



جامعة الشام برنامج الماجستير في كلية الهندسة قسم المعلوماتية



التعرف على الصوت باستخدام خوارزمية MFCC وشبكات الذكاء الصناعي

إعداد الطالب: م. مصطفى عفارة
إشراف: د.م محمد أمين الزبداني

٢٠٢٠-٢٠٢١

نظام التعرف على الصوت باستخدام خوارزمية MFCC

الخلاصة:

يلعب الصوت البشري دور مهم جدا في مجال التحقق والمصادقة كبصمة حيوية مميزة لكل شخص بشكل فريد حول العالم، التعرف على الصوت هي تقنية حيوية مستخدمة للتحقق من أشخاص محددين؛ بحيث تقدم تحسين لنظام الحماية بدون تكلفة مالية بسيطة نسبيا مقارنة مع باقي الوسائل الشائعة، من الممكن تطبيق عدة خوارزميات وتقنيات لعمل هذا النظام مثل نماذج ماركوف المخفية (HMM) Hidden Markov Model و خوارزمية معامل ميل التردد النغمي للطيف المعكوس Mel Frequency Cepstral Coefficient (MFCC) المستخدمة في هذا البحث والتي تستخدم تقنية استخراج الميزات Features extracting لتميز الصوت عن طريق توليد معاملات خاصة لأصوات المستخدمين بشكل فريد لكل مستخدم، قمنا بهذا البحث بعرض الخوارزمية المذكورة وتحليل نظام عمل للخوارزمية لعملية تفقد الأشخاص في الدوائر الإدارية والمؤسسات كمثال عليه بيئة جامعة ونظام تفقد للطلاب وقمنا برسم مخططاته وتطبيقه باستخدام بيئة الفيجوال استوديو بلغة البايثون واستعراض النتائج.

الكلمات المفتاحية: الصوت، البصمة الحيوية، استخراج الميزات، التحقق، MFCC.

المقدمة:

يتضمن الصوت ما يقوله الناس وكيفية التحدث أيضا وهذان العاملان يستخدمان لعملية التحقق، إن تقنيات التحقق الأخرى كالتحقق عن طريق بصمة اليد أو الوجه أو القزحية أو شبكية العين أو طريقة المشي أو الكتابة إضافة إلى كونها وسائل للتحقق فريدة لكل بشري مميز وغير مكرر إلا أنها تفتقر إلى المصادقة مقارنة بالتحقق عن طريق الصوت بشكل تطبيقي؛ فبشكل مبسط من الممكن استخدام التعرف على الصوت لكتابة بحث أو تفريغ صوتي أو ترجمة أو إصدار أوامر وما إلى ذلك وهو الأمر المميز لهذه التطبيقات مقارنة مع تطبيقات التحقق الأخرى، عوضا عن كونه أسهل تطبيقيا وأقل تكلفة مقارنة مع أقرانه من وسائل التعرف إذ أنه لا يحتاج سوا إلى مايكروفون وحاسب مبرمج بالخوارزمية المذكورة.

إن معالجة الكلام الرقمي هو أحد فروع علوم الحاسب الواسعة ويمكننا مجازيا تقسيم هذا العلم إلى عدة أقسام:

1. نماذج إنتاج الصوت وبنائها إلكترونيا: هذا المجال يهتم بدراسة صوت الإنسان وكيفية بنائه ومحاكاته إلكترونيا.

2. تشفير الكلام: الغرض من التشفير هنا هو تقليل المساحة التخزينية أو عرض حزمة النقل للبيانات.
3. تحويل النص إلى صوت أو بالعكس: الأمر المعني بعمليات التفاعل بين الحاسب والمستخدم لتحقيق غرض معين وخاصة للمكفوفين.

4. التحقق من المتكلم: يتم تعريف الحاسب بالمتكلم عن طريق دراسة عينة كلامه.

إن عملية التعرف على الشخص بعلم الإشارة يعتمد على استحصال الخواص الطيفية مثل التردد الأساسي وطاقة الطيف، وتتم هذه العمليات عن طريق مرشحات الإشارة بكميات معينة وتموضعات هذه المرشحات ضمن دائرة المعالجة لتنتج الخرج المقصود بحيث يتم توزيعها عن طريق معاملات التنبؤ الخطي المشفر أو التنبؤ غير الخطي.

استخدامات الخوارزمية:

من الممكن تطبيق الخوارزمية في عدد لا حصر له من التطبيقات والمشاريع (الأمر الذي يلعب دورا مهما في العالم الذكي المعاش)؛ مثل:

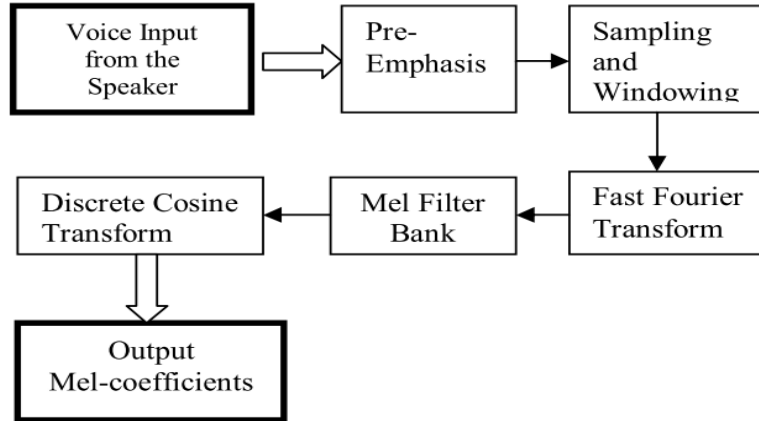
- 1- التعرف على صوت الآلات في المصانع لمعرفة تشخيص العطل الفني.
- 2- في المجال الأمني التعرف على المجرمين أو المواطنين في الحدود.
- 3- في المجال الطبي الذي يساعد على تشخيص حالة المريض بناء على صوته.
- 4- في المجال العسكري يساعد على التعرف على نوع الأسلحة المستخدمة.
- 5- في المجال الإداري وخاصة للشركات يساعد على التفاعل مع الزبون والعملاء عن طريق بوت الاستقبال.
- 6- في المجال الجيولوجي يساعد على التعرف على أنماط الظواهر الطبيعية للأرض.
- 7- في المجال الخدمي يساعد على قيادة السيارات الذكية والتحكم بها بالأوامر الصوتية للسائق.



الشكل (1): يوضح الاستخدامات المختلفة لتقنية التعرف على الصوت

خوارزمية معامل ميل التردد النغمي للطيف المعكوس MFCC:

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 كافي 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.



الشكل (2): يوضح آلية عمل الخوارزمية بشكل مبسط

خطوات عمل الخوارزمية:

1. استقبال إشارة الصوت:

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

2. الترشيح الأولي Pre Emphasis / Pre Processing:

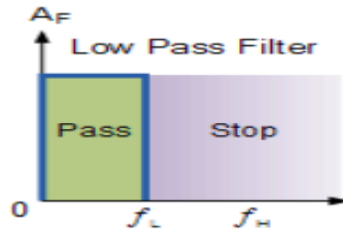
نقوم باستقبال الإشارة وتصفيته بمشرح التمرير المنخفض بناء على العلاقة أدناه:

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

بحيث يكون خرج الإشارة y هو دخل المرحلة التالية.

وقيمة a تكون عادة بين $0.9 - 1.0$ وهي ثابت تحول للإشارة.

