مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (38) العدد (2) 2016
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (38) No. (2) 2016

دراسة تحليلية لخوارزميتي (MFCC و Endpoint) ومدى تأثيرهما في نسب التعرف على الصوت

دعد يوسف الكعدي*

(تاريخ الإيداع 18 / 11 / 2015. قُبِل للنشر في 31/ 3 / 2016)

□ ملخّص □

يشتمل التعرف على الصوت قسمين أساسيين وهما التعرف على الكلام والتعرف على المتكلم، حيث تعد عمليات التعرف هذه من أهم التقنيات الحديثة وقد تم تطوير العديد من الأنظمة التي تختلف بالطرق المستخدمة في استخراج السمات وطرق التصنيف لتدعم أنظمة تعرف من هذا النوع.

اشتملت الدراسة في هذا البحث على القسمين السابقين، حيث تم تصميم نظام تعرف على المتكلم وأوامره الصوتية واستخدام عدة خوارزميات متكاملة لإنجاز البحث. قمنا بإجراء دراسة تحليلية لخوارزمية Mel Frequency Cepstral Coefficients (MFCC) المستخدمة في استخراج السمات، وتمت دراسة بارامترين خاصين بهذه الخوارزمية هما عدد المرشحات في بنك المرشحات وعدد السمات المأخوذة من كل إطار وعلاقة هذين البارامترين ببعضهما ومدى تأثير قيمتهما على نسب التعرف. وتم استخدام الشبكات العصبية ذات التغذية الأمامية والانتشار الخلفي للخطأ Feed Forwarding back propagation Neural Networks (FFBPNN) كمصنف وحللنا أداء الشبكة للوصول إلى أفضل خصائص ومكونات محققة عملية التعرف. كما تمت دراسة خوارزمية Endpoint المستخدمة لإزالة فترات الصمت وتأثيرها في نسب التعرف على الصوت.

الكلمات المفتاحية: المتكلم، الكلام، السمات، الشبكات العصبية.

*قائم بالأعمال/ معاون- قسم هندسة الحاسبات والتحكم الآلي- كلية الهندسة الميكانيكية والكهربائية- جامعة تشرين- اللاذقية- سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (38) العدد (2) 2016
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (38) No. (2) 2016

Analysis study about (MFCC and Endpoint) algorithms and the extent of their impact in voice recognition rates

Daed Youssef Alkody *

(Received 18 / 11 / 2015. Accepted 31 / 3 / 2016)

□ ABSTRACT □

Voice recognition includes two basic parts: speech and speaker recognition. These recognition processes consider as the most important processes of modern technologies, many systems has been developed that differ in the methods used to extract features and classification ways to support recognition systems of this type.

The study was conducted in this research on the previous subject, where the system is designed to recognize the speaker and his voice orders and focus on several complementary algorithms to carry out the research. we conducted an analytical study on MFCC algorithm used in the extraction of features, and it has been studying two parameters the number of filters in the filters bank and the number of features that taken from each frame and the impact of these two parameters in the recognition rate and the relationship of these two parameters on each other. It was the use of feed forwarding back propagation neural networks performance analysis as characteristics and we analyze the performance of the network to gain access to the best features and components to the process of achieving recognition. And it has been studying Endpoint algorithm that used to remove periods of silence and its impact on voice recognition rates.

Key Words: speaker, speech, feature, neural network.

* **Academic Assistant, Department of computer & control Engineering, Faculty of Mechanical & Electrical Engineering, Tishreen University, Lattakia, Syria**

مقدمة:

الكلام هو عبارة عن موجة صوتية تحمل معلومات من المتكلم إلى المستمع، لذا فإن معظم التطبيقات المطبقة على الصوت هي عبارة عن تطبيقات إشارة رقمية، حيث يتم تحويل الموجة الصوتية (إشارة تشابهية) إلى إشارة رقمية يتم معالجتها فيما بعد، وتكمن عادة معظم التطبيقات على الصوت في المجالات التالية:[3][2][1]

1 - التعرف على الكلمة.

2 - التعرف على المتكلم.

تحاول خوارزميات التعرف على الكلمة تمييز الكلمات بشكل منفرد، أما خوارزميات التعرف على المتكلم فتعمل على تحديد هوية الشخص المتكلم حيث أن صوت الشخص يستخدم بعد معالجته كبصمة صوتية مميزة لكل شخص.

يقسم التعرف على الكلام إلى المجالات الثلاثة التالية:

1 - كلمة واحدة في كل مرة isolated-word speech.

2 - مجموعة من الكلمات بفواصل زمنية connected word recognition.

3 - كلام مستمر continuous speech.

تم استخدام المجال الأول في هذا البحث.

أهمية البحث وأهدافه:

• دراسة خوارزمية MFCC ومدى تأثير تغير قيم بارامترين خاصين بهذه الخوارزمية (عدد المرشحات في بنك المرشحات وعدد السمات المأخوذة من كل إطار)على نسب التعرف.

• تحديد العلاقة بين البارامترين السابقين.

• معرفة فعالية خوارزمية Mfcc على كل من مرحلتي التعرف على المتكلم والتعرف على الكلام.

• تحليل أداء شبكات FFBPNN باعتماد تقنيتين مختلفتين في التدريب وهما طريقة انحدار الميل gradient descent algorithmgd) وطريقة الميل الموحد المتدرج Scaled Conjugate Gradient (scg).

• تحديد أهمية خوارزمية Endpoint في تحسين نسبة التعرف.

طرائق البحث ومواده:

1-خوارزمية إزالة فترات الصمت Endpoint Algorithm:[6][5]

من المشاكل التي تواجه عملية معالجة إشارة الكلام هي تحديد بداية النطق الحقيقي ونهايته وذلك لتقليل حجم الملفات الصوتية التي تؤدي إلى تقليل المعالجة المطلوبة لإشارة الكلام والحصول على كفاءة أكبر في التمييز. تتلخص خوارزمية endpoint بالمرحلتين الآتيتين:

1. حساب معدل السعة لإشارة الكلمة.
2. حساب العتبة المناسبة لتحديد إشارة الكلام.

1-1-حساب معدل السعة لإشارة الكلام:

إن السعة لإشارة الكلام تتغير مع الزمن ،وتمثيل إشارة الكلمة بالاعتماد على معدل السعة لوقت قصير يعطي تمثيلاً ملائماً لإشارة الكلمة لأنه يعكس هذه التغييرات بشكل واضح ويتم إنجاز هذه المرحلة من خلال الخطوات الآتية:

1 - تقسيم مصفوفة الصوت إلى أطر Framming:حيث تم اعتماد حجم ثابت للإطار (Frame size=220بمقدار تداخل (overlapping=110) وتكمن أهمية حدوث التداخل في تسهيل الانتقال بين الأطر، نرمز للأطر الناتجة بالرمز S(n).

2 - معالجة الأطر بنافذة هامينغ Hamming window:

تخضع نافذة هامينغ للعلاقة الآتية:

$$W_H(n) = 0.54 - 0.46\cos\left(\frac{2n\pi}{N-1}\right) \quad \begin{matrix}\mathbf{L=N+1}\\ \mathbf{0<=n<=N}\end{matrix} \qquad (1)$$

حيث أن:

$W_H(n)$: مطال العينة الجديد.

n: ترتيب العينة في النافذة.

N: الطول الكلي للنافذة ويساوي هنا 220. أي أن النافذة بحجم الإطار.

