Isolated Arabic Digit Recognition Using Genetic Algorithm

Dr.Yusra Faisal Al-Irhaim **Ali Insaf Jasim**
Yusrafaisalcs@uomosul.edu.iq
College of Computer Sciences and Mathematics
University of Mosul

Received on : 25/3/2013 **Accepted on :24/6/2013**

Abstract

This study aims at constructing an intelligent system for recognizing the single Arabic numbers. It consists of two basic stages: the stage of features extraction and the stage of recognition. In the first stage, the technology of (Mel-Frequency cestrum coefficient (MFCC)) was employed. But in the second stage, the genetic algorithm was used.

The results of the test showed that words recognition percentage was (100%) for the words used with training, and it was (97%) for the words used with no training. The proposed system was constructed using the MatLab version (0.7) program, and the data used in the system are the following numbers: (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9). Also, six speakers (four males and two females) performed the voice recording.

Key words: Speech recognition, Arabic digits, Genetic algorithms.

تمييز الأرقام العربية المفردة النطق باستخدام الخوارزمية الجينية

د.يسرى فيصل الارحيم **علي انصاف جاسم**
كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2013/03/25 **تاريخ قبول البحث: 2013/06/24**

الملخص

يهدف هذا البحث الى بناء نظام ذكائي لتميز الارقام العربية المفردة ، حيث يتكون النظام من مرحلتين أساسيتين ، مرحلة استخلاص الصفات ومرحلة التمييز . في مرحلة استخلاص الصفات تم الاعتماد على تقنية درجة النغم (Mel-Frequency cestrum coefficient (MFCC)) ، اما في مرحلة التمييز فقد استخدمت الخوارزمية الجينية .

اوضحت نتائج الاختبار أن نسبة تمييز الكلمات وصلت الى 100% لكلمات المدرب عليها، و 97% للكلمات غير المدربة تم بناء النظام المقترح باستخدام برنامج ماتلاب والبيانات المستخدمة هي الارقام التالية (صفر ، واحد ، اثنان ، ثلاثة ، اربعة، خمسة ، ستة ، سبعة ، ثمانية ، تسعة) لستة متكلمين ,اربع من الذكور واثنان من الاناث.

الكلمات المفتاحية: تمييز الكلام, الارقام العربية, الخوارزمية الجينية.

1. المقدمة

لقد بدأ اهتمام خبراء الحاسوب والباحثين في التعرف على الكلام منذ أكثر من أربعة عقود, وذلك لكي يصل الإنسان إلى مرحلة تجعله قادرا على التخاطب مع الكمبيوتر وإعطاؤه الأوامر والتعليمات صوتيا وبدون الحاجة إلى استخدام لوحة المفاتيح والماوس وغيرها من الوسائل, وذلك توفيرا للوقت والجهد. وفي السنوات الأخيرة تطور التعرف على الكلام تطورا واضحا وكبيرا, بحيث أصبحت برامج التعرف الآلي تدخل في أغلب مجالات الحياة, ووصلت إلى دقة مرضية نوعا ما [9] .

الكلام هو عملية استحداث موجات صوتية بواسطة الحركة الارادية للتركيب التشريحي في نظام توليد الكلام لدى الانسان لنقل المعلومات من المتكلم الى السامع , فالكلام هو ديناميكية الإشارة التي تحتوي على المعلومات

الأساسية والتي تدعى بالموجة الصوتية. هذه الموجة التي تنتج خلال عملية الضغط داخل الفم للمتكلم وهي نتيجة لسلسلة حركات متناسقة ومتتالية لتراكيب النظام اللفظي لدى الإنسان [6] .

ومع ازدياد انتشار استخدام الحاسوب في العقود الأخيرة اصبح استخدام الكلام كواجهة بين الإنسان والحاسوب يقع ضمن الأهداف الكبيرة لعلماء الهندسة والحاسوب ذلك لجعل تصرف الحاسوب قريباً أو مشابهاً لحد ما من تصرفات البشر [7] .

إن أهمية معالجة اللغة العربية بالحاسوب لم تعد أمر رفاه أو أمراً ثانوياً بل هو أمر في غاية الأهمية وعليه يعتمد مستقبل اللغة ومكانة العرب في الحضارة الحالية بل ومستقبلهم الاقتصادي والعلمي .

وكان من نتيجة هذا الاهتمام المكثف في هذا الميدان تطور تطبيقات عديدة في معالجة إشارة الكلام ، فقد نما العلم كثيرا في مجالات تطبيقية تشمل تصميم شبكات الهاتف ومكبرات الصوت بأنواعها وشبكات المعلومات وتطبيقاتها كثيرة لا حصر لها وفي جميع المجالات ، والتي يمكن تصنيفها إلى تمييز الكلام (Speech recognition) ، تركيب الكلام (Speech synthesis) وتشفير الكلام (Speech coding) [1] . إن تمييز الكلام (Speech recognition) يهتم بتمييز كلمات معينة أو تشخيص المتكلم، وان تمييز الكلام هو محور بحثنا .

تعد الخوارزمية الجينية واحدة من علوم الحاسب الالي الحديثة التي تبحث عن اساليب متطورة لبرمجته للقيام بأعمال واستنتاجات تشابه ولو في حدود ضيقة تلك الاساليب التي تنسب لذكاء الانسان , حيث ان الخوارزمية الجينية لها دور الكبير في حل مسائل متعددة في الحياة العملية ، ومنها ما يتعلق باللغات الطبيعة ، وهي حقل جديد بدأ الدخول مؤخرا في معالجة اللغات الطبيعية ومن هذا المنطلق يتبين لنا اهمية الخوارزمية الجينية في تمييز الكلام [10] .

1.1. دراسات سابقة

- الدكتورة سارة عزيز وسبا عبد الواحد وميسا عبد الكريم (2006) استخدموا الخوارزمية الجينية والشبكة العصبية للتعرف على بعض الصوتيات العربية التي سجلت لأكثر من شخص وبمختلف الاعمار ، حيث اعتمد في هذا البحث على مرحلتين اساسيتين الاولى تضمنت عملية تحويل الاشارة الصوتية الى مقاطع وتحليلها وتحديد افضل العينات التي تمثل الصفة الاساسية للإشارة الصوتية اما الثانية تضمنت عملية التمييز باستخدام الخوارزمية الجينية والشبكات العصبية [3].
- Miguel Lopez, and Oscar Castillo (2006) قاما الباحثان باستخدام الشبكات العصبية والخوارزميات الجينية والمنطق الضبابي للتعرف على الصوت على وجه الخصوص، ويتم تمييز صوت المتكلم من خلال تحليل الإشارات الصوتية مع مساعدة من التقنيات الذكية، مثل الشبكات العصبية , حيث تم الاعتماد على الشبكات العصبية لتحليل إشارة الصوت من المتكلم غير المعروف، ثم استخدم المنطق الضبابي ونظرا لعدم اليقين من عملية اتخاذ القرار, تم استخدام الخوارزميات الجينية لتحسين بنية الشبكات العصبية [17].
- صبا عبدالواحد صدام (2008) استخدمت الخوارزمية الجينية في تحديد افضل معمارية شبكة عصبية لتمييز بعض الفونيمات العربية حيث يهدف هذا البحث الى تحديد افضل معمارية شبكة عصبية باستخدام الخوارزمية الجينية لتمييز بعض الفونيمات العربية ، حيث يتضمن البحث ثلاث مراحل اساسية شملت الاولى والثانية معالجة الاشارات الصوتية المسجلة من خلال اجراء الحسابات اللازمة لتحديد افضل المواقع التي يمكن

اختيارها لتقطيع سلسلة طويلة من العينات واستخلاص قيم طيف الترددات التي تحتويها الاشارة الصوتية ، اما المرحلة الثالثة تضمنت استخدام الشبكة العصبية ذات الانسياب الخلفي المعدلة بمعامل متعلم متكيف والمحددة معماريتها باستخدام الخوارزمية الجينية لتمييز مجموعة الانماط التي تمثل كل منها صفة مميزة من صفات الفونيمات العربية [5].

