مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (1) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (1) 2017

تحسين أنظمة التعرف على الكلام عن طريق جمع خوارزميتين لاستخلاص السمات

د. جعفر الخير•

(تاريخ الإيداع 9 / 10 / 2016. قُبِل للنشر في 7/ 2 / 2017)

□ ملخّص □

تعد تقنيات التعرف على الكلام من أهم التقنيات الحديثة التي دخلت بقوة في مجالات الحياة المختلفة سواء الطبية أو الأمنية أو الصناعية. وبناءً عليه تم تطوير العديد من الأنظمة المعتمدة على طرق مختلفة في استخلاص السمات و التصنيف.

في هذا البحث تم إنشاء ثلاثة أنظمة للتعرف على الكلام، تختلف عن بعضها البعض بالطرق المستخدمة في مرحلة استخلاص السمات،حيث استخدم النظام الأول خوارزمية MFCC بينما استخدم النظام الثاني خوارزمية LPCCأما النظام الثالث فاستخدم خوارزمية PLP.تشترك هذه الأنظمة بطريقة التصنيف حيث استخدمت خوارزمية الـHMM كمصنف.

في البداية تم دراسة وتقييم أداء عملية التعرف على الكلام للأنظمة الثلاثة السابقة المقترحة منفردةً. بعد ذلك تم تطبيق خوارزمية الجمع على كل زوج من الأنظمة المدروسة وذلك لدراسة أثر خوارزمية الجمع في تحسين التعرف على الكلام.

تم اعتماد نوعين من الأخطاء، الأخطاء التزامنية (simultaneous errors)والأخطاء الاعتمادية (dependent errors)، كوحدة مقارنة لدراسة فعالية خوارزمية الجمع في تحسين أداء عملية التعرف على الكلام. يتبين من نتائج المقارنة أن أفضل نسبة تعرف على الكلام تم الحصول عليه ا في حالة جمع الخوارزميتان MFCC وPLPحيث تم الحصول على معدل تعرف 93.4%.

الكلمات المفتاحية التعرف على الكلام، استخراج السمات، نماذج ماركوف المخفية.

• أستاذ مساعد، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (39) العدد (1) 2017
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (39) No. (1) 2017

Improvement of Speech Recognition by Merging Two Features Extraction Algorithms

Dr. Jaffar Alkhier •

(Received 9 / 10 / 2016. Accepted 7 / 2 / 2017)

□ ABSTRACT □

The speech recognition is one of the most modern technologies, which entered force in various fields of life, whether medical or security or industrial techniques. Accordingly, many related systems were developed, which differ from each otherin feature extraction methods and classification methods.

In this research,three systems have been created for speech recognition.They differ from each other in the used methods during the stage of features extraction.While the first system used MFCC algorithm, the second system used LPCC algorithm, and the third system used PLP algorithm.All these three systems used HMM as classifier.

At the first, the performance of the speechrecognitionprocesswas studied and evaluatedfor all the proposedsystems separately. After that, the combination algorithm was applied separately on eachpair of the studied system algorithmsin order to study the effect of using the combination algorithm onthe improvement of the speech recognition process.

Twokinds of errors(simultaneous errors and dependent errors) were usedto evaluate the complementaryof each pair of the studied systems, and to study the effectiveness of the combination on improving the performance of speech recognition process. It can be seen from the results of the comparison that the best improvement ratio of speech recognition has been obtained in the case of collection MFCC and PLP algorithms with recognition ratio of 93.4%.

Key words: Speech recognition, features extraction, Markov Hidden models

• **Assistant DeProfessor, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria,**

مقدمة:

بدأ اهتمام خبراء الحاسب والباحثين في مجال التعرف على الكلام منذ أكثر من أربعة عقود، وذلك لكي يصل الإنسان إلى مرحلة تجعله قادرا على التخاطب مع الكمبيوتر وإعطاءه الأوامر والتعليمات صوتيا وبدون الحاجة إلى الكتابة وغيرها من الطرق، وذلك توفيرا للوقت والجهد.

وفي السنوات الأخيرة تطورت نظم التعرف على الكلام تطورا" واضحا" وكبيرا"، بحيث أصبحت برامج التعرف الآلي تدخل في أغلب مجالات الحياة، ووصلت إلى دقة مرضية نوعا" ما [1]. يمكن تصنيف أنظمة التعرف على الكلام اعتماداً على الطرق المستخدمة في استخراج السمات (feature extraction methods) وطرق التصنيف (classification methods) التي تعتمدها[2].

تؤدي عملية استخراج بعض المعلومات من إشارة الكلام إلى خسارة معلومات أخرى، حيث أن طرق استخراج السمات المطلوبة للتعرف على الكلام تختلف عن بعضها البعض بالسمات التي تعتمدها، وبالتالي فإن لكل طريقة نسبة تعرف صحيحة محدودة وغير كاملة [3] [4]. بناءاً على ذلك ظهرت فكرة جمع ميزات نظامين أو أكثر من أجل تحسين عملية التعرف بالاستفادة من ميزات كل نظام لتعويض النقص أو الضعف في الأنظمة الأخرى. ومنه كانهدفهذه الدراسةتحسيننتائجالتعرفعلىالكلاممنخلالالجمعبينأكثرمنتقنيةمنالتقنياتالمستخدمةفياستخراجالسمات (feature extraction) ومقارنتها للوصول إلى نظام تعرف بأفضل أداء.

في بداية هذاالبحث سيتم تقديم موجز مختصر عن مهام معالجة الكلام وأهم ميزاته ليتم بعدها شرح خوارزميات استخلاص السمات المستخدمة في أنظمة التعرف على الكلام. خوارزمية الجمع المقترحة سيتم شرحها في الفقرة الرابعة يليها طرق البحث ومناقشة النتائج ليتم بعدها تلخيص البحث بخاتمة تلخص أهم النتائج التي تم الوصول إليها.

مقدمة في معالجة الكلامIntroduction to Speech Processing:

يمكن تقسيم معالجة الكلام اعتماداً على المهام الموكلة إليهاإلى عددمن المجالات الرئيسية، وهي:

-التعرف على الكلام (Speech recognition): حيث تُحوّلإشارة الكلام إلى تدفق من الرموز (الفونيمات، والكلمات) التي تمثل المعلومات في الكلام.

-التعرف على المتكلم (Speaker recognition): لمعرفة المتكلم الذي قام باصدار اشارة الكلام من مجموعة من المتكلمين ذات سمات صوتية معروفة

-التحقق من المتكلم (Speaker verification): للتأكد من أن المتكلم الذي قام باصدار اشارة الكلام هو نفسه الشخص المراد التأكد منه أم لا.

-تركيب الكلام (Speech synthesis): وذلكبتوليد اشارات الكلام اصطناعياً بحيث تكون سمات هذا الكلام مختلفة ولم يتم اصدارها من أي متكلم قبل ذلك.

-ترميز الكلام (Speech coding): يتم تمثيل اشارة الكلام بصيغة فعالة والتي تستخدم لأغراض النقل والتخزين بحيث يمكن استعادة الاشارة الاصلية فيما بعد.

