وزارة التعليـــم العالي و البحث العلمـــي

جامعة الجزائر-**2**- أبو القاسم سعد الله

كلية اللغة العربية و آدابها و اللغات الشرقية قسم علوم اللسان

أطروحة لنيل دكتوراه العلوم

التعرف الآلي على الكلام العربي المتصل

إعداد: داود مسعـــود إشراف: د. مراد عباس

2016

وزارة التعليـــم العالي و البحث العلمـــي

جامعة الجزائر-**2**- أبو القاسم سعد الله

كلية اللغة العربية و آدابها و اللغات الشرقية قسم علوم اللسان

أطروحة لنيل دكتوراه العلوم

إعداد: **داود مسعـــود**

التعرف الآلي على الكلام العربي المتصل

أعضاء لجنـــــة المناقشـــــــة

أ.د حركات مصطفى	.....	رئيسا
د. مراد عباس	.....	مقررا
د. سلامي سعدي	.....	عضوا
د. غريبي عارف	.....	عضوا

2016

إهداء

اهدي هذا العمل إلى أبي و أمي

و إخوتي و عائلتي الكبيرة

و زوجتي و عائلتها

و إلى ابني محمد و بنيتي عائشة

و إلى كل أصدقائي

داود مسعود

شكر

أشكر الله تعالى و أحمده على أن وفقنا لهذا

ثم أتوجه بالشكر الجزيل إلى الأستاذ عبد المجيد سالمي الذي طالما كان لنا عونا و سندا إذا فترت الهمم

كما أشكر الأستاذ مصطفى حركات على احتواء هذا العمل

و لا يفوتني أن أتقدم بالشكر الخاص إلى كل موظفي قسم علوم اللسان بكلية اللغة العربية و آدابها

و اللغات الشرقية بجامعة الجزائر 2

و إلى كل من أعانني من قريب أو بعيد

جزاكم الله خيرا

ملخص

قمنا في هذا العمل بانجاز نظام للتعرف الآلي على الكلام العربي المتصل، يعتمد على تقنية (MFCC) Mel-Frequency Cepstrum Coefficients كوسيلة لتحليل الإشارة الصوتية و استخلاص معاملاتها، كما استعملنا نماذج ماركوف المخفية(HMM) كتقنية للتعرف، و لتحقيق هذا النظام استخدمنا مجموعة أدوات نماذج ماركوف المخفية (HTK).

لأجل تقييم قدرة النظام المنجز على التعرف قمنا بمجموعة من الاختبارات و التجارب التي توصلنا من خلالها إلى نتائج جيدة و مشجعة باعتبار المشاكل و الصعوبات التي تعترض أنظمة التعرف عموما كالتداخل بين الأصوات و الضوضاء و غيرها، و قد تحصلنا و قد تحصلنا نسبة تعرف قدرت بـ:75.21 %

كلمات المفاتيح: العلاج الآلي للكلام, التعرف الآلي على الكلام, التحليل الصوتي, معاملات MFCC، نماذج ماركوف المخفية HMM.

Résumé

Nous avons fait dans ce travail une étude sur la reconnaissance automatique de la parole arabe continue.

Pour effectuer ce travail nous avons utilisé la technique (MFCC) Coefficients Cepstre Mel-fréquence comme une technique d'analyse de la parole et l'utilisation des modèles de Markov cachés (HMM) comme une technique de reconnaissance, et pour réaliser ce système, nous avons utilisé la plateforme (HTK).

Puis, Nous avons utilisé le système de reconnaissance qui nous avons présenté la pour estimer la prononciation et indiquer le taux de réussite de la prononciation des phonèmes, qui nous donne Le taux de reconnaissance 75.21% .

Mots clé: Le Traitement Automatique de la Parole, la reconnaissance automatique de la Parole, l'analyse de la parole.

Abstract

In this work we did a study about arabic Speech Recognition.

In order to use MFCC (Coefficients Cepstre Mel-fréquence) as a technique for analyzing speech and also using Hidden Markov models (HMM) as recognition technique.

We used the system for recognizing that, we presented it to estimate the pronunciation and to indicate the success rate of the pronunciation of phonemes. It gives us the recognition rate 75.21%

Key word: Speech Automatic Recognition, , acoustic analysis, Hidden Markov Model (HMM).