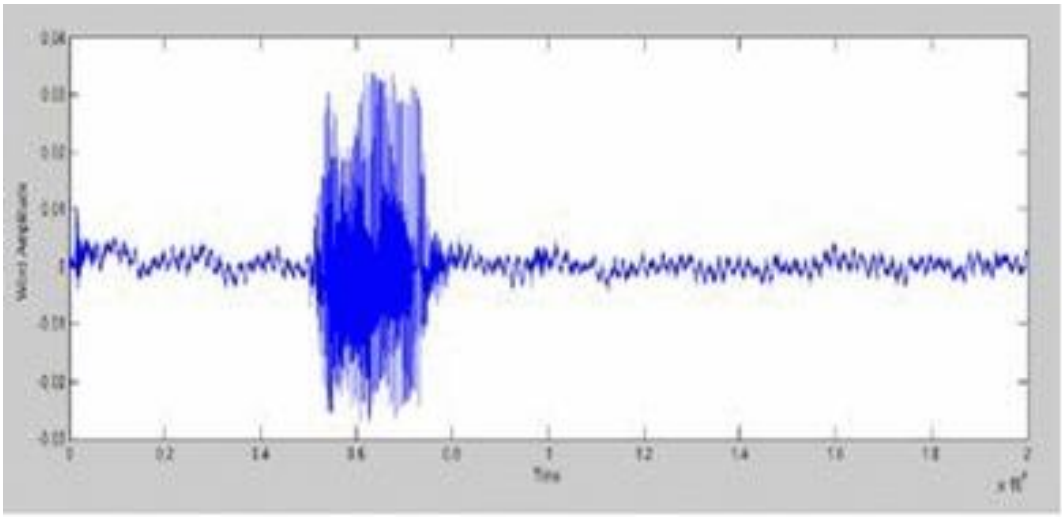
- Fadila Maouche, Mohamed Benmohamed (2008) تم استخدام الخوارزمية الجينية للتعرف على الكلمات العربية المفصولة حيث تم استخدام خوارزمية الـ (MFCC) لتمثيل الاشارة والهدف من هذا البحث لتحسين اداء لنظام التعرف على الكلام (ASR)[13] .
- José M. Villegas, Alejandra Mancilla, Patricia Melin (2008) ، استخدم الباحثان الخوارزمية الجينية لتحسين مشكلات (وحدات، والطبقات والخلايا العصبية) وافضل تدريب للشبكة العصبية الاصطناعية (ANN)، حيث يتم استخدام خوارزمية الجينية والتي لها القدرة على تحقيق بنية أفضل ونتيجة تدريب ANN افضل ، لمهمة معينة، وفي هذه الحالة يتم التمييز للشخص عن طريق الصوت والوجه[15].
- Marta wroniszewske and Jacek Dziedzic (2010) تم استخدم الخوارزمية الجينية لتمييز الكلمات بعد عملية تحويل الاشارات الصوتية الى مجموعة من الكلمات للتعرف على الاوامر الصوتية أي بناء نموذج نظام الأوامر الصوتية التي تستخدم في نطاق واسع للتعرف على الصوت في الأجهزة الالكترونية مثل الهواتف الخلوية او في الملاحة وغيرها [16].

2. معالجة الاشارة الصوتية

حتى يَحْدث التعامل مع الصوت الكلامي عبر الحاسوب لا بدّ من تحويل الصوت إلى أرقام ثنائية، وذلك عن طريق تحويل إشارة الصوت الكهربائية إلى النظام الثنائي، واللّجوء إلى أخذ عينات من الإشارة التشابهية في لحظات تفصلها مُدَدٌ متساوية ، وذلك بتردد ثابت يدعى بتردد أخذ العينات، ثم نحْصَل على عينات متقطعة للإشارة التشابهية، ثم يُقَسَّم المجال الواقع بين أعلى وأدنى سعة نبضة إلى عدة مستويات، وبعد ذلك تحوَّل الإشارة التشابهية إلى إشارة رقمية. ثم تجري المعالجة الأولية للإشارة الصوتية واعتماد معاملات هذه الإشارة في المجالين الزمني والترددي يتم تحليلها والتمييز بينها .

1.2. الاشارة الصوتية في المجال الزمني

يمكن ملاحظة المعاملات الإشارة الصوتية في المجال الزمني مباشرة من خلال البرمجيات الصوتية مثل Sound Forge أو Cool Edit، أو من البيئات البرمجية مثل Matlab أو Visual Prolog , حيث أسهل طريقة لتمثيل الإشارة الكلامية هي تمثيل سعة النبضة بدلالة الزمن حيث تتغير طبيعتها وشكلها بدلالة الزمن، ومن الضروري ملاحظة أن معاملات الإشارة الصوتية في المجال الزمني تتغير حسب الضجيج (المايكروفون والحاسوب، وبُعد المتكلم عن المصوات) المايكروفون (أثناء تسجيل الكلمات والارقام، والحالة الصحية والنفسية للإنسان وغير ذلك ,كما في الشكل (1) [2].



شكل (1) الاشارة الصوتية في المجال الزمني

2.2. استخلاص الصفات

تعد تقنية استخلاص الصفات من إحدى التقنيات ذوات التطبيق الواسع في مجالات عديدة من أهمها تمييز الأنماط (Pattern Recognition)، وتعد إحدى الركائز المهمة التي يقوم عليها هذا المجال، فهي من المراحل الأولى في عملية التمييز إذ تهدف إلى إيجاد سلسلة من متجهات الصفات (Feature Vector) تجهز المميز (Recognizer) بتمثيل محكم للإشارة الداخلة مما يسهل على النظام مهمة تمييز النمط الداخل مستخدماً تلك المعاملات الثابتة المستخلصة من إشارة الإدخال. يتم في مرحلة استخلاص الصفات اقتباس صفات الشيء لغرض المقدرة على تمييزه فيما بعد من خلال تلك الصفات التي تصف ذلك الشيء وتميزه عن غيره، ولم تقتصر عملية استخلاص الصفات على نوع محدد من البيانات فقد شملت الصورة والكلام والفيديو .ومما ينبغي ذكره أن خواص إشارة الكلام تتغير بصورة مستمرة مع الزمن نتيجة لعدد من العوامل الطبيعية، وهذا ما يجعل من المستحيل استخلاص الصفات إشارة الكلام بصورة كاملة دون تقسيمها إلى عدد من المقاطع(Frames) واستخلاص صفات كل مقطع، وذلك يرجع إلى كون إشارة الكلام مستقرة نوعا ما في فترة محددة قصيرة، وبذلك يكون الناتج عدد من متجهات الخواص(Feature Vector) بعدد المقاطع المكونة في الإشارة [11].

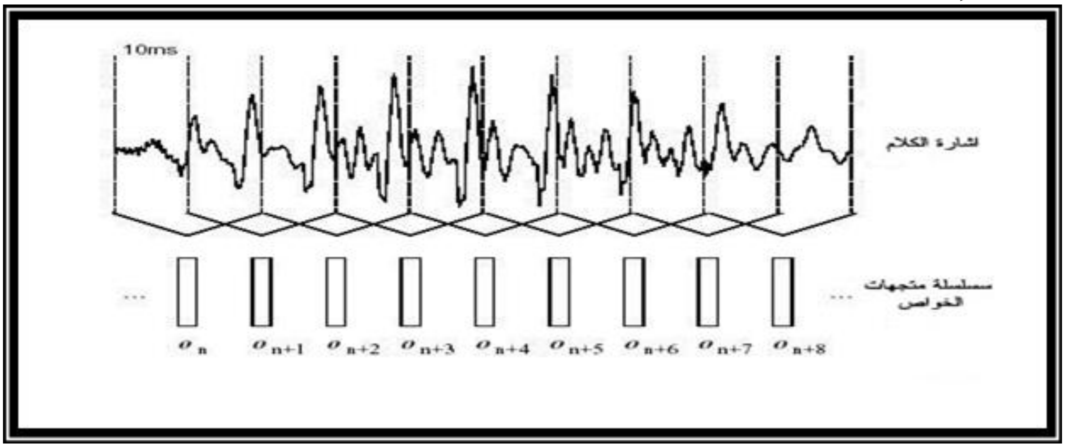
وبالرغم من انه لا يوجد اتفاق على الخواص المثالية لإشارة الكلام، لكن يفضل أن تكون الخواص المستخلصة للكلام متميزة بما يأتي :

1. يجب أن تسمح للأنظمة الأوتوماتيكية أن تمييز بين الأصوات المتشابه لإشارة الكلام.
2. يجب أن تعطي إحصائيات تختلف باختلاف المتحدثين وبيئة الكلام.