يركز هذا البحث بشكل اساسي على مهمة التعرف على الكلام، حيث سيتم دراسة وتقييم أثر جمع خوارزميتي استخلاص سمات على تحسين عملية التعرف على الكلام.

أ.أنواع الكلام:

في البداية وقبل البدء بشرح التقنيات والخوارزميات المستخدمة في مجال التعرف على الكلام سنتكلم بشكل موجز عن أهم الميزات والمصطلحات المستخدمة في هذا المجال.

يتألف الكلام المراد التعرف عليه بواسطة أنظمة التعرف على الكلام من مجموعة من الكلمات التي يمكن تصنيفها بالشكل التالي:

1 - الكلمة المعزولة (isolated word): تعني كلمة واحده يستقبلها ويحللها النظام في وقت واحد. أي مايعني أن النظام يفرق بين الكلمات بواسطة سكوت المستخدم بين كل كلمة وكلمة ويكون هذا النظام له حالتين إما استماع أو صمت.

2 - الكلمة المتصلة (connected word):تعني ايضاً كلمة واحده يستقبلها ويحللها النظام في وقت واحد بنفس طريقة الكلمات المعزولةلكن هنا يتم تقليل فتره الصمت بين الكلمات بحيث تبدو كأنها متصلة أو كأنها جملة متقطعة.

3 - الكلامالمستمر(continuous speech): يتم استخدام جمل كاملة بنطق طبيعي بدون أية شروط على طريقة النطق للمستخدم. حيث يعتبر هذا النظام الذي يستخدم هذا النوع من الكلام من أصعب الأنظمة تطبيقافي التعرفعلى الكلام لصعوبة تحديد حدود كل كلمة على حدوأيضا قلة الدقة في نطق الكلمات عندما تكون ضمن الجملة، وعدم ثبات طول الكلمة من مستخدم إلى آخر[6].

في هذا البحث تم استخدام كلمات معزولة isolated wordفي عملية المقارنةودراسة فعالية خوارزمية الجمع اعتماداً على كلمات معزولة تم تسجيلها بواسطة عدة أشخاص من أجل عملية المقارنةوالتي سيتم شرحها لاحقاً.

ب. تطبيقات التعرف على الكلام:

يوجد العديد من التطبيقات العملية لأنظمة التعرف على الكلام منها:

- التعرف على الكلام الملفوظ وتحويله إلى نص حاسوبي، أومعرفة من هو الشخص الذي يتكلم في مقطع صوتي من ضمن مجموعة من الأشخاص الذين يحتفظ النظام بمقاطع صوتية لكلامهم[7].
- التعليم والتعلم: لعلم الأصوات دور تطبيقي كبير يتمثل في تعلم اللغات وتعلمها.
- والعديد من التطبيقات التي في مجالات الحماية والأمان مثل البصمة الصوتية وتطبيقات مختلفة لمساعدة المعوقين إلخ ....

ج. نظام التعرف على الكلام:

تتكون معظم الأنظمة الحديثةللتعرف على الكلام والمعتمدة على نموذج ماركوفمن ثلاثةمراحلأساسية :

1 - مرحلة استخلاصالسمات (Feature extraction):يتم في هذه المرحلة تحويل إشارة الكلام إلى تسلسل من أشعة السمات(feature vectors)التي تمثل المعلومات المخزنة في الكلام المنطوق. يتم في هذه المرحلة تقليل أبعاد إشارة الكلام الأصلية (dimensionality reduction) لإعداد هذه الإشارة في صيغة المتطلبات الأساسية لمرحلة التصنيف التالية.من الخصائص الهامة لمرحلة استخراج السمات هو كبت المعلومات التي ليس لها أهمية (irrelevant) من أجل تصنيف صحيح مثل المعلومات حول المتحدث(التردد الأساسي) والمعلومات التي تخص قناة النقل (مميزات المايكروفون).

2 - مرحلة التصنيف (Acoustic classification):وظيفة المصنف هو إيجاد الرسم التخطيطي(mapping) بين تسلسل أشعة السماتوبين عنصر الكلام المتعرف عليها. تم استخدام المصنف Hidden Markov Model (HMM)في هذا البحث وهو من أشهر المصنفات المستخدمة في مجال التعرف على الكلام في الوقت الحالي.

3 - نماذج اللغة (Language models):وظيفة نماذج اللغة هو اختيار الفرضيات التي هي على الأرجح التسلسل الصحيح لعناصر الكلام للغة معطاة [8]. تعتمد درجة تعقيد نموذج اللغة المستخدم على درجة تعقيد المشكلة المطلوب حلها فمثلاً يكون نموذج اللغة للتعرف على الكلام المستمر أكثر تعقيداً منه عند معالجة عدد محدد من الأوامر المنطوقة.

خوارزميات استخلاص السمات:

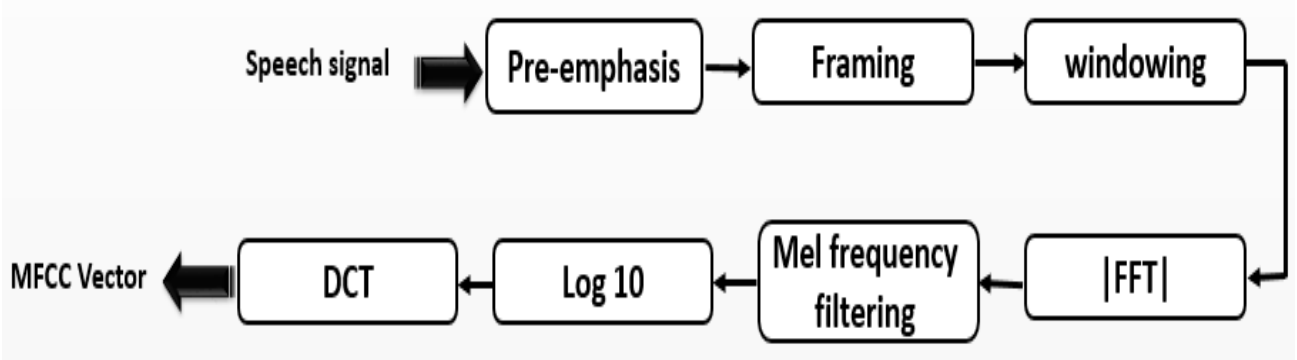
تماستخداماالخوارزمياتالتالية(MFCC) Mel Frequency Cepstral Coefficients و Linear Prediction Cepstral coefficients (LPCC) و PerceptualLinear Prediction (PLP)[15] في مرحلة استخلاص السمات لأنظمة التعرف على الكلام المدروسة في هذا البحث. في هذه الفقرة سيتم شرح مبدأ وآلية عمل كل خوارزمية على حدى.

1 1 خوارزمية MFCC:

تعتبر خوارزمية الـ MFCC (Mel Frequency Cepstral Coefficients)من الخوارزميات السائدة والمهيمنة المستخدمة في استخلاصالسمات في نظم التعرف على الكلام وذلك بسبب حساسية مرشحاتها لخواص إشارة الصوت البشرية[9].تستخدم معاملات الـMFCC بشكل كبير في التعرف على الكلام، حيث تم تقديم هذه المعاملات من قبل العالمين Davis and Mermelstein، في عام 1980 ومازالت متقدمة في هذا المجال منذ ذلك الوقت.