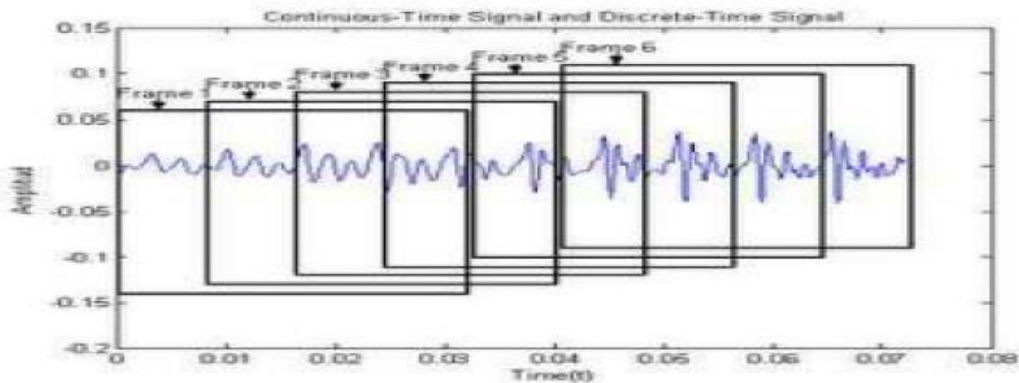
الغرض من عملية الترشيح هذه هي تضخيم الإشارة وتعويض الجزء المنخفض منها الناتج عن ضعف دقة اللفظ أو تدني مستوى الصوت للمتكلم.



الشكل (3): يوضح الرسم البياني لمرشح التمرير المنخفض

3. تقطيع الإشارة:

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: $16 \text{ ms} = 0.016 \text{ sec} = 256 / 16000$ 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: $125 \text{ f per sec} = 16000 / (256 - 128)$ ، ومنه نستنتج أن عملية التداخل تستخدم لإنتاج إطارات مستمرة والمغزى منها هو الحفاظ على تراكب البيانات بشكلها الصحيح لضمان دقة عملية التعرف وتقليل زمن التنفيذ.



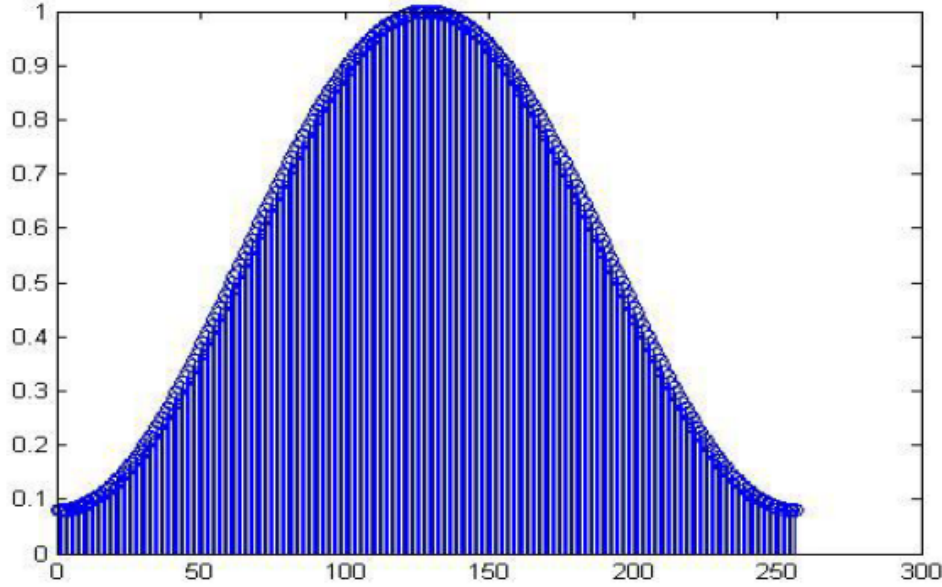
الشكل (4): يوضح عمليتي تقطيع وتداخل الإشارة

4. نافذة هامينغ Hamming Window:

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

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

بحيث أن: $0 \leq n \leq N-1$.



الشكل (5): يوضح بيانيا نتائج مخرجات نافذة هامينغ

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

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:

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

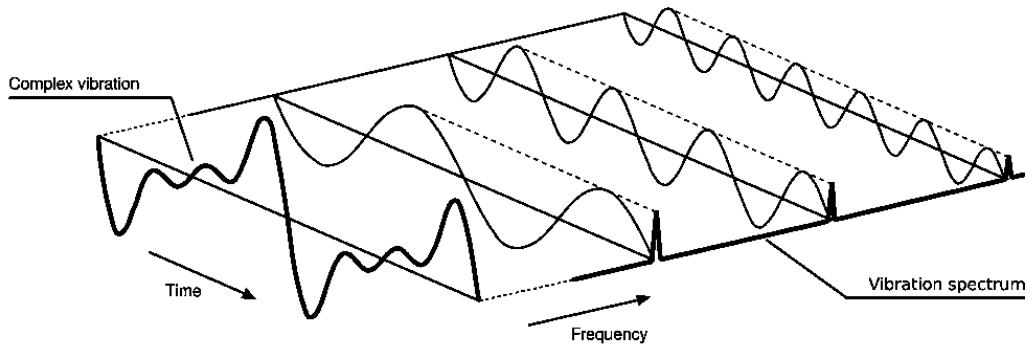
عندما يتم تطبيق تحويل فورييه السريع على إطار محدد فإنه بالافتراض تتم العملية على أن هذا الإطار هو دوري ومستمر لخرجها الناتج عن عملية الالتفاف convolution، ولكن لدينا مشكلة هنا وهي أن النقطة الأولى والأخيرة لا تدخلان ضمن شروط تحويل فورييه السريع للإشارة المستمرة إذ أنه من المحتمل بعد تطبيق العملية أن 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.

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

تعطى القدرة الطيفية للإطار (السعة الترددية) بالعلاقة التالية:

تذكير مهم:

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



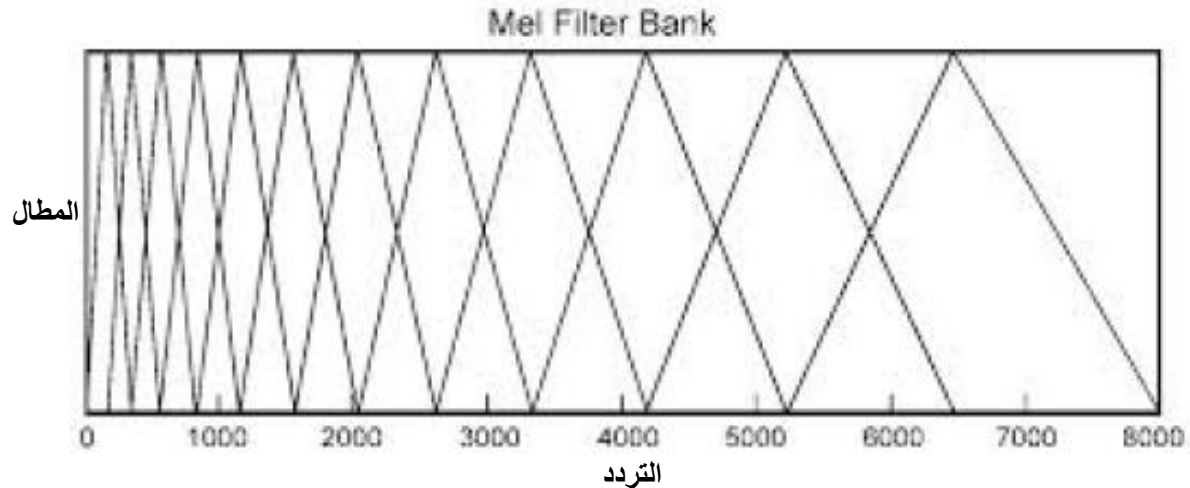
الشكل (6): تمثيل صوري لتحويل فورييه السريع

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:

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

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

وإن معادلة مرشح البنك هي كالتالي:



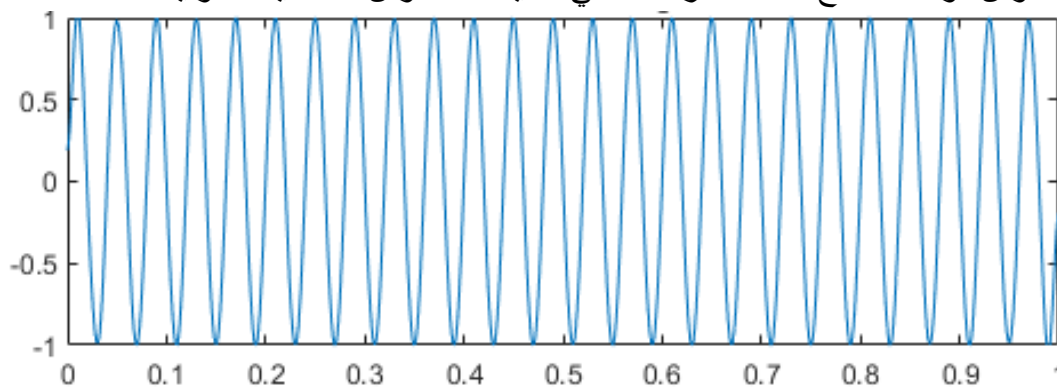
الشكل (7): يوضح الرسم البياني لخرج مرشح البنك

7. تحويل فورييه المتقطع العكسي (Discrete Fourier Transform (DFT) / (تحويل التجيب المتقطع): 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:

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

N: هي عدد مرشحات البنك المستخدمة في التطبيق.

نظرا لأننا أجرينا FFT، فإن DCT تقوم بتحويل مجال التردد للإشارة إلى النطاق الزمني وهذه العملية يطلق عليها الطيف العكسي أو querfrequency، وفي مصطلحات أخرى cepstrum؛ بحيث تم أخذ أول أربعة حروف من الكلمة frequency أو الكلمة spectrum وقلبا تعبيراً عن العملية الرياضية المطبقة حقيقة بمعنى المعكوس، ومنه نستنتج معامل التردد النغمي للطيف المعكوس لكل عينة صوتية مدخلة.



الشكل (8): يوضح خرج عملية تحويل تجيب التمام القطعي DCT

يمكن شرح عملية الطيف المعكوس كالتالي:

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

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 (4).
نقوم بحساب تابع الاستجابة النبضية للإشارة من المعادلة:

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

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

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

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

$$S(n) = E(n) * d(n)$$

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

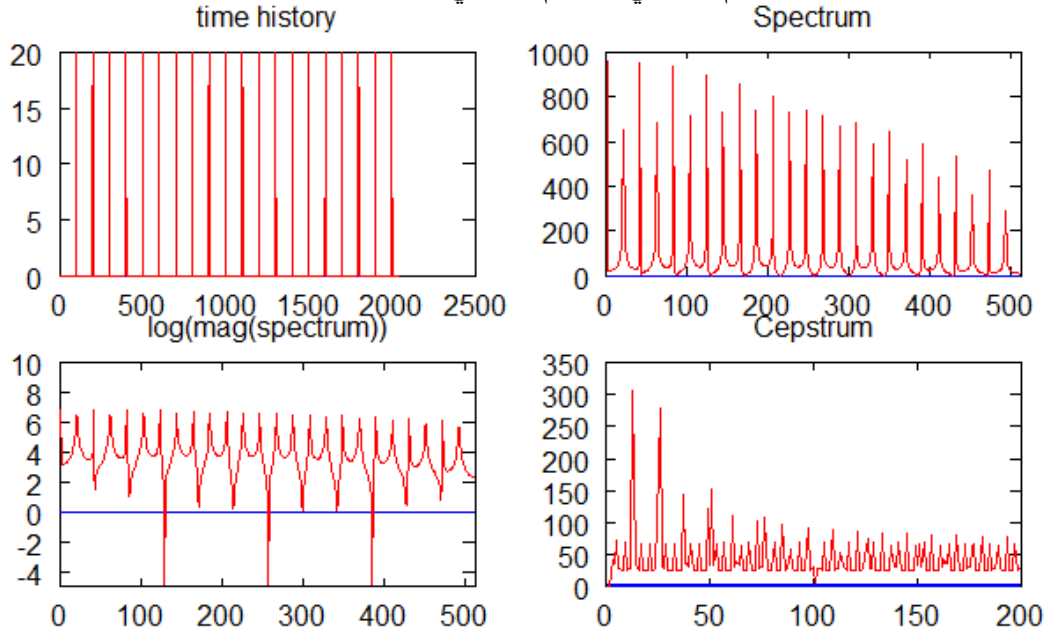
$$\log[S(n)] = \log[E(n) * d(n)] = \log[E(n)] + \log[d(n)] = \text{Ceps}_e + \text{Ceps}_d$$

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

$$C = FT^{-1}(\text{Ceps}_e + \text{Ceps}_d) = FT^{-1}(\text{Ceps}_e) + FT^{-1}(\text{Ceps}_d) = C_e + C_d$$

حيث تنتج الإشارة بالمجال الترددي لها.

تنويه (1): إن الميزات المستحصلة من الخوارزمية شبيهة بالميزات الناتجة عن الطيف المعكوس cepstral ولذلك تم تسمية الخوارزمية بهذا الاسم (Mel Frequency Cepstrum Coefficients (MFCC).
تنويه (2): 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 \sin, \cos وبالتالي فإن تحويل التجيب المنقطع سيكون للقسم الحقيقي وتحويل فورييه سيكون للأشمل للقسم الحقيقي والقسم التخيلي.



الشكل (9): يوضح الرسوم البيانية لخرج الإشارة لكل من المجال الزمني والترددية واللوغاريتمي والطيف المعكوس

شرح تفصيلي لتابع MFCC بلغة البايثون:

Parameters:

- **signal** – the audio signal from which to compute features. Should be an N*1 array
- **samplerate** – the samplerate of the signal we are working with.
- **winlen** – the length of the analysis window in seconds. Default is 0.025s (25 milliseconds)
- **winstep** – the step between successive windows in seconds. Default is 0.01s (10 milliseconds)
- **numcep** – the number of cepstrum to return, default 13
- **nfilt** – the number of filters in the filterbank, default 26.
- **nfft** – the FFT size. Default is 512.
- **lowfreq** – lowest band edge of mel filters. In Hz, default is 0.
- **highfreq** – highest band edge of mel filters. In Hz, default is samplerate/2

- **preemph** – apply preemphasis filter with preemph as coefficient. 0 is no filter. Default is 0.97.
- **ceplifter** – apply a lifter to final cepstral coefficients. 0 is no lifter. Default is 22.
- **appendEnergy** – if this is true, the zeroth cepstral coefficient is replaced with the log of the total frame energy.
- **winfunc** – the analysis window to apply to each frame. By default no window is applied. You can use numpy window functions here e.g.
winfunc=numpy.hamming

Returns: A numpy array of size (NUMFRAMES by numcep) containing features. Each row holds 1 feature vector.

