

الخلاصة:

يلعب الصوت البشري دوراً مهماً في مجال التحقق والمصادقة كبصمة حيوية مميزة لكل شخص بشكل فريد حول العالم، إن التعرف على الصوت هي تقنية حيوية مستخدمة للتحقق من أشخاص محددين؛ بحيث تقدم تحسين لنظام الحماية بدون تكلفة مالية كبيرة بل بسيطة نسبياً مقارنةً مع باقي الوسائل الشائعة، من الممكن تطبيق عدة خوارزميات وتقنيات لعمل هذا النظام مثل نماذج ماركوف المخفية (HMM) Hidden Markov Model وخوارزمية تزييف الوقت الديناميكي (DTW) Dynamic Time Warping وخوارزمية معاملات ميل التردد النغمي للطيف المعكوس (MFCC) Mel Frequency Cepstrum Coefficients والمستخدمة في هذا البحث والتي تعمل على توليد معاملات خاصة لأصوات المستخدمين بشكل فريد ولكل مستخدم، قمنا بهذا العمل بعرض الخوارزمية المذكورة كنظام محاكاة عملية التقيد والتحقق من الأشخاص في الدوائر الإدارية والمؤسسات كمثال عليه بيئة الجامعات ونظام تقيد للطلاب من خلال التحدث إلى هذا النظام والذي سيقوم بدوره بتمييزهم، وتمت عملية التدريب على قاعدة بيانات افتراضية من الشبكة عن طريق نماذج الخليط الغاوسي (GMMs) Gaussian Mixture Models باستخدام لغة Python على بيئة VS Code وقمنا باستعراض النتائج.

Abstract:

The human voice plays an important role in the field of verification and authentication as a unique biometric fingerprint for each person around the world. Voice recognition is a vital technology used to verify specific people; In order to provide an improvement to the security system without a large financial cost, but relatively simple compared to the rest of the common methods. It is possible to apply several algorithms and techniques to make this system such as **Hidden Markov Model (HMM)**, **Dynamic Time Warping (DTW)** and **Mel Frequency Cepstrum Coefficients (MFCC)** algorithm that used in this work. We work to generate special coefficients for the voices of users uniquely and for each user. In this work, we presented the mentioned algorithm as a simulation system for the work of the algorithm in the process of checking and verifying people in administrative departments and institutions as an example of the university environment and a presence system for students by talking to this system, which in turn will recognize them. The training was done with random dataset we brought from the internet using **Gaussian Mixture Models (GMMs)** in **Python** language at **VS Code** environment and we reviewed the results.

KeyWords: MFCC, FFT, DCT, Bank Filter, HMM, DTW, GMMs, Speech Recognition, Wave Spectrum.

مقدمة:

تعد معالجة اللغة الطبيعية (NLP (Natural Language Processing مجالاً فرعياً من اللغويات وعلوم الكمبيوتر والذكاء الاصطناعي المعني بالتفاعلات بين أجهزة الكمبيوتر واللغة البشرية، ولا سيما كيفية برمجة أجهزة الكمبيوتر لمعالجة وتحليل كميات كبيرة من بيانات اللغة الطبيعية، ويعتبر الهدف هو جعل جهاز الكمبيوتر قادر على تحليل محتويات الأمواج المسموعة المستقبلية بأجهزة الاستقبال لهذا النوع من الإشارات، بما في ذلك الفروق الدقيقة للغة داخله، وذلك لتمكين الآلة بعد ذلك من استخراج المعلومات والأفكار الواردة في الإشارات بدقة وكذلك تصنيف وتنظيم هذه الإشارات نفسها، ويمكن تقسيم هذا المجال مجازياً إلى عدة أقسام:

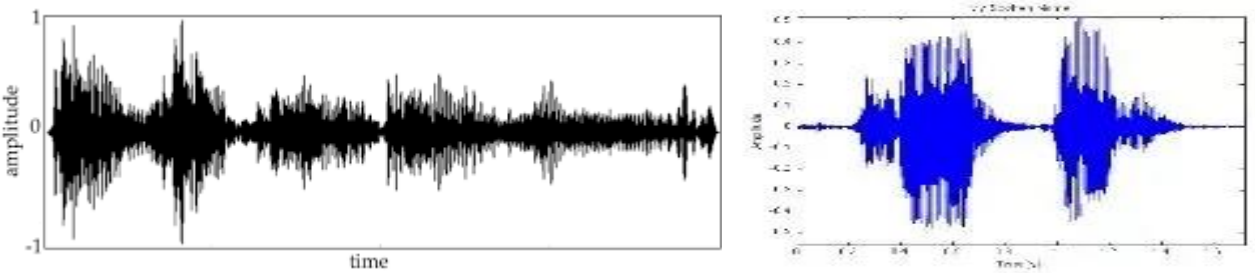
- نماذج إنتاج الصوت وبناءه إلكترونياً.
- تحويل النص إلى صوت والعكس.
- عمليات التحقق من خلال الصوت.
- التعرف على اللغات بشكل عام.

ويعتبر التعلم العميق Deep Learning والذي هو أحد فروع الذكاء الاصطناعي Artificial Intelligence من المنهجيات الرائدة في تحليل الأمواج الصوتية لغاية معالجة اللغات الطبيعية، إذ من الممكن أن نقوم من خلال خوارزميات التعلم العميق بعمل العديد من المشاريع الهادفة، منها:

- التعرف على صوت الآلات في المصانع لمعرفة تشخيص العطل الفني.
- في المجال الأمني التعرف على المجرمين أو المواطنين في الحدود.
- في المجال الطبي الذي يساعد على تشخيص حالة المريض بناء على صوته.
- في المجال العسكري يساعد على التعرف على نوع الأسلحة المستخدمة.
- في المجال الإداري وخاصة للشركات يساعد على التفاعل مع الزبون والعملاء عن طريق بوت الاستقبال.
- في المجال الجيولوجي يساعد على التعرف على أنماط الظواهر الطبيعية للأرض.
- في المجال الخدمي يساعد على قيادة السيارات الذكية والتحكم بها بالأوامر الصوتية للسائق.
- في المجال التقني يساعد على التحقق من أصحاب الأجهزة المستخدمة كالهواتف النقالة والحواسيب وغيرها.

مشكلة البحث وأهدافه:

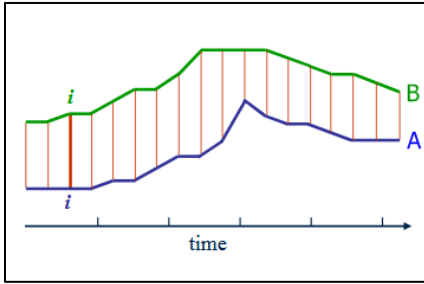
تكمن المشكلة الرئيسية لهذا العمل في أن أنظمة الرقابة والتحقق اليدوية لطالما كانت تعاني من بعض الأخطاء البشرية جراء ضغوطات العمل وغيرها من المشاكل، ففي الجامعات على سبيل المثال يحتاج المحاضر لتسجيل تفقد الطلاب الحاضرين في المساق التعليمي في كل جلسة على حدة، مما يعاني من بعض الإجهاد قبيل البدء بالمحاضرة العلمية، بالإضافة إلى أنه من الممكن لبعض الطلاب التملص من المحاضرة عن طريق بعض زملائهم الذين سيتولون تسجيل حضورهم بطريقة قد لا يستطيع المدرس التأكد منها، كما أن عملية التحقق هذه تأخذ من الوقت الكثير في المحاضرات التي يحضرها عدد كبير من الطلاب، علاوة على أننا في عصر السرعة والتكنولوجيا الرقمية، وحيث أن الأتمتة الحاسوبية قد غزت جميع الدوائر الإدارية والعنصر البشري والطرق اليدوية أصبحت لا تؤدي الغرض الرئيسي من عملها على الوجه المطلوب، بالإضافة إلى الأخطاء الناجمة عنها والوقت المستغرق، فإننا نرى أنفسنا بحاجة لمثل هذا النوع من الأعمال الذي يساعد على تخطي الكثير مما ذكر آنفاً، ولهذا فإننا نحتاج أولاً إلى مواكبة الركب الحضاري الغربي بشكل خاص والعالمي بشكل عام من خلال هذا النوع من الأعمال، ويهدف البحث إلى تطوير النهج الإداري المؤسسي وإدخال الآلة بشكل حضوري أكثر مما هو معمول به حالياً، كما يهدف البحث إلى عمل محاكاة لأنظمة التعرف من خلال تمييز أسماء الأشخاص المراد التعرف عليهم، إذ أننا اعتبرنا أنه لا يجب تمييز اسم الشخص فقط من خلال تحليل الأحرف والكلمات وسياق الكلام، بل يجب تلافي مشكلة انتحال الشخصية التي من الممكن أن تحصل، ولهذا فإننا بحاجة للفظ اسم الشخص من قبل الشخص نفسه ومن ثم التحقق من ذلك، أي أننا نقوم برفض اسم الشخص بدون مطابقة نغمات الموجة الصوتية التي تنتمي للكلمات المذكورة والتي قمنا بتدريبها من خلال مجموعة بيانات فريدة لكلمات اللغة الإنجليزية وبعض أصوات النغمات الموسيقية كالبوب والجاز وغيرها من خلال محاكاة للأصوات البشرية والتي تعتبر عملية أعقد من تحليل الأصوات البشرية كون النغمات الموسيقية أعقد من موجياً من الأصوات البشرية لتداخل عدة منابع إنتاج للموجات الصادرة.



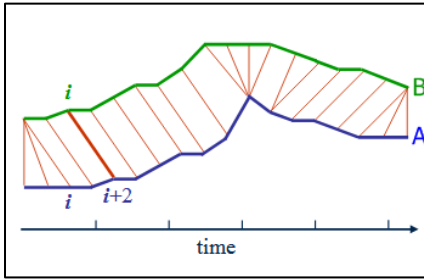
الشكل (1): مقارنة بين موجات النغمات الموسيقية (اليسار) وموجات صوت الإنسان (اليمين)

خوارزمية DTW:

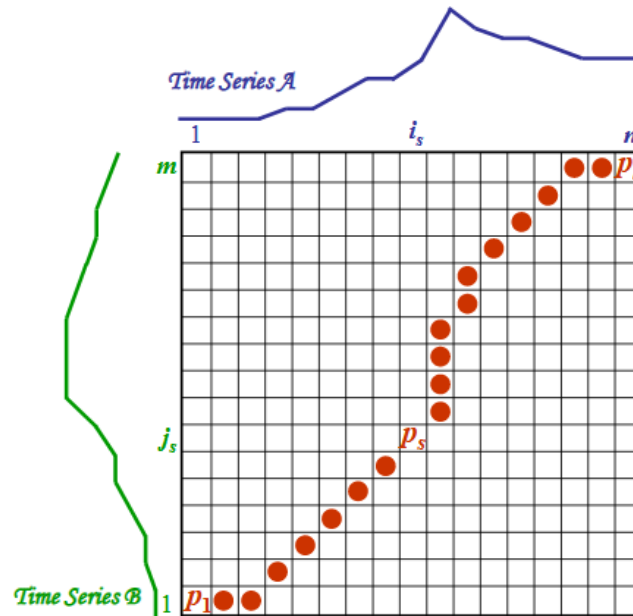
يعد التزييف الزمني الديناميكي (DTW (Dynamic Time Warping أحد خوارزميات قياس التشابه بين سلسلتين زمنيتين مؤقتتين والتي قد تختلف في السرعة، والهدف من طرق مقارنة السلاسل الزمنية هو إنتاج قياس المسافة بين سلسلتين زمنيتين للدخل، وعادةً ما يتم حساب التشابه أو الاختلاف بين السلاسل ثنائية الوقت عن طريق تحويل البيانات إلى متجهات وحساب المسافة الإقليدية بين تلك النقاط في مساحة المتجه، وبالتالي فإن DTW هي خوارزمية حساب التطابق الأمثل بين تسلسلين معينين مع الأخذ بالاعتبار قيود معينة للحساب، وذلك



من خلال قياس المسافة بين السلاسل ذات الوقتين، ويمكن استخدام أي مسافة مثل المسافة الإقليدية أو مسافة مانهاتن أو غيرها بين السلسلة ذات الوقتين للمقارنة ويوضح الشكل المجاور هذا التعبير لهذين المفهومين إذ يتم محاذاة النقطة i في سلسلة زمنية واحدة مع النقطة i على السلسلة الأخرى، وذلك سيؤدي إلى نتيجة تشابه ضعيف بين هذين التسلسلين.