بهدف تنعيم الإشارة وتقليل الانقطاعات بين الأطر، تتم عملية معالجة الأطر بنافذة هامينغ S`(n) وفقاً للعلاقة الآتية:

$$S`(n) = (2) \quad W(n))$$

3 - حساب معدل السعة لكل إطار (احتمالية الكلام):

وتشتمل الخطوات التالية:

- نقل قيم الإطار من المجال الزمني إلى المجال الترددي(xi) بهدف الحصول على مجموعة الترددات الأساسية.
- حساب اللوغاريتم النيبري لقيم الإطار(xi)logمن ثم حساب متوسط اللوغاريتمات لقيم الإطار lgg.
- يتم مقارنة القيمة lgg لكل إطار مع قيمة الصفر ليتم حساب احتمالية الكلام prsp وفق علاقتين مختلفتين[6]:

Lgg<0	Lgg>0
gg=exp(lgg); prsp(i)=gg/(1+gg);	prsp(i)=1/(1+exp(-lgg));

(3)

1-2-حساب العتبة المناسبة لتحديد إشارة الكلام:

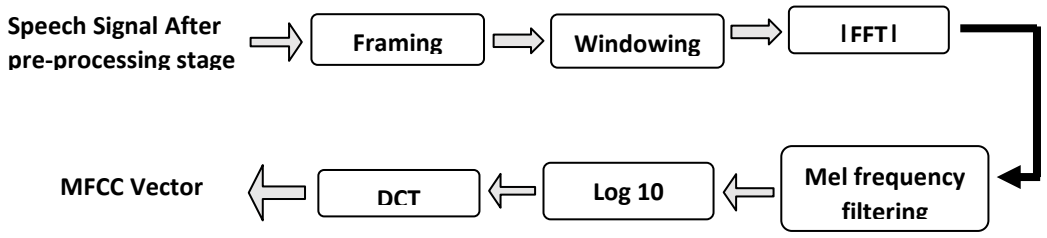
يتم تحديد قيمة العتبة threshold بالاعتماد على أكبر قيمة احتمالية prspmax وأقل قيمة احتمالية prspmin للأطر وفق جملة المعادلات التالية [6]:

$$I1 = 0.03 * (prspmax - prspmin) + prspmin$$
$$I2 = 4 * prspmin$$
$$threshold = MIN(I1, I2) + 0.4 \qquad (4)$$

ومن ثم يتم إزالة إطارات أو المحافظة عليها بمقارنة قيمة معدل السعة لكل إطار مع قيمة العتبة، الأطر التي لديها قيمة معدل سعة أصغر من قيمة العتبة سيتم حذفها.

1-3-خوارزمية MFCC:[8][7][4]

تعد من الطرق السائدة المستخدمة في استخراج السمات وذلك بسبب حساسية مرشحاتها لخواص إشارة الصوت البشرية. تعتمد خوارزمية MFCC على التغيرات المعروفة في عرض حزمة الترددات للأذن البشرية، حيث أن لمرشحاتها تباعدا خطيا على الترددات المنخفضة (الأقل من 1000 هرتز) ولوغاريتميا على الترددات المرتفعة(أكبر من 1000 هرتز) الأمر الذي يجعلها تستخدم بشكل كبير في التعرف على الصوت والتقاط الصفات الرئيسية للكلام .



الشكل (1) المخطط الصندوقي لخوارزمية MFCC[4]

يبين الشكل (1) المخطط الصندوقي لخوارزمية MFCCويتضمن:

- تقسيم مصفوفة الصوت إلى أطر ومعالجتها بنافذة هامينغ: هما مرحلتين تم توضيح مفهومهما سابقا.
- بعد عملية النوفذة سوف يتم تطبيق تحويل فورييه السريع (FFT) fast fourier transform من أجل كل إطار وذلك من أجل استخراج مركبات التردد للإشارة.
- ترشيح الترددات بمقياس الميلي Mel frequency filtering :تعمل MFCC على ترشيح طيف الإشارة الصوتية عن طريق مجموعة من المرشحات المثلثية المتباعدة بانتظام وفقا لمقياس mel الترددي(اختصار لكلمة melody) الذي يعبر عن علاقة تربط التردد الملاحظ لنغمة صافية m بترددها المقاس الأصليfوالموضحة بالعلاقة (5) [7].

$$m = 2595\, log\left(\frac{f}{700} + 1\right) \qquad (5)$$

يستطيع الإنسان أن يميز التغيرات الصغيرة في طبقة الصوت (pitch) بشكل أفضل عند الترددات الصغيرة من الترددات الكبيرة، بالتالي فإن تضمين هذا المقياس يجعل سماتنا تتطابق بشكل أقرب من سمع الإنسان[14].

- يتم بعد ذلك حساب اللوغاريتم mel scale spectrum ومن ثم يستخدم تحويل جيب التمام المتقطع DCT لإعادة تحويل mel scale spectrum اللوغاريتمي إلى مجال الزمن، ونتيجة هذا التحويل هو الحصول على شعاع السمات.

اعتماداً على مراحل خوارزمية MFCC تم تحديد مجموعة من البارامترات اللازمة لتنفيذ الخوارزمية مبينة بالجدول (1):

جدول (1) البارامترات الخاصة لخوارزمية [8]MFCC

SN	property	Value
1	Frame size	220 sample
2	Frame length	220/11025=20ms
3	overlapping	110 frame (50%)
4	Number of filters	----------
5	Filter type	traingular
6	Number features	----------

تحتاج هذه الخوارزمية 6 بارامترات لتنفيذها، تم تثبيت 4 بارمترات منها والعمل على تغيير البارامترين الباقيين ومناقشة تأثير تغير قيمتهما على نسبة التعرف .

1 -خوارزمية K-means:[12]

وهي خوارزمية عنقدة تستخدم لتجميع البيانات وفق خصائصها إلى k تجمع، ووفقا للبحث سيتم انتخاب 10 أطر من كل لفظ صوتي (كل إطار يتضمن مجموعة من عينات الصوت) ليتم استخدامها لاحقاً كأشعة دخل للشبكة العصبية، تم استخدام هذه الخوارزمية لأن خوارزميةendpoint تسبب إزالة فترات الصمت لكل لفظ صوتي بحجم بتات مختلف عن بقية الألفاظ الصوتية الأمر الذي ينتج أشعة سمات بحجوم مختلفة.

2- الشبكات العصبية ذات التغذية الأمامية والانتشار الخلفي للخطأ Feed Forwarding back propagation Neural Networks (FFBPNN):[9][19]

إن الشبكات العصبية الصنعية قادرة على حل الكثير من مهام التمييز المعقدة وهي تستخدم في تطبيقات تمييز الكلام لأهداف عامة حيث أنها تعالج الجودة المنخفضة ومعطيات الضجيج واستقلالية المتحدث.

تمتلك الشبكة العصبية مجموعة من المكونات تبدأ بتحديد حجم شعاع الدخل والطبقات (بما فيها عدد العصبونات لكل طبقة وتابع التفعيل للعصبونات وعدد الطبقات)وتابع التدريب(معبراً عن التقنية المتبعة لتغيير الأوزان والانحيازات للوصول بالخرج إلى الهدف المرغوب) وبارامترات تابع التدريب بما فيها معامل التعلم وعدد التكرارات وقيمة الخطأ.

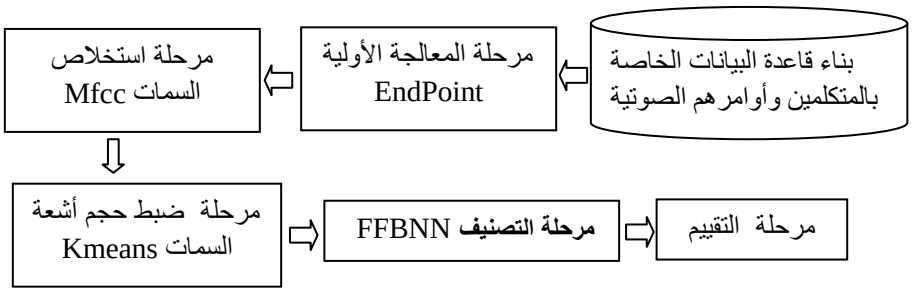
تم تحليل أداء شبكات FFBPNN باستخدام تقنيتين مختلفتين في التدريب وهما طريقة انحدار الميل gd وطريقة الميل الموحد المتدرج scg وسيتم في كل تقنية دراسة تحليلية لاختيار قيم خصائص الشبكة للوصول إلى شبكة عصبية بخصائص مناسبة لعملية التعرف.

تستخدم شبكات FFBPNN تقنية انحدار الميل بشكل افتراضي وهي تقنية تعتمد على تحديث أوزان الشبكة والانحياز في الاتجاه الذي تتناقص فيه قيمة تابع الكلفة ، وللقيام بذلك يتم استخدام تابع التدريب traingd في حين أن تقنية scg تعتمد على الدمج بين تقنيتي الميل الموحد وليفربيرغ ماركورس(Levenberg-Marquardt) بحيث تقلل كمية الحسابات المطلوبة في كل تكرار للحصول على الأوزان والانحيازات النهائية ولتطبيق هذه التقنية يتم استخدام تابع التدريب trainscg [10][11].

النتائج والمناقشة:

تم العمل وفق المراحل الآتية:

- بناء قاعدة البيانات.
- مرحلة بناء النظام : يوضح الشكل (2)المخطط الصندوقي للنظام المطبق.
- مرحلة التقييم.



الشكل (2) يوضح المخطط الصندوقي للنظام المطبق

1-بناء قاعدة البيانات:

تم تسجيل *576* لفظ صوتي تعود لأربع متكلمين تتراوح أعمارهم بين (19-28 عاماً)، لفظ كل متكلم *144* لفظ صوتي يعود كل 16 لفظ فيها لكلمة (أمر صوتي) أي لفظ كل متكلم 9 أوامر صوتية باستخدام برنامج *Matlab2011* مشكلة قاعدة بيانات، تم استخدام *360* لفظ منها كعينات تدريب و *216* كعينات اختبار. يبين الجدول (2) خصائص عينات الصوت المسجلة.

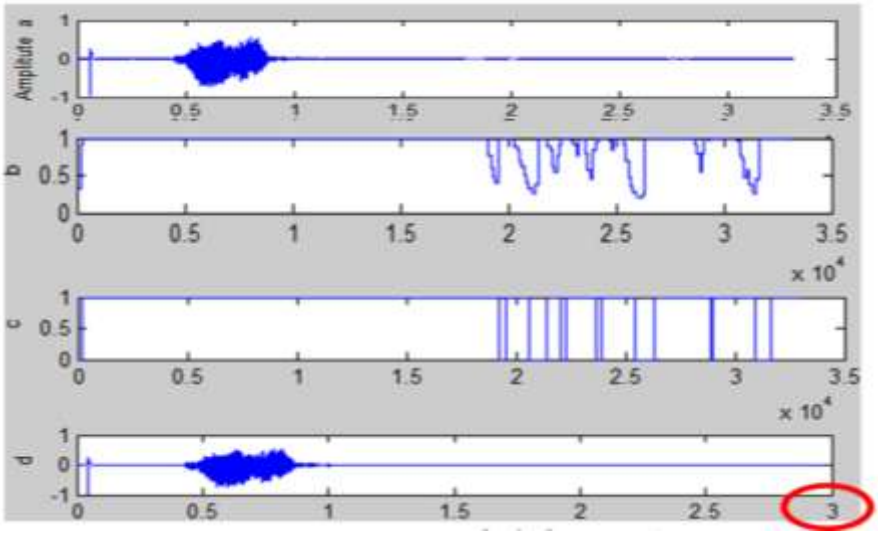
جدول (2) خصائص عينات الصوت المسجلة

SN	property	Value
1	Audio Sample rate	11025hz
2	Audio Sample Size	16bit
3	Audio Sample type	double
4	Audio Sample range	-1<=S<=1
5	Channel	1 Mono
6	Length	3sec

تم تخزين هذه العينات الصوتية بملفات ذات صيغة(wav)علماً أن اختيار16بت بدقة مضاعفة ومجال [-1،+1] تعادل مرحلة التقييس *Normalization*.

2-مرحلة المعالجة الأولية:

تم اعتماد خوارزمية endpoint في هذه المرحلة للحصول على إشارة الصوت المعالجة، يبين الشكل(3) مراحل خوارزمية endpoint مطبقة على اللفظ الصوتي down من قبل أحد المتكلمين، موضحة شكل الإشارة الناتجة لكل مرحلة تفصيلية لهذه الخوارزمية:



الشكل(3) مراحل خوارزمية *Endpoint*

***(a)*إشارة الصوت قبل معالجتها في المجال الزمني (*b*)مصفوفة معدل السعة لكل إطار**

(*c*)مصفوفة معدل السعة بعد مقارنتها بجهد العتبة (*d*) إشارة الصوت بعد إزالة فترات الصمت

حيث نتج أن المتوسط الحسابي لعدد العينات التي تم إزالتها من كل لفظ صوتي يساوي *3000* عينة (الحجم الكلي لإشارة الصوت33065 عينة) كما يظهر الشكل(3).

3-خوارزمية *MFCC*:

اعتماداً على مراحل الخوارزمية وجدنا أنها تمتلك *6* بارامترات أساسية ، في البحث سنقوم بتثبيت 4 بارمترات منها والعمل على تغيير البارامترين الباقيين وهما عدد المرشحات في بنك المرشحات وعدد السمات المأخوذة من كل إطار ودراسة علاقتهما ببعضهما ومدى تأثير تغير قيمتهما على تغيير نسبة التعرف حيث تم استخدام عينات الاختبار وحساب نسبة التعرف على اختلاف عدد المرشحات وعدد السمات وفق العلاقتين الآتيتين:

$$\begin{cases} \text{نسبة التعرف على المتكلم} = \dfrac{\text{عدد العينات الصحيحة (المصنفة أنها تعود لأحد المتكلمين)}}{\text{عدد العينات الكلية (216)}} \\ \text{نسبة التعرف على الكلام} = \dfrac{\text{عدد الأوامر الصحيحة}}{\text{عدد الأوامر الكلية (216)}} \times 100 \end{cases} \quad (6)$$

الجدول (3) يبين تأثير تغير عدد المرشحات وعدد السمات على نسب التعرف للمتكلم

N.Of MFCC For Each frame	Id rate								
	Numbers of filters in filter- bank								
	15	19	24	27	30	35	40	45	50
6	75.92	76.85	77.77	77.77	77.77	78.24	78.24	78.24	78.24
12	76.85	77.31	77.77	78.7	78.7	79.62	79.62	79.62	79.62

42	77.77	80.09	81.94	86.57	86.57	87.50	87.50	87.50	87.50
55	77.77	80.09	81.94	86.57	86.57	87.50	87.50	87.50	87.50

يبين الجدول (3) أنه عند زيادة عدد المرشحات في بنك المرشحات ابتداء من 15 وصولاً إلى 35 نلاحظ زيادة في نسب التعرف على اختلاف عدد السمات المأخوذة من كل إطار، وعند زيادتها أكثر من 35 وصولاً إلى 50 نلاحظ ثبات نسب التعرف، وعند زيادة عدد السمات في كل إطار ابتداء من 6 وصولاً إلى 42 نلاحظ زيادة نسب التعرف وعند الزيادة القصوى في عدد السمات إلى 55 نلاحظ ثبات نسب التعرف على أعظم قيمة تعرف وصلنا إليها عند عدد سمات 42.