تحليل وتصميم النظام

مقدمة:

إن تطور الحاسب الآلي وتنوع مهامه واستخداماته لفت الأنظار إلى توسيع تطبيقاته لتشمل (قدر الإمكان) كافة مجالات الحياة الواقعية ومنها الأنظمة الإدارية، ومما يعوق التقدم الإداري سابقا هو تعقيد الهرميات وهيكلية التنظيم بتعداد أفراده وتنوع قراراته، الأمر الذي أصبح سهلا مرنا سريعا باستخدام الحاسب الآلي، وإن من أهم عوامل نجاح العمل الإداري الالتزام بالحضور من قبل أعضاء التنظيم، ولهذا ولتسهيل عملية تفقد وجود الأعضاء وباستخدام الحاسب الآلي وخوارزميات الذكاء الصناعي المتطورة لجأ الباحثون إلى توظيف هذه التقنيات الحديثة لعمل نظام تفقد ذكي ذو كفاءة عالية وتكلفة رمزية لتسهيل هذه العملية الهامة في النظم الإدارية.

سنقوم بعمل تحليل لهذا النظام على أنه نظام إداري لجامعة يقوم بتفقد حضور الطلاب بما سيتقدم ذكره لاحقا بما يساعد إدارة الكليات على تفقد حضور الطلاب بشكل سهل وبزمن قياسي ودون تكلفة كبيرة.

مشكلة البحث:

تكمن مشكلة البحث أن عدم التزام حضور الطلاب للمناهج العلمية في الجامعة سيضر بسير العملية التعليمية للمؤسسة وبالتالي اتبعت المؤسسات منهجية التفقد الورقي للحضور لضبط هذه العملية، بالإضافة إلى حضور من لا يرغب بتواجده ضمن الحصص الدراسية المتخصصة سيؤثر على مكانة المدرس على أقل تقدير كونه رئيس الجلسة، فضلا عن صعوبة عملية التفقد ورقيا في حال تواجد عدد كبير من الطلاب أثناء الحصة الأمر الذي سيسبب ضياع في وقت الدرس وعدم إكماله على الوجه المطلوب كما سيسبب صعوبة التنسيق مع باقي الإداريين لضبط العملية التعليمية لهذه الحصة طوال فترة الدراسة، عوضا عن كون التفقد الورقي من الأساليب الكلاسيكية القديمة التي لا تلبي متطلبات العصر وحاجة التطور في المنطقة.

دوافع تبني البحث:

- للتخفيف من زخم العمل الإداري وتسهيل عملية التفقد.
- ضمان سير العملية التعليمية في الجامعات.
- مواكبة التطور ومتطلبات العصر.

1. الأهداف:

يهدف هذا المشروع إلى ضمان سير العملية التعليمية وضمن التزام الطلاب بضوابط الدوام وشروطه.

2. أهمية البحث:

تكمُن أهميته من أهمية النظم الإدارية في تنظيم المؤسسة والاستفادة من خططها قدر الإمكان، بالإضافة إلى أن الخطط التقليدية لا تلبي حاجات الإنسان الحديث وبالتالي الحاجة إلى التحسن التكنولوجي، عوضاً عن أهميته التقنية كونه أحد خدمات العصر المتطورة حديثاً والمستخدمة في كثير من المؤسسات في دول العالم المتطور مما يتيح لنا اللحاق بركب التقدم الحضاري.

3. آلية عمل النظام:

يقوم الطالب بالتحدث باسمه (مثلاً) أمام مستقبل صوتي أمام باب القاعة ليتم أخذ تفقده بالحضور بعد التعرف عليه، بحيث يكون المستقبل موصولاً بحاسب مبرمج على الخوارزمية المذكورة ومتصل بنظام مؤتمت خاص بالجامعة ويتم تدفق البيانات إليه كاستقبال / إرسال.

4. المنهجية المتبعة:

النموذج الحزوني؛ بحيث تم اختيار هذا النموذج لما يتمتع به من سهولة التعديل وتصحيح الأخطاء بالإضافة إلى مرونة البناء البرمجي من خلال هذا السياق.

5. الطرق المتبعة لجمع البيانات:

1- البحث على الشبكة العنكبوتية:

تم البحث على النظام المقصود عبر شبكة الحاسب العالمية GOOGLE وتم التوصل إلى أن النظام معمول به في غالبية الدول المتقدمة وهو نظام Standard لا يمكن الاستغناء عنه بالمؤسسات الكبيرة.

2- الملاحظة والمراقبة:

من خلال الملاحظة تم التوصل إلى أن الجامعات في المنطقة بحاجة كبيرة لهذا النوع من الأنظمة لعدة أسباب وهي:

(1) لا يحتاج إلى تكلفة كبيرة سوا بعض المعدات المتاحة للجميع كالحاسب ومسجل الصوت بالإضافة إلى بعض البرمجيات الموجودة في اللازمة لعمل الجامعات مثل أنظمة الأتمتة على أقل تقدير.

(2) سهولة تنفيذ النظام.

(3) صعوبة التقليد بالأساليب القديمة وعدم كفايتها للنضج الإداري المطلوب.

(4) متابعة مسيرة التطوير التقني اللازم للرقى بالمستوى الحضاري للمنطقة.

تحليل النظام:

تعتبر هذه المرحلة أساسية في بناء أي نظام برمجي والتي سيعتمد نجاح المراحل اللاحقة لها على مدى دقتها وصحتها، إذ يتم في هذه المرحلة تجميع كافة البيانات اللازمة لهذا النظام وتحديد آلية عمل النظام القديم لمعرفة المشاكل التي يستم تجنبها في النظام الجديد ومن هنا سنقوم بتحديد المتطلبات الخاصة بتطوير النظام الجديد لمعرفة خصائصه وميزاته ووظائفه المراد إنجازها.

تحديد متطلبات النظام:

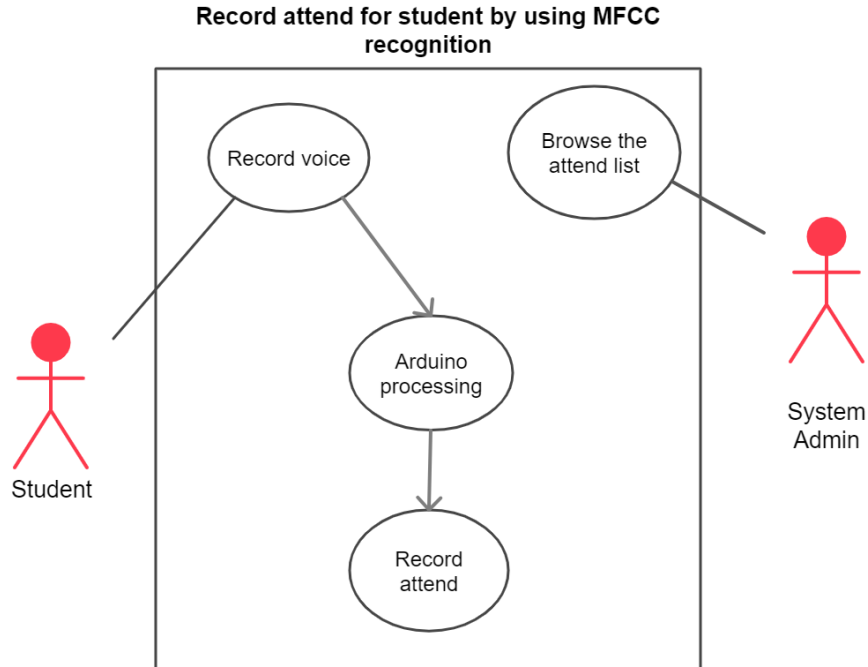
1. متطلبات وظيفية:

- (1) يتم تحديد الوظائف الفعلية والمتمثلة في المهام المختلفة للنظام من حيث عملية الإدخال والمعالجة وفقاً للتالي:
(1) الإدخال: يجب أن يكون النظام قادراً على إدخال الأصوات إلى جهاز الحاسب المحقون بخوارزمية العمل لمعالجته وتسجيل المخرجات النهائية.
- (2) الحفظ والمعالجة: يجب على النظام أن يوفر مهام معالجة تضمن حضور الطالب للحصص بشكل غير قابل للخطأ الذي يؤدي إلى خلخلة عمل النظام الإداري.

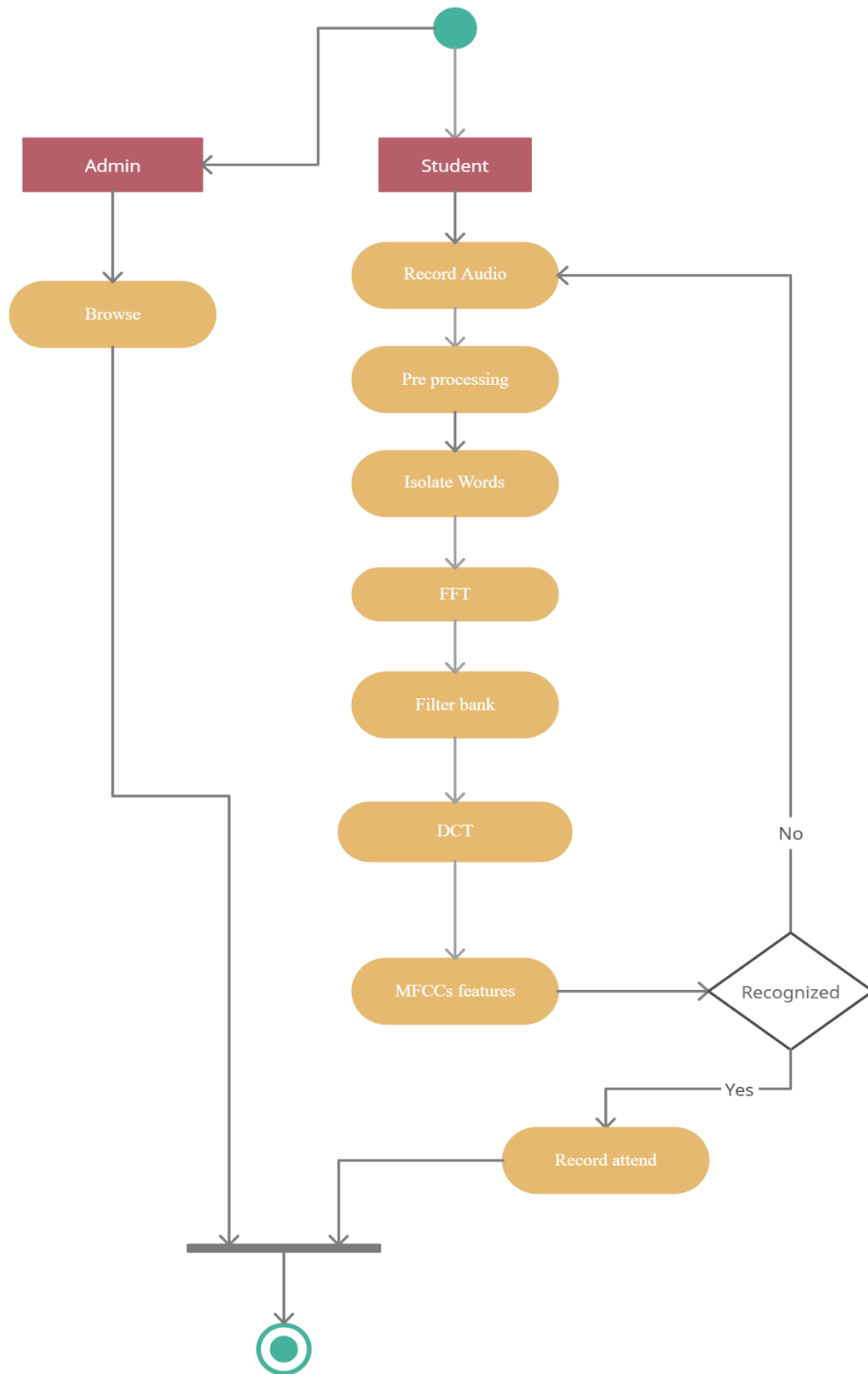
2. متطلبات غير وظيفية:

- (1) سهولة الاستخدام: لا يحتاج النظام إلى كثير استخدام من الطلاب إذ أنه يقوم بأخذ التفقد من خلال ضغط زر للمستقبل الصوتي ليتم تسجيل الصوت بفترة زمنية لا تتجاوز الـ 3 ثوان، كما أنه لا يتطلب من الإداري المشرف على برمجيات الجامعة إلى كثير عمل إذ أنه يسجل بيانات الحضور لكل طالب بعد التعرف عليه ويعرضها في سجل الحضور الفصلي للطلاب في هذا الصف.
- (2) السرعة: النظام المعتمد يتميز بالسرعة لتسجيل حضور الطلاب بفترة زمنية قصيرة إذ أنه يقوم بتسجيل اسم الطالب ويرسل البيانات على الفور إلى الحاسب بعد ضغط الزر ليتم تسجيل الطالب الذي يليه مباشرة.
- (3) البساطة: إذ أنه لا يحتاج سوا إلى ضغط زر من الطالب ليتم تسجيل حضوره، ومن الإداري المشرف على نظام الجامعة لا يحتاج شيء سوا عرض سجل الحضور للطلاب عند الطلب.
- (4) الوثوقية: تتمتع الخوارزمية المطروحة بدقة جيدة إذ أنها تقوم بتسجيل الطلاب بكفاءة عالية.
- (5) الأمان: إن أمان النظام يعتمد على أمان النظام المؤتمت في الجامعة إذ أنه من غير المسموح لأي أحد آخر غير المصرح بهم بالدخول إلى سجل الحضور.