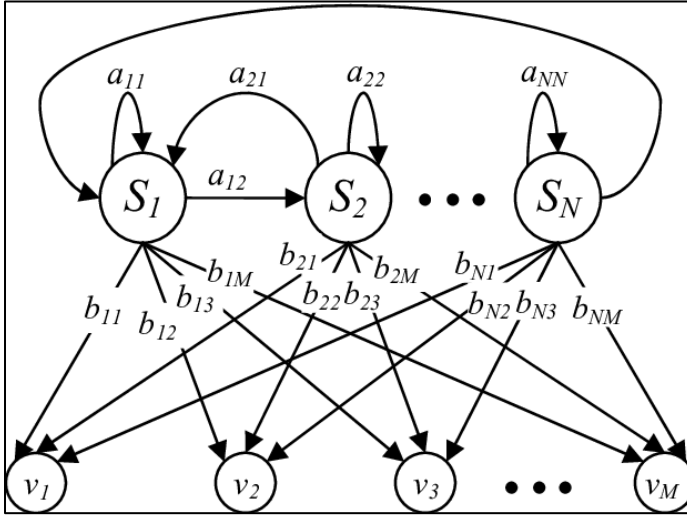
إن DTW تنتج محاذاة غير خطية بين سلسلتين زمنيتين، أي أنها تبحث عن أفضل محاذاة بين المتسلسلة ذات الوقتين بحيث ينتج عن هذا المقياس تشابه أكثر سهولة مما يسمح بمطابقة الأشكال المتشابهة حتى لو كانت خارج الطور في محور الوقت، وهذا ما يوضحه الشكل المجاور.



الشكل (2): يوضح حسابات مصفوفة المسافة لـ DTW

خوارزمية HMMs:

تعتبر نماذج ماركوف المخفية (HMMs (Hidden Markov Models من الخوارزميات الهامة في علوم التعرف على الأنماط، فهي فئة من النماذج الرسومية الاحتمالية التي تسمح لنا بالتنبؤ بسلسلة من المتغيرات (المخفية) من مجموعة من المتغيرات المدروسة، وهي شبكة إحصائية بايزية Statical Bayesian Network غير متغيرة مع الزمن بذاتها وإنما بخطوات الخوارزمية نفسها وما ذلك إلا للتنبؤ بأفضل تسلسل للحالات المخفية، ويعود سبب تسمية حالات الدراسة بالمخفية هو أننا ندرس الحالة الحالية للنموذج بغض النظر عن



تطوراته الماضية لنتنبأ بحالته المستقبلية الغير معروفة (المخفية)، ويوضح الشكل المجاور نموذج تسلسلي احتمالي للنماذج، بحيث نقوم بحساب احتمال وقوع عدة حالات في المرحلة الواحدة والحالات المنبثقة عنها بمرحلة أخرى بغض النظر عن الحالات التي تسبقها بتوزيع احتمالي مجاله $[0,1]$ لكل حالة مرحلية ومن ثم نختار أفضل حالة للنموذج بناءً على ما هو مطلوب من خلال احتمالية الانتقال لحالة جديدة في المرحلة اللاحقة بشرط تحقق الحالة الحالية، واحتمالية الانتقال إلى حالة مرصودة مرتبطة بحالة مخفية، والاحتمال الأولي للانتقال إلى حالة مخفية، وتوضح المعادلات التالية المفاهيم السابقة رياضياً كما يلي:

The probability of observing a sequence

$$Y = y(0), y(1), \dots, y(L-1)$$

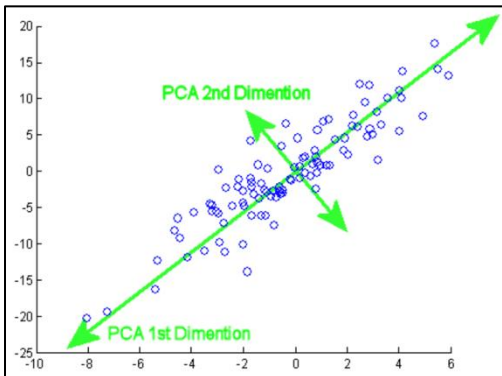
of length L is given by

$$P(Y) = \sum_X P(Y | X)P(X),$$

where the sum runs over all possible hidden-node sequences

$$X = x(0), x(1), \dots, x(L-1).$$

كما أن هناك العديد من الخوارزميات الأخرى التي تستخدم في تحليل المدخلات إلى عينات ولكل منها خصائصه



وميزاته ومساوئه كخوارزمية تحليل المكونات الرئيسية PCA

(Principle Components Analysis) التي تعتمد بشكل

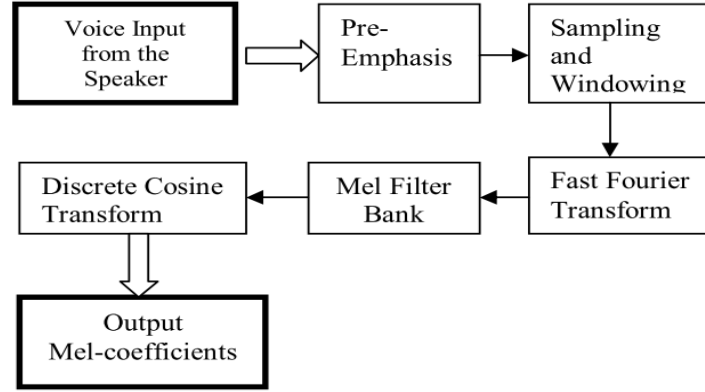
رئيسي على تقليل أبعاد المدخلات مع مراعاة الحد الأدنى من فقدان

المعلومات والميزات الهامة وذلك لتقليل زمن تحليل المدخلات

ومساحة تخزين النتائج المرحلية لها كما يظهر الشكل المجاور.

خوارزمية MFCC:

إن خوارزمية معاملات الميل الترددي للطيف المعكوس **Mel Frequency Cepstrum Coefficients (MFCC)** تعتبر خوارزمية تحليلية جيدة للإشارات؛ إذ تتعامل الخوارزمية مع الصوت المدخل كعينات للمعالجة، وهي تقوم بحساب خاص لكل عينة صوتية فريدة لاستخراج ميزاتها؛ بحيث نحصل على 39 ميزة لكل عينة صوتية من هذه التقنية، وعدد الميزات هذا كافٍ لمعرفة بيانات العينة الفريدة بحيث نأخذ 13 ميزة منها مرتبطة بمطال (سعة) التردد، ويتم تجميع هذه الميزات ضمن تابع (معامل) واحد يكون تابع ميل التردد النغمي للطيف المعكوس للعينة ومن ثم نقوم بتجميع هذه العينات بتردد واحد لنحصل على المعاملات للصوت الكامل المدخل على العملية، وقد تم تسمية الخوارزمية بهذا الاسم نظراً لأن الميزات المستحصلة من الخوارزمية شبيهة بالميزات الناتجة عن الطيف المعكوس cepstral.



الشكل (3): يوضح آلية عمل الخوارزمية بشكل مبسط

خطوات عمل الخوارزمية:

1. استقبال إشارة الصوت Input:

تقوم الخوارزمية باستقبال الصوت من المستخدم بتقنيات الاستقبال المعروفة كالميكروفون أو كملف جاهز من الشبكة أو الحاسب.

2. الترشيح الأولي Pre Emphasis/Pre Processing:

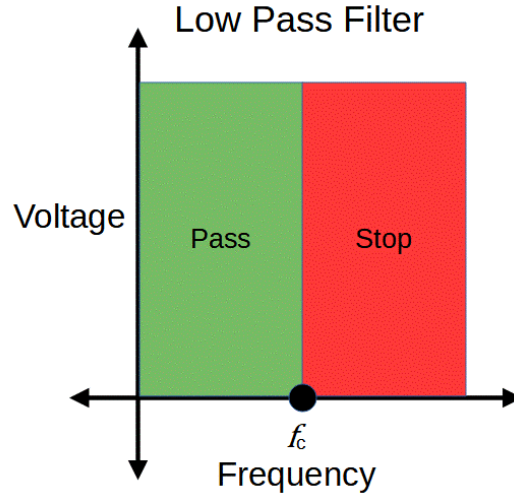
نقوم باستقبال الإشارة وتصفيتهَا بمرشح التمرير المنخفض (Low Pass Filter) بناءً على العلاقة أدناه:

$$y(n) = x(n) - a * x(n - 1)$$

بحيث يكون خرج الإشارة y هو دخل المرحلة التالية.

وقد تكون a عادةً بين $0.9 - 1.0$ وهي ثابت تحول للإشارة.

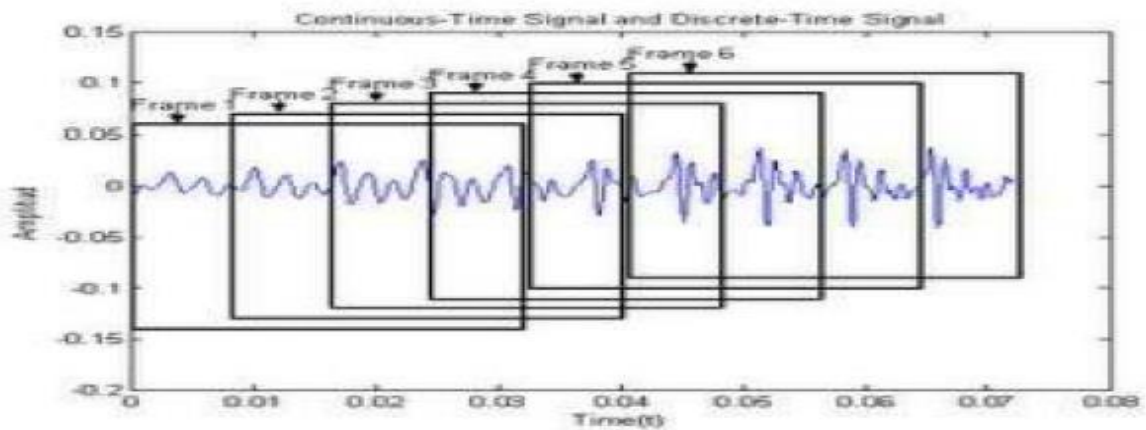
الغرض من عملية الترشيح هذه هي تضخيم الإشارة وتعويض الجزء المنخفض منها الناتج عن ضعف دقة اللفظ أو تدني مستوى الصوت للمتكلم.



الشكل (4): يوضح الرسم البياني لمرشح التمرير المنخفض

3. تقطيع الإشارة (Sampling):

يتم تقطيع الإشارة المدخلة إلى إطار واحد بمعدل 15-20 ms مع تداخل هذه الإطارات مع بعضها البعض بنسبة 50% من حجم الإطار (أي أن الإطار يتشارك بالبيانات مع من يليه بمقدار النصف)، في حالة الإشارات الضعيفة يكون عادة حجم الإطار مرفوعاً للقوة 2 لتسهيل عملية تحويل فورييه السريع وفي حالة الإشارات القوية لسنا بحاجة لتقوية الإشارة بهذه العملية، مثال: ليكن لدينا إشارة بتردد 16 kHz وحجم الإطار 256 نقطة تقطيع عندئذ زمن التقطيع هو: $256/16000 = 0.016 \text{ sec} = 16 \text{ ms}$ بالإضافة إلى أن نسبة التداخل هي 50% أي أن 128 نقطة ضمن الإطار الواحدة مشتركة مع الإطار الذي يليه فعندئذ يكون عدد الإطارات الناتجة في الثانية الواحدة هي: $16000/(256-128) = 125 \text{ FPs}$ ، ومنه نستنتج أن عملية التداخل تستخدم لإنتاج إطارات مستمرة والمغزى منها هو الحفاظ على تراكم البيانات بشكلها الصحيح لضمان دقة عملية التعرف وتقلل من زمن التنفيذ.

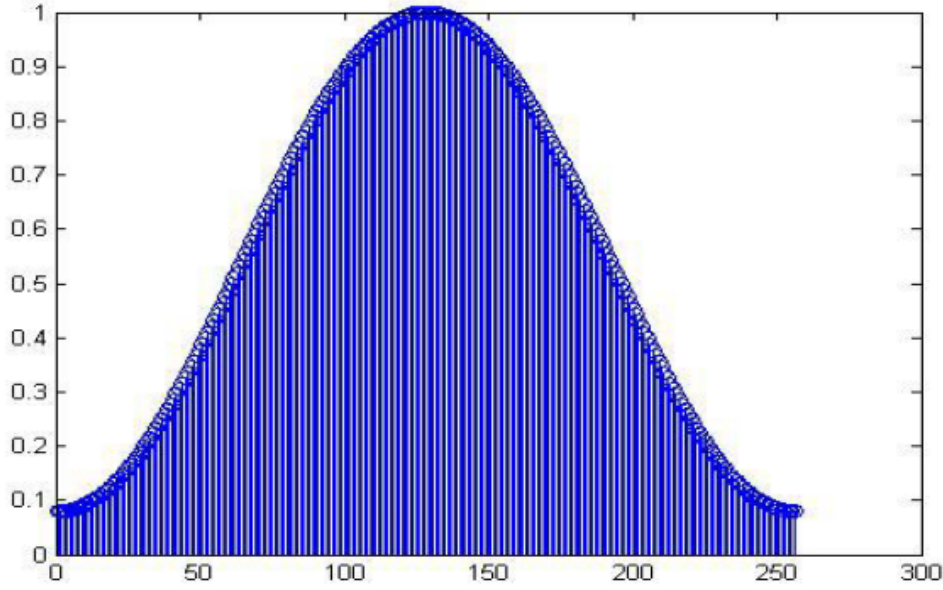


الشكل (5): يوضح عمليتي تقطيع وتداخل الإشارة

4. نافذة هامينغ Hamming Window:

لحفظ الاستمرارية للإطارات الناتجة من النقطة الأولى إلى النقطة الأخيرة في الإطار الواحد نستخدم نافذة هامينغ والتي تضمن أيضاً مضاعفة الإطارات وتكرارها كما شرح مسبقاً، التوصيف الرياضي لهذه العملية هو الانطواء للإشارة المدخلة مع نافذة هامينغ كالتالي:

$$x(n) * w(n); w(n) = 0.54 - 0.46 \cos[2\pi/(N-1)]; 0 \leq n \leq N-1.$$



الشكل (6): يوضح بيانياً نتائج مخرجات نافذة هامينغ

5. تحويل فورييه السريع (FFT): Fast Fourier Transform:

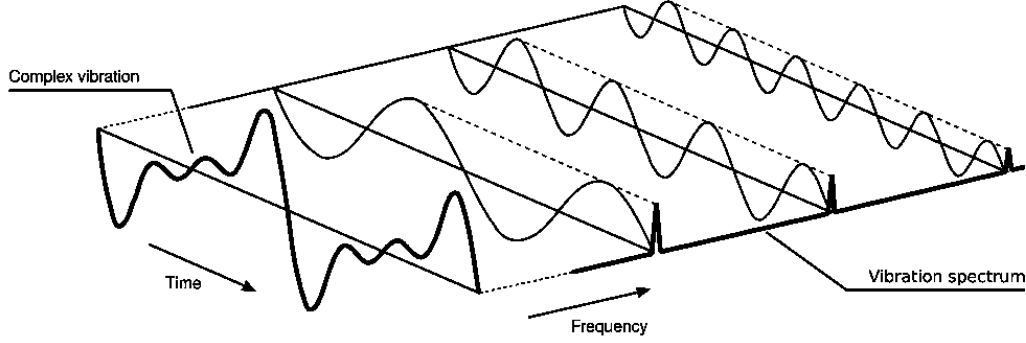
إن التحليل الطيفي للإشارة يظهر مدى توافق النغمات المختلفة مع توزيعات الطاقة المختلفة في التردد، لذلك فإن تحويل فورييه السريع كما هو معروف يظهر الاستجابة الترددية لمطال كل إطار على حدة، وتكون علاقة فورييه بالشكل التالي:

$$S_i(k) = \sum_{n=1}^N s_i(n)h(n)e^{-j2\pi kn/N} \quad 1 \leq k \leq K$$

عندما يتم تطبيق تحويل فورييه السريع على إطار محدد فإنه بالافتراض تتم العملية على أن هذا الإطار هو دوري ومستمر لخرجها الناتج عن عملية الالتفاف convolution، ولكن لدينا مشكلة هنا وهي أن النقطة الأولى والأخيرة لا تدخلان ضمن شروط تحويل فورييه السريع للإشارة المستمرة إذ أنه من المحتمل بعد تطبيق العملية أن تحدث تأثيرات غير مرغوب بها في استجابة التردد؛ ولهذا نلجأ لحل هذه المشكلة إلى تطبيق نافذة هامينغ على النقطة الأولى والأخيرة لضمان استمرارية تدفق الإطارات وتُعطى القدرة الطيفية للإطار (السعة الترددية) بالعلاقة التالية:

$$P_i(k) = \frac{1}{N} |S_i(k)|^2$$

تذكير مهم: مقياس الميل هو علاقة ربط بين الصوت المدرك مع ترددها الفعلي المقاس؛ بحيث يجعل ميزاتنا المتحصلة تتطابق بشكل وثيق مع ما يسمعه البشر؛ وذلك لأن البشر أفضل بكثير في تمييز التغيرات الصغيرة في النغمة عند الترددات المنخفضة مقارنة بالترددات العالية.



الشكل (7): تمثيل صوري لتحويل فورييه السريع

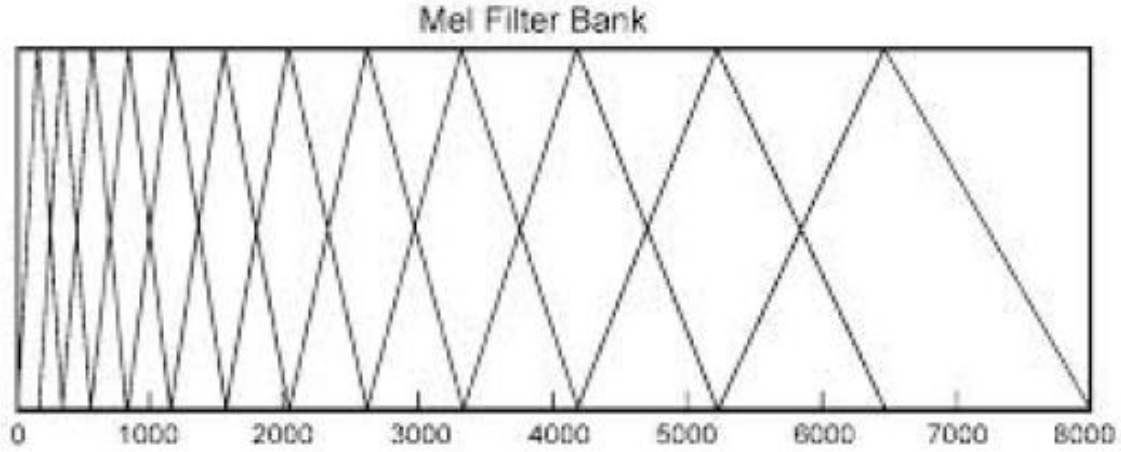
6. مرشحات تمرير النطاق المثلثية /Triangular Band Pass Filters/ مرشح البنك Bank Filter:

يتم في هذه المرحلة ترشيح تابع الاستجابة الترددية 39 مرة بمرشح تمرير النطاق المثلثي للحصول على الطاقة اللوغاريتمية لهذه الإطارات الناتجة تحضيراً للعملية اللاحقة، تتوضع هذه المرشحات على طول تردد الميل للإشارة، وبشكل أوضح إن موضع الترددات المركزية (قبل المركز) بين 1 – 133.33 kHz يكون فيها 13 مرشح خطي متداخل بنفس النسبة المذكورة (50%) بينما بالنسبة للمركز الترددات بين 1 – 8 kHz يكون لدينا 26 مرشح متداخل ولكن متوضع بنسب لوغاريتمية وذو طاقة عالية تقلل من دقة الخوارزمية، ويمكن توصيف تابع توضع المرشحات بحيث تحقق أكبر دقة ممكنة للخوارزمية على الشكل التالي:

$$f(i) = \text{floor}((nfft+1)*h(i)/\text{samplerate})$$

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

وتُوضح معادلة مرشح البنك كما هو مجاور:



الشكل (8): يوضح خرج مرشح البنك

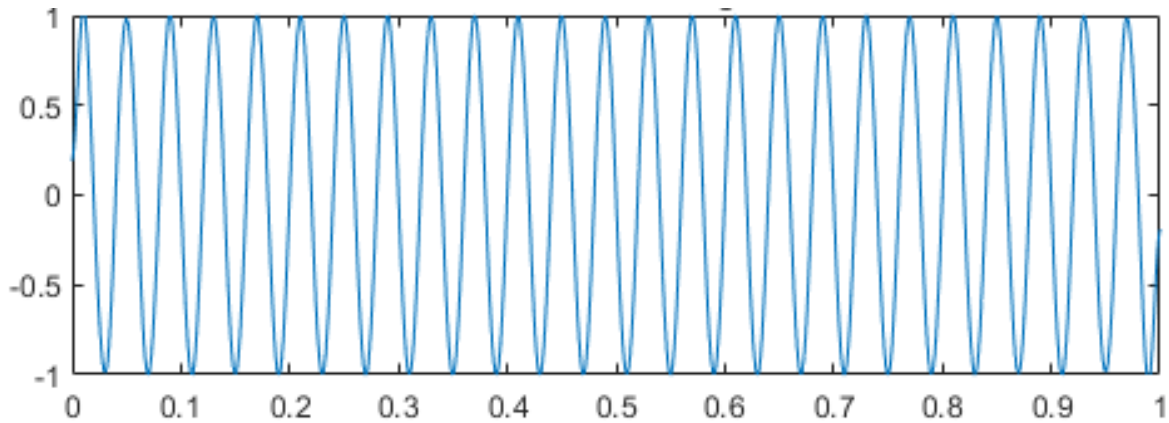
7. تحويل فورييه المتقطع العكسي (IDFT):

في هذه الخطوة سوف نقوم بتطبيق تحويل التجيب المتقطع (Discrete Cosine Transform (DCT)) على خرج المرشحات في العملية السابقة للحصول على معاملات الطيف المعكوس لتدرج الميل، تعطى علاقة التحويل بالصيغة التالية:

$$C(n) = \sum E_k \cos [n * (k - 0.5) * \pi/40]; n = 0, 1, 2, \dots, N.$$

N: هي عدد مرشحات البنك المستخدمة في التطبيق.

نظرا لأننا أجرينا تحويل فورييه السريع (FFT (Fast Fourier Transform)، فإن DCT تقوم بتحويل مجال التردد للإشارة إلى النطاق الزمني وهذه العملية يطلق عليها الطيف العكسي cepstrum؛ بحيث تم أخذ أول أربعة حروف من الكلمة spectrum وقلبها تعبيراً عن العملية الرياضية المطبقة حقيقة بمعنى المعكوس، ومنه نستنتج معامل التردد النغمي للطيف المعكوس لكل عينة صوتية مدخلة.



الشكل (9): يوضح خرج عملية تحويل تجيب التمام القطعي DCT

يمكن شرح عملية الطيف المعكوس كالتالي:

$Ceps = FT^{-1} [\log (FT)the\ signal + 2jIIM]; M:$ القسم التخيلي لتابع اللوغاريتم

القسم الحقيقي للتابع Ceps يعبر عن شدة وطور إشارة الطيف الرئيسية، أما القسم التخيلي يعبر عن قيمة الزاوية لمثلث الإشارة الناتج عن العملية السابقة في الخطوة (6).

نقوم بحساب تابع استجابة النبضة للإشارة من المعادلة:

$$s(n) = e(n) * d(n).$$

ولكن عند حساب جداء الانطواء يصبح من الصعب جداً فصل هذين الجزأين وخاصة بعد أخذ تحويل فورييه لهما واستنتاج طيفهما؛ ولكن بأخذ الطيف المعكوس يصبح التعامل معهما أسهل كما يلي:

$$Ceps = FT^{-1} \{ \log (FT[s(n)]) \}.$$

لكن عند تحويل الإشارة من المجال الزمني إلى الترددي يصبح جداء الانطواء كالتالي:

$$S(n) = E(n) * q(n).$$

عندئذ يصبح الانطواء بعد استخدام خاصية اللوغاريتم بتحويل ما داخله من جداء إلى مجموع اللوغاريتمات:

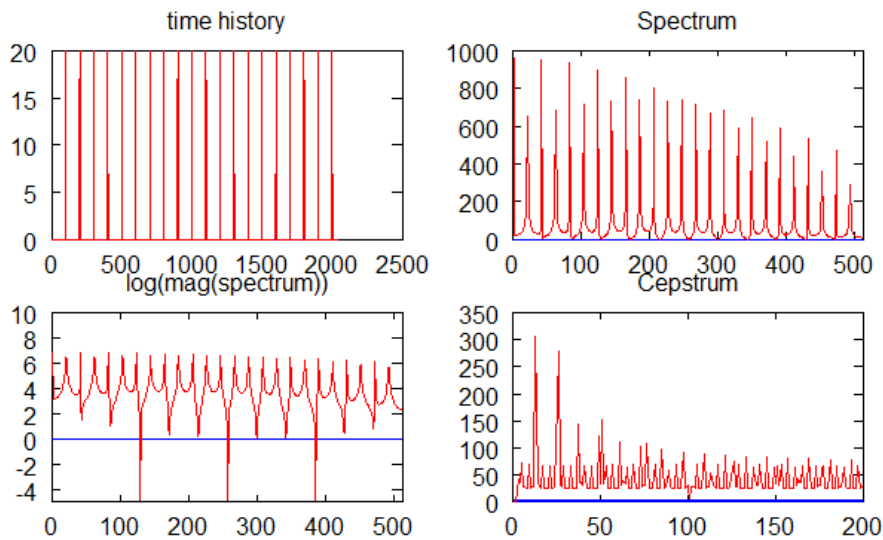
$$\log[S(n)] = \log [E(n) * q(n)] = \log[E(n)] + \log[q(n)] = Ceps_e + Ceps_d.$$

عندها يقوم تحويل فورييه العكسي بالتعامل مع كل متغير على حدة كالتالي:

$$C' = FT^{-1} (Ceps_e + Ceps_d) = FT^{-1} (Ceps_e) + FT^{-1} (Ceps_d) = C'_e + C'_d.$$

حيث تنتج الإشارة بالمجال الترددي لها.

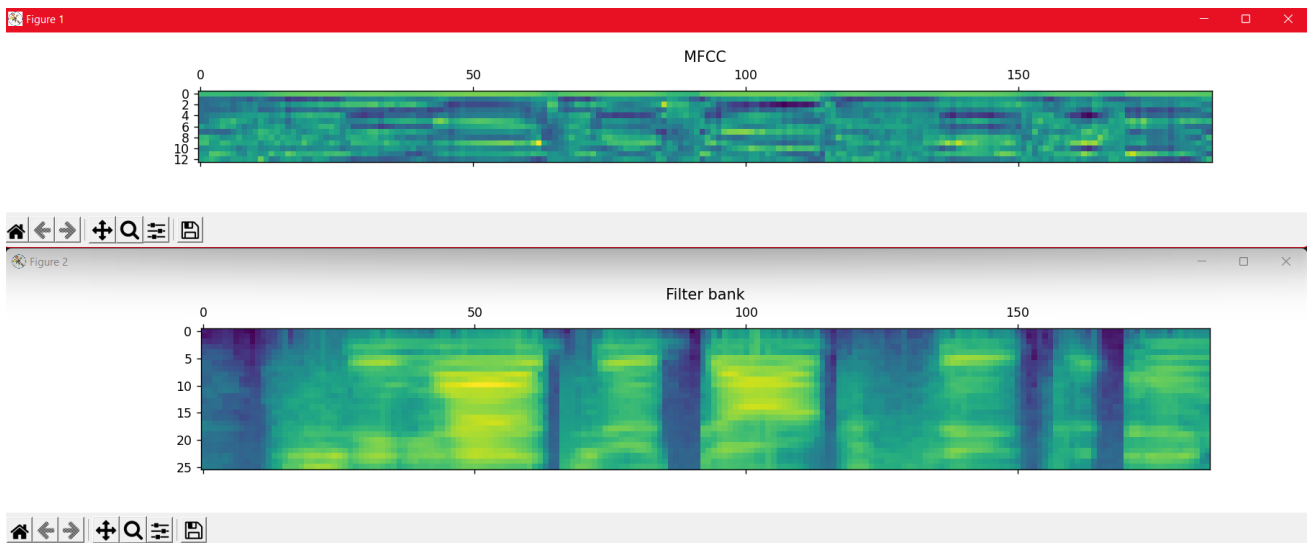
تنويه: قد يتساءل البعض ما علاقة تحويل التجيب المنقطع بتحويل فورييه المنقطع، والجواب هو أن تابع تحويل فورييه المنقطع هو ذاته التجيب المنقطع بعد تحويل علاقة أولر الشهيرة على المعادلة بتحويل التابع الأسّي e إلى تابعين مثلثيين هما Sin, Cos وبالتالي فإن تحويل التجيب المنقطع سيكون للقسم الحقيقي وتحويل فورييه سيكون الأشمل للقسم الحقيقي والقسم التخيلي.



الشكل (10): يوضح الرسوم البيانية لخرج الإشارة لكل من المجال الزمني والترددي واللوغاريتمي والطيف المعكوس

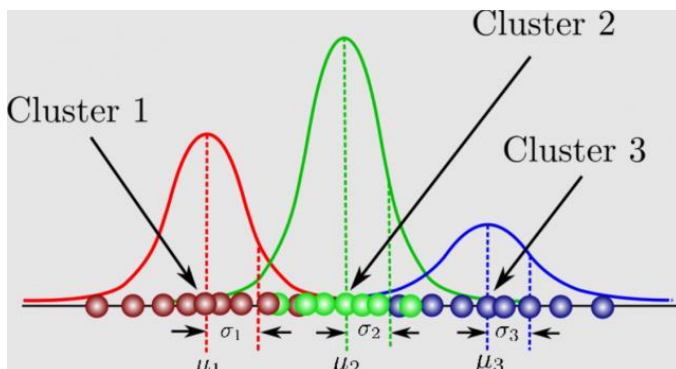
النتائج:

قمنا باستخراج ميزات النموذج باستخدام خوارزمية MFCC من عدة أصوات موسومة بأسماء افتراضية، كما قمنا بتدريب النموذج باستخدام خوارزمية GMM، واستغرقت فترة التدريب والاختبار مدة خمس ساعات متواصلة، حيث قمنا بتدريب النموذج على مجموعة البيانات بمقدار 70% واختبارها على 30% المتبقية، وكانت النتائج أن قام النظام بتسجيل بعض الفروقات بين الاختبار والتدريب بالنسبة لصوتيتين فقط؛ حيث تم تسجيل درجتي خطأ من عشر درجات (أي 20%) كأعلى درجة خطأ في النموذج، أي أن النموذج يحتوي على نسبة دقة 80% تقريباً، وفيما يلي رسم بياني توضيحي لمعاملات الطيف المعكوس ومرشح البنك لصوتية واحدة فقط.



خوارزمية GMMs:

تعد نماذج الخليط الغاوسي (GMMs (Gaussian Mixture Models الشكل الأعم لنماذج التوزيع الاحتمالي الغاوسية، ويستخدم لنمذجة مجموعات البيانات في العالم الحقيقي، ويمكن استخدامها لتمثيل أي مجموعة بيانات في توزيعات غاوسية متعددة، ويفترض هذا النموذج أن جميع نقاط البيانات يتم إنشاؤها من توزيعات غاوس مع معاملات غير معروفة، وتقوم هذه النماذج بتجميع مجموعة من نقاط البيانات في مجموعات حيث قد لا تكون هذه



المجموعات محددة بشكل واضح في مجموعة البيانات، كما أنها تستخدم لعملية التصنيف وتقدير احتمالي لمجموعة بيانات مدروسة، وتتميز بالدقة بالنسبة للبيانات الشاذة، ويوضح الشكل المجاور مفهوم التوزيع الطبيعي لنماذج خليط غاوسي.

