفة الثنائية لأجل إدخال عناصر إشارة الكلام إليها.
3. حساب تحويل فورير للمصفوفة الثنائية ومن ثم اعتماد القيمة المطلقة (الطيف spectrum).
4. حساب العزوم السبعة .
5. اعتماد العزوم السبعة كإدخال لشبكة الانتشار العكسي المدربة سابقا لأجل التمييز.

ويوضح الشكل (2) المخطط الصندوقي لسير عمليات المرحلة الأولى للخوارزمية المقترحة:



الشكل (2) المخطط الصندوقي لسير عمليات المرحلة الأولى

إما الشكل (3) فيمثل المخطط الصندوقي لسير عمليات المرحلة الثانية للخوارزمية المقترحة:



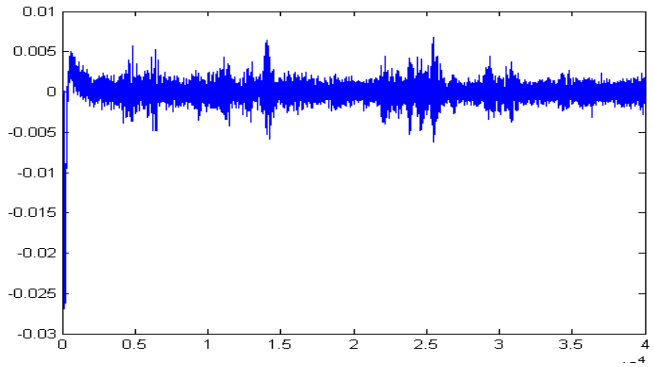
الشكل (3) المخطط الصندوقي لسير عمليات المرحلة الثانية

10- مناقشة النتائج:

من خلال التطبيق العملي للخوارزمية المقترحة تم تدريب الشبكة على عينة تتكون من 35 صوت من الذكور و30 صوت من الإناث قاموا جميعهم بنطق مقطع واحد لفترة زمنية قدرها 5 ثواني . و أثبتت تجارب التدريب للشبكة التوصل إلى استقرارية عالية في التدريب مما دعم مرحلة التمييز عند اختبار صوت مجهول.

عند إجراء الاختبار لعشرة أصوات مختلفة قاموا بنطق ذات المقطع المعتمد في التدريب ولفئات عمرية مختلفة تبين أن الشبكة نجحت في تحديد جنس8 عينات وأخفقت في عينتين أي أن نسبة التمييز تقارب 80%.

وفيما يلي تطبيق مبسط لإشارة الكلام:-



شكل (4) يمثل إشارة الكلام

يعطي الشكل (4) صيغة لمصفوفة مربعة لإشارة الكلام لأجل اعتمادها في حساب العزوم السبعة وتكون كالآتي:

مصفوفة مربعة لإشارة الكلام

.....	2	2	2	1	1	1	1	0	0	0]
	1	1	1	1	1	1	1	0	1	1
	0	0	0	0	0	0	0	0	0	1
	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
[.....	0	0	0	0	0	0	0	0	0	0

أما مصفوفة العزوم السبعة التي تم حسابها من المصفوفة المربعة لإشارة الكلام هي:

العزم الأول	العزم الثاني	العزم الثالث	العزم الرابع	العزم الخامس	لعزم السادس	العزم السابع
1.2495]	2.0145	1.5792	2.2672	4.5087	1.8483	[4.0642

يتم إدخال مصفوفة العزوم السبعة السابقة كإدخال على الشبكة العصبية (الانتشار العكسي) التي تم تدريبها سابقاً لأجل تمييز جنس المتكلم.

11- الأعمال المستقبلية:

يمكن توظيف أفكار الخوارزمية المقترحة في إمكانية حساب العلاقة بين جنس المتكلم وعمر المتكلم لتحديد كثير من العوامل والمؤشرات على جهاز النطق لدى المتكلم ، إضافة إلى إمكانية الاستفادة من أفكار الخوارزمية المعتمدة في البحث بعد تطويرها باعتماد الخوارزميات المبينة بدلاً من الشبكة العصبية أو استخدام المنطق المضبب الذي من المتوقع أن يقدم تمييز أفضل بسبب ما يمتلك من إمكانية التوفيق في النتائج المتقاربة (أنثى / ذكر) للمتكلمين.

المصادر

[1]. البدراني، هدية صالح عبدا لله ،(2001)، "تمميز الحروف العربية المكتوبة بنظام (7 word) "، بحث ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

[2]. Abod L.K.,(1998), Classification of satellite image using neural network, Ph.D. Thesis, Department of physics, college of science, university of Baghdad.

[3]. AL_Yaseen Sh., 2000 Face Recognition using Neocognitron Neural Network, M.Sc. Thesis , Department of Computer Engineering , College Of Engineering , University of Mosul.

[4]. Chitra, S.P.(1993) Use neural networks for problem solving, Chemical engineering progress, pp. 44-52.

[5]. Flusser, J., 2000. On the independence of rotation moment invariant, Pattern recognition, Vol.33, pp.1405-1410.

[6]. Gonzales, R. C. and Wintz, 1987. Digital Image processing, Addison Wesley publishing company.

[7]. Lippmann R.P.,(1987), Introduction to computing with neural Nets, IEEE Assp Magazine, Vol. 4, No. 2, pp. 4-22.

[8]. Rao, V.B. and Rao,H.V.(1993), C++ Neural Networks and Fuzzy Logic, Henry Holt and Company, Ins., New York.

[9]. Wassermann, P.D.,(1989), Neural Computing theory and practic van nostrand reinhold:New York.