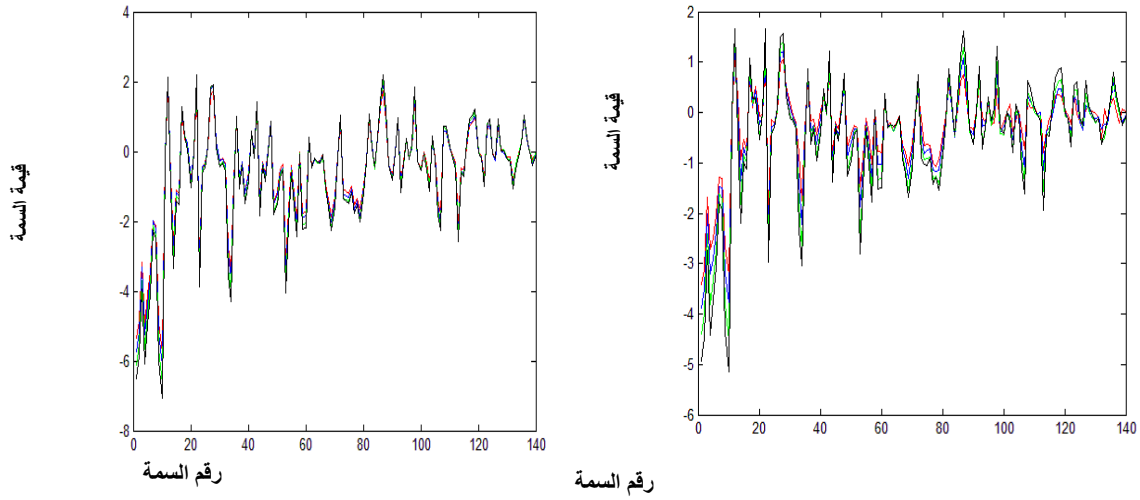
الجدول (4) يبين تأثير تغير عدد المرشحات وعدد السمات على نسب التعرف على الكلام

N.Of MFCC For Each frame	Id rate								
	Number of filters in filter- bank								
	15	19	24	27	30	35	40	45	50
6	78.24	79.16	80.55	80.55	80.55	79.62	79.62	79.62	79.62
12	82.40	83.33	85.18	84.72	84.72	83.79	83.79	83.79	83.79
42	83.33	84.25	90.74	88.88	88.88	87.50	87.50	87.50	87.50
55	83.33	84.25	90.74	88.88	88.88	87.50	87.50	87.50	87.30

يبين الجدول (4) بأنه عند زيادة عدد المرشحات في بنك المرشحات ابتداء من 15 وصولاً إلى 24 نلاحظ زيادة في نسب التعرف على اختلاف عدد السمات المأخوذة من كل إطار، وعند زيادتها أكثر من 24 وصولاً إلى 35 نلاحظ ثبات نسب التعرف عند عدد سمات يساوي 6 وانخفاضها قليلاً عندما عدد السمات (12,42)، وعند زيادتها أكثر من 35 وصولاً إلى 50 نلاحظ ثبات نسب التعرف، وعند زيادة عدد السمات في كل إطار ابتداء من 6 وصولاً إلى 42 نلاحظ زيادة نسب التعرف وعند الزيادة القصوى في عدد السمات إلى 55 نلاحظ ثبات نسب التعرف على أعظم قيمة تعرف وصلنا إليها عند عدد سمات 42.

وفق الدراسة حصلنا على أعلى نسبة تعرف قدرها *90.74%* عند التعرف على الأوامر الصوتية و *87.50%* عند التعرف على المتكلم.

لتحديد أهمية اختيار عدد المرشحات في بنك المرشحات ولتوضيح سبب ثبات نسب التعرف في بعض النتائج قمنا بتحليل مصفوفة السمات وتبين أن مصفوفة السمات الناتجة عن بنك مرشحات مكون من (15,19,24,30 مرشح) تظهر تباين قيم السمات الأمر الذي سبب اختلاف نسب التعرف، أما مصفوفة السمات الناتجة عن بنك مرشحات مكون من (35,40,45,50) مرشح تظهر تقارب في قيم السمات كثيرا الأمر الذي أعطى نسب تعرف متقاربة جداً الشكل (4).



a) مصفوفة السمات الناتجة عن بنك مرشحات(15-19-24-30)

b) مصفوفة السمات الناتجة عن بنك مرشحات(35-40-45-50)

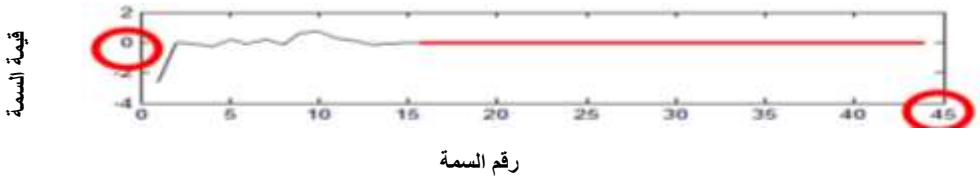
الشكل (4) اختلاف قيم مصفوفة السمات باختلاف بنك المرشحات في خوارزمية *MFCC*

لإيجاد العلاقة بين عدد السمات وعدد المرشحات قمنا بتغيير عدد السمات تدريجيا من أجل كل بنك مرشحات على حدا للوصول إلى علاقة بين عدد السمات وعدد المرشحات فتبين لدينا مايلي:

جدول (5) تغير نسبة التعرف مع تغير عدد السمات وتأثير زيادة عدد السمات > عدد المرشحات

Number of filters 15		Number of filters 19		Number of filters 24		Number of filters 27	
Recognition rate	Number of features	Recognition rate	Number of features	Recognition rate	Number of features	Recognition rate	Number of features
76.19	6	76.55	6	77.77	6	77.77	6
76.90	12	76.65	12	77.77	12	78.57	12
77.77	14	80.15	18	81.74	23	86.50	26
77.77	19	80.15	25	81.74	25	86.50	35
77.77	25	80.15	35	81.74	35	86.50	40
77.77	42	80.15	42	81.74	42	86.50	50

يتبين من الجدول السابق أنه عند اختيار عدد سمات يقارب عدد المرشحات فإننا نحصل على أكبر نسبة تعرف وبعد ذلك مهما زدنا عدد السمات لتصبح أكبر من عدد المرشحات فإنه تثبت نسبة التعرف، وعند تحليل مصفوفة السمات الناتجة عن اختيار 42 سمة من الإطار وعدد مرشحات 15 مرشح نتج الشكل(4):

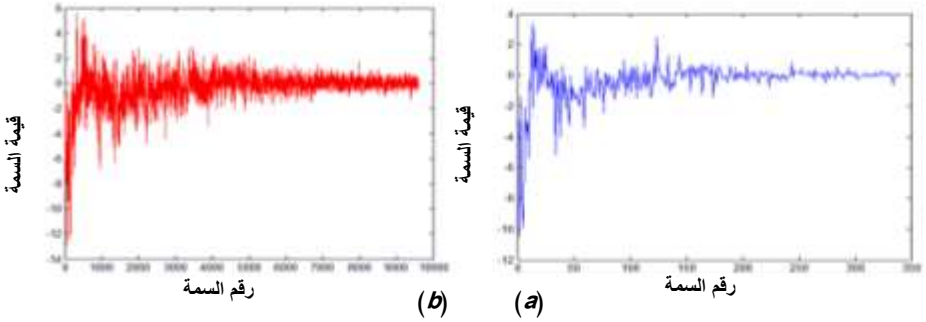


الشكل (5) مصفوفة السمات الناتجة عن اختيار *15* مرشح و*42* سمة

يبين الشكل (5) أن مصفوفة السمات الناتجة هي مصفوفة بقيم مخالفة للصفر بعدد يساوي عدد المرشحات وقيم مساوية للصفر لباقي قيم المصفوفة، الأمر الذي يعني عدم أهمية اختيار عدد سمات أكبر من عدد المرشحات.

4-4 مرحلة توحيد أشعة السمات:

يوضح الشكل (6) مصفوفة السمات الناتجة عن خوارزمية MFCC، وشعاع السمات المستخدم في مرحلة التصنيف بعد انتخاب 10 أطر وفق خوارزمية k-means ليتم استخدامها كدخل للشبكة العصبية.