الخاتمة:

إن إدخال خوارزميات الذكاء الصناعي وتعلم الآلة على المؤسسات الإدارية من شأنه رفع المستوى الإداري وتعزيز الجانب الأمني لهذه الدائرة فباستخدام هذا النوع من الخوارزميات يمكننا تلافي العديد من المشاكل والأخطاء الغير مرغوب بها، وبناءً على ما تم ذكره سابقاً فإن خوارزمية MFCC تعتبر من الخوارزميات التحليلية العميقة وتستخدم لتحليل العديد من أنواع الإشارات كالاhtزازات الناجمة عن الحركة الداخلية لمكونات الأرض، والإشارات البيولوجية والإشارات المنبعثة من الفضاء العميق والمستقبلية بأجهزة الرصد الفلكي، ولها عدة نماذج استخدام؛ كاستخدام مرشح التمرير العالي فقط بمرحلة الترشيح في حال كانت الإشارات المدروسة عالية التردد، أو مرشحات التمرير ذات النطاقين أو تكرار استخدام كل من مرشحات التمرير العالي والمنخفض والمتوسط ضمن كافة مراحل الترشيح بعدد من طبقات الترشيح، وأثبت هذا النوع من المشاريع فعاليته الكبيرة في تلبية متطلبات التحقق، كما عمل على تحقيق التطور التكنولوجي المطلوب لمواكبة التقدم التقني العالمي.

المراجع:

- https://www.researchgate.net/publication/224217606_A_novel_approach_for_MFCC_feature_extraction
- <https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9>
- [https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab#:~:text=Hidden%20Markov%20Models%20\(HMMs\)%20are,that%20someone%20wears%20\(observed\)](https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab#:~:text=Hidden%20Markov%20Models%20(HMMs)%20are,that%20someone%20wears%20(observed))
- <https://medium.datadriveninvestor.com/dynamic-time-warping-dtw-d51d1a1e4afc>
- <https://github.com/AnushreeJoshi/Speech-Recognition>
- <https://medium.com/swlh/ml-gmm-em-algorithm-647cf373cd5a>

الملحق:

plot_MFCC_features:

```
import matplotlib.pyplot as plt          # استيراد مكتبة الرسم
from scipy.io import wavfile             # استيراد مكتبة معالجة البيانات الصوتية
from python_speech_features import mfcc, logfbank # استيراد مكتبة الخوارزمية المستخدمة

# تحديد ملف القراءة للرسم وضبط التقطيع
frequency_sampling, audio_signal = wavfile.read("C:/Users/moust/OneDrive/Desktop/أحمد المنصور/Code/Dataset/train_wavdata/fajw0/1.wav")

audio_signal = audio_signal[:15000]

# رسم معاملات MFCC
features_mfcc = mfcc(audio_signal, frequency_sampling)

print ("\nMFCC:\nNumbers of windows = ', features_mfcc.shape[0])
print ('Length of each feature = ', features_mfcc.shape[1])

features_mfcc = features_mfcc.T

plt.matshow (features_mfcc)
plt.title ('MFCC')

# خرج مرشح البنك في المجال الترددي
filterbank_features = logfbank (audio_signal, frequency_sampling)

print ("\nFilter bank:\nNumber of windows = ', filterbank_features.shape[0])
print ('Length if each feature = ', filterbank_features.shape[1])

filterbank_features = filterbank_features.T

plt.matshow (filterbank_features)
plt.title ('Filter bank')
plt.show()
```

spk_rec:

```
from python_speech_features import mfcc
import os
from os import listdir
from os.path import isfile, join, isdir
import numpy as np
from sklearn.mixture import GaussianMixture # استيراد مكتبة خوارزمية GMM
from sklearn.linear_model import LogisticRegression
import scipy.io.wavfile as wav
from scipy import stats

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

# تحديد ملفات التدريب والاختبار
train_path="pathe of training files"
test_path="path of testing files"
vec_length=13 # تحديد عدد معاملات MFC

# تابع إنشاء ميزات الخوارزمية
def createMFCC(path):
    (rate,sig)=wav.read(path)
    mfcc_feat=mfcc(sig,rate, winlen=0.025, winstep=0.010)

    return mfcc_feat.reshape(-1, 1)

# إنشاء نموذج GMM
gmm= []

#-----تدريب النموذج-----
person_files=[f for f in listdir(train_path) if isdir(join(train_path,f))]
print (person_files)

for i in range(0,len(person_files)):
    audioPath=join(train_path,person_files[i])
    audioFiles=[l for l in listdir(audioPath) if isfile(join(audioPath,l))]

    mfcc_temp=np.empty((0,vec_length)).reshape(-1, 1)
    for j in range(0,len(audioFiles)):
        filePath=join(audioPath,audioFiles[j])
        temp_feat=createMFCC(filePath)
```


mfcc_temp=np.append(mfcc_temp,temp_feat,axis=None) # تحضير جميع ميزات
واحد GMM الخوارزمية لتدريبها على نموذج

gmm.append(GaussianMixture(n_components=10,covariance_type='full',max_iter=
20, random_state=None)) # تدريب النموذج
gmm[i].fit(mfcc_temp.reshape(-1, 1))

#-----اختبار النموذج-----
test_files=[f for f in listdir(test_path) if isdir(join(test_path,f))]

print ("The number of labels identified correctly are:")
for i in range(0,len(test_files)):
audioPath=join(test_path,test_files[i])
audioFiles=[l for l in listdir(audioPath) if isfile(join(audioPath,l))]

res=[]
print ("current label",i)
mfcc_temp=np.empty((0,vec_length))
for j in range(0,len(audioFiles)):
filePath=join(audioPath,audioFiles[j])
temp_feat=createMFCC(filePath) #إنشاء اختبار لميزات الخوارزمية

num_features=temp_feat.shape[0]

labels=np.empty((num_features,1))
for k in range(0,num_features):
y=[None]*22
for l in range(0,22):
y[l]=np.exp(gmm[l].score_samples(temp_feat[k].reshape(-1, 1))) # تسجيل
درجة التحقق من كل ميزة

max_class=y.index(max(y)) # وضع أعلى درجة تحقق كخرج للاختبار
labels[k]=max_class

res.append(stats.mode(labels)[0]) # mfcc ملف واحد يتم إعطاء فئة الوضع لجميع ميزات
على أنها الفئة المتوقعة لهذا الملف

رقم الخرج لكل توقع صحيح بين الأشخاص#
correct=0
for x in range(0,2):
if(res[x]==i):
correct=correct+1
print (correct)



جامعة الشام
SHAM UNIVERSITY

تم بفضل الله