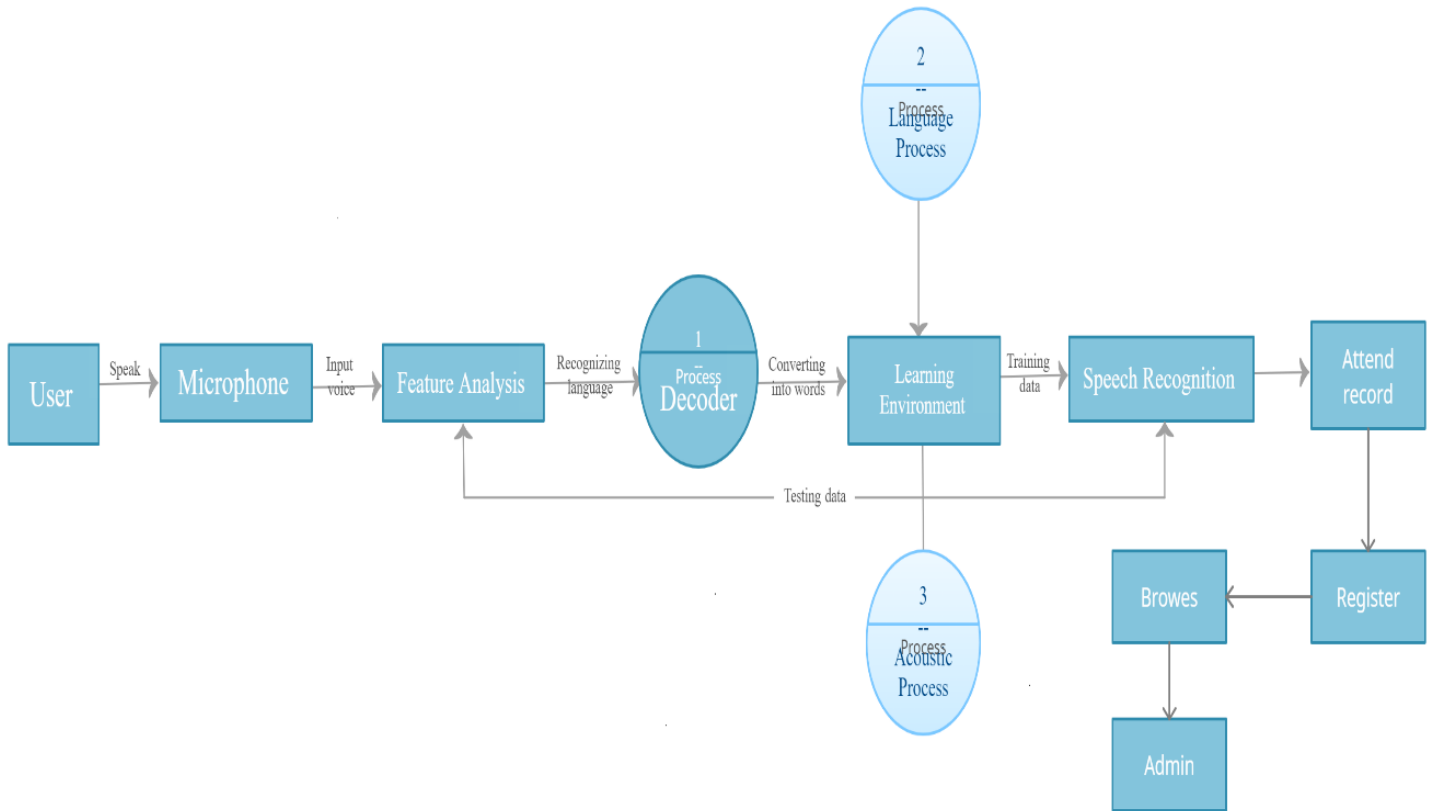
مخططات تحليل النظام:



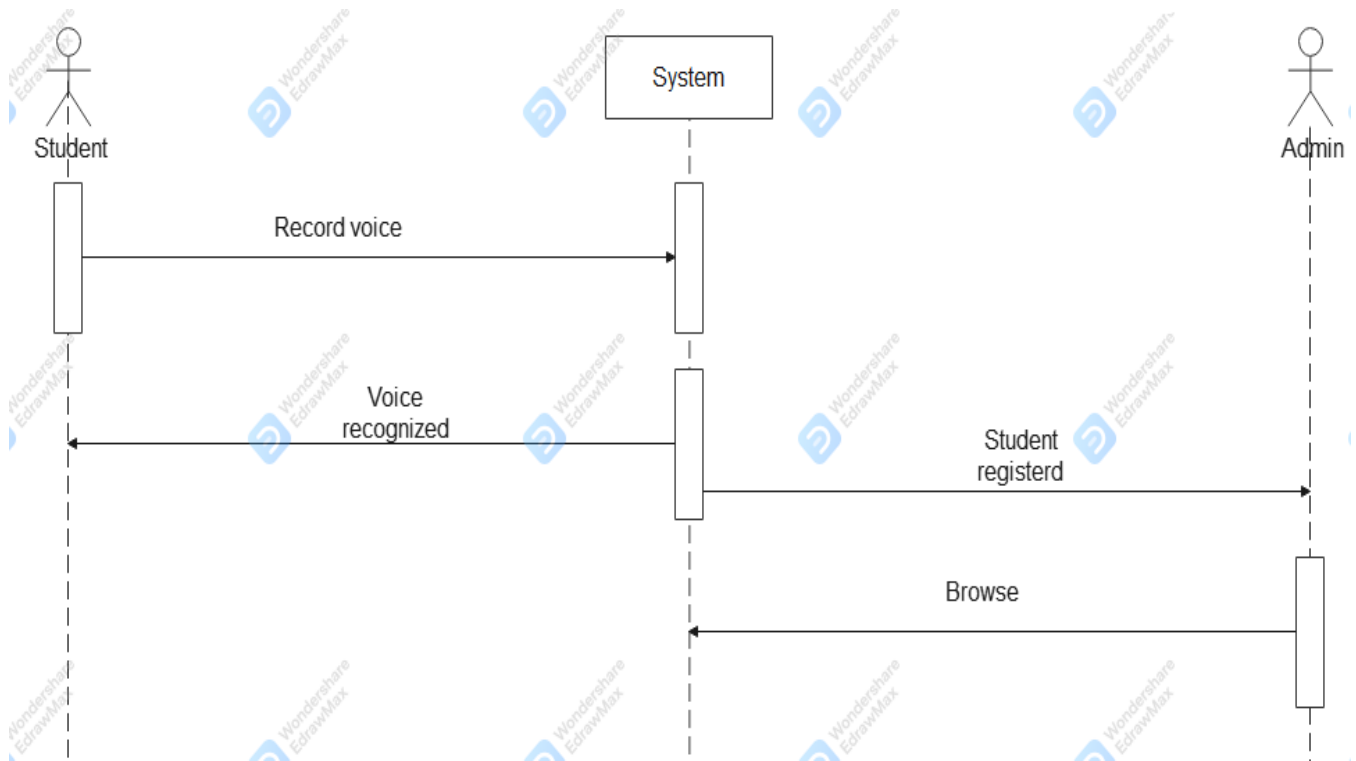
الشكل (10): مخطط حالات استخدام النظام



الشكل (11): مخطط النشاط للنظام



الشكل (14): مخطط تدفق البيانات للنظام



الشكل (15): مخطط التسلسل للنظام

تصميم النظام:

إن مرحلة تحليل النظام تعتبر مدخلات لمرحلة تصميم النظام بمكوناته للحصول على المهام والوظائف الملموسة لإجراء تركيب وبناء النظام، وتم اتباع منهج شراء الحزمة البرمجية وتخصيصها بحيث تتلاءم مع النظام المؤتمت في الجامعة، وإن من مكونات النظام المعتمدة لبناء هذا التطبيق:

1- المكونات البرمجية:

وهي إجراءات العمل البرمجي لهذا النظام إذ أنه يتطلب خوارزمية التعرف على الصوت، بالإضافة إلى قاعدة البيانات الخاصة بالبيانات الصوتية للطلاب، وبرمجيات الربط بين خوارزمية العمل والنظام المؤتمت في الجامعة، وقد اعتمدنا نموذج الطبقات بحيث تعمل على عزل بين منطق البرنامج عن الواجهات التفاعلية للمستخدم والذي يؤدي إلى مرونة عالية في تصميم الواجهات وإمكانية التحديث مستقبلاً وهي كالتالي:

1. طبقة الواجهات:

تحتوي على قوائم الطلاب بالإضافة إلى أزرار معنية بأشهر الفصل وقائمة بأسماء المواد وقوائم كليات الجامعة؛ بحيث تكون سهلة الفهم والتعامل من قبل الإداري وغير معقدة وتمتاز بالجاذبية والتناسق.

2. طبقة قاعدة البيانات:

تحتوي على قاعدة بيانات صوتيات الطلاب في الفصل وهي على شكل ملف EXCEL.

3. طبقة الأكواد:

وهي الطبقة المخفية الغير ظاهرة لكل مستخدم النظام بحيث تحتوي على خوارزمية العمل المكتوبة بلغة الـ Python ومكتباتها المستخدمة بحيث أنشأنا شبكة عصبية من نوع الأمامي لتدريبها على معاملات خوارزمية العمل، كما سنعمد إلى كتابتها بلغة أردوينو لتنفيذ المشروع. كما تم رسم مخططات النظام بشكل كامل في مرحلتى التحليل والتصميم لمعرفة تفاصيله والمساعدة على بناء أو تعديل أو تطوير أو تصحيح الأخطاء فيه مستقبلاً.

أدوات البرمجة اللازمة لإنجاز النظام:

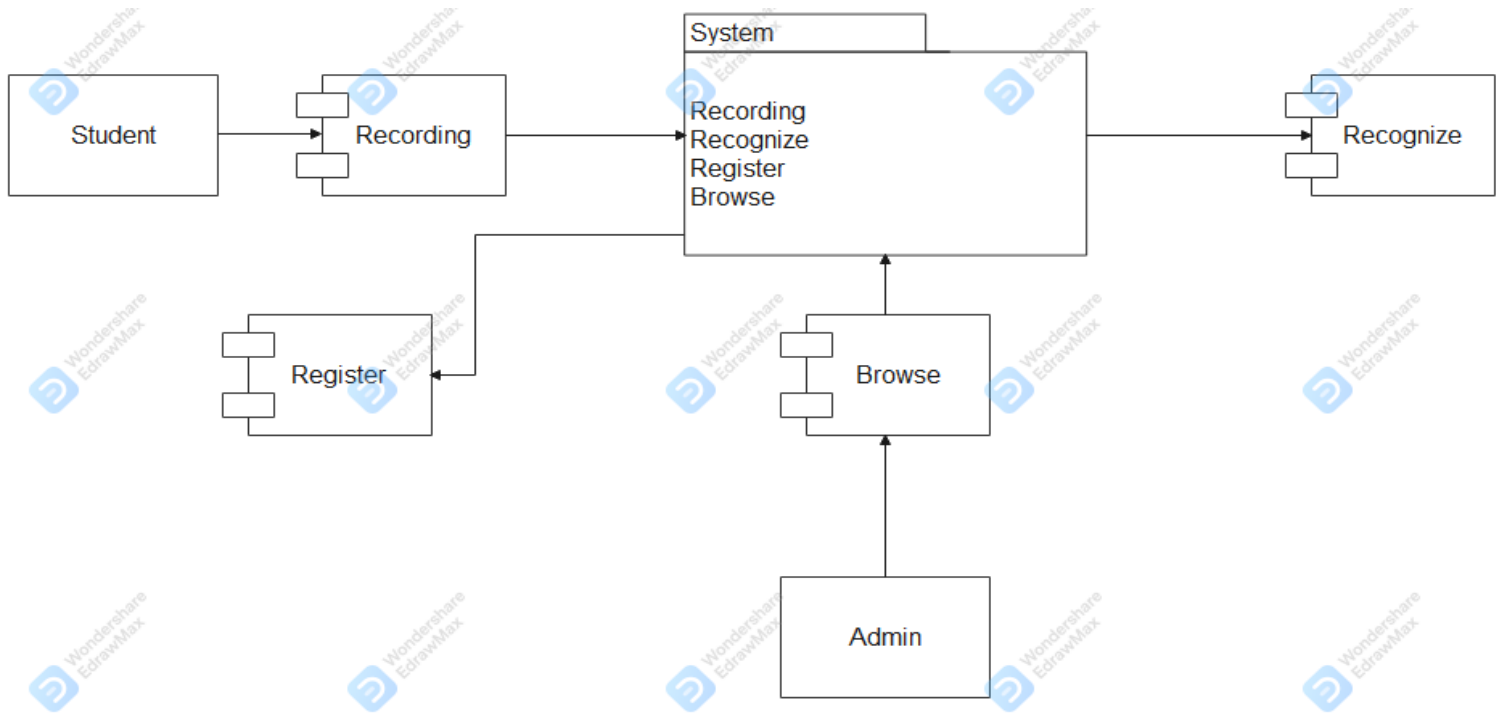
اعتمدنا في بناء النظام على بيئة Microsoft Visual Studio 2019 المدعومة بحزمة Python 3.9 لبناء المشروع بلغة Python العالمية سهلة الفهم والتنفيذ، بالإضافة إلى إنشاء قاعدة بيانات على بيئة Microsoft Excel لبناء قاعدة بيانات متميزة بالمرونة وسهولة عرض والتعامل مع البيانات المراد معالجتها، وتم رسم مخططات النظام على بيئة Wondershare EdrawMax آخر إصدار له من الشركة المنتجة، وموقع create.ly.

2- المكونات المادية:

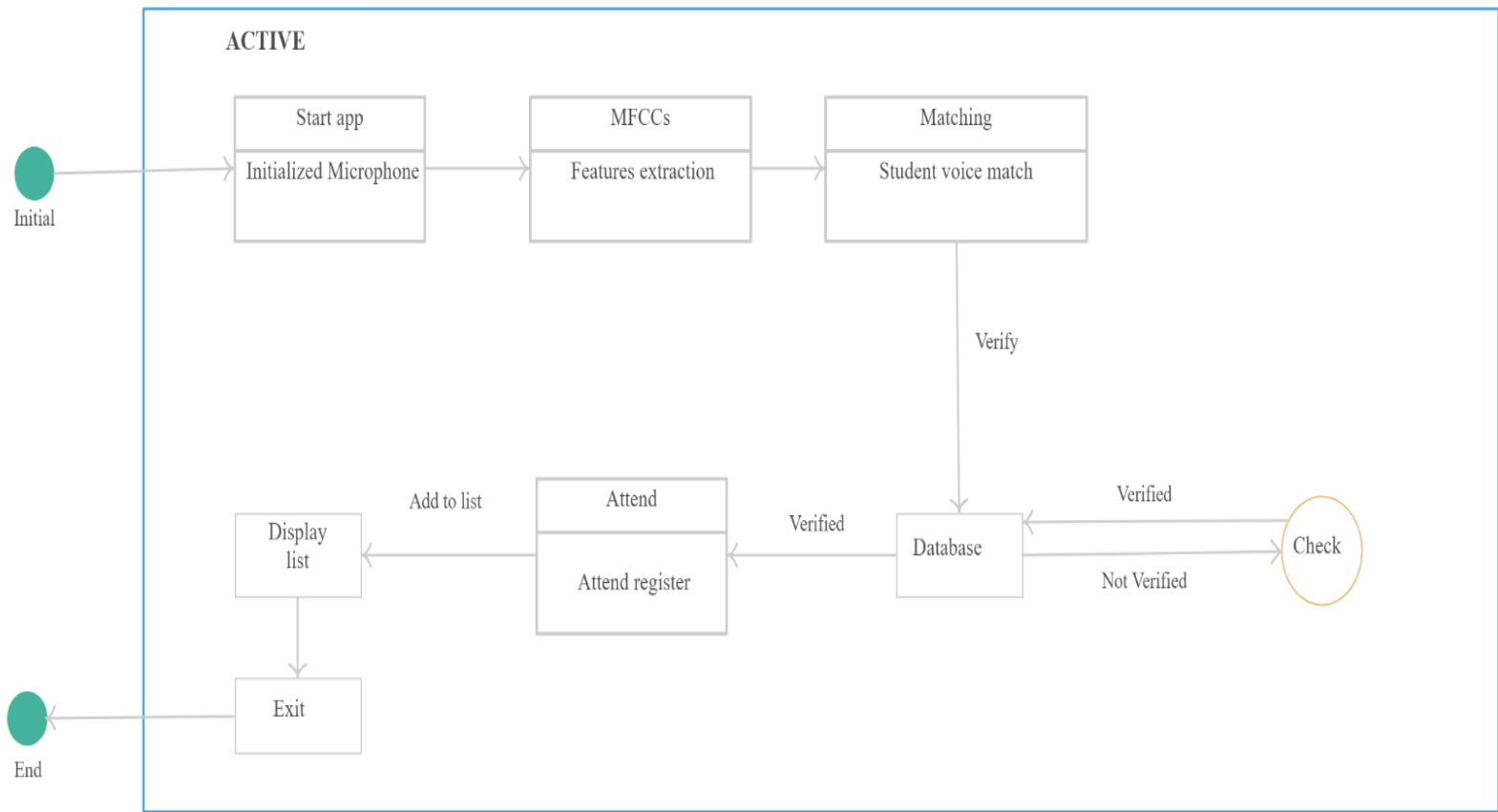
إذ يلزم لتطبيق هذا المشروع مستقبل صوتي (مايكروفون) مثبت على أبواب القاعات إضافة إلى شريحة أردوينو وكابلات التوصيل مع المستقبل ودائرة تغذية كهربائية وحاسب.

مخططات تصميم النظام:

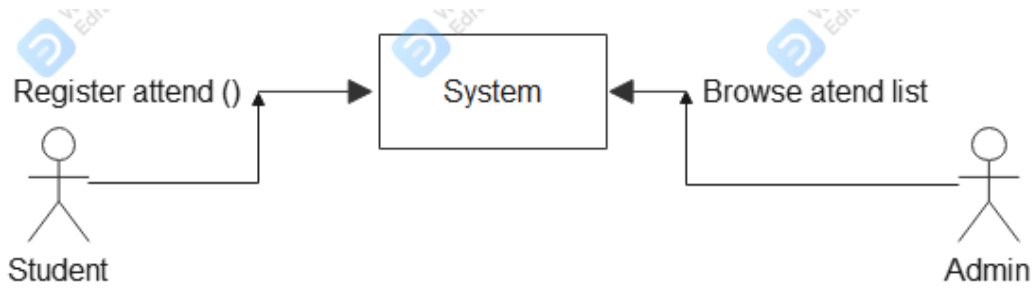
إن مخططات التصميم إلى جانب مخططات التحليل تعتبر البنية الأساسية لنظام العمل الذي سوف نقوم ببنائه، وتعمل هذه المخططات دور المعمارية الموضحة لعمليات البناء والتعديل والتصحيح، سوف نقوم برسم مخططات تصميم النظام مع عدم تكرار بعض المخططات كمخطط الأصناف والتسلسلي، فقط مخططات الحالة والتعاون والتواصل ومخططات التحريم والمكونات والتسليم بمخطط واحد.



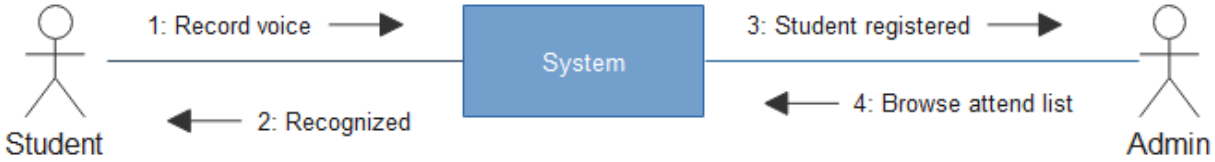
الشكل (17): مخططات التحزيم والمكونات والتسليم للنظام



الشكل (18): مخطط الحالة للنظام



الشكل (19): مخطط التواصل للنظام



الشكل (20): مخطط التعاون للنظام

التنفيذ والاختبار:

قمنا بعمل محاكاة للمشروع على الحاسب فقط ونظرا لعدم توفر بيانات الطلاب اللازمة لتنفيذ المشروع قمنا بجلب قاعدة بيانات جاهزة من الشبكة العنكبوتية تتضمن النغمات الصوتية الموسيقية المعروفة كالجاز والبوب والروك وغيرها ضمن مجموعة صوتيات بعدد 10 أنواع نغمية و 100 صوتية لكل نغمة لتنفيذ البرنامج وتدريب الشبكة العصبونية الأمامية التسلسلية المنشأة عليها واستخراج النتائج بناء على تصنيف النغمات العشرة؛ بحيث حصلنا على دقة تمييز تتراوح بين 60 – 70 بالمئة وفي بعض الأحيان 80% وأحيان أخرى 90% تقريبا، بحيث تم كتابة كود الخوارزمية على بيئة VS 2019 وكود الاختبار والتنفيذ أيضا على نفس البيئة بلغة البايثون باستعمال عدة مكتبات شهيرة مثل Keras، وقمنا بهذه العمليات على حاسب محمول بالمواصفات التالية:

Asus GL552VW (ROG), Core I7 6700hq (4cores, 8 th, 2.6GHz), Ram 16G 2400Hz, Hard 120 SSD M.2 (1T HDD), GPU GTX 960m 4G until 12G.

واستغرقت عملية تدريب الشبكة ما يقارب دقيقتان مع توليد النتائج على شكل ملف ECXEL وهي تعني الميزات المستخرجة من كل ملف صوتي (1000 ملف) بعدد معاملات 20 للتردد النغمي للطيف المعكوس MFCC.

الجدوى الاقتصادية:

لا يحتاج تطبيق هذا المشروع إلى كثير معدات وتجهيزات وإنما فقط إلى مستقبل صوتي مثبت على أبواب القاعات الدراسية بالإضافة إلى كابلات توصيل مع شريحة الأردوينو التي ستقوم بخوارزمية العمل وأيضا إلى حاسب الآدمن القائم على نظام الجامعة لمعالجة البيانات وإرفاقها بسجلات الطلاب في النظام للاستفادة منها لاحقا، والقطع المطلوبة للتنفيذ متوفرة في أسواق الإلكترونيات بمبالغ زهيدة لا تتجاوز الـ \$17 لشريحة الأردوينو Mega2565 وكابلاتها بسعر \$2 والمستقبل الصوتي بسعر لا يتجاوز \$3 من إنتاج شركات صينية.

المراجع:

مقرر الذكاء الصناعي للدكتور المهندس محمد أمين الزبداني.
مقرر هندسة البرمجيات للدكتور المهندس محمد أمين الزبداني.

<https://www.simplifiedpython.net/speech-recognition-python/>

https://www.researchgate.net/figure/BLOCK-DIAGRAMS-FOR-MFCC-FEATURE-EXTRACTION_fig1_350467478

<http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

<http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

<https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9>

<https://nasainarabic.net/main/articles/view/high-tech-gadgets-speech-recognition>

<https://www.kaggle.com/seriousran/mfcc-feature-extraction-for-sound-classification>

`python_speech_features.base.mfcc(signal, samplerate=16000, winlen=0.025, winstep=0.01, numcep=13, nfilt=26, nfft=512, lowfreq=0, highfreq=None, preemph=0.97, ceplifter=22, appendEnergy=True, winfunc=<function <lambda>>)`
Compute MFCC features from an audio signal.

https://www.mathworks.com/help/dsp/ug/filter-banks.html?s_tid=srchtitle

<https://www.kaggle.com/carlthome/gtzan-genre-collection>

The background of the entire image is a dark blue field with a complex, light blue circuit-like pattern. This pattern consists of numerous thin, interconnected lines that form a dense, web-like structure, reminiscent of a printed circuit board or a neural network. The lines vary in thickness and orientation, creating a sense of depth and complexity.

الحمد لله الذي بنعمته تتم الصالحات