(*b*) (*a*)

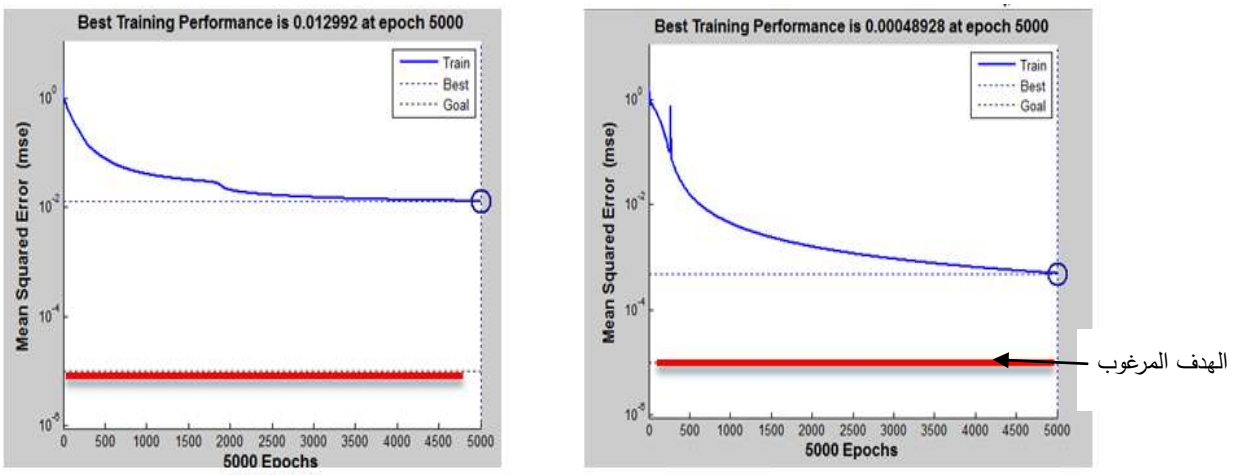
الشكل(6) يوضح (*a*) مصفوفة السمات بعد *K-means* **(*b*)مصفوفة السمات الناتجة عن مرحلة *MFCC***

4-5 مرحلة التصنيف:

4-5-1 تقنية انحدار الميل gd:

1 - تحديد قيمة معامل التعلم a:تمت المقارنة بين قيمتين هما (0.01-0,035)

يبين الشكل(7) أن اختيار معامل التعلم 0.035 سبب اقتراب الخرج من الهدف المرغوب أكثر من معامل التعلم 0.01.

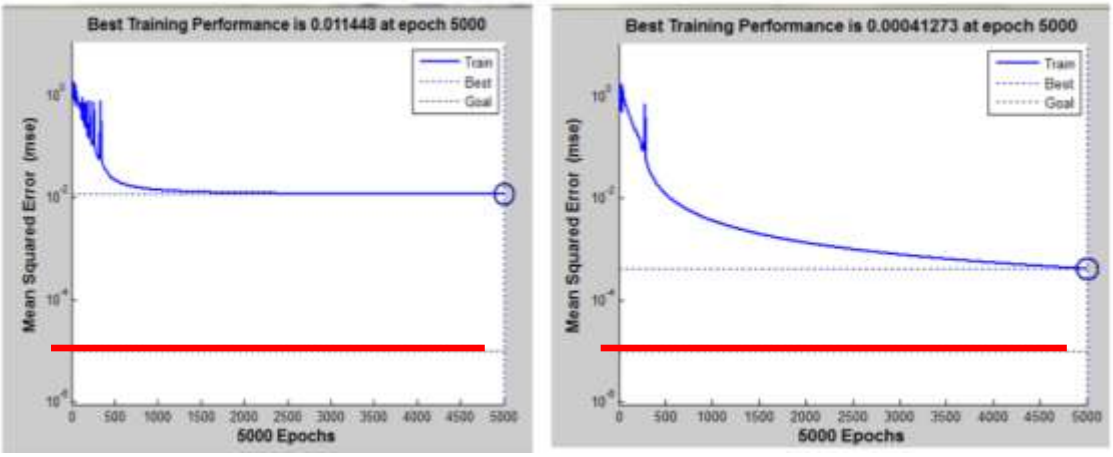


a) معدل التعلم 0.035 **(b) معامل التعلم 0.01**

الشكل (7) تحليل أداء الشبكة في الحالة الاولى 100 عصبون للطبقة الخفية

2 - اختلاف عدد عصبونات الطبقة الخفية:

يبين الشكل (8) أنه مهما زدنا عدد عصبونات الطبقة الخفية لم يؤدي إلى تدريب الشبكة وإنما مازال الخرج الفعلي بعيد عن الهدف المرغوب .

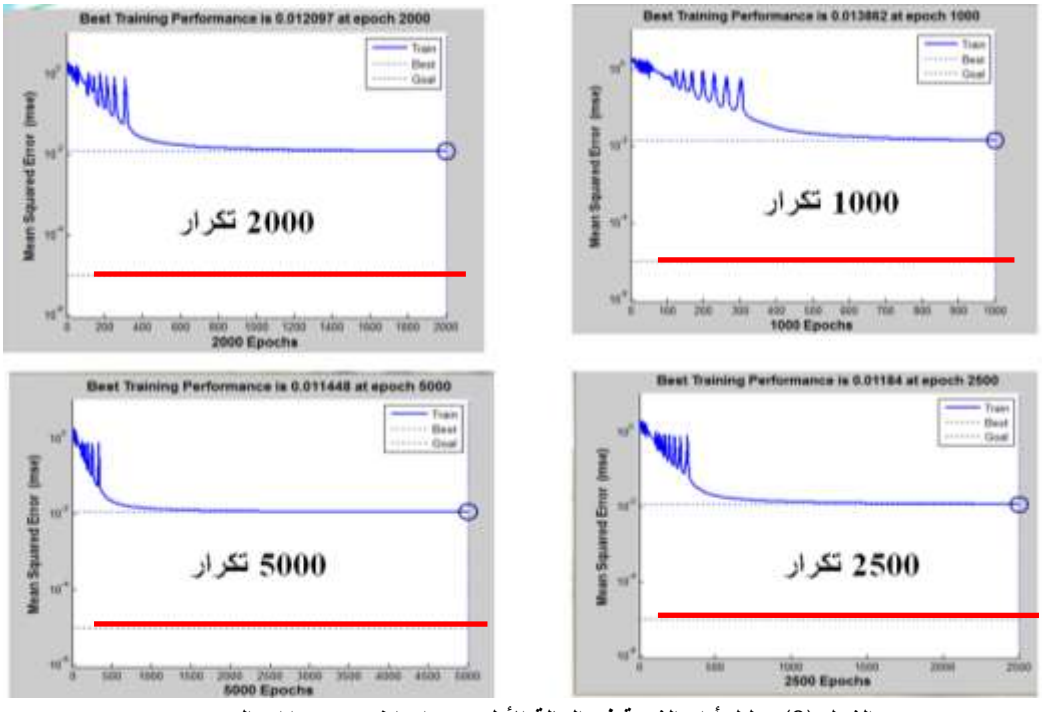


(a 100 عصبون للطبقة الخفية (b)300 عصبون للطبقة الخفية

الشكل(8) تحليل أداء الشبكة في الحالة الأولى معدل التعلم 0.035

3-زيادة عدد التكرارات طالما أن الخرج لم يصل إلى الهدف المرغوب.

يبين الشكل (9) مهما زدنا عدد التكرارات لن يتم الوصول لقيمة خرج فعلي تقارب الهدف.

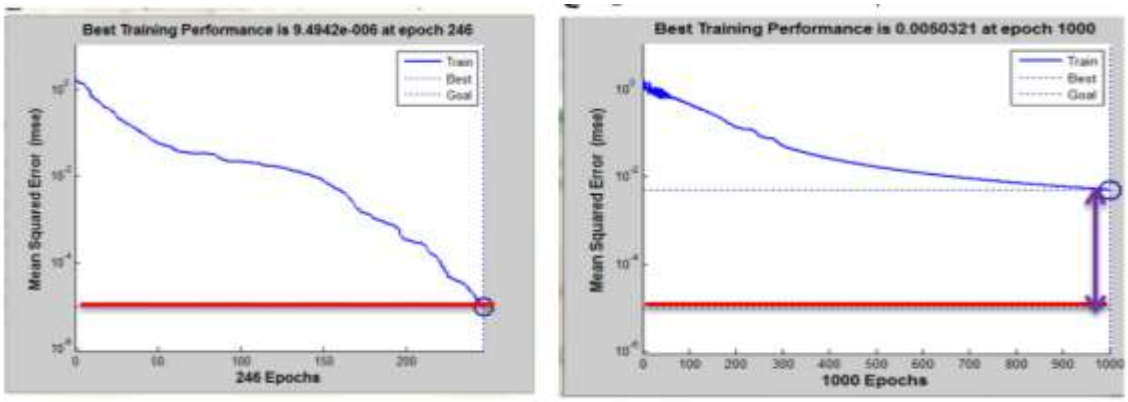


الشكل (9) تحليل أداء الشبكة في الحالة الأولى - اختلاف عدد تكرارات التدريب

4-5-2 تقنية الميل الموحد المتدرج *scg*:

1 - يبين الشكل (10) مقارنة بين شبكتين عصبيتين لهما نفس الخصائص (عدد الطبقات، عدد العصبونات، توابع التفعيل، معامل التعلم، عدد التكرارات، قيمة الخطأ) ولكن باستخدام تقنيتين مختلفتين في التدريب

gd,sgd وتبين لدينا أن تقنية *scg* أدت إلى تدريب الشبكة وبلوغ الهدف على عكس تقنية *gd* التي لم تحقق ذلك مهما زدنا التكرارات أو غيرنا معامل التعلم أو عدد عصبونات الطبقة الخفية لذلك لا يفضل استخدام تقنية *gd* في التدريب.

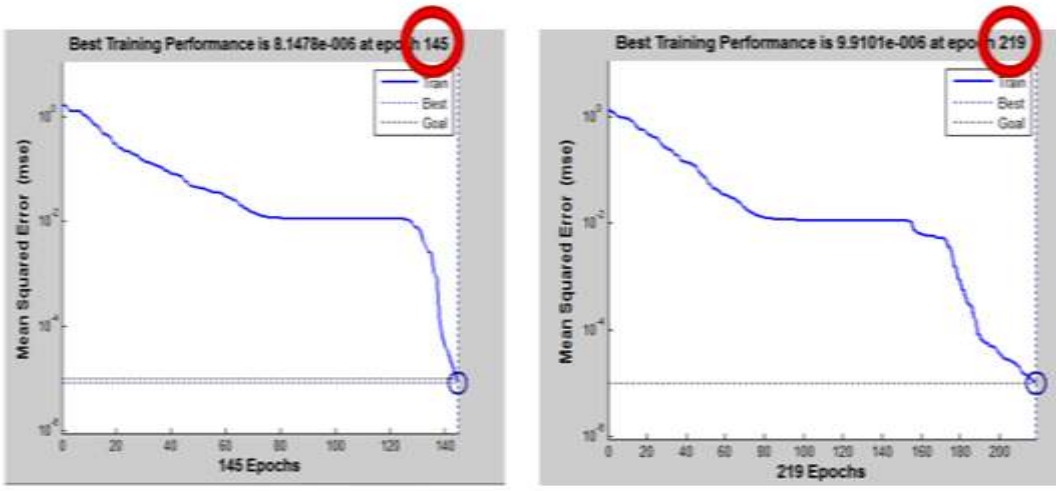


trainscg(*b*) *Traingd* (*a*

الشكل (10) مقارنة بين استخدام التقنيتين المختلفتين في التدريب وتوابع تفعيل ثابتة (*tansig*)

2-اختلاف عدد عصبونات الطبقة الخفية.

يبين الشكل(11) أنه تم الوصول إلى الهدف باستخدام عدد مختلف من العصبونات للطبقة الخفية ولكن عند عدد عصبونات أقل استغرق عدد تكرارات أكثر للوصول إلى الهدف المرغوب.

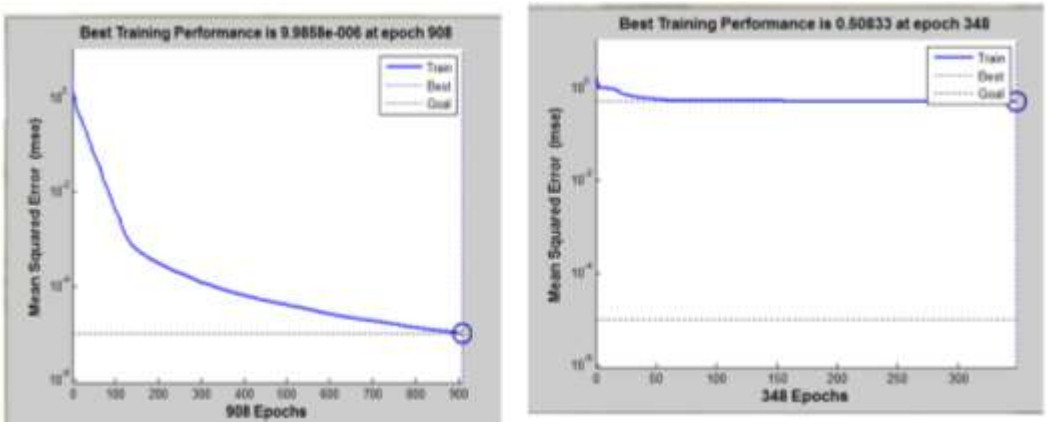


90(*b*)*10* (*a*)

الشكل (11) مقارنة بين استخدام عدد عصبونات مختلف للطبقة الخفية

3-اختلاف تابع التفعيل لطبقة الخرج.

يبين الشكل(12) أن التابع *logsig* غير قادر على فصل العينات في حين أن التابع *tansig* حقق ذلك.

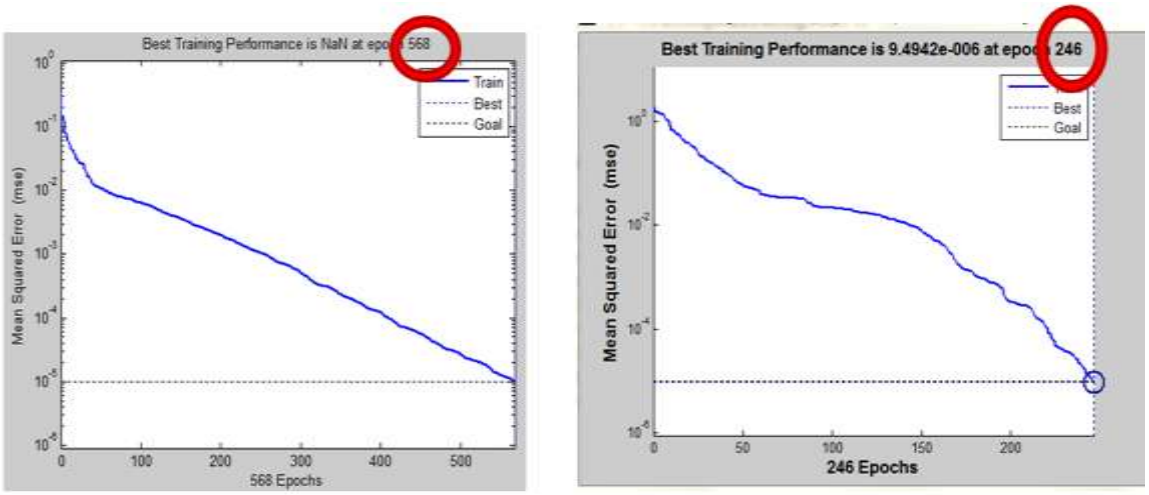


tansig(*b*) *logsig* (*a*)

الشكل (12) مقارنة بين استخدام توابع تفعيل مختلفة لطبقة الخرج

4-اختلاف عدد الطبقات:

يبين الشكل(13) أن اختيار عدد طبقات خفية مختلف لن يؤثر على القدرة على فصل العينات وإنما زيادة عدد الطبقات الخفية سبب في زيادة عدد مرات التدريب.

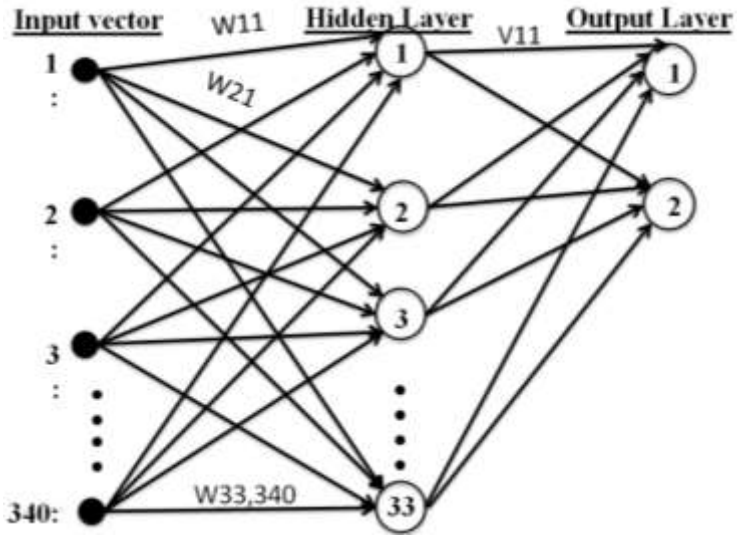


(*a*)طبقة واحدة (*b*)طبقتان

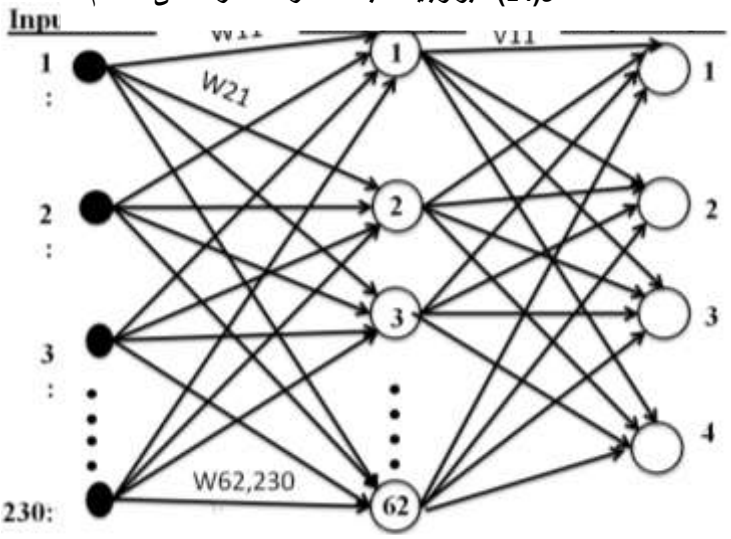
الشكل (13) مقارنة بين استخدام عدد مختلف من الطبقات الخفية

وفقاً للدراسة التحليلية تم اقتراح البنية الآتية للشبكة العصبية ذات التغذية الأمامية والانتشار الخلفي للخطأ *FFBPNN*:

- ✓ طبقة خفية واحدة.
- ✓ تابع الـ *'tansig'* للطبقة الخفية ، تابع *'tansig'* للطبقة الثانية (طبقة الخرج) ، تابع *'trainscg'* كتابع تدريب
- ✓ تم اختيار معدل تعلم القيمة 0.035 كقيمة نهائية، قيمة الخطأ 0.00001
- ✓ اختلفت عدد عصبونات الطبقة الخفية وطبقة الخرج حسب مرحلة التعرف، يظهر الشكلين(14)و(*15*) طبولوجيا الشبكة المستخدمة في كل من مرحلتي التعرف على المتكلم والتعرف على الكلام على التوالي.



الشكل(14) طبولوجيا الشبكة المقترحة للتعرف على المتكلم



الشكل(15) طبولوجيا الشبكة المقترحة للتعرف على الكلام

4-5 مرحلة التقييم:

تم التعرف على*189* عينة اختبار من أصل *216* عينة أي بما يعادل نسبة قدرها *87.50%* في مرحلة التعرف على المتكلم، و*196*من أصل *216* عينة أي بما يعادل *90.74%* في مرحلة التعرف على الكلام .

جدول (6) نتائج التعرف على المتكلمين

Expected	Recognized			
	Speaker1	Speaker2	Speaker3	Speaker4
Speaker1	46	3	5	0
Speaker2	1	49	1	3
Speaker3	4	1	49	0
Speaker4	1	7	1	45

جدول (7) نتائج التعرف على الكلام

expected	recognized								
	Open	close	left	right	up	down	thirty	sixty	forty
open	19	4	0	0	0	0	0	0	1
close	3	20	0	0	0	0	0	0	1
left	0	0	24	0	0	0	0	0	0
Right	0	0	0	24	0	0	0	0	0
Up	0	0	0	0	24	0	0	0	0
down	0	0	0	0	0	24	0	0	0
Thirty	0	0	0	0	0	0	19	5	0
Sixty	0	0	0	0	0	0	4	20	0
forty	2	0	0	0	0	0	0	0	22

الاستنتاجات والتوصيات:

بالنسبة لمرحلة المعالجة الأولية:

- باستخدام خوارزمية تحديد بداية الكلام ونهايته كان المتوسط الحسابي لعدد البتات التي تم الاستغناء عنها من مصفوفة كل ملف صوتي هي *5492* أي بنسبة قدرها *16.60%*
- إلغاء تفعيل خوارزمية *Endpoint* على أشعة الاختبار وتطبيقها على الشبكة العصبية المقترحة أدى إلى انخفاض نسبة التعرف على الكلمات المنطوقة من *90.74%* إلى *71.42%* ونسبة التعرف على المتكلم من *87.50%* إلى *75.40%*

في مرحلة استخلاص السمات:

بالنسبة لتحديد عدد المرشحات استنتجنا أن :

• بنك المرشحات المكون من عدد مرشحات(15–19): أدت إلى نسب تعرف على اختلاف السمات أقل من باقي بنوك المرشحات لذلك لا يحبذ استخدامها.

• بنك المرشحات المكون من عدد مرشحات (24–35):أدت إلى نسب تعرف متزايدة (تعرف على المتكلم) و متقاربة (في حالة التعرف على الكلام)على اختلاف السمات المستخدمة.

• بنك المرشحات المكون من عدد مرشحات (36–50):ثبتت نسبة التعرف.

بالنسبة لعلاقة عدد المرشحات بعدد السمات استنتجنا أنه:

• عندما يكون عدد السمات أصغر من عدد المرشحات: زيادة عدد السمات على اختلاف عدد المرشحات تؤدي إلى زيادة نسبة التعرف.

• عندما يكون عدد السمات أكبر من عدد المرشحات: ثبتت نسبة التعرف على أعلى نسبة تعرف حصلنا عليها عندما عدد السمات أقل بقليل من عدد المرشحات.

وبالتالي يحبذ اختيار بنك مرشحات مكون من عدد مرشحات بين(24–35)، ويحبذ اختيار عدد سمات أصغر من عدد المرشحات بقليل.

وفقاً للدراسة نستطيع تحديد (عدد مرشحات *35* عدد سمات *34*) من أجل مرحلة التعرف على المتكلم و(عدد مرشحات *24* عدد سمات *23*) من أجل مرحلة التعرف على الكلام.