إن الأصوات التي تولد من قبل الإنسان يتم ترشيحها حسب شكل المسلك الصوتي (vocal tract)، فإذا تمكنا من 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) حيث أن هدف الـ MFCC هو تمثيل هذا الغلاف بدقة.

تعتمد الـMFCC على التغيرات المعروفة في عرض حزمة الترددات للأذن البشرية، حيث أن لمرشحاتها تباعدا خطيا ضمن مجال الترددات المنخفضة (الأقل من 1000 هرتز)ولوغاريتمياًضمن مجال الترددات المرتفعة (أكبر من 1000 هرتز) وهي تستخدم من أجل التقاط الصفات الرئيسية للكلام [9].يوضح الشكل (1) خطوات عمل خوارزمية الـ MFCC.



الشكل (1) المخطط الصندوقي لعمل الخوارزمية MFCC

- **pre-emphasis**

يتم تطبيق عملية pre-emphasis (وهي عملياً مرشح تردد عالي high pass filter) على الإشارة وذلك من أجل تعويض جزء التردد العالي الذي تم فقده أثناء آلية إنتاج الكلام (زيادة الطاقة النسبية للطيف عالي التردد)، حيث يتم إعادة تقييم كل قيمة في إشارة الكلام باستخدام الصيغة (1) ،

$$s2(n) = s(n) - a*s(n-1) \quad (1)$$

حيث: s(n): إشارة الكلام

s2(n): إشارة الخرج بعد عملية الـ pre-emphasis

a: ثابت تتراوح قيمته بين 0.9 و 0.1

- **التأطير (framing)**

إشارة الكلام هي إشارة متغيرة باستمرار لذلك من أجل تبسيط الدراسة نعتبر أنه من أجل نطاق زمني قصير (short time scale) فإن إشارة الصوت لا تتغير كثيراً لهذا السبب يتم تقطيع الإشارة إلى عدد من الإطارات (frames)، زمن كل إطار من 20 إلى 40 ميلي ثانية مع وجود تداخل اختياري يساوي نصف أو ثلث حجم الإطار وذلك من أجل تسهيل الانتقال من إطار إلى آخر[10].

- **النوفذة (windowing)**

كل إطار(frame) سوف يخضع لعمية النوفذة (windowing) باستخدام نافذة هامينغ (Hamming window)، وذلك منأجل القضاء على الانقطاعات عند الحواف. تعطى نافذة هامينغ بالعلاقة(2):

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right), \quad 0 \le n \le N \quad (2)$$

حيث w(n) هو مطال العينة الجديد.

n هو ترتيب العينة في النافذة.

N هو الطول الكلي للنافذة

بعد عملية النوفذة windowing سوف يتم تطبيق تحويل فوريه السريع FFT من أجل كل إطار وذلك من أجل استخراج مركبات التردد للإشارة في مجال الزمن[11].

- **ترشيح الإشارة وفقاً لتردد ميل (Mel frequency filtering)**

تعمل خوارزمية الـMFCC على ترشيح طيف الإشارة الصوتية (short time power spectrum) عن طريق مجموعة من المرشحات المثلثية (Mel filter bank) التي صممت كمحاكاة لمرشحات تمرير الحزمة band pass filtering التي تظهر في النظام السمعي.تكون مجموعة المرشحات المثلثية السابقة متباعدة بانتظام وفقا لمقياس ميل الترددي(Mel frequency scale)الذي يعطى بالعلاقة (3):

$$m = 2595 \log\left(\frac{f}{700} + 1\right) \quad (3)$$

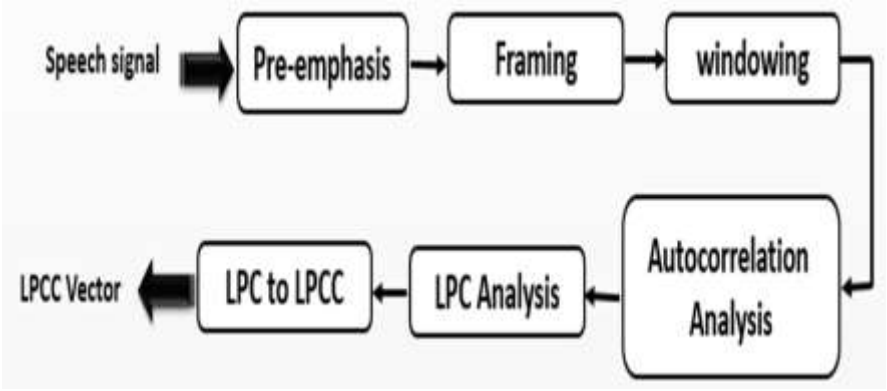
حيث m: التردد المقاس لنغمة صافية (perceived frequency of pure tone)وf:ترددها المقاس الأصلي

يستطيع الإنسان أن يميز التغيرات الصغيرة في الـpitch(وهي الارتفاع أو الانخفاض النسبي لنغمة (tone) كما تدركها الأذن، والتي تعتمد على عدد الاهتزازات التي تنتجها الحبال الصوتية في الثانية) وبشكل أفضل عند الترددات الصغيرة من الترددات الكبيرة.بالتالي فإن تضمين هذا المقياس يجعل سماتنا أقرب إلى سمع الإنسان[12].

يتم بعد ذلك حساب اللوغاريتم لطيف مجال ميل (Mel scale spectrum)، ومن ثم يستخدم تحويل جيب التمام المتقطعDCTلإعادة تحويل طيف مجال ميل اللوغاريتمي إلى مجال الزمن حيث نحصل نتيجة هذا التحويل على شعاع الـMFCC.

2 1 خوارزمية LPCC:

يبين الشكل (2)المخطط الصندوقي لعمل خوارزمية الـ LPCC.إن الفكرة الرئيسية لخوارزمية الـ LPCCالمعتمدة على تحليل التنبؤ الخطي Linear Predictive Analysisوالذي بدوره يعتمد على آلية إنتاج الكلام (أي أنه يستخدم نموذج مرشح–مصدر source–filter التقليدي)،هي أن عينة محددة من الكلام في الوقت الحالي يتم تقريبها كمزيج خطي من عينات الكلام السابقة[13].



الشكل (2) المخطط الصندوقي لعمل الخوارزمية LPCC

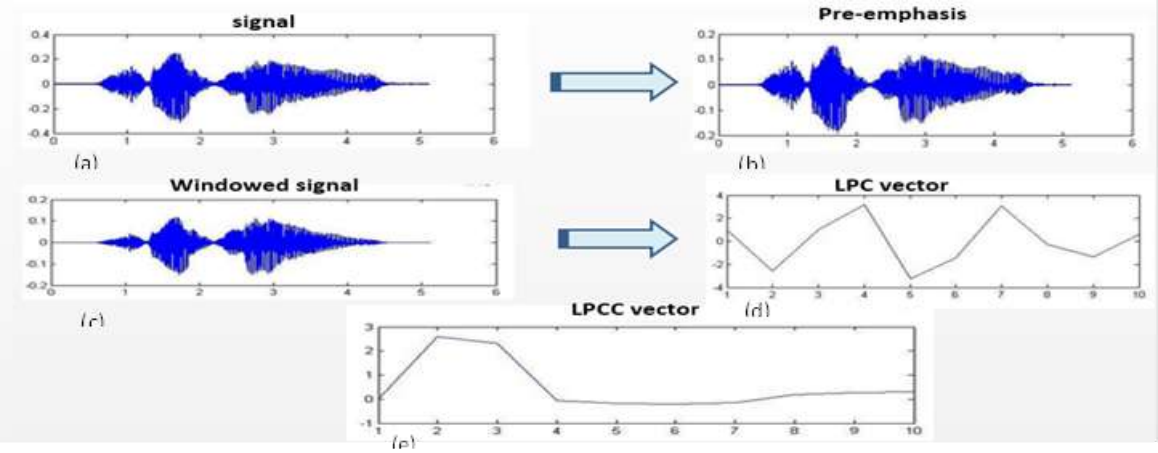
من خلال تقليص مجموع مربعات الاختلافات (على فترة زمنية محدودة) بين عينات الكلام الفعلية وقيم التوقع (التنبؤ) الخطية، سوف يتم تحديد مجموعة فريدة من البارامترات (معاملات التنبؤ الخطية)، هذه المعاملات تشكل أساسا لتحليلات التنبؤ الخطي للكلام.في الواقع إن عوامل التنبؤ الفعلية لا تستخدم في التعرف على الكلام لأنها نموذجية تظهر التباينالعالي، لذلك يتم تحويل معاملات التنبؤ هذه إلى مجموعة أخرىمن البارامترات هي Cepstral Coefficients بواسطة المعادلات الرياضية التالية (4).

$$c_{\mathrm{LPC}}(m)=\begin{cases}-a(m)-\sum_{i=1}^{m-1}(1-\frac{i}{m})a(i)c_{\mathrm{LPC}}(m-i), & 1\le m\\ \sum_{i=1}^{m-1}(1-\frac{i}{m})a(i)c_{\mathrm{LPC}}(m-i), & m>P\end{cases} \quad (4)$$

حيث m=1،...،P ؛ a(m) تعبر عن معاملات التنبؤ الخطي LPC coefficients.

P : يعبر عن ترتيب النموذج model order.

يبين الشكل (3) خطوات الخوارزمية للأمر الصوتي shutdown، والتي تستخدم نفسعمليات pre–emphasis،التأطيرframing، النوفذة windowing المطبقة في خوارزمية الـMFCC.



الشكل (3) خطوات الخوارزمية LPCC للأمر الصوتي "shutdown"

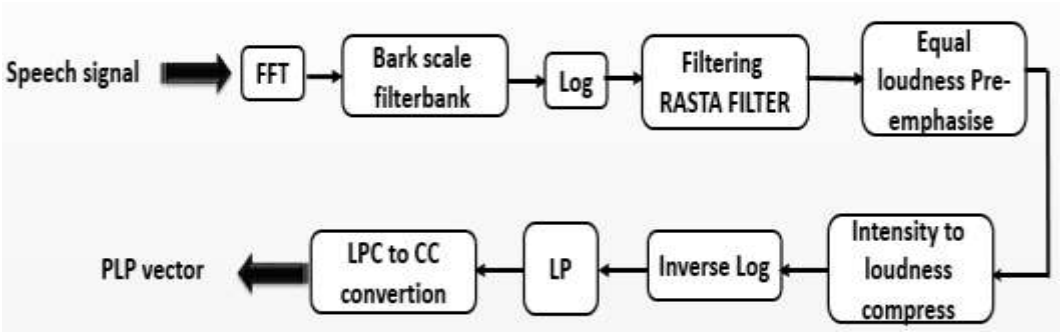
3 1 خوارزمية PLP:

تعتمد هذه التقنية علىالعلاقة بين مؤثر فيزيائي والإدراكات المؤثرة (psychophysics of hearing)، حيث تقوم باستبعاد المعلومات التي ليس لها صلة بالكلام وبالتالي تحسين عملية التعرف .

تعتبر خوارزمية الـ PLPمماثلة لـLPC إلا أن خصائصها الطيفية تم تحويلها لتتناسب مع خصائص النظام السمعي عند الأنسان ، حيث أن PLP تقارب ثلاث جوانب رئيسية [14]:

- منحني تمييزالحزمة الحدية (The critical-band resolution curve)
- منحني تساوي الجهارة(الشدة الصوتية) (The equal loudness curve)
- علاقة قانون الطاقة بشدة الجهارة (The intensity-loudness power-law relation)

يبين الشكل (4)المخطط الصندوقي لعمل خوارزمية الـ PLP



الشكل (4) المخطط الصندوقي لعمل الخوارزمية PLP

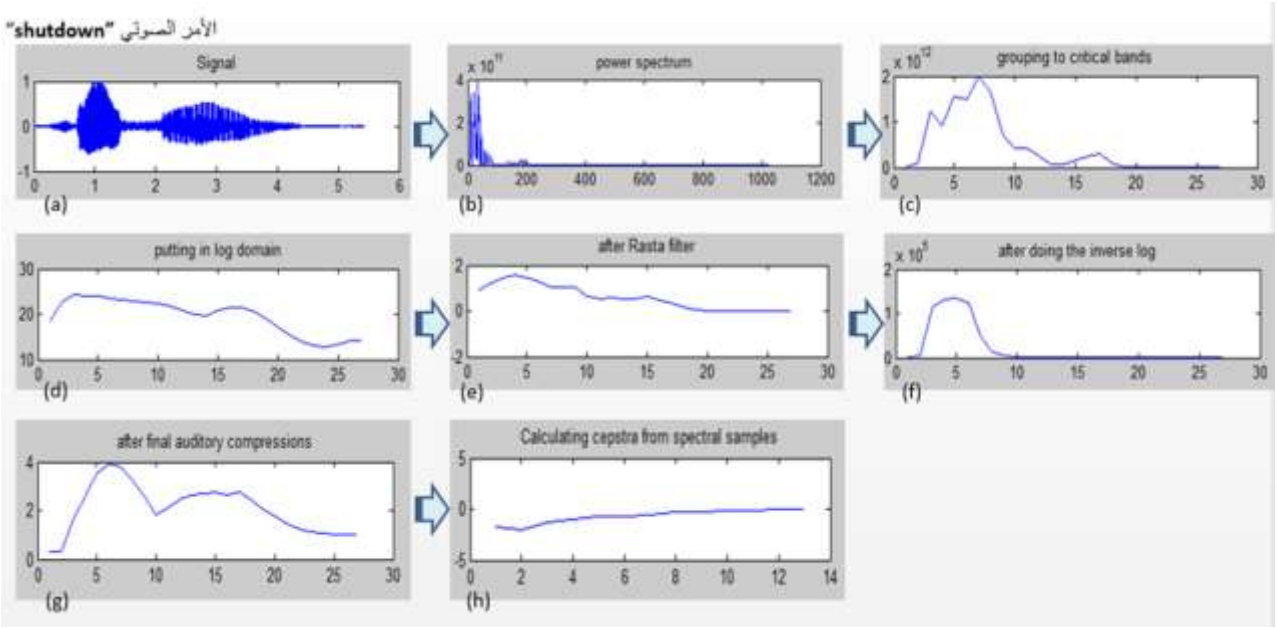
بعد معالجة إشارة الكلام يتم حساب تحويل فورييه السريع FFT للحصول على طيف الإشارة (power spectrum)، ومن ثم يتم استخدام مرشحات بشكل شبه منحرف تعتمد على نطاق barkلتتم بعدها عملية pre-emphasis لطاقة الطيف (power spectrum) بواسطة منحني قياس ارتفاع الصوت (equal-loudness curve)، والتي تقارب الحساسية غير المتساوية لسمع الإنسان عند ترددات مختلفة، عند حوالي 40 db،حيث أن كل معامل لطاقة الطيف power spectrum coefficient سوف يتم ضربه بالوزن E الذي يعطى بالعلاقة [5]:

$$E(w)=\frac{(w^2+56.8\times10^6)w^4}{(w^2+6.3\times10^6)^2(w^2+0.38\times10^9)} \qquad (5)$$

ممايسبب انخفاضاً في الحساسية في نطاق التردد العالي وهذا ما يجعل ملائمة الطيف للتنبؤ الخطي (LP) أكثر تجانسا على نطاق الترددات العالية.

تتم بعد ذلك عملية ضغط لطيف الكلام ويتم تنفيذ هذه العملية وفق (power–law of hearing) والذي ينمذج العلاقة غير الخطية بين الكثافة intensity والشدة loudnessالمقاسةلها. حيث يتم في هذه المرحلة تقليل التغيرات الديناميكية وتسطيح قمم الطيف ليكون خرج هذه المرحلة طيف مسطح أكثر (smoother) مع قمم (peaks) أقل وضوحاً. في المرحلة التالية يتم حساب المعاملات التنبؤية، وتحويلها إلى معاملات cepstral ومن ثم الحصول على شعاعPLP .

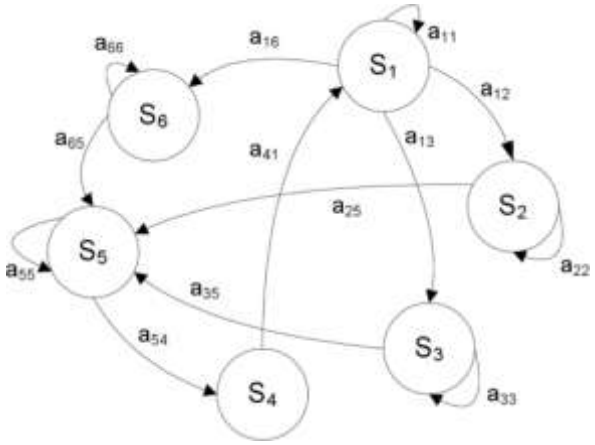
يبين الشكل (5) نتائج الخوارزمية PLP من أجل الأمر الصوتي " shutdown ".



الشكل (5) خطوات الخوارزمية PLPللأمر الصوتي "shutdown"

سلاسل ماركوفMarkov Chain:

يمكن تصنيف النماذج الرياضية إلى محددة(Deterministic) أو تصادفية (Stochastic). وفي الحياة العملية توجد عدة حالات تتضمن ظواهر ذات سلوك غير قطعي لا يمكن السيطرة عليها بشكل تام أو التنبؤ بسلوكها المستقبلي بشكل مؤكد والتي يطلق عليها مصطلح العمليات التصادفية [16].فيصبح هنا النموذج التصادفي هو الأكثر ملاءمة لتمثيلها رياضياً.



الشكل (6) سلسلة ماركوف لـ6 حالات مع انتقالاتها

يمكن للمنظومة الموضحة بالشكل (6) أن توصف خلال أي فترة زمنية،كأن تكون موصوفة في واحدة من مجموعة الحالات المتقطعة(N) (Discrete states) (S1،S2،.....،SN).وبالتالي خلال تواجد هذه المنظومة ضمن تلك الحالات المتقطعة، فإنها تخضع إلى تغيرات في الحالة (من الممكن الرجوع إلى الحالة نفسها) وفقاً لمجموعة من الاحتمالات المرتبطة بالحالة. ويرمز إلى الزمن المرتبط بتغير الحالة بـ (t=1,2, ... N)، ويرمز للحالة الحقيقية خلال الزمن (t) بـ (Qt).

إن وصف الاحتمالية بصورة كاملة للمنظومة أعلاه يتطلب وصف الحالة الحالية عند الزمن (t)، فضلاً عن كل الحالات السابقة لها. فينظر إلى سلسلة ماركوف كنوع من مخطط الاحتمالات (Probabilistic Graphical Model) أو طريق لتمثيل الفرضيات الاحتمالية. يمكن القول أن سلسلة ماركوف محددة بالمكونات التالية:

1 - مجموعة N من الحالات وتمثل بـQ={q1,q2,....qN}

2 - المصفوفة الاحتمالية الانتقالية A (transition probability matrix) وتمثل بـ

$$A=\begin{pmatrix} a_{11} & \dots & a_{1N} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{N1} & \dots & a_{NN} \end{pmatrix}$$

حيث أن كل aij تمثل احتمالية الانتقال من الحالة i إلى الحالة j بحيث تحقق الشرط التالي:

$$\sum_{j=1}^{N} a_{ij} = 1 \qquad \forall i$$

3 - حالات خاصة هي حالات البداية q0 وحالة النهاية qf التي لا ترتبط مع أية مشاهدات (Observations).

4 - التوزيع الاحتمالي الابتدائي للحالات (Initial probability distribution)

$$\pi = \pi_1, \pi_2, \dots, \pi_N$$

$$\sum_{i=1}^{N} \pi_i = 1$$

وكذلك تكون الاحتمالية(probability) التي تبدأ بها سلسة ماركوف عند الحالة i في بعض الحالات πi = 0 مما يعني أنه لا يمكن أن تكون الحالة ابتدائية (initial state).

وتعرف فرضية ماركوف بالعلاقة التالية:

$$P(qi\backslash q1 \ldots qi-1) = P(qi\backslash qi-1)$$

حيث أن:

P(qi\q1…qi−1) تمثل احتمالية حدوث الحالة qi عند توفر الحالات q1…qi−1

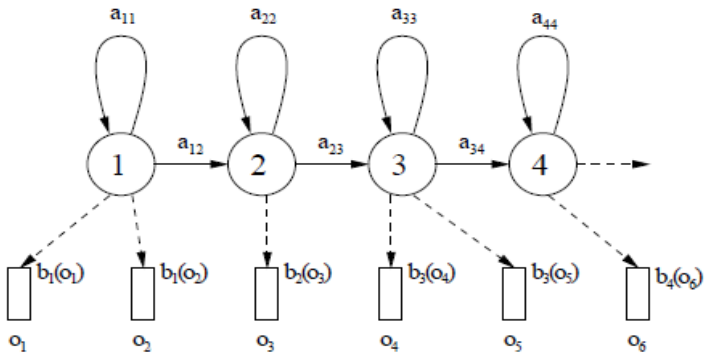
و P(qi\qi−1) تمثل احتمالية حدوث الحالة qi عند توفر الحالة qi−1 التي تسبقها فقط[17]:.

1 4 نموذج ماركوف المخفي Hidden Markov Model

نموذج ماركوف المخفي (HMM) عبارة عن نظام محطات الآلة المحدودة (finite state machine) القادر على توليد مشاهدات باحتمالية انتقال الحالة عند الزمن t التي تعتمد فقط على الحالة السابقة لها عند الزمن t− 1. علماً أن تسلسل الحالة التي تنتج المشاهدة المعطاة مجهول. لذا ففي نموذج ماركوف المخفي تكون الحالة ليست مرئية، لذلك سمي بنموذج ماركوف المخفي والانتقالات بين الحالات تحكمها مجموعة من الاحتمالات يطلع عليها احتمالات الانتقال من حالة معينة والتي يمكن أن تنتج نتيجة أو مشاهدة بحسب توزيع الاحتمالية المرتبط بتلك الحالة. إن الاختلاف بين نموذج ماركوف المخفي ونموذج ماركوف هو وجود الاحتمالات الإضافية والذي يمثل الجزء المخفي للنموذج ويرتبط بالمشاهدة الناتجة من كل حالة. فنموذج ماركوف المخفي هو نموذج تصادفي قادر على التصنيف الإحصائي.

1 5 آلية عمل: HMM

يظهر الشكل (7) نموذج ماركوف المولد حيث يتم اختيار الحالة الأولى وهي الأعلى احتمالية في احتمالية الحالة الابتدائية عند بداية دخول الملاحظات إلى الحالة. يتم حساب احتمالية الانتقال من الحالة الحالية إلى كل الحالات الممكن الانتقال اليها بناء على الملاحظة التي تمت قراءتها حيث يتم الانتقال إلى الحالة التي لديها احتمالية aij أعلى من الحالات الأخريات بحيث نحصل في نهاية هذه العملية على احتمالية انتماء هذه الملاحظة إلى هذه الحالة. و بمعنى آخر احتمالية انتماء هذه الإشارة الصوتية إلى الكلمة. وبذلك يؤخذ القالب ذو الاحتمالية الأعلى على أنه هو الكلمة الصحيحة.

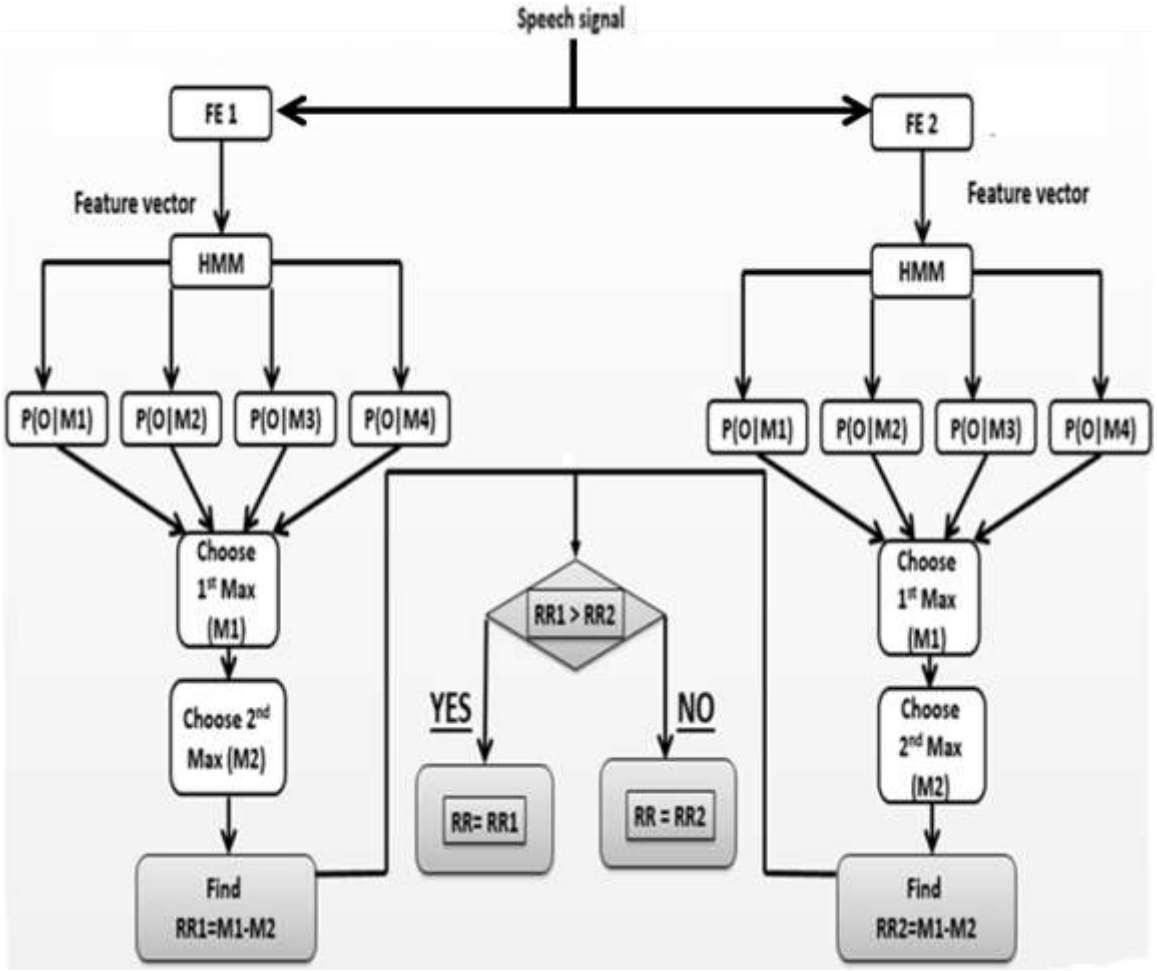


الشكل (7) نموذجماركوفالمولد

خوارزمية الدمج المقترحة بين نظامي تعرف على الكلام:

يمكن تلخيص خطوات الخوارزمية المقترحة للجمع بين نظامين كما هو مبين في الشكل (8) حيث:

- FE (Feature Extraction): استخلاص السمات.
- HMM: المصنف المستخدم مع أنظمة التعرف وهو نموذج ماركوف الخفي.
- P(O|M1)، P(O|M2)، P(O|M3)، P(O|M4) : خرج نموذج ماركوف الخفي وهو عبارة عن مجموعة من التسلسلات الممكنة لعناصر الكلام واحتمالاتها.
- M_1: الاحتمال الأكبر من قيم الاحتمالات الناتجة عن خرج المصنف HMM.
- M_2 : الاحتمال الأكبر الثاني من قيم الاحتمالات الناتجة عن خرج المصنف HMM.
- RR1 (Recognition Rating1): معامل التعرف لنظام التعرف الأول والذي تم اقتراحه ليكون الفرق بين قيمة الاحتمال الأكبر والاحتمال الأكبر الثاني من قيم الاحتمالات الناتجة عن خرج المصنف HMM للنظام الأول.
- RR2 (Recognition Rating2): معامل التعرف لنظام التعرف الثاني والذي تم اقتراحه ليكون الفرق بين قيمة الاحتمال الأكبر والاحتمال الأكبر الثاني من قيم الاحتمالات الناتجة عن خرج المصنف HMM للنظام الثاني.
- RR (Recognition Rate): معدل التعرف النهائي الناتج عن تطبيق خوارزمية الجمع.



الشكل (8)مخطط خوارزمية الجمع بين نظامي تعرف

يتم تنفيذ هذه الخطوات على نظامي تعرف على التوازي:

أولا: مرحلة استخلاصالسمات: يتم على التوازي استخلاصالسمات باستخدام نوعين من الخوارزميات

ثانيا: مرحلة التصنيف، حيث يدخل شعاع السمات المستخرج من المرحلة الأولى ومن كل خوارزمية على حدى إلى المصنف المستخدم وهو HMM.خرج هذا المصنف عبارة عن تسلسل من الكلمات واحتمالاتها.

ثالثا: يتم اختيار القيمة الاحتمالية الكبرى لتسلسل الاحتمالات M1 من خرج مصنف كل خوارزمية على حدى.

رابعا: يتم اختيار ثاني أكبر قيمة احتمالية لتسلسل الاحتمالات M2 من خرج مصنف كل خوارزمية على حدى.

خامسا: يتم حساب الفرق بين أكبر قيمة احتمالية وثاني أكبر قيمة احتمالية لنظامي التعرف:

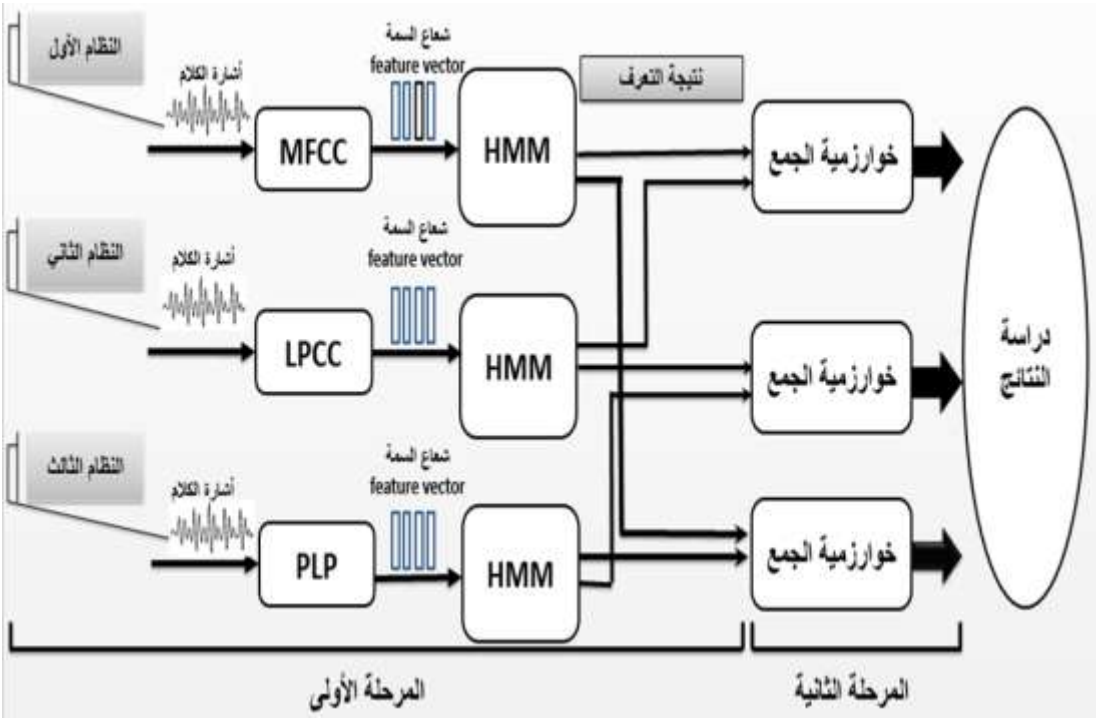
$$RR1=M1-M2, RR2=M1-M2$$

سادسا: تؤخذ قيمة الفرق الأكبر كناتج خرج خوارزمية الجمع.

إن ناتج خوارزمية الجمع (الذي اعتمد نتيجة خرج النظام الذي امتلك قيمة الفرق الأكبر بين أكبر احتمال والاحتمال الأكبر الثاني الذي يليه من سلسلة احتمالات خرج نموذج ماركوف الخفي) يعبر عن زيادة في موثوقية قرار هذا النظاممقارنة مع موثوقية النظام الآخر لذلك تم اعتماد خرجه كخرج لخوارزمية الجمع.

النتائج والمناقشة:

يبين الشكل (9) مخطط صندوقي لمراحل العمل:



الشكل (9) المخطط الصندوقي لمراحل العمل

كما ذكرنا سابقاً تم بناء ثلاثة أنظمة للتعرف على الكلام.

- النظام الأول يعتمد على MFCC في استخراج السمات.
- النظام الثاني يعتمد على LPCC في استخراج السمات.
- النظام الثالث يعتمد على PLP في استخراج السمات.

حيث استخدمت الأنظمة الثلاثة الـHMM في مرحلة التصنيف.

في البداية تم تدريب الـHMM على مجموعة من معطيات التدريبDatasetللأوامر الصوتية التالية:

Shutdown، Restart، Net ، Documents ، المسجلة بواسطة برنامج Audacity، وكل أمر بصوت 33شخص من مختلف الأعمار، حيث تم حساب معدل التعرف لكل نظام على حدى، على الشكل التالي:

$$\text{معدل التعرف} = \frac{\text{عدد العينات الصحيحية}}{\text{عدد العينات الكلية}}$$

ولدراسة أداء خوارزمية الجمع ومدى تكامل زوج من الانظمة تم حساب نوعين من الاخطاء المعتمدة في المراجع العلمية وهي [17]:

1 – الأخطاء التزامنية Simultaneous error: تحدث عندما يرتكب النظامين i و j خطأًفي نفس الوقت

$$Simultaneous\ error = \frac{N_{sim\ (i,j)}}{N_{ref}} \times 100$$

حيث $N_{sim(I,j)}$: عدد الأخطاء التزامنية و N_{ref} : عدد الأوامر الصوتية الكلي.

2 – الأخطاء الاعتمادية Dependent errors: تحدث عندما يرتكب النظامين نفس الخطأ.

$$Dependent\ error = \frac{N_{dep\ (i,j)}}{N_{ref}} \times 100$$

حيث $N_{dep(I,j)}$: عدد الأخطاء الاعتمادي.

تم تطبيق خوارزمية الجمع المقترحة على 132 عينة صوتية (كلامية) دون تحديد فئة عمرية (وهي ناتج نطق الأوامر الصوتية الأربعة السابقة بصوت 33 شخص)، والتي تم تسجيلها بوساطة برنامج audacity، بينما استخدم برنامج الـ Matlab2012aمع المكتبات (voicebox , signal processing)كبيئة عمل لدراسة خوارزمية الجمع المقترحة ومقارنة النتائج.