المقدمـة

المقدمـــــــــــة

يعتبر العلاج الآلي للكلام من أوسع مجالات البحث حيث يتعلق بأهم وسيلة اتصال يعتمد عليها الإنسان في التواصل و هو الكلام، لما يمتاز به من سهولة في الإنتاج و فاعلية في التعبير عن مراد المتكلم، فنجد أن الإنسان يحاول استعماله في عدة وظائف غير التواصل مع بني جنسه، فهو مثلا يحاول التحكم في بعض أصناف الحيوانات بتدريبها على ألفاظ معينة، و بعد ظهور التكنولوجيا الحديثة أصبح من الممكن تحويل الإشارة الصوتية إلى إشارة تماثلية و رقمية يمكن للوسائل التكنولوجية الحديثة التعامل معها و الاستفادة منها، حيث ظهرت عدة مجالات رئيسية للعلاج الآلي للكلام و هي: التعرف الآلي على الكلام و فهم و تركيب الكلام، و هذه الأنظمة لا تتعلق بالعلوم التقنية فحسب بل هناك مجموعة من المفاهيم و التخصصات الأخرى التي لابد من معرفتها حول الكلام و السمع و اللسانيات و علم الأصوات و الالكترونيات و الذكاء الاصطناعي و الرياضيات و غيرها.

و قد حضي التعرف الآلي على الكلام العربي بالكثير من الاهتمام نظرا لمكانة اللغة العربية مما أدى إلى ظهور عدد من أنظمة التعرف التي قد تختلف عن بعضها البعض في بعض العمليات و المعالجات التي تتبعها أثناء عملها على الرغم من أن النتيجة التي يجب أن تحصل عليها واحدة، فهي تعتمد على عملية تحويل الإشارة الصوتية، الملتقطة إلى مجموعة معطيات و التي يجب أن تعالج فيما بعد لأجل الوصول إلى معلومات قد تكون هي النتيجة المراد التوصل إليها، أو قد تستخدم كمدخل إلى أنظمة معالجة اخرى للتعرف على الكلام أو التحكم للتصرف على أساسه.

إن الشروع في انجاز نظام للتعرف الآلي على الكلام العربي لابد أن يبدأ بمرحلة تحضير قاعدة للبيانات تحتوي على تسجيلات إشارة الكلام المراد التعرف عليها و التي هي الحروف العربية فهذه المرحلة تعد مرحلة هامة و تتطلب دقة عالية و معرفة بمختلف الحروف و خصائصها حيث أن دقة النتائج المتحصل عليها تتعلق بجودة المعطيات التي توفرها قاعدة البيانات و لتحضير هذه الأخيرة نقوم بإنشاء مدونة تحتوي على عدد كاف من الجمل

باللغة العربية تكون الحروف فيها في مختلف المواضع (في بداية و وسط و نهاية الكلمة) و بعدها نقوم بتسجيل هذه الجمل بواسطة عدة أشخاص ثم عملية تقطيع الإشارة المتحصل عليها إلى مقاطع تعالج لأجل الحصول على قاعدة بيانات تستعمل لاحقا في عمليتي التدريب و الاختبار لنظام التعرف الآلي على الكلام المقترح.

و بعد تحضير قاعدة البيانات تبدأ مرحلتي التدريب ثم الاختبار لنظام التعرف حيث يعتبر فيها التعرف علي مقاطع الصوت من الإشارة الصوتية المدخلة مشكلة كبيرة ناتجة عن التداخل بين الأصوات عند الانتقال من حرف إلى آخر، و هنا تظهر أهمية التقنيات التي تستخدم في التعرف الآلي على الكلام مثل نموذج ماركوف المخفي HMM، لما تمتاز به من قدرة على التمييز بين الأصوات المختلفة في الإشارة الصوتية من خلال معاملاتها.

هناك جملة من التحديات تظهر عند انجاز نظام للتعرف الآلي على الكلام، من بينها أن الأصوات بطبيعتها تتداخل و تؤثر إحداها على الأخرى إضافة إلى وجود الضوضاء مما قد يؤثر على دقة التعرف.

لتحقيق هذا الهدف، فقد نظمنا عملنا على النحو التالي:

المقدمة: فيها بيان تعريف بالبحث.

الفصل الأول: يقدم دراسة حول خصائص حروف اللغة العربية

أما الفصل الثاني: فيعطي لمحة عامة عن التعرف الآلي على الكلام و التقنيات الرئيسية لتحليل الإشارة الصوتية و استخراج المعاملات لتمثيل الإشارة الصوتية.

الفصل الثالث: شرح و تقديم لنماذج ماركوف المخفية.

أما الفصل الأخير: فنناقش فيه بتفصيل أكبر منهجية العمل, باستخراج المعاملات من الإشارة الصوتية, و استعماله من أجل التعرف الآلي على الكلام بتطبيقه على مدونة أصواتها باللغة العربية.