بالنسبة لمرحلة التعرف:

- باستخدام عينات التدريب وعينات الاختبار ذاتها ونفس الخوارزميات أبدت خوارزمية *MFCC* إمكانية التعرف على الكلام بنسبة قدرها % 90.74ونسبة تعرف على المتكلم% 87.50،أي أبدت الخوارزمية إمكانية التعرف على الكلام أفضل من المتكلم بنسبة قدرها *3.24%*.

مقارنة مع الأبحاث السابقة:

يوضح الجدول(8) مقارنة الدراسة المقترحة مع الأبحاث السابقة.

جدول(8) مقارنة الدراسة مع الدراسات السابقة

الباحث	قاعدة البيانات	الطرق المستخدمة للبحث	ملاحظات
S. S. Wali , S. M. Hatture and S. Nandyal [15]	(10-50مستخدم) كل مستخدم لفظ العبارة 10مرات	Silience part removal *Algorithm,Mfcc,vq,ffbpnn*	درس مرحلة التعرف على المتكلم فقط واستخدم عبارة واحدة فقط في مرحلة التعرف اهتم بنسبة التعرف فوجد أنها %(70-92) من أجل(10-50) مستخدم ،لم يدرس بارامترات خوارزمية *mfcc* ، لم يستخدم خوارزمية*endpoint*
Mayur R Gamit, KinnalDhameliya[16]	10 كلمات من قبل 28 متكلم	Energy and zero crossing rates(zcr), Mfcc&Lpc, ffbpnn	درس مرحلة التعرف على الكلام فقط ووجد نسبة تعرف 82% عند استخدام mfcc ونسبة 86% عند الدمج (mfcc&lpcc) ،لم يدرس بارامترات خوارزمية mfcc، لم يستخدم خوارزمية*endpoint*
KirandeepKaur,Neelu Jain[17]	-----	*Vq,gmm,svm,dtw,hmm Lpcc,lpc,wavelete,mfcc,*	درس مرحلة التعرف على المتكلم فقط ووجد أن خوارزمية *mfcc* أفضل خوارزمية لاستخلاص السمات و*gmm* أفضل مصنف، ،لم يدرس بارامترات خوارزمية mfcc، لم يستخدم خوارزمية*endpoint* .
Namrata Dave1[7]	-------	*Lpc,plp,mfcc,ffbpnn*	درس مرحلة التعرف على الكلام فقط ووجد أن خوارزميتي

			mfcc,plp أفضل من*Lpc*، لم يدرس بارامترات خوارزمية *mfcc*، لم يذكر خوارزميةلازالة فترات الصمت.
GOYANI, M, DAVE, N.[8]	*750* عينة يعود كل 150عينة فيها للفظمعين(5ألفاظ صوتية)	*Lpc,plp,mfcc,ffbpnn*	درس مرحلة التعرف على الكلام فقط ووجد أن خوارزمية*mfcc* أفضل من و*plp,Lpc*، لم يدرس بارامترات خوارزمية *mfcc*وانما حدد عدد المرشحات يساوي 26و عدد السمات12أو 42،لم يذكر خوارزميةلازالة فترات الصمت.
AshishB.Ingale , Dr.D.S.Chaudhari[18]	--------	*Mfcc,svm,hmm*	درس التعرف على مشاعر المتكلم حيث ميز *5* حالات باستخدام مصنفين مختلفين ووجد أن مصنف *hmm* أفضل من مصنف *svm*، لم يدرس بارامترات خوارزمية *mfcc*، لم يذكر خوارزمية لازالة فترات الصمت.
الدراسة المقترحة	576عينة ، من قبل 4متكلمين لفظ كل منهم *9* كلمات	*Endpoint ,mfcc,k-means,ffbpnn*	درس مرحلتي التعرف على الكلام والمتكلم ، وناقش تأثير تغير قيمة بارامترين خاصين بخوارزمية *mfcc* على نسبة التعرف ،وفعالية خوارزمية *mfcc* على كل مرحلة ،ودراسة خوارزمية *endpoint* وتأثيرها على نسبة التعرف

التوصيات:

1 - تسجيل الأصوات في استوديوهات مخصصة.

2 - دراسة بارامترات أخرى لخوارزمية *MFCC* وتأثيرها على نسبة التعرف.

3 - دراسة خوارزميات مختلفة لإزالة فترات الصمت وتأثيرها على نسبة التعرف.

المراجع:

[1]- CARROLL, T.;COLANGELO, R.;STROTT, T.*"Bird Call Identifier –Identifying Songs of Bird Species through Digital Signal Processing Techniques"*. 2010,118.

[2]-Xue, X.*"Joint Speech and Speaker Recognition Using Neural Networks"*.NOVIA-University of applied science. 2013,60.

[3] – CHOUDHARY, A.;KSHIRSAGAR,R.*"Process Speech Recognition System using Artificial Intelligence Technique"*.(IJSCE) ,ISSN: 2231-2307, Volume-2, Issue-5, 2012,PP(239-242).

[4]- TAN, L.;KARNJANADECHA, M. *"Modified Mel-Frequency Cepstrum Coefficient Department of Computer Engineering"*. Faculty of Engineering Prince of Songkhla University Hat Yai,Songkhla Thaiiland,90112,2001,PP(1-4).

[5]- GIANNAKOPOULOS, *T. A." method for silence removal and segementation of speed signals"*. report, University of Athens, Greece and NCSR DEMOKRITOS, Greece.1975,PP(207-215).

[6]-Rabiner, L.R. ;Sambur,M.R.*"An algorithm fordetermining the end point of isolated utterances"*.bell labs technical journal,2013,pp(297-315).

[7]- Dave1,N.*"Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition"*. 1G H Patel College of Engineering, Gujarat Technology University,INDIA,Volume 1, Issue VI, 2013,pp(1-5).

[8]- GOYANI, M.;DAVE, N.*"Performance Analysis of LPC, PLP and MFCC ParametersIn Speech Recognition"*. GCET, SPU, V.V.Nagar, Gujarat, INDIA,2013,pp(174-178).

[9]- DEMUTH, H.; BEALE, M.*"Neural Network Toolbox For Use with MATLAB"*. version 3, 1997, 700.[10]- D.E. Rumelhart and J. McClelland, editors, Parallel Data Processing, Vol.1, Chapter 8, The M.I.T. Press, Cambridge, MA, 1986, pp(318-362).

[11]- Elsevier experts.*"Pattern Recognition"*. Fourth Edition, Elsevier Inc.,2009,510.

[12]-Kmeans produc document. "http://www.mathworks.se/help/toolbox/stats/kmeans.html", Download Date 11/2/2015.

[14]-http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs, Download Date 7/5/2015.

[15]-Wali,S. S.;Hatture,S. M.;Nandyal,S.*"MFCC Based Text-Dependent Speaker Identification Using BPNN"*.International Journal of Signal Processing Systems Vol. *3, No. 1, June 2015,pp(30-34).*

[16]-Gamit,M. R.;Dhameliya,K.*"ISOLATED WORDS RECOGNITION USING MFCC, LPC AND NEURAL NETWORK"*. IJRET: International Journal of Research in Engineering and Technology, eISSN: 2319-1163 | pISSN: 2321-7308, Volume: 04 Issue: 06 | June-2015,pp(146-149).

[17]- Kaur,K.;Jain,N.*"Feature Extraction and Classification for Automatic Speaker Recognition System – A Review"*. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015 ISSN: 2277 128X,pp(1-6).

[18]-Ingale,A.B.;Chaudhari,D.S.*"Speech Emotion Recognition Using Hidden Markov Model and Support Vector Machine "*. International Journal of Advanced Engineering Research and Studies ,IJAES/Vol .I/Issue III/April-June,2012,pp(316-318).

[19]–د.مريم ساعي؛ م.علي ميا *"التعرف على الأشخاص باستخدام الأذن"* . رسالة ماجستير، جامعة تشرين، الصفحة(50–55)، 2013.