معدل التعرف

يبين الجدول (1) قيم نسبة التعرف الناتجة بالنسبة للأنظمة المدروسة منفردةً قبل تنفيذ خوارزمية الجمع بالاضافة إلى الانظمة الناتجة عن خوارزمية الجمع .

الجدول (1) نسبة التعرف للأنظمة المدروسة المفردة والمجمعة

الخوارزمية	معدل التعرف
MFCC	85.4426 %
LPCC	78.6885 %
PLP	82.1639 %
MFCC & LPCC	88.5246 %
MFCC & PLP	93.4426 %
LPCC & PLP	86.8852 %

نلاحظ من النتائج الحاصلة تحسن نسبة التعرف بعد تطبيق خوارزمية الجمع بنسب مختلفة أفضلها عند الدمج بين الخوارزميتين MFCC و PLP.

الأخطاء الأعتمادية:

يبين

الجدول (2) قيم الأخطاء الاعتمادية بين كل زوج من الأنظمة المدروسة، حيث يتضح من هذه النتائج أن أفضل تكامل يحصل بين الخوارزميتين MFCC و PLP وذلك لعدم الاشتراك بالأخطاء في نفس الوقت.

الجدول (2) الأخطاءالاعتمادية

	MFCC	PLP	LPCC
MFCC	--	0%	4.92%
PLP	0%	--	3.28%
LPCC	4.92%	3.28%	--

الأخطاء التزامنية:

يبين

الجدول (3) قيم الأخطاء التزامنية بين كل زوج من الأنظمة المدروسة، حيث يتضح من هذه النتائج أن أفضل تكامل يحصل بين الخوارزمية MFCC من جهة وكلاً من الخوارزميتان PLP و LPCC حيث تبين النتائج أن الخوارزمية MFCC تعطي أخطاءً مختلفة مقارنة مع الخوارزميتان PLP و LPCC مما يجعلها مناسبة لعملية الجمع.

الجدول (3) الأخطاء التزامنية

	MFCC	PLP	LPCC
MFCC	--	6.56%	6.56%
PLP	6.56%	--	9.84%
LPCC	6.56%	9.84%	--

الخاتمة:

تعتبر ميزة التعرف على الكلام من المواضيع الهامة والتي لاقت الكثير من الاهتمام لدى الباحثين لما لها من استخدامات واسعة في مجالات الحياة المختلفة. فقد تم تطوير العديد من الخوارزميات والتي أظهرت نتائج الدراسة والمقارنة وجود محاسن ومساوئ لكل خوارزمية. في هذا البحث تم اقتراح خوارزمية تقوم بالاستفادة من محاسن الخوارزميات المقترحة في المراجع العلمية بهدف تحسين معدل التعرف قدر الامكان عن طريق الجمع بين هذه الخوارزميات والاستفادة من اختلافها في النتائج. حيث أظهرت النتائج تحسناً ملحوظاً في نتائج التعرف لكل زوج من الأنظمة المجمعة combined systems على نتائج الأنظمة المفردة. بالإضافة إلى ذلك فقد بينت النتائج تفوق خوارزمية الـ MFCC على الخوارزميتين PLP و LPCC من ناحية تكاملها في عملية الجمع.

لزيادة موثوقية خوارزمية الجمع المقترحة وإعطاءها صفة المعيارية سيتم توسيع هذا البحث بحيث يتم زيادة عدد الطرق المستخدمة في مرحلة استخراج السمات وبالتالي زيادة دائرة المقارنة لاختيار أفضل زوج. بالإضافة إلى زيادة عدد الأوامر الصوتية المختبرة واستخدام قواعد بيانات صوتية عالمية مستخدمة في دراسة الخوارزميات المستخدمة في مجال التعرف على الصوت.

المراجع:

[1] Marius Zbancioc, MihaelaCostin :*using neural networks and LPCC to improve speech recognition*, International IEEE SCS Conference, Proceedings, Vol. 1, 2003 EX 720, pp. 445 – 448.

[2] Levy, C., Linares, G., Nocera, P., Bonastre, J.-F. : *Reducing computational and memory cost for cellular phone embedded speech recognition system*, Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on (Volume:5) , pages(309-12(vol.5 , Print ISBN:9-8484-7803-0.

[3] Dimitriadis, Maragos, P. Potamianos*:Robust AM-FM Features for Speech Recognition*, IEEE signal processing letters, VOL. 12, NO. 9, 2005.

[4] Takami Yoshida, Kazuhiro Nakadai, and Hiroshi G. Okuno: *Automatic Speech Recognition Improved by Two-Layered Audio-Visual Integration For Robot Audition*, 9th IEEE-RAS International Conference on Humanoid Robots December 7-10, 2009 Paris, France.

[5] Lavneet Singh, GirijaChetty*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.

[6] Aleksander Pohl, BartoszZiółko*:Using Part of Speech N-Grams for Improving Automatic Speech Recognition of Polish*, Machine Learning and Data Mining in Pattern Recognition Lecture Notes in Computer Science Volume 7988, 2013, pp 492-504.

[7] BEN FRED , Kaïs OUN *: Phoneme Recognition using Hidden Markov Models* , International Journal of Control, Energy and Electrical Engineering (CEEE) , vol.1, pp.57-61, 2014.

[8] DeividasEringis, GintautasTamulevičius*: Improving Speech Recognition Rate through Analysis Parameters***,** Electrical, Control and Communication Engineering. Volume 5, Issue 1, Pages 61–66, ISSN (Online) 2255-9159, May 2014.

[9] Annika Hämäläinen, Hugo Meinedo, Michael Tjalve, Thomas Pellegrini, Isabel Trancoso, Miguel Sales Dias*: Improving Speech Recognition through Automatic Selection of Age Group – Specific Acoustic Models*, Computational Processing of the Portuguese Language Lecture Notes in Computer Science Volume 8775, pp 12-23, 2014.

[10] HomayoonBeigi: *Fundamentals of speaker Recognition*-Springer Sience 2011-ISBN: 978-0-387-77591-3.

[11] Neustein, Amy; Patil, Hemant: Forensic Speaker Recognition –A. (Eds) – Springer 2012-ISBN 10: 146140262X/ ISBN 13:9781461402626.

[12] K.K PaliWal: *Advances in speech, Hearing and Language Processing*, Volume1, pages 1-78, 1990, ISBN:1-55938-210-4.

[13] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar: *A Review on Speech Recognition Technique*, International Journal of Computer Applications (0975 – 8887), Volume 10– No.3, November 2010.

[14] 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(s): 73 - 76 vol.1

[15] Namrata Dave, *Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition,* international journal for advance research in engineering and technology,Volume 1, Issue VI, July 2013.

[16] H. Hermansky, N. Morgan, A. Bayya, P. Kohn: *RASTA-PLP speech analysis techniqe* ,IEEE International Conference on , 1992 , pp: 121-124.

[17] Lukas Burget :*Measurement of complementarity of Recognition Systems* , Springer-Verlag Berlin Heidelberg ,ISBN 3-540-230421,pages(283-288),2004.