و في الخاتمة العامة، نقدم النتائج المحققة في التعرف الآلي على الكلام العربي المتصل، و آفاق هذا العمل.

الفصل الاول

أصوات اللغة العربية و خصائصها

1. مقدمة:

يتكون جهاز النطق عند الإنسان من عدة أعضاء تكون منظومة متكاملة يستطيع من خلالها إصدار أصوات مختلفة و لهذا كان من الضروري معرفة هذه الأعضاء و دورها في إصدار الأصوات للتوصل إلى خصائص كل صوت من الأصوات اللغوية و دراسة كيفية إنتاجها للتوصل إلى السيطرة على جهاز النطق السيطرة اللازمة لنطق الصوت بالصورة الصحيحة و التوصل إلى تصنيف الأصوات تصنيفا يساعدنا على معرفة الظواهر البيولوجية و الفيزيائية التي تؤثر على كل صوت بالإضافة إلى المساعدة في إيجاد الطرق الملائمة لمعالجة و اختبار كل صوت.

2. الصوت:

لغة:الصوت الجرس، والجمع أصوات قال ابن السكّيت: الصوت صوت الإنسان وغيره، و الصائت:الصائح [1].

اصطلاحا: هو الأثر السمعي الذي تحدثه موجات ناشئة من اهتزاز جسم ما و هو اضطراب طبيعي خارجي يعرض لجميع الأجسام و خاصة الهواء[2].

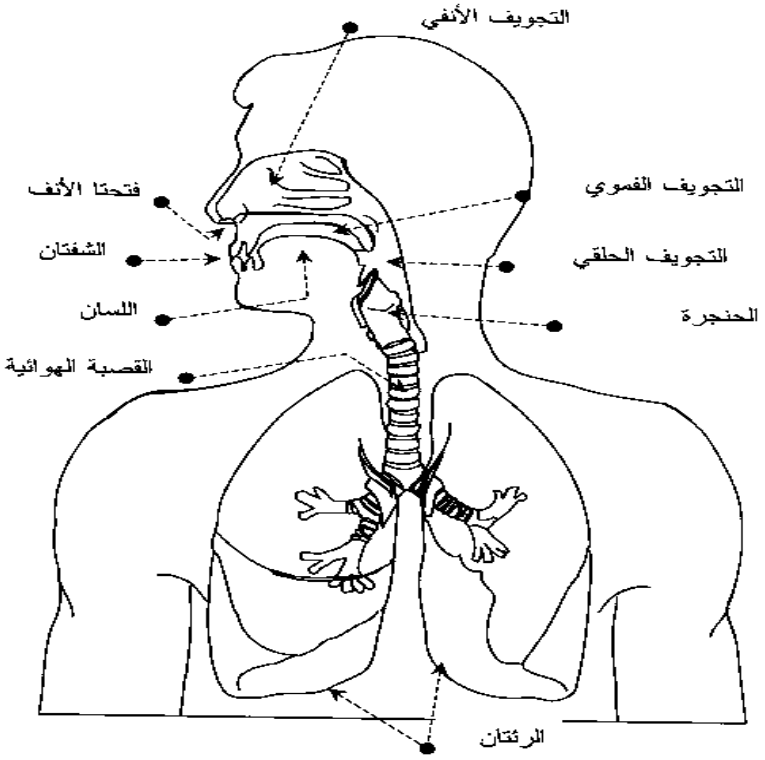
إن السبب القريب للصوت تموج الهواء بسرعة و قوة من أي سبب كان ثم إن سبب هذا التموج قد يكون [3]:

- قرعا لجسم كثيف بجسم كثيف- كما في صوت المطارق يضرب بها الحديد.
- قلعا لجسم من جسم رطب تكون فيه كثافة مع لزوجة كما في قلع الرجل من الطين.
- قرعا للهواء نفسه بجسم يمخره (يخترقه) كما يسمع من دوي الصوت إذا حفزت به الهواء حفزا بقوة و على هذا تزحم الأوتار فترعد، فتصوت.
- تسريبا للهواء الكثير المجتمع في مسلك ضيق يتخلص منه يخرج منه متموجا فيصوت كما في المزامير.

