

تحسين أداء الخوارزميات المستخدمة في تمييز الكلام

د. حسن الأحمد *

طارق فارس علي **

(تاريخ الإيداع 23 / 7 / 2018. قبل للنشر في 24 / 12 / 2018)

□ ملخص □

تعد تقنيات التعرف على الكلام من أهم التقنيات الحديثة وقد تم تطوير العديد من الأنظمة المختلفة من حيث الطرق المستخدمة في استخراج السمات وطرق التصنيف.

يتضمن التعرف على الصوت مجالين هما: التعرف على الكلام والتعرف على المتكلم، حيث اقتصر البحث على مجال التعرف على الكلام.

يقدم البحث مقترحاً لتحسين أداء أنظمة التعرف على الكلمات المفردة عن طريق خوارزمية للجمع بين أكثر من تقنية من التقنيات المستخدمة في استخلاص السمات وتعديل الشبكة العصبونية لدراسة تأثيرها على عملية التعرف ودراسة تأثير الضجيج على النظام المقترح.

تم في هذا البحث دراسة أربع أنظمة لتمييز الكلام، حيث اعتمد النظام الأول خوارزمية MFCC لاستخلاص السمات واعتمد النظام الثاني خوارزمية PLP، في حين اعتمد النظام الثالث على الدمج بين سمات الخوارزميتين السابقتين إضافة إلى معدل تخطي الصفر، وفي النظام الرابع تم تعديل الشبكة العصبونية المستخدمة في عملية التمييز وتقليل نسبة الخطأ فيها، كما قمنا بدراسة أثر الضجيج على هذه الأنظمة السابقة.

تمت مقارنة النتائج من حيث معدل التعرف وزمن تدريب الشبكة العصبونية لكل نظام على حدة، لنحصل على نسبة تعرف وصلت حتى 98% باستخدام النظام المقترح.

الكلمات المفتاحية: تمييز الكلام، MFCC، PLP، الشبكات العصبونية الاصطناعية (ANN).

* أستاذ مساعد - قسم هندسة الحاسبات والتحكم الآلي - كلية الهندسة الميكانيكية والكهربائية - جامعة تشرين - اللاذقية - سوريا.

البريد الإلكتروني: has_zav@yahoo.com

** طالب دراسات عليا (ماجستير) - قسم هندسة الحاسبات والتحكم الآلي - كلية الهندسة الميكانيكية والكهربائية - جامعة تشرين -

اللاذقية - سوريا.

البريد الإلكتروني: taemtito@gmail.com

Enhancing The Performance Of Speech Recognition Algorithms

Dr.Hassan Alahmad*
Tarek Ali**

(Received 23 / 7 / 2018. Accepted 24 / 12 / 2018)

□ ABSTRACT □

The speech recognition is one of the most modern techniques, many related systems were developed, which they differ in feature extraction methods and classification methods.

Voice recognition is divided into two areas: speech recognition and speaker recognition, however the research was limited to focus on the field of speech recognition.

The research offers a proposal to improve the performance of single word recognition systems by an algorithm to combine more than one of the techniques used in features extraction, and modify the neural network to study its effect on speech recognition, and to study the effect of noise on the proposed system.

Four systems were studied for speech recognition, first one used MFCC algorithm for features extraction, second one used PLP algorithm, third one merged MFCC, PLP, and zero crossing rate features ,in the last system we modified the neural network with less error rate, We have studied the impact of noise on these previous systems.

The research provided a comparative study for the recognition ratio, and training time for each system, to obtain a recognition ratio reached up to 98% using the proposed system.

Key words: Speech recognition, MFCC, PLP, and artificial neural networks(ANN).

*Assistant Professor, Department Of Computer and Automation control- Faculty of mechanical and electrical Engineering - Tishreen University - Lattakia – Syria
E-mail:has_zay@yahoo.com

**Master student, Department Of Computer and Automation control- Faculty of mechanical and electrical Engineering - Tishreen University - Lattakia – Syria
E-mail: taemtitoo@gmail.com

مقدمة:

بدأ اهتمام خبراء الحاسب والباحثين في مجال التعرف على الكلام منذ أكثر من أربعة عقود، وذلك لكي نصل إلى مرحلة نجعل الآلة قادرة على فهم كلام الانسان وتلقي الأوامر والتعليمات صوتياً وبدون الحاجة إلى وسائل الادخال التقليدية، وذلك توفيراً للوقت والجهد.

أجريت العديد من البحوث والدراسات في مجال التعرف على الكلام خلال العقود الماضية ففي عام 2015 قام الدكتور جعفر الخير والمهندسة راما حسن بتصميم خوارزمية لدمج عدة أنظمة لاستخلاص السمات وتم اعتماد /4/ كلمات في عملية التعرف وهي Net Documents, Shutdown, Restart, تم تسجيلها بصوت /33/ شخص من مختلف الأعمار [1].

في هذا البحث استخدمت خوارزميات MFCC و PLP و LPCC في استخلاص السمات و HMM كمصنف وتوصلت الدراسة إلى معدل التعرف الأعلى /93.44/ عند تكامل نظامي MFCC و PLP ولكن ذلك كان على حساب زيادة زمن التعرف وزمن التدريب.

في عام 2015 قام كل من Inge Gavant و Diana Mulitaru بدراسة المصنفات المختلفة وارتباطها بحجم المفردات المستخدمة في عملية التعرف [2].

تم استخدام خوارزميتي PLP و MFCC في استخلاص السمات وأظهرت نتائج هذا البحث تفوق الشبكات العصبونية على نموذج ماركوف المخفي بنسبة 10% حيث أصبحت الشبكات العصبونية هي المصنف القياسي المستخدم في عملية التعرف على الكلام منذ عام 2011.

في عام 2015 أيضاً قام كل من Veton Z. Këpuska و Hussien A. Elharati بدراسة للخوارزميات المختلفة في استخلاص السمات حيث تم استخلاص /13/ سمة باستخدام كل خوارزمية ودمج السمات ليصبح العدد الكلي للسمات في كل عملية التعرف /39/ سمة حيث اعتمد البحث أصوات /208/ متحدت من الذكور والإناث الراشدين والكلمات المستخدمة هي (zero to nine and the letter o) وكان عدد عينات التدريب /2072/ وعدد عينات الاختبار /2486/، حيث تم اختبار النظام في شروط ضجيج مختلفة وقد توصلت هذه الدراسة إلى أن خوارزمية PLP مع عينات دلّتا ودلّتا الدلتا هي الأفضل في شروط الضجيج المرتفع بنسبة تصل إلى /93.52% وفي التعرف على الكلام بدون وجود ضجيج كانت خوارزمية LPCC هي الأفضل بنسبة تعرف /99.95% [4].

تفيد نتائج هذه التجربة بأن عملية دمج السمات لم تعطي نتائج جيدة مقارنة بالخوارزميات المنفردة بل كانت سيئة في شروط الضجيج المتوسطة والمرتفعة.

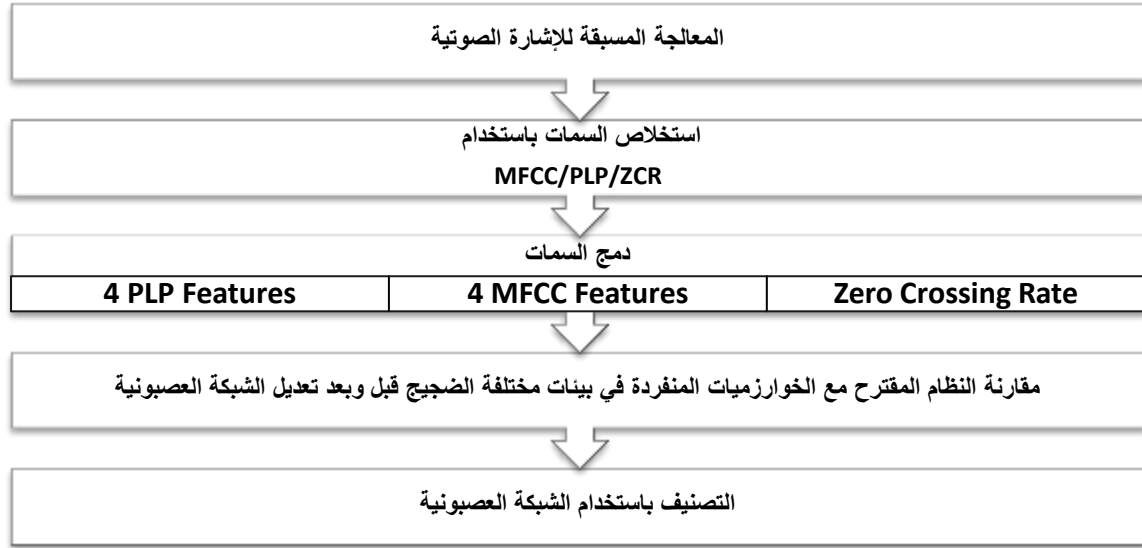
في عام 2016 قام كل من Poonam Sharma و Angali Garg بدراسة تبين معدل التعرف على كلمات اللغة الهندية وتأثير الجنس ونوع الشبكة العصبونية وخوارزمية استخلاص السمات على نسبة التعرف [3].

تبين في الدراسة أن نسبة التعرف لكلام الإناث تفوق مقابلتها للذكور بنسبة 2% كما أن عملية دمج سمات PLP مع MFCC أدت إلى تحسين نسبة التعرف بمقدار 19% مقارنة مع PLP و 1% مقارنة مع MFCC.

في عام 2017 قامت كل من يسرا فيصل الرحيم ولجين عبد القادر بدراسة لتمييز الكلام المعتمد على المتحدث استخدمت خوارزميات MFCC و PLP و RASRA-PLP حيث تم تسجيل أصوات 4 ذكور و 5 إناث اعمارهم بين (30-16) وتضمنت عينات التدريب 864 عينة تدريب واستخدم النموذج الرياضي للتكميم الشعاعي كمصنف مع تغيير بارامترات حيث كانت خوارزمية MFCC هي الأفضل بمعدل تعرف /98.5% [5].

أثبتت الدراسات السابقة أن عملية دمج السمات بطريقة أو بأخرى قد تؤدي إلى نتائج أفضل في عملية التعرف ولا يمكن الجزم بوجود خوارزمية أفضل من غيرها لأن دقة التعرف ترتبط بطبيعة اللغة وجنس المتحدث وعمره والكثير من العوامل الأخرى، كما أن المصنف يلعب دوراً كبيراً في معدل التعرف وزمنه، وحتى بارامترات المصنف تؤثر بشكل كبير على عملية التعرف لذلك لا يمكن اعتماد نتائج بحث ما كنتائج شاملة وذلك بسبب التنوع الكبير جداً في ظروف عملية التعرف ومحدداتها.

بناءً على الدراسات السابقة تم اختيار الشبكة العصبونية كمصنف، وتم اختيار عملية دمج السمات المستخلصة باستخدام PLP و MFCC كونها قدمت الأداء الأفضل بشكل عام. والشكل (1) يبين المخطط الصندوقي العام للنظام المقترح وعملية الدمج فيه.



الشكل (1): المخطط العام للنظام المقترح

أهمية البحث وأهدافه :

لا تزال التطبيقات المستخدمة للتفاعل صوتياً مع المستخدمين تعاني من انخفاض في دقة التعرف على الكلام، لذلك برزت الحاجة إلى تطبيقات ذات نسبة تعرف عالية وزمن تعرف صغير بنفس الوقت لتتناسب مع احتياجات المستخدمين المختلفين.

يهدف هذا البحث لتطوير تطبيق يدمج بين التقنيات المختلفة لاستخلاص السمات من الكلام قبل تمريرها على المصنف، وذلك من أجل دراسة تأثير عملية الدمج على دقة التعرف والحصول على أعلى دقة ممكنة وبأقل زمن ممكن.

منهجية البحث:

تم إجراء هذه الدراسة بالاعتماد على منهجين:

المنهج الوصفي حيث تم دراسة تأثير نوع الشبكة العصبونية المستخدمة في تمييز الكلام على نسبة التعرف وزمن تدريب المصنف (الشبكة العصبونية)، إضافة إلى دراسة تأثير دمج السمات باستخدام خوارزميات مختلفة على نسبة

التعرف، والمنهج التجريبي من خلال دمج خوارزميتي PLP و MFCC ومعدل تخطي الصفر على مستوى السمات وقياس نسبة التعرف وتأثير الضجيج على هذه النسبة، ثم قياس نسبة التعرف عند تجريب خوارزمية الدمج المقترحة على مستخدمين مختلفين.

ملاحظة : جميع التجارب جرت في غرفة منزلية باستخدام مايكروفون منخفض الجودة .

النتائج والمناقشة:

1- الشبكة العصبونية الاصطناعية:

تعتبر الشبكات العصبونية ذات التغذية الأمامية واحدة من أهم الطرق الحديثة التي لها كفاءة عالية في إعطاء نتائج مرضية وجيدة في التعرف، وجزء هام من بناء الشبكة العصبونية هو استخدام خوارزمية دقيقة وقوية في التعلم، والأكثر انتشاراً هي الشبكة العصبونية ذات التغذية الأمامية والانتشار الخلفي للخطأ، وفيها يكون اتجاه الإشارات الداخلة في الشبكة دوماً إلى الأمام، وبذلك تكون الإشارة الخارجة من أي عصبون تعتمد على الإشارات الداخلة فقط، وتحتاج الشبكات العصبونية ذات التغذية الأمامية إلى وجود زوجين من المتجهات هما متجه الإدخال ومتجه الإخراج المطلوب، تبدأ عملية التدريب بمتجه الإدخال حيث يطبق على الشبكة فينتج الإخراج الحقيقي، ويقارن مع ما يقابله من متجه الإخراج المتوقع والفرق بينهما يمثل الخطأ الذي يستخدم لتعديل الأوزان طبقاً لخوارزمية التعليم، ويستمر التدريب إلى أن يصل الخطأ إلى أقل ما يمكن [6].

2- استخراج السمات (features extraction)

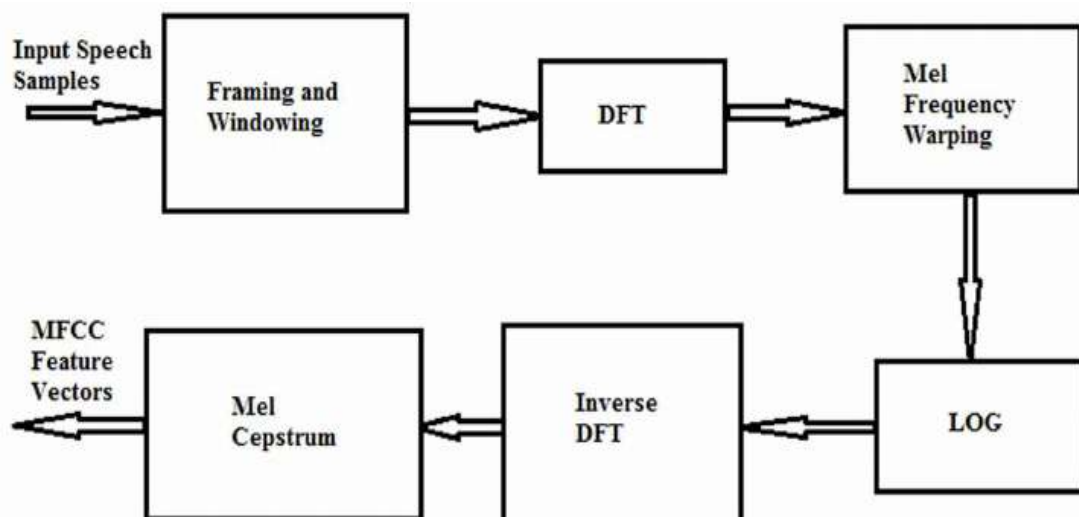
يطلق على العملية التي يتم فيها تحويل البيانات إلى مميزات اسم عملية استخراج السمات، وقد تنوعت الطرق المستخدمة في استخراج السمات، اعتمدنا في هذا البحث خوارزمية معاملات تردد الميل (MFCC) وخوارزمية (PLP) في مرحلة استخراج السمات، حيث يكون دخل هذه المرحلة هو الإشارة الصوتية وخرجها شعاع من السمات التي تمثلها ترددات توافقيات الرنين للمجرى الصوتي الممتد من الحبال الصوتية إلى الشفاه.

1-2- خوارزمية (MFCC) :

تعتبر خوارزمية (MFCC) من الطرق السائدة المستخدمة في استخراج السمات وذلك بسبب حساسية مرشحاتها لخواص إشارة الصوت البشرية، وتستخدم معاملات (MFCC) بشكل كبير في مجال التعرف على الكلام، حيث قُدمت معاملات تردد الميل في عام 1980 ومازالت رائدة في هذا المجال منذ ذلك الوقت.

تُرشح الأصوات المولدة من قبل الإنسان حسب شكل المسلك الصوتي (Vocal Tract)، وبالتالي إذا أمكن تحديد شكل المسلك الصوتي بدقة فإنه يمكن تحديد الصوت (Phoneme) الذي يتم إنتاجه، يتجلى شكل المسلك الصوتي في غلاف طيف طاقة الزمن القصير (Short Time Power Spectrum) والغاية من الخوارزمية (MFCC) هو تمثيل هذا الغلاف بشكل دقيق.

تعتمد الخوارزمية (MFCC) على التغيرات في عرض الحزمة الترددية للأذن البشرية، وتستخدم من أجل التقاط الصفات الرئيسية للكلام، حيث تمتلك (MFCC) تباعداً خطياً على الترددات الأقل من 1000 هرتز وتباعداً لوغاريتمياً على الترددات الأعلى من 1000 هرتز، ويبين الشكل (2) خطوات عمل الخوارزمية (MFCC).



الشكل(2): خطوات عمل الخوارزمية (MFCC)[4].

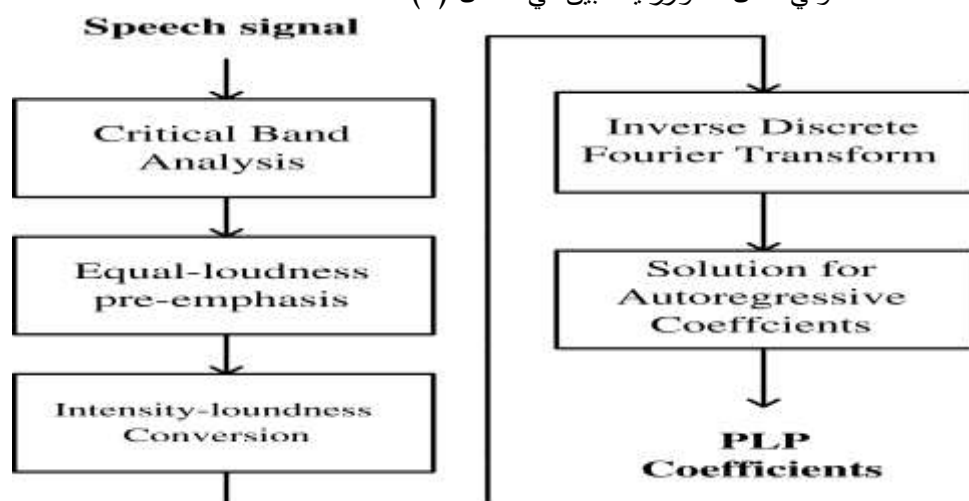
2-2- خوارزمية PLP

تعتمد هذه التقنية على (psychophysics of hearing) العلاقة بين مؤثر فيزيائي و الإدراكات المؤثرة حيث تقوم باستبعاد المعلومات التي ليس لها صلة بالكلام بالتالي تحسين عملية التعرف.

تعتبر خوارزمية PLP مماثلة لخوارزمية LPCC إلا أن خصائصها الطيفية تم تحويلها لنتناسب مع خصائص النظام السمعي عند الانسان ،فهي تقارب ثلاث جوانب رئيسية:

- منحنى تمييز الحزمة الحدية (The critical-band resolution curve).
- منحنى تساوي الجهارة (الشدة الصوتية) (The equal loudness curve).
- علاقة قانون الطاقة بشدة الجهارة (The intensity-loudness power-law relation).

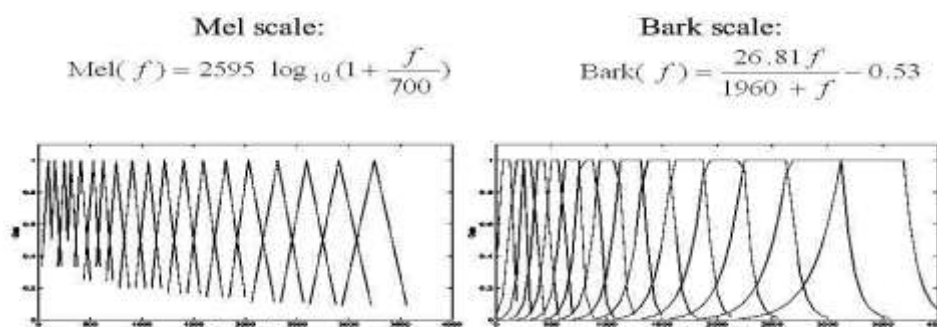
المخطط الصندوقي لعمل الخوارزمية مبين في الشكل (3):



الشكل(3) المخطط الصندوقي لعمل الخوارزمية PLP [4]

وفي الشكل (4) نجد مقارنة بين مجال Mel المستخدم في خوارزمية MFCC ومجال Bark المستخدم في خوارزمية PLP.

Bark and Mel Scales



الشكل (4): مقارنة بين نطاق بارك ونطاق ميل [10]

3- القسم العملي:

في هذا القسم سنقوم ببناء الشبكة العصبونية في البداية، ثم استخدامها في عملية التعرف ومقارنة النتائج قبل وبعد التغيير، ودراسة أثر الضجيج على نسبة التعرف.

3-1- بناء الشبكة العصبونية المقترحة:

استخدمت أغلب الأبحاث السابقة في هذا المجال شبكة التعقب الخلفي وتابع التدريب هو (trainscg) ونسبة الخطأ هي 0.00001 والكود الخاص بها في بيئة ماتلاب هو الكود التالي :

```
net=newff(minmax(p),[1000,200],{'tansig','logsig'},'trainscg');
net.trainParam.show=50;
net.trainParam.lr=0.035;
net.trainParam.epochs=1000;
net.trainParam.goal=0.00001;
```

استغرق تدريب هذه الشبكة 4 دقائق و 48 ثانية وهو زمن طويل نسبياً.

لذلك تم استخدام الشبكة التالية بناءً على دراسة تجريبية مفصلة حيث تم الوصول إلى دقة مضاعفة وبزمن /5 ثواني فقط من خلال تغيير تابع التدريب إلى (trainlm) ونسبة الخطأ إلى 0.00000001، عدد عصبونات الخرج تم اختياره /5 عصبونات كي نستخدم عملية التقريب في التعرف على الخرج، والكود التالي يوضح الشبكة في بيئة ماتلاب :

```
net=newff(minmax(p),[20,15,5],{'tansig','tansig','purelin'},'trainlm');
%net.trainParam.mem_reduc=2;
net.trainParam.show=50;
net.trainParam.lr=0.07;
net.trainParam.epochs=50;
net.trainParam.goal=0.00000001;
```

3-2 سيناريوهات العمل:

اعتمدنا في ها البحث 5 كلمات في عملية التعرف (Go,Hello,No,Stop,Yes) وتمت التجارب في غرفة منزلية باستخدام مايكروفون متوسط الجودة بتردد تسجيل 11025hz ومدة التسجيل 3 ثواني.

تتألف قاعدة البيانات الصوتية من /150/ عينة صوتية فقط وتم اعتماد /12/ سمة في النظام الاول باستخدام MFCC و /12/ سمة في النظام الثاني باستخدام خوارزمية PLP أما في النظام المعتمد على عملية الدمج المقترحة تم دمج /4/ سمات باستخدام كل من الخوارزميتين السابقتين وسمة أخرى هي معدل تخطي الصفر ليصبح عدد السمات /9/ أي أقل بثلاث سمات من الخوارزميات التقليدية، وقد تم اختيار /9/ سمات بدلاً من /12/ لتقليل مصادر الحساب بالتالي تقليل الذاكرة المستخدمة في المعالجة وتقليل زمن التعرف والتدريب.

في السيناريو الأول تمت عملية التعرف بدون إضافة ضجيج أو إجراء أي تعديل على الأصوات المسجلة أما في السيناريو الثاني تمت عملية التعرف بعد إضافة ضجيج للإشارة لتصبح نسبة الإشارة للضجيج (snr=30) وفي السيناريو الثالث تمت إضافة ضجيج للإشارة لتصبح نسبة الإشارة للضجيج (snr=25) والجدول التالي يبين نتائج السيناريوهات السابقة.

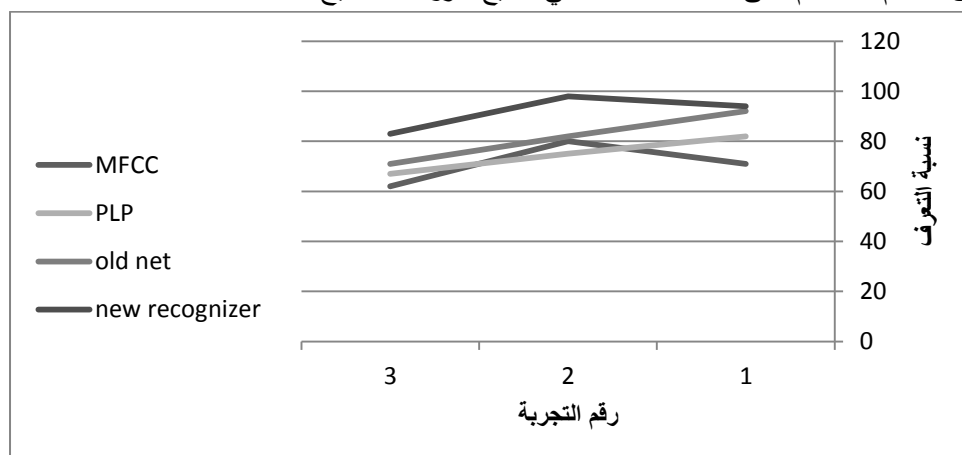
الجدول (1): نسب التعرف في جميع السيناريوهات

MFCC+PLP+ZCR+ new network	MFCC+PLP+ZCR+ old network	PLP	MFCC	الخوارزمية المستخدمة		نسبة التعرف
20%	20%	13%	4%	Go	SNR=unknown	
16%	18%	8%	16%	Hello		
19%	20%	20%	10%	No		
19%	16%	20%	17%	Stop		
20%	19%	20%	20%	Yes		
94%	93%	82%	67%	Total		
20%	19%	10%	10%	Go	SNR=30	
19%	19%	11%	17%	Hello		
20%	17%	16%	15%	No		
20%	7%	18%	18%	Stop		
19%	20%	20%	20%	Yes		
98%	82%	75%	80%	Total		
14%	19%	7%	6%	Go	SNR=25	
15%	6%	7%	12%	Hello		
15%	15%	14%	7%	No		
20%	11%	19%	17%	Stop		
20%	20%	20%	20%	Yes		
84%	71%	67%	62%	Total		

نلاحظ أن نسبة التعرف الأعلى كانت في نسبة إشارة إلى ضجيج (SNR=30) في النظام الذي اعتمد على خوارزمية الدمج المقترحة وهذه النسبة كانت /98%.

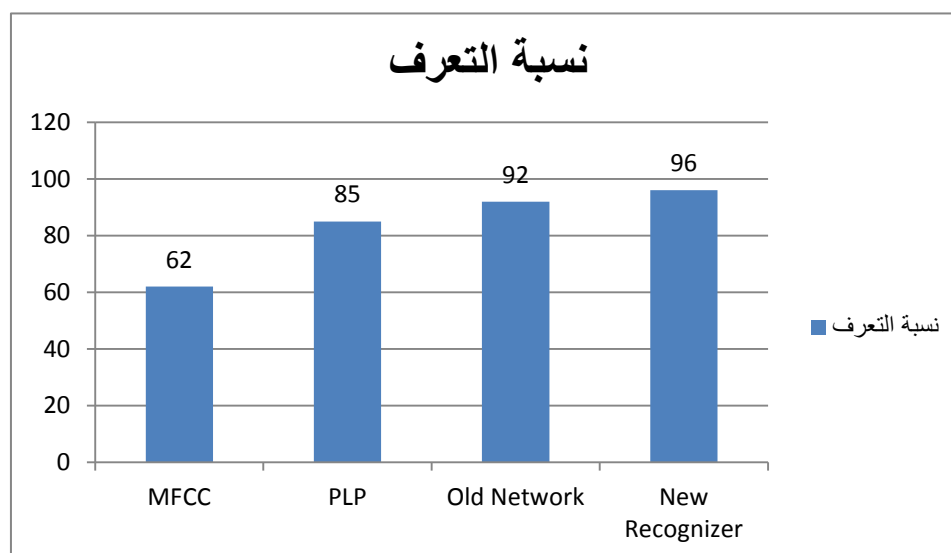
نلاحظ من الجدول السابق التحسن الملحوظ في نسبة التعرف في النظامين اللذين اعتمدا على دمج السمات مقارنة بالأنظمة المفردة على الرغم من أن عدد السمات المستخدمة هو /9/ سمات في النظامين الثالث والرابع بينما تم استخدام /12/ سمة باستخدام MFCC في النظام الأول و/12/ سمة باستخدام PLP في النظام الثاني. كما نلاحظ أن نسبة الخطأ تزداد عند التعرف على الكلمتين GO و NO بسبب التشابه بينهما فكلاهما يتكون من حرف ساكن يليه حرف صوتي مماثل وهما بنفس الطول.

كما أن عملية الدمج أعطت نسبة تعرف أفضل في حالات الضجيج المختلفة لذلك فإن هذا النظام أكثر استقراراً يوضح المخطط المبين في الشكل (5) مقارنة للخوارزميات السابقة مع اختلاف الضجيج المؤثر على الإشارة ليبيّن بوضوح تفوق النظام المصمم على الأنظمة السابقة في جميع شروط الضجيج



الشكل (5): مقارنة الخوارزميات الأربعة المدروسة

في السيناريو الرابع قمنا بتعديل قاعدة البيانات لتصبح /500/ ملف صوتي بأصوات أشخاص مختلفين لنحصل على نسبة التعرف التالية المبينة في الشكل (6).



الشكل (6): نسبة التعرف بعد زيادة عينات التدريب إلى 500 عينة

تم اختيار سيناريوهات العمل في هذا البحث من قبل الباحث، لكن يتقاطع مع الأبحاث السابقة في بعض النقاط، والجدول التالي يبين مقارنة بين النظام المقترح والأبحاث السابقة.

الجدول(2): مقارنة النظام المقترح مع الدراسات المرجعية.

النتيجة	عدد السمات	المصنف	خوارزمية التعرف	عام الدراسة	اسم الدراسة
93.44%	25+ 12+ 10	HMM	MFCC&PLP	2015	تحسين نتائج التعرف على الصوت باستخدام تكامل أنظمة مختلفة
تفوق ANN /10%/	-	HMM ANN	PLP MFCC	2015	New trends in machine learning for speech recognition
99.95%	39		PLP+ $\Delta+\Delta\Delta$	2015	Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC,LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions
التحسن 1% 19%	-	ANN	MFCC PLP MFCC+PLP	2016	Feature extraction and recognition of hindi spoken words using neural networks
98.5%	-	VQ	MFCC	2017	Speaker dependent speech recognition in computer game cotrol
98%	9	ANN	MFCC+PLP+ZCR	2018	النظام المقترح : تحسين أداء الخوارزميات المستخدمة في تمييز الكلام

الاستنتاجات والتوصيات:

- حجم قاعدة البيانات (عينات التدريب) الصوتية ونسبة التشابه (الإرتباك) بين الكلمات يلعب دوراً هاماً في نسبة التعرف فكلما زادت عينات التدريب تزيد نسبة التعرف وكلما زاد الإرتباك تقل هذه النسبة.
- هنالك تضارب بين نسبة التحسين في معدل التعرف وعدد السمات مع الزمن حيث أن زيادة عدد السمات يحسن نسبة التعرف ولكن على حساب زيادة زمن التعرف.
- من الأفضل تصميم الشبكة العصبونية المناسبة لكل مجموعة بيانات حيث أن أداء الشبكة قد يكون سيئاً إذا تم تغيير قاعدة البيانات الصوتية.
- إن خوارزمية التعلم للشبكة العصبونية تؤثر بشكل كبير على زمن التدريب.
- نسبة الخطأ للشبكة العصبونية تؤثر بشكل كبير على معدل التعرف.

- عملية دمج السمات في ظروف الضجيج المنخفض تزيد من نسبة التعرف بشكل كبير من 12 حتى 23 بالمائة وحتى في شروط الضجيج المتوسط تعطي نتائج أفضل من خوارزميات PLP و MFCC من 16 إلى 21 بالمائة لنسنتنج أن عملية الدمج تعطي استقرار أكبر في عملية التعرف ضمن الشروط المختلفة.
- التوصيات :**
- عدم زيادة عدد السمات عن 12 سمة مما قد يؤثر سلباً على نسبة التعرف
 - توحيد البارامترات المستخدمة في الأنظمة المختلفة كالتردد و نسبة الاشارة إلى الصوت وبيئة التسجيل لتكون عملية المقارنة دقيقة وذات جدوى.
 - العمل على إزالة الضجيج بطرق أكثر فاعلية.

المراجع :

- 1- م.راما غسان حسن، تحسين نتائج التعرف على الصوت بالاعتماد على نتائج تكامل أنظمة مختلفة، جامعة تشرين، 2015، 85.
- [2] INGE GAVAT, DIANA MILITARU, *New trends in machine learning in speech recognition* , SISOM 2015 Bucharest 21-22 May, pp 276.
- [3] POONAM SHARMA, ANGALI GARG, *Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks*, International Journal of Computer Applications (0975 – 8887) Volume 142 – No.7, May 2016, pp 17.
- [4] VETON Z. KĚPUSKA, HUSSIEN A. ELHARATI, *Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC,LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions*, Journal of Computer and Communications, 2015, 3, 1-9, 9.
- [5] YUSRA FAISAL AL-IRAHYIM, LUJAIN YOUNIS ABDULKADER, *Speaker Dependent Speech Recognition in Computer Game Control*, International Journal of Computer Applications (0975–8887)Volume 158–No 4,January 2017, 37.
- [6] DIAMANTARAS K. AND KUNG S, *Principle Component Neural Networks Theory and Applications*, New York, John Wiley & Sons Inc, 2006, 255.
- [7] LAVNEET SINGH, GIRIJA CHETTY, *A Comparative Study of Recognition of Speech Using Improved MFCC Algorithms and Rasta Filters*, Information Systems, Technology and Management Communications in Computer and Information Science Volume 285, 2012, pp 304-314.
- [8] BHAVNA SHARMA, K. VENUGOPALAN, *Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images*. IOSR Journal of Computer Engineering, Volume 16, Issue 1. Jan 2014, pp 35.
- [9] PITZ, M, SCHLUTER R, NEY H, MOLAU S, *Computing Mel-frequency cepstral coefficients on the power spectrum*, Print ISBN: 0-7803-7041-4 INSPEC Accession Number: 7120280 Acoustics, Speech, and Signal Processing, 2001, Proceedings. (ICASSP '01). 2001 IEEE International Conference on (Volume: 1) Page 73 - 76 vol.1, pp12.

- [10] H. HERMANSKY, *Perceptual linear predictive (PLP) analysis of speech*, Speech Technology Laboratory, Division of Panasonic Technologies, Inc. 3888 State Street, Santa Barbara, California 93105. 1989, pp 1752.
- [11] H. DEMUTH, M. BEALE, *Neural Networks Toolbox User's Guide*. The MathWorks, Inc. 1992-2002, pp 826.
- [12] P. RANI, S. KAKKAR, S. RANI, *Speech recognition using neural networks*. International conference on advancement in engineering and technology. 2015, pp 14.
- [13] N. DAVE, *Features Extraction Methods LPC, PLP and MFCC in Speech Recognition*, International Journal for Advanced Research in Engineering and Technology. Vol.1, Issue VI, July 2013, pp5.
- [14] BHUSHAN C. KAMLE, *Speech recognition using artificial neural networks*, Int'l journal of Computing, Communication & Instrumentation Engg, (IJCCIE), Vol 3, Issue 1, 2016, pp 4.
- [15] A. MANSOUR, G. SALH, H. Z. ALABDEN, *Speech recognition using back propagation algorithm in neural network*, International Journal of Computer Trends and Technology(IJCTT), Vol 23,Number 3, May 2015, pp21.