3.2. مرحلة استخلاص الصفات

تمثل مرحلة تحليل الإشارة لاستخلاص الصفات الصوتية بدقة والتي تمثل المعلومات الصوتية (Acoustic information) لتلك الإشارة حيث يتم معالجة الإشارة من البداية إلى النهاية وذلك بتقسيم الإشارة إلى إطارات متساوية في الطول ثم يتم استخلاص الصفات لكل اطار لتكون النتيجة النهائية متجه من الصفات يمثل ذلك الاطار. هناك عدة طرق تستخدم لاستخلاص الصفات لأية إشارة صوتية فيمكن استخدام الطاقة (Energy) ، معدل السعة (Average magnetite) ، معدل تقاطع الصفر(Zero crossing) ، معاملات درجة النغم

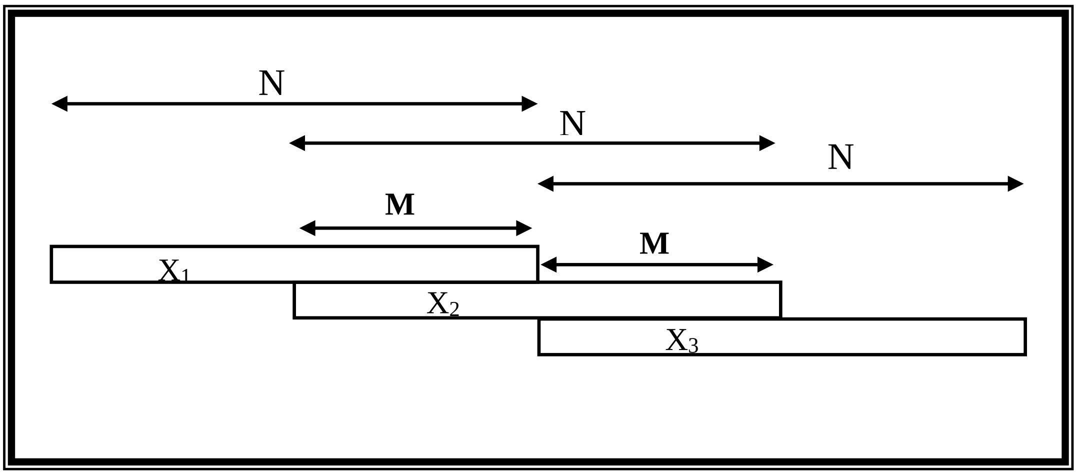
(Mel – Frequency cepstrum coefficients (MFCC)) ومعاملات التنبؤ الخطي (Linear predictive coefficients (LPC))، وفي هذا البحث تم استخدام طريقة درجة النغم (MFCC) لهذا الغرض.[1] الشكل (2) يوضح ما سبق ذكره على جزء مقتطع من إشارة كلام تم تقسيمه إلى تسعة مقاطع (Frames) وإيجاد متجه الخواص لكل مقطع على حدة: [4]



شكل (2): نظام جزئي لاستخلاص الصفات مقطع من إشارة كلام

4.2. تهيئة الاطارات

تقسم إشارة الكلام الرقمية إلى عدد من الإطارات كل إطار يتكون من N من العينات وهذه الاطارات تكون بشكل متجاور مفصولة عن بعضها بـ M من العينات فإذا كانت $M \le N$ فان الاطارات المتجاورة ستكون متداخلة (Adjacent frame overlap) من هذا المبدأ يتم تقسيم الاشارة الصوتية الى عدة إطارات متداخلة وبمقدار M من العينات . [4] كما في الشكل(3)



الشكل(3) تهيئة الاطارات بصورة متداخلة

5.2. المعالجة بالنافذة

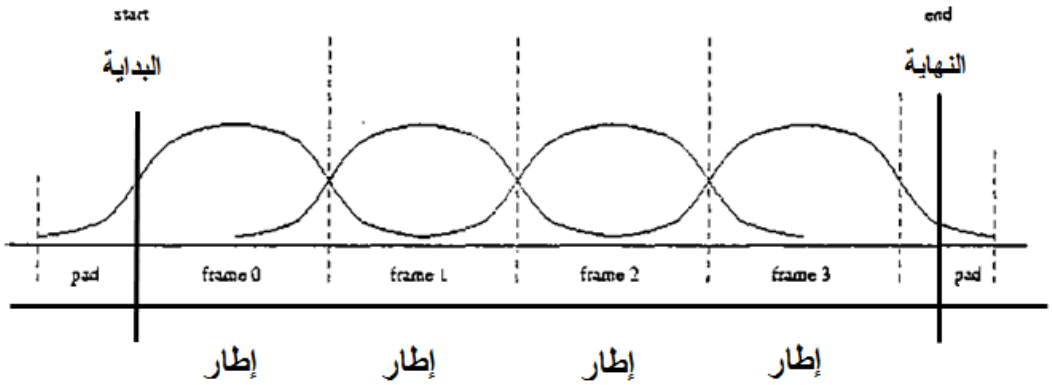
يتم في هذه الخطوة تمرير كل إطار من خلال نافذة بفترة معينة وذلك لتقليل عدم استمرارية الإشارة في بداية ونهاية كل اطار، [18] [19] كما موضح في الشكل (4) .

إذا عرفنا النافذة بـ W(n) والاشارة الداخلة بـ $X_l(n)$ فان الإشارة الخارجة هي :

$$X'_l(n) = X_l(n) * W(n) \qquad 0 \le n \le N-1 \quad \ldots\ldots\ldots\ldots \quad (1)$$

النافذة الأكثر شيوعا في أنظمة تمييز الكلام هي نافذة هامينك (Hamming window) والمعرفة بالمعادلة :

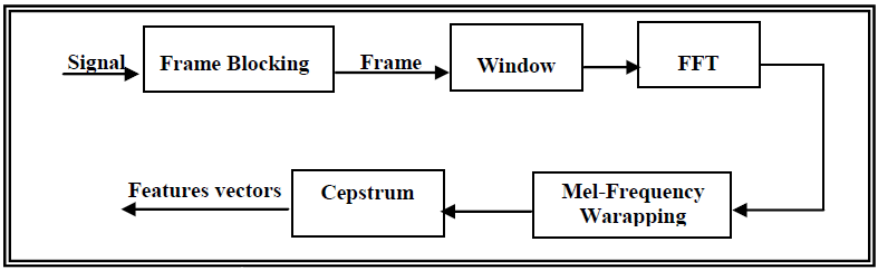
$$W(n) = 0.54 - 0.46 \cos [2\pi / N-1] \qquad 0 \le n \le N-1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$



الشكل(4) هيكلية النافذة بصورة متداخلة

6.2. تقنية درجة النغم

تعد خوارزمية الـ MFCC طريقة لتحليل الإشارة الصوتية ذات المواصفات العالية لتثمين مواصفات الكلام مثل طبقة الصوت ونغمته حيث تعتمد بالأساس على ترشيح الترددات الواطئة خطياً , والعالية لوغارتيما لأستخلاص الموصفات المهمة للصوت [18] وفي ما يأتي شرح مفصل للمراحل التي يمر بها الصوت لأستخلاص الصفات منه . الشكل (5) يوضح المخطط الكتلي لمعالجة (MFCC) حيث تم الذكر مسبقاً عن اول خطوتين لمخطط الـ MFCC .



الشكل (5) يوضح المخطط الكتلي لمعالجة (MFCC)

الخطوة الثالثة تحويل فورير السريع وهو تطبيق سريع لتحويل فورير المنفصل (DFT) الذي يحول N عينات التي تمثل إطار الى الطيف الترددي الاطار كما في المعادلة الاتية :

$$X_n = \frac{1}{N}\sum_{k=0}^{N-1}\left(X_k * e^{\frac{-2\pi jkn}{N}} \right) \qquad n = 0,1,2,\ldots,N-1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

حيث ان N : تمثل حجم الاطار

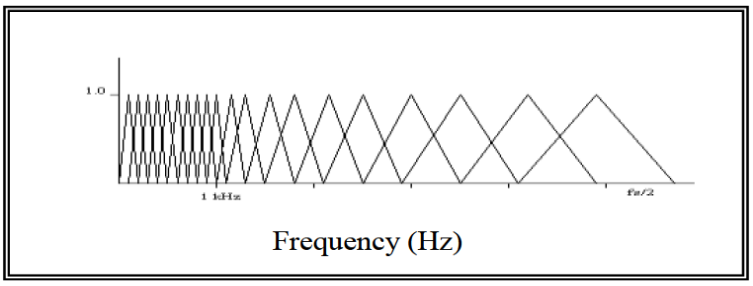
Xk : قيمة العينة قبل FFT

Xn : قيمة العينة بعد FFT

j : تمثل جذرا خياليا

الخطوة الرابعة: ان الاستفادة من مقياس الترددات (mel) هو عمل بنك ترشيح (filter bank) الذي يحوي مرشحا واحداً مكون من مكونات الـ mel المطلوبة حيث يتم بموجبه تجميع مرشحات

تمرير حزمة مثلثة الشكل (triangular band pass filter) . يتم بواسطة هذه الطريقة تغطية جميع الترددات بدون اي فقدان للمعلومات عند ادخال الاشارة الناتجة عن fft على المرشح تظهر مواقع القوة لتلك الاشارة [12][18] . وكما موضح بالشكل (6) لبنك الترشيح .



الشكل (6) بنك الترشيح –mel

Cepstrum : في الخطوة الاخيرة يتم تحويل نتيجة بنك الترشيح الى المنطلق الزمني وهي MFCC والتي تجهزنا بتمثيل جيد لمواصفات الطيف لجزء من اشارة الرقم الصوتي للإطار المحلل [12] [14] ، وتستخدم المعادلة الاتية لحساب الـ MFCC :

$$MFCC_i = \sum_{k=1}^{m} X_k COS \frac{\pi i (k-0.5)}{M} \qquad i = 1,2,...,P \quad \text{.....(4)}$$

حيث ان m : عدد مرشحات في بنك الترشيح

Xk : هي اشارة الإطار الواحد mel cepstrum

P : عدد معاملات المتجه الواحد للإشارة المحللة .

1.3. مقدمة

تم بناء نظام لتمييز الارقام العربية المنطوقة ، من خلال إدخال مجموعة الارقام واستخلاص الصفات وتمييزها اعتمادا على الخوارزمية الجينية في ذلك ، وهذا النظام يتضمن ثلاث مراحل هي:

- تهيئة البيانات .
- استخلاص الصفات .
- واخيرا مرحلة التمييز باستخدام الخوارزمية الجينية .

2.3. تهيئة البيانات

تم التسجيل لستة أشخاص اربعة من الذكور واثنان من الإناث تتراوح أعمارهم بين (22-30) عاما ، قام كل واحد منهم بتسجيل الارقام من(0-9) حيث ان كل رقم تم تسجيله 20 مرة وبذلك اصبح عدد الملفات(900) , التسجيل تم بظروف ملائمة وباستخدام لاقط الصوت (ميكروفون) ، في هذا البحث تم تسجيل الرقم بمعدل عينات يساوي (11025 HZ) وتمثيل كل عينة بـ (bit16) واستخدام الأسلوب الأحادي (Mono) في التسجيل. إن الاصوات التي تم تسجيلها تمثل الأرقام من العدد صفر إلى العدد التسعة باللغة العربية الفصحى ولعدد من المتكلمين والتسجيل تم في ظروف متشابهه لكافة الأشخاص وتم حفظ هذه المقاطع بالامتداد (wav) وهي من الصيغ المستخدمة بشكل واسع في بيئة النوافذ (Windows) لذا تعد من اكثر صيغ الصوت الشائعة حاليا .

1.2.3. مرحلة إزالة الصمت

من المشكلات التي تواجه عملية معالجة إشارة الكلام الرقمية هي تحديد بداية ونهاية النطق الحقيقي وذلك لتقليل المعالجة المطلوبة لإشارة الكلام ولتصبح عملية التمييز اكثر كفاءة ووثوقية . إزالة الصمت في هذا البحث تتلخص بخطوتين أساسيتين هما:

1. حساب معدل السعة لإشارة الكلمة الرقمية .
2. إزالة الصمت بتطبيق خوارزمية تحديد بداية ونهاية الكلام.

ان السعة لإشارة الكلام تتغير مع الزمن ، وتمثيل إشارة الكلمة بالاعتماد على معدل السعة لوقت قصير يعطي تمثيل ملائم لإشارة الكلمة لأنه يعكس هذه التغييرات بشكل واضح وجيد ، حيث يتم إنجاز هذه العملية من خلال الخطوات التالية:

1. معالجة إشارة الكلمة من البداية إلى النهاية بنافذة هامينك وكالآتي

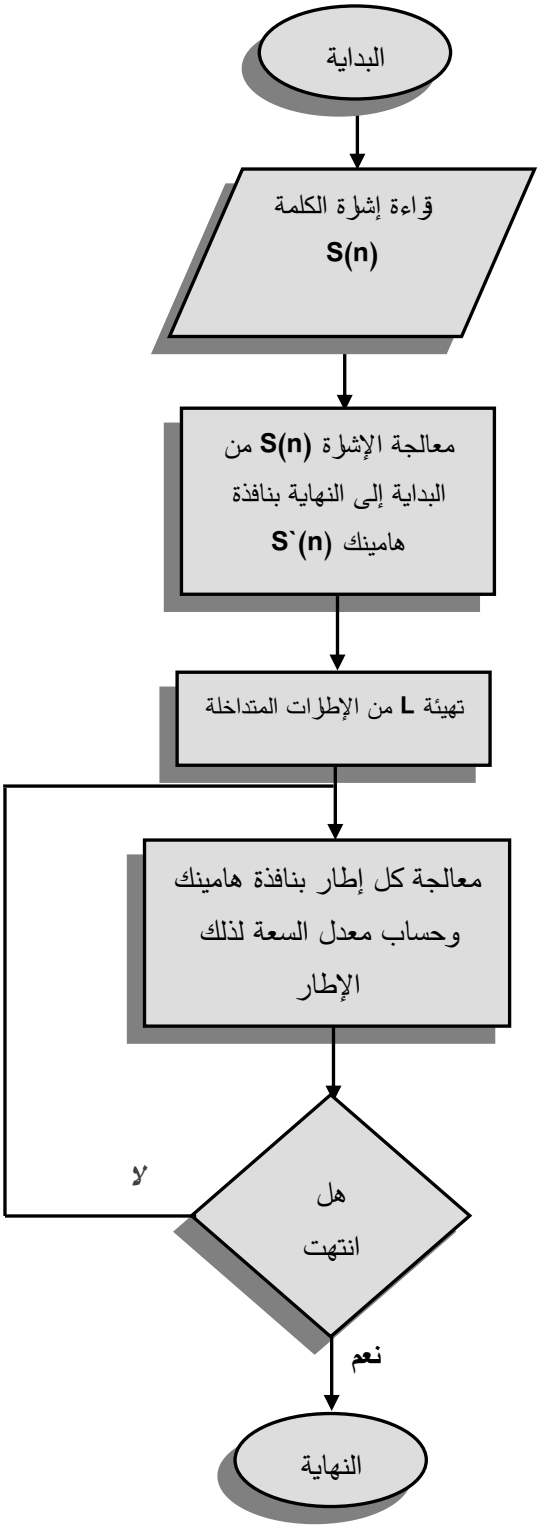
$$S`(n) = S(n) * W(n) \quad n = 1,2,\ldots, N \quad \ldots\ldots (5)$$

$S(n)$: إشارة الكلمة الأصلية .

$W(n)$: نافذة هامينك .

(N) : تمثل عدد العينات في كل إطار .

2. تقطيع الإشارة $S`(n)$ من البداية إلى النهاية إلى اطارات متداخلة ومتساوية في الحجم كل إطار يتكون من $(N = 512)$ من العينات وبتداخل مقداره $(M=256)$ من العينات، وكل إطار يتم معالجته بشكل منفصل بأخذ نافذة هامينك ثم حساب معدل السعة لذلك الإطار بتطبيق المعادلة لنافذة هامينك ، الشكل (7) يوضح عملية حساب السعة للإشارة الرقمية .



الشكل (7) المخطط الانسيابي لعملية حساب معدل السعة لاشارة الرقمية

2.2.3. استخدام خوارزمية تحديد بداية ونهاية الكلمة الرقمية

– لتكن A تمثل مصفوفة حساب معدل السعة للإشارة الداخلة فنأخذ أعلى قيمة واقل قيمة من قيم معدل السعة والتي يمكن تمثيلها بـ Amax و Amin على الترتيب.

– نحسب قيمة العتبة (Threshold) للكلام وكالآتي:

$I_1 = 0.03 * (Amax – Amin) + Amin$ .....(6)

$I_2 = 4 * Amin$

$I_3 = MIN (I_1, I_2) + 0.4$

– يتم حساب الإطارات التي سيتم إزالتها من بداية الإشارة بالاعتماد على قيمة العتبة ، الإطارات المحذوفة من قيم معدل السعة للإشارة تمثل الإطارات التي معدل سعتها اقل من قيمة العتبة فإذا فرضنا أن F تشير إلى عدد الإطارات التي سيتم إزالتها من بداية الإشارة بالتالي يمكن حساب عدد العينات التي يتم إزالتها على النحو الاتي:

عدد العينات المزالة = (((F-1)*M) + N)

– وفي حالة حذف السكوت من نهاية الكلمة تطبق نفس الخطوات السابقة ، لإزالة السكوت من بداية الإشارة مع الأخذ بنظر الاعتبار ان : I=1 طول المصفوفة (A) =J

وعليه نستمر بحذف الإطارات مادامت قيمة السعة اقل من العتبة وان I < J ثم يتم إزالة العينات التي تمثل السكوت من نهاية الإشارة بنفس الطريقة السابقة [1] [8].

4. مرحلة استخلاص الصفات

تعد مرحلة استخلاص الصفات من اهم المراحل التي تؤخذ بنظر الاعتبار عند تصميم أي نظام تمييز ، ولقد استخدمت خوارزمية الـ (MFCC)) في عملية استخلاص الصفات.

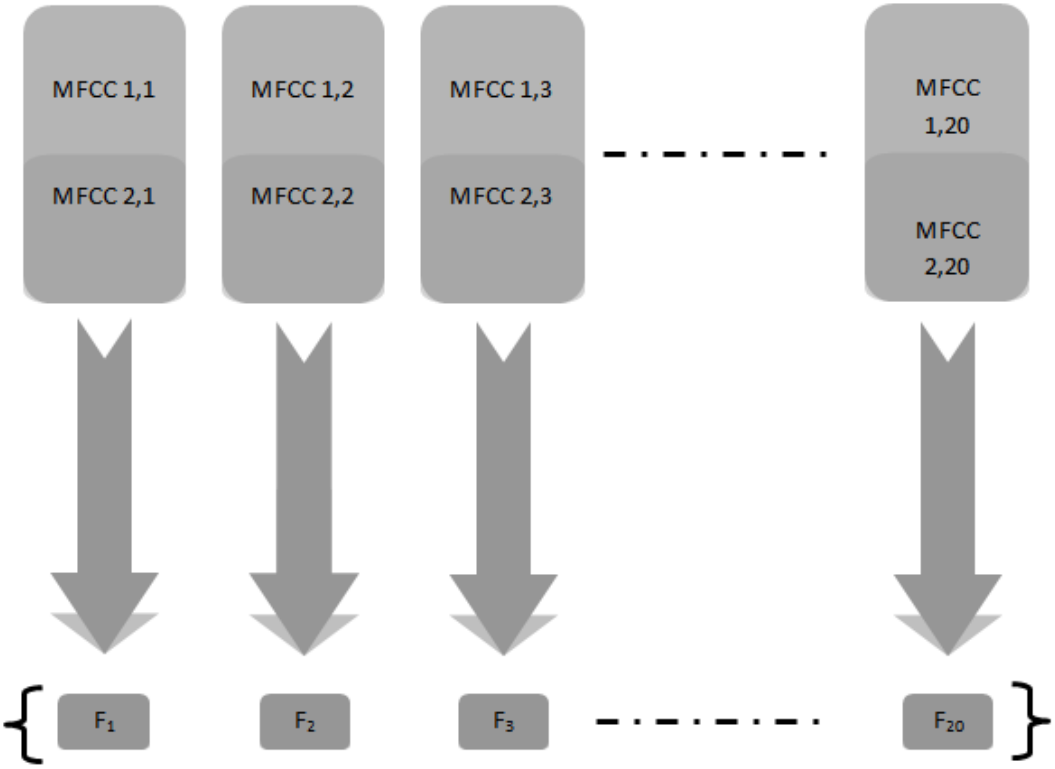
1. تطبق مرحلة المعالجة الاولية للاشارة الصوتية وتقسم الاشارة الى عدة اطارات ، كل اطار عبارة عن (256 عينة) وبالتالي تنتج مصفوفة عبارة عن (m*n) حيث ان :

   m : طول الاطار n : عدد الاطارات

2. تطبق تقنية الـ (MFCC) كما موضح في الفصل الثاني ، لكل صف من المصفوفة للحصول على (24) معامل لكل اطار .
3. تطبق معادلة الالتواء (skewness) للحصول على مصفوفات احادية تستخدم فيما بعد في عملية التمييز باستخدام الخوارزمية الجينية ، فائدة المعادلة لتجاوز المشكلة الناتجة عن عدم تحديد زمن ثابت لتسجيل الكلمات والتي تمثل اختلاف في عدد المقاطع الناتجة من تسجيل الى اخر وهذا بدوره يؤدي الى اختلاف في عدد متجهات الصفات التي يمكن الحصول عليها من كل كلمه, تكرر عملية استخلاص الصفات لجميع الملفات الصوتية الخاصة بالتدريب لكل رقم وتخزن في ملف ليتم استخدامه في عملية التمييز . كما في الشكل (8) :

   σ^3 : تمثل الانحراف المعياري

$$Skewness = \frac{1}{N}\left(\frac{\sum_{i=1}^{N}(Xi - \overline{X})^3}{\sigma^3}\right) \quad \text{.....(7)}$$



شكل (8) مخطط تطبيق معادلة الالتواء لاستخراج متجه الصفات

5. تمييز الأرقام العربية باستخدام الخوارزمية الجينية

تم تخصيص 900 ملف صوتي لـ ستة اشخاص لعشرة ارقام (0-9) لخمسة عشر مقطعاً صوتياً لكل رقم لتكوين قاعدة بيانات وذلك عن طريق تكرار عملية استخلاص الصفات لخمسة عشر تسجيلاً لكل رقم ولجميع الاشخاص وخزنها في ملف ليتم استخدامها فيما بعد في عملية التمييز, فيما ياتي خطوات توضح الخوارزمية المقترحة :

1. قراءة الملفات الخاصة بالارقام من 0 – 9 حيث يكون الملف الخاص بالرقم 0 بمثابة الجيل الاول للرقم صفر , والملف الخاص بالرقم 1 بمثابة الجيل الاول للرقم واحد وكذلك باقي الارقام .
2. قراءة الملف الصوتي المراد تمييزه واستخلاص الصفات منه .
3. حساب لياقة كل الافراد للرقم (0) عن طريق حساب الفرق الاقليدي لادنى مسافة (Euclidean minimum distance) وتخزن اللياقة بالمتجه (Vector) خاص مثل(اللياقة 1).
4. اعادة ترتيب افراد الجيل بالاعتماد على لياقة كل فرد تصاعدياً من الفرد ذي افضل لياقة وحتى الافراد السيئين اللياقة فالأسوأ .
5. تحديد نسبة الاستبدال والبقاء التي تمثل معياراً لعدد من الافراد في الجيل الحالي ، الذي سيتم نسخ موروثاتهم الى الجيل الجديد يهدف الحفاظ على افضل الصفات .
6. نسخ عدد محدد من الاباء من بداية مصفوفة جيل الاباء للرقم (0) بالاعتماد على نسبة الاستبدال والبقاء وخزنهم في مصفوفة جديدة (اقوى الاباء 1) وكذلك خزن قيم اللياقة لهؤلاء الاباء في متجه اللياقة (اللياقة المثلى 1) .

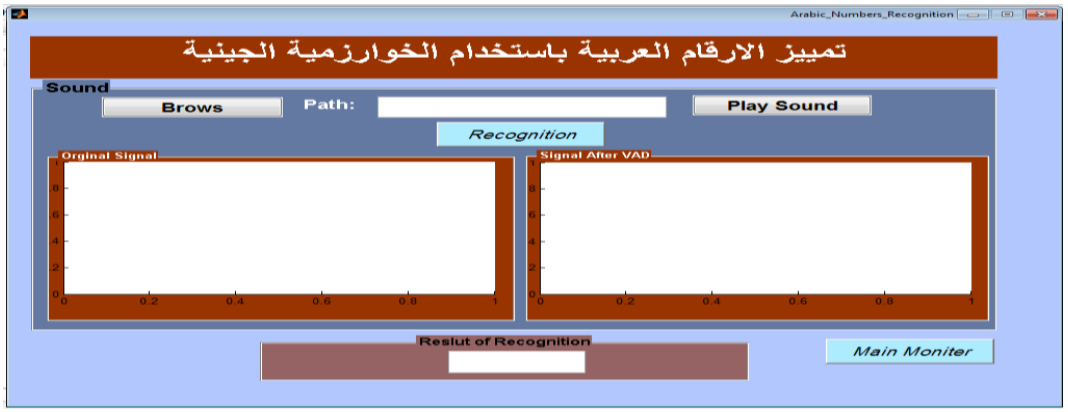
7. اقصاء عدد محدد من اسوء الاباء من نهاية مصفوفة الجيل الاباء للرقم (0) بالاعتماد على نسبة الاستبدال والبقاء , وحذف قيم اللياقة لهؤلاء الاباء (الاسوأ من متجه اللياقة (اللياقة 1)).
8. تكوين جيل من الرقم(0) بالاعتماد على اخر توليد للجيل للرقم (0) ويتم اختيار الكروموسومين اثنين ذوي اعلى لياقة من الجيل الاخر اللذين لم يتم اختيارهما سابقاً لتوليد افراد جدد.
9. يستخدم التزاوج الرياضي (arithmetic crossover) لتوليد اطفال جدد ويعد هذا النوع ملائماً جداً مع القيم الحقيقية .
10. يتم حساب لياقة الاطفال الجدد باستخدام حساب الفرق الاقليدي لأدنى مسافة ما بين الاطفال الجدد ، وخزن اللياقة في متغير مثل لياقة اثنين الذي سيحمل جميع قيم لياقة الجيل الجديد من الرقم (0) .
11. تطبيق مفهوم الطفرة على الكروموسوم الجديد وذلك بتولد قيم عشوائية (في هذه الرسالة تم الاعتماد على دالة rand) بلغة ماتلاب لتوليد هذه القيم العشوائية تمثل نسبة احتمال حدوث الطفرة ، فإذا كانت القيمة العشوائية اقل من (0.001) قيم استدعاء دالة الطفرة التي تكون من نوع جمع ، طرح (add – sub mutation) وهي ملائمة جداً مع القيم الحقيقة .
12. خزن الافراد الجدد (الاطفال) في متجه يرمز الى الجيل الجديد المتولد (الجيل الجديد للرقم (0)) .
13. حذف أخر كروموسومين تم استخدامهما في عملية التزاوج في الجيل الحالي ليتسنى إجراء عملية التزاوج على بقية افراد الجيل , ومن دون تكرار لأحد الكروموسومات , مع حذف قيم لياقتهم من متجه اللياقة (اللياقة 2)
14. تكرار الخطوات من (8) الى (14) للكروموسومات الباقية في الجيل السابق .
15. إضافة أفضل الآباء من الجيل السابق (أقوى الاباء 1) الى الجيل المتولد (الجيل الجديد 1) , وكذلك إضافة قيم اللياقة لهؤلاء الآباء (اللياقة المثلى 1) إلى قيم اللياقة للجيل الجديد (اللياقة 2) , وبذلك يتم الحصول على العدد نفسه من افراد الجيل الأصلي في الجيل المتولد مع المحافظة على افضل موروثات الأفراد التي قد تقود إلى الحل الأفضل.
16. بعد انتهاء مجمل العمليات الجينية لتوليد جيل جديد يتم خزن الجيل الجديد (الجيل الجديد 1) للرقم (0) في مصفوفة الجيل الحالي للرقم (0) , وخزن قيم اللياقة من مصفوفة اللياقة (اللياقة 2) الى مصفوفة (اللياقة 1) , وإحلال كل من مصفوفة لجيل الجديد (الجيل الجديد 1) ومصفوفة اللياقة (اللياقة 2) من اجل الاستعداد لعملية توليد الجيل الجديد وحساب قيم لياقاته .
17. الاحتفاظ بأفضل (أقل) قيمة لياقة للجيل الجديد في متغير (الأفضل 1) الذي يمثل متجهآ يحوي على افضل قيمة محتملة (أفضل حل optimal solution) , والمأخوذة من الجيل كله .
18. شرط التوقف المقترح هو اعتماد عدم التغير في اللياقة لأجيال وإلا فإن عملية توليد الاجيال الجديدة تستمر حتى مئة جيل , وكل هذه الاجيال يتم مقارنتها مع كروموسوم للرقم (0) , ومجموع الخواص الرقم (0) .
19. بعد الانتهاء من توليد اجيال الرقم (0) يتم اعتماد الاسلوب السابق نفسه في التوليد للرقم التالي , الرقم (1) , الرقم (2) وكذلك بقية الأرقام .

20. بعد توقف توليد الاجيال (لجميع الارقام) بالاعتماد على المقارنة مع الكروموسوم المجهول لكل رقم , وحساب افضل لياقة لكل جيل , فأن الرقم المجهول يكون من الرقم الذي يكون له اقل قيمة لياقة من بين هذه القيم .
21. طباعة نتيجة التمييز .
22. تكرار الخطوات من (3) الى (22) .
23. نهاية خوارزمية التمييز الجينية .

6. كيفيــة استخدام النظـام

لغرض توفير إمكانية تحقيق المهام الرئيسة للنظام من (إزالة السكوت من بداية ونهاية الإشارة الرقمية ، تمييز هذه الارقام بالاعتماد على كيفية نطقها) فانه يتم التعامل مع النظام من خلال عدد من الواجهات التي تم بناؤها لتسهيل استخدام النظام المقترح . وفيما يأتي شرح لكيفية استخدام النظام :

في البداية تشغيل برنامج الـ (matlab) سوف تظهر لنا النافذة الرئيسة للنظام المبنى للتميز الارقام العربية كما مبين في الشكل(9) :

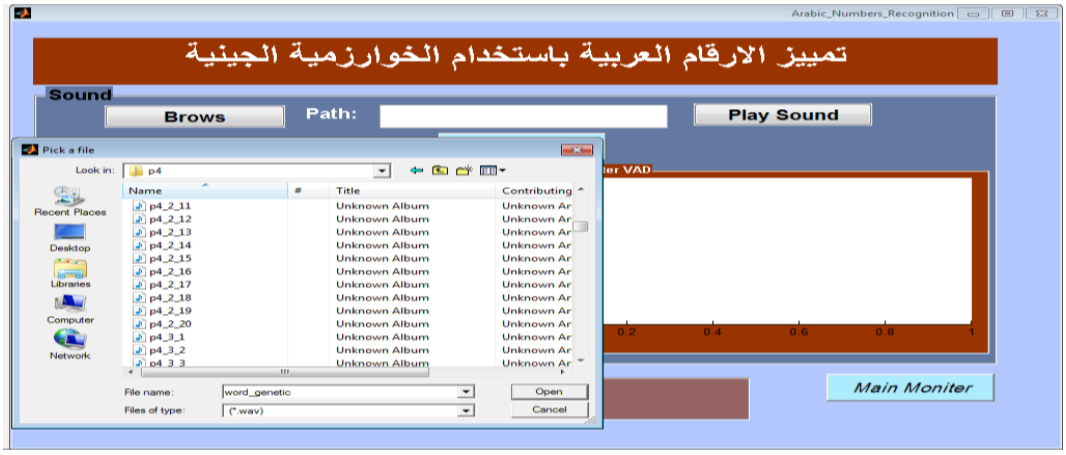


الشكل(9) الواجهـة الرئيسيـة

حيث تحتوي النافذة على مايلي :

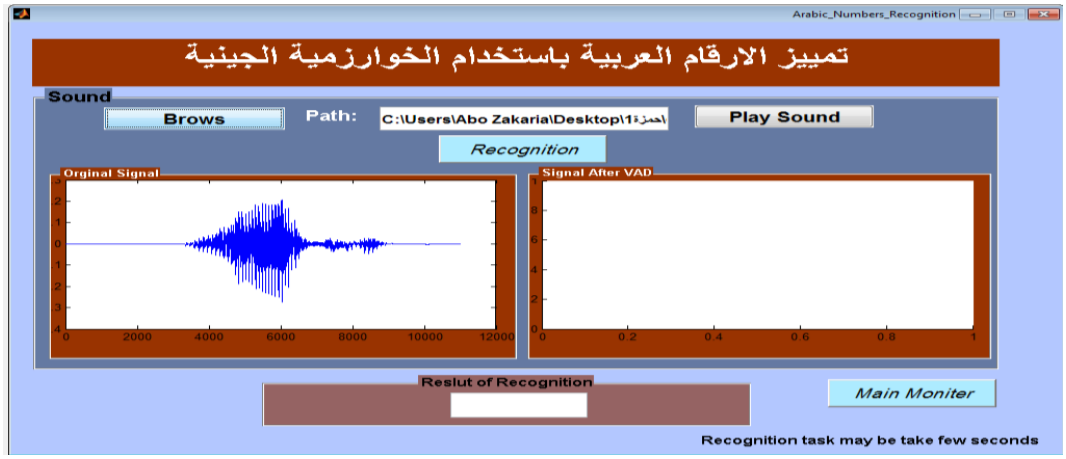
1. زر الامر (Brows) : يتم من خلال هذا الزر اختيار الصوت الرقمي المقطع مسبقاً ، حيث يظهر لنا مسار الصوت المختار في نافذة الحوار (Path) ، وتحويله الى شكل اشارة صوتية في نافذة الحوار (Orginal Signal) .
2. زر الامر (Play Sound) : يتم من خلال هذا الزر تشغيل الصوت المختار من قائمة (Brows) وسماعه للتأكيد .
3. زر الامر (Recognition) : يتم من خلال هذا الزر تمييز الاشارة الرقمية الممثلة في نافذة الحوار (Orginal Signal) ، مع ازالة لحظات الصمت من الاشارة الرقمية الممثلة في المجال الزمني وتمثيلها في نافذة الحوار (Signal after VAD) ، يتم اظهار ناتج التمييز للإشارة الرقمية في نافذة الحوار (Result of Recognition) .

بدايةً يتم تحديد احد الأرقام المقطعة من خلال مفتاح الـ (Brows) ،التي تظهر عند بداية الدخول للبرنامج ،حيث تظهر لنا قائمة بجميع الاصوات المقطعة كما مبين في الشكل (10)



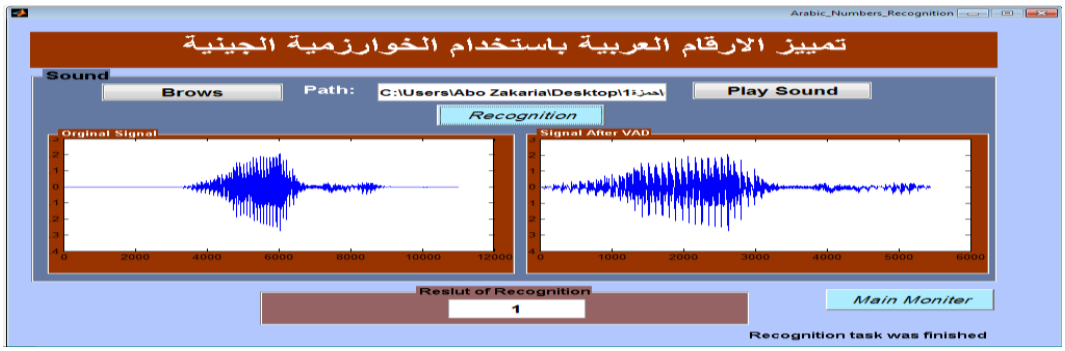
الشكل (10) نافذة اختيار الصوت المقطع

سيتم بعد اختيار احد الاصوات المقطعة ، ظهور الصوت كأشارة رقمية في نافذة الحوار لل (Original Signal) الممثلة للمجال الزمني للاشارة الرقمية ، كما في مبين في الشكل (11)



الشكل (11) نافذة تظهر اختيار الصوت وتمثيله الى اشارة رقمية

بعد ذلك يتم اختيار زر الأمر (Recognition) ليتم تمييز الإشارة الرقمية الداخلة مع إزالة الصمت منها ، ويتم عرضها في نافذة الحوار (Signal after VAD) وإظهار ناتج التمييز في نافذة الحوار (Result of Recognition) كما مبين في الشكل (12) :



الشكل (12)نافذة تظهر الصوت بعد تمييزه و إزالة لحظات الصمت منه

7. النتائج

بعد تطبيق الخوارزمية الجينية على الاشارة الرقمية ، تم الحصول على النتائج الاتية كما موضح في الجداول الاتية لكل من البيانات المدربة والبيانات غير المدربة وبتطبيق المعادلة التالية :

$$\text{نسبة التمييز المئوية} = \frac{\text{عدد الكلمات المميزة بشكل صحيح}}{\text{العدد الكلي للكلمات}} * 100\%$$

الجدول (1) يوضح نسب التمييز المئوية للبيانات المدربة

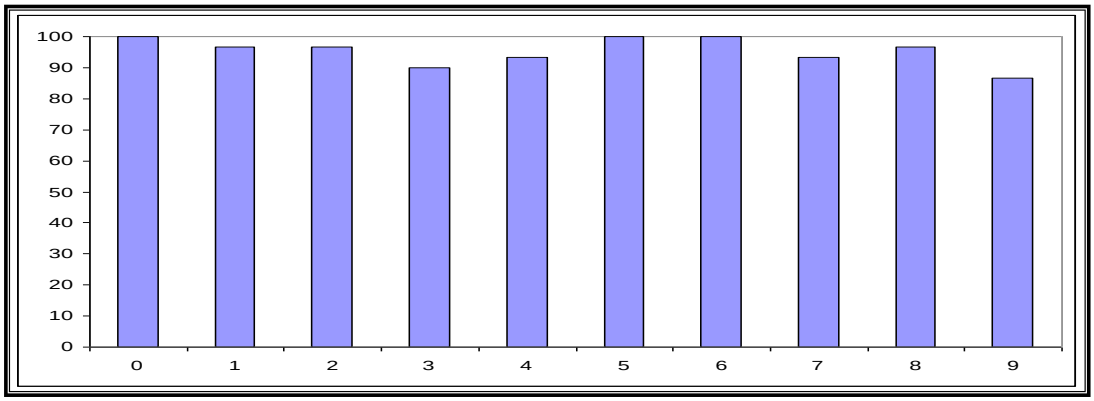
الرقم	0	1	2	3	4	5	6	7	8	9
النسبة %	100	100	100	100	100	100	100	100	100	100

وكانت نسبة التمييز الكلية (100%) .

الجدول (2) يوضح نسب التمييز المئوية للبيانات غير المدربة

الرقم	0	1	2	3	4	5	6	7	8	9
النسبة %	100	96.6	96.6	90	93.3	100	100	93.3	96.6	86.6

وكانت نسبة التمييز الكلية (95%) .



الشكل (13) مخطط يوضح النسبة المئوية للبيانات الغير مدربة

الجدول (3) يوضح عدد التسجيلات الصحيحة للأشخاص

الأشخاص / الأرقام	0	1	2	3	4	5	6	7	8	9
Male 1	5	5	5	5	5	5	5	4	5	5
Male 2	5	5	5	5	5	5	5	5	5	5
Male 3	5	5	5	5	5	5	5	5	5	5

Male 4	5	4	4	4	4	5	5	5	4	5
Female 1	5	5	5	4	5	5	5	5	5	5
Female 2	5	5	5	5	5	5	5	5	4	1

8. الاستنتاجات

تم من خلال هذه الدراسة التوصل إلى الاستنتاجات الآتية :

1. لإزالة السكوت (Silence) من بداية ونهاية الإشارة تأثير كبير على دقة النتائج سواء في عملية التقطيع أو التمييز.
2. لنسبة التعيان تأثير كبير على عملية التمييز ، فزيادة نسبة التعيان لها تأثير فعال بزيادة دقة التمييز على الرغم من ان تأثيرها يكون واضحاً بزيادة الزمن المطلوب لإتمام عملية التمييز.
3. للضوضاء تأثير سلبي على نتائج عملية التمييز ولهذا يفضل تسجيل الارقام في غرفة مغلقة بدون وجود أي نوع من الضوضاء ، فضلاً عن تأثير الحالة النفسية والصحية للشخص المتكلم على عملية التمييز.
4. ان معالجة الإشارة باستخدام طريقة التحليل (MFCC) فضلاً عن سرعتها ودقة نتائجها فإنها تزودنا بنموذج جيد لإشارة الرقمية وذلك بإعطائها افضل تمثيل ممكن للإشارة وهذا يقود إلى دقة أعلى في أداء التمييز .

8. المقترحات المستقبلية

هناك عدة مقترحات مستقبلية نوصي بتطبيقها :

1. يمكن استخدام طرق تحليل أخرى الإشارة الصوتية مثل (LFCC) أو (LPC) لغرض استخلاص الصفات بدلا من (MFCC) .
2. استخدام طرق أخرى في عمليات التمييز مثل نموذج ماركوف المخفي (HMM) والشبكة العصبية بدلا من الخوارزمية الجينية .
3. تهجين الخوارزمية الجينية بالمنطق المضبب لزيادة نسبة التمييز.

المصادر

[1] الطاهر ، انعام غانم سعيد ، (2004) ، " تمييز الفونيمات العربية " ، بحث ماجستير ، علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل .

[2] بسام على مصطفى ، (2003) ، " نظام قاعدة بيانات لتشخيص المتكلمين " ، بحث دكتوراه ، علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل .

[3] سارة عزيز وسبا عبدالواحد وميسا عبدالكريم, (2006)," خوارزمية تدريب السريعة والخوارزمية الجينية لتمييز بعض الفونيمات العربية", بحث علوم الحاسوب, كلية علوم الحاسوب والرياضيات، جامعة البصرة.

[4] شهلة عبدالوهاب عبدالقادر ، (2010) ، " تمييز اصوات الارقام العربية " قسم انظمة الحاسوب ، المعهد التقني الموصل ، جامعة الموصل .

[5] صدام, صبا عبدالواحد, (2008),"استخدام الخوارزمية الجينية لتحديد معمارية شبكة انسياب الخطأ خلفا المعدلة بمعامل تعلم متكيف لتمييز بعض الفونيمات العربية", بحث ماجستير,علوم الحاسوب والرياضيات, جامعة البصرة.

[6] عربو ، ايمان فتحي احمد ، (2004) ، "معالجة الكلام باستخدام المرشحات المكيفة " ، بحث ماجستير ، علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل.

[7] غصون سالم بشير ، (2003) ، "استخدام الخوارزمية الجينية في مطابقة الصور" ، بحث ماجستير ، علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل.

[8] فاتن بشير عبدالاحد ، انعام غانم سعيد ، (2005) ، " تقطيع الكلمة العربية الى احرف وتمييزها " ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، مجلة الرافدين لعلوم الحاسوب والرياضيات المجلد (2) العدد (2) .

[9] فهد العتيبي ، (2002) رسالة ماجستير ، " التعرف على الكلام العربي لمتحدث واحد باللغة العربية " .

[10] قتيبة مازن عبد المجيد ، (2009) ، " استخدام الذكاء الصناعي في تطبيقات الهندسة الكهربائية " ، بحث ماجستير ، رسالة مقدمة الى الاكاديمية العربية في الدنمارك في نظم المعلومات الادارية .

[11] متراس ، بان احمد حسن ، همسة معن ثابت ، (2007) ، " استخدام الخوارزمية الجينية في حل مسألة النقل " ، المجلة العراقية للعلوم الاحصائية (11) ، ص[134-152].

[12] Bisrat Girma , (2012),' Non-uniform sampling based feature extraction for automatic speech recognition ' , master of science in electrical engineering , addis ababa university .

[13] Fadila Maouche, Mohamed Benmohamed,(2008)," Automatic

Recognition of Arabic words by Genetic Algorithm and MFCC Modeling", degree of master of science in Engineering Letters.

[14] Gaurar , Deranesamoni shaking Deiv , Gopal krishna sharme , and Mahna Bhuttacharya , (2012) , ' Development of Application specific continous speech recognition system in hindi ' , journal of signal and information processing ,3 , 394 – 401 .

[15] Jose M. Villegas, Alejandra Mancilla, Patricia Melin,(2008)," Optimization of Modular Neural Network, Using Genetic Algorithms: The case of face and voice Recognition", soft computing for hybrid intelligent system.

[16] Marat wroniszewske and Jacek Dziedzic,(2010)," Voice Command Recognition using Hybrid Genetic Algorithm", Facutly of Technical Physics and Applied Mathematics, Gdansk University of Technology.

[17] Miguel Lopez and Oscar Castillo,(2006)," Voice Recognition with Neural Network, Type-2 Fuzzy Logic and Genetic Algorithms", thesis of science in Engineering Letters.

[18] Noelia Alcaraz mesegner , (2009) , ' Speech Analysis for Amtomatic speech recognition ' master of science in electronics , Norwegian university of science and technology .

[19] Yuan Meng , (2011) , ' Speech Recognition on DSP : algorithm optimization and performance analysis ' , master of philosophy in electronic engineering , Chinese university of hong kong .