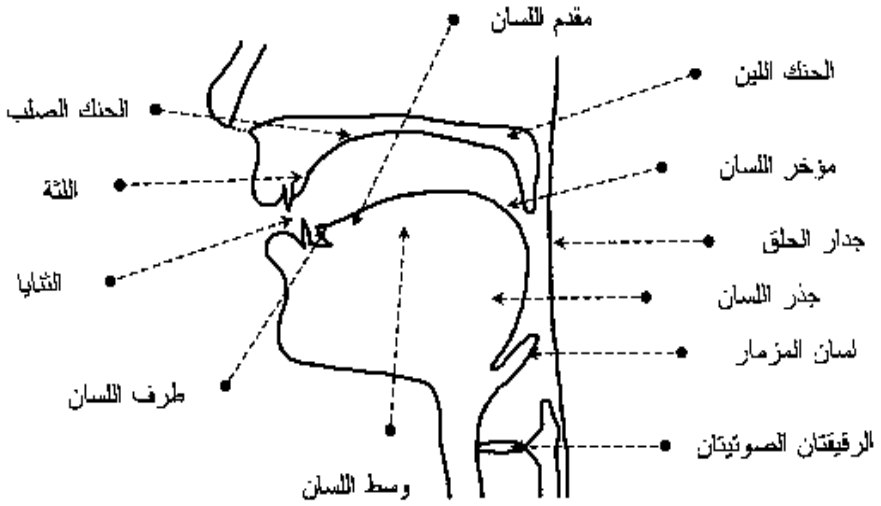
3. الصوت الإنساني و جهاز النطق:

إن جهاز النطق عند الإنسان هو نفسه جهاز التنفس و الذي له دور أساسي و هو دور بيولوجي يتمثل في عملية التنفس التي يضمن بها الإنسان الحفاظ على حياته بالإضافة إلى دور اللسان في التذوق و الأنف في

الشم و لهذا تعتبر أهمية هذه الأعضاء في النطق ثانوية بالمقارنة بأهميتها البيولوجية و التي تتمثل في الحفاظ على حياة الإنسان[4].



الشكل 1.1 : الجهاز الصوتي و الجهاز التنفسي[5]



الشكل 2.1: أعضاء النطق في الجهاز الصوتي[5]

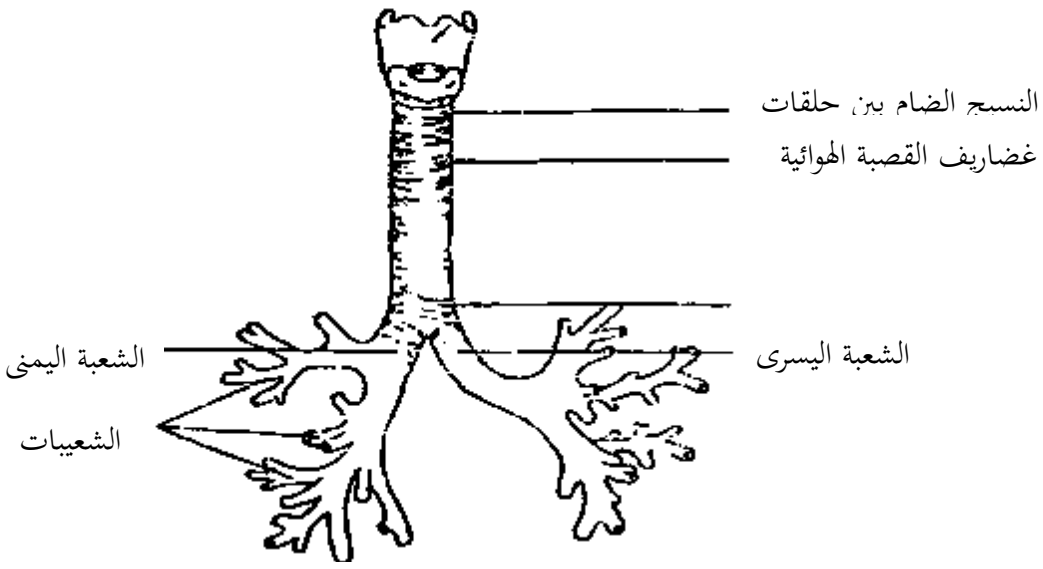
و يتكون جهاز النطق من الأعضاء التالية: الرئتان، القصبة الهوائية، الحنجرة، الوتران (الحبلان) الصوتيان، الحلق، تجويف الفم و يحتوي على (اللسان ، اللهاة، الحنك الأعلى ، الأسنان، اللثة، الشفتان)، التجويف الأنفي (الخيشوم) و في ما يلي تفصيل الكلام عن هذه الأعضاء:

1.3. الرئتان:

و هما كتلتان مخروطيتان قابلتان للتمدد و الانكماش، و يهمنا من عملهما أنهما تدفعان الهواء بعد دخوله إليهما في عملية التنفس فيخرج من تجويف جهاز النطق و تتكون الحروف في أثناء خروجه[7].

2.3. القصبة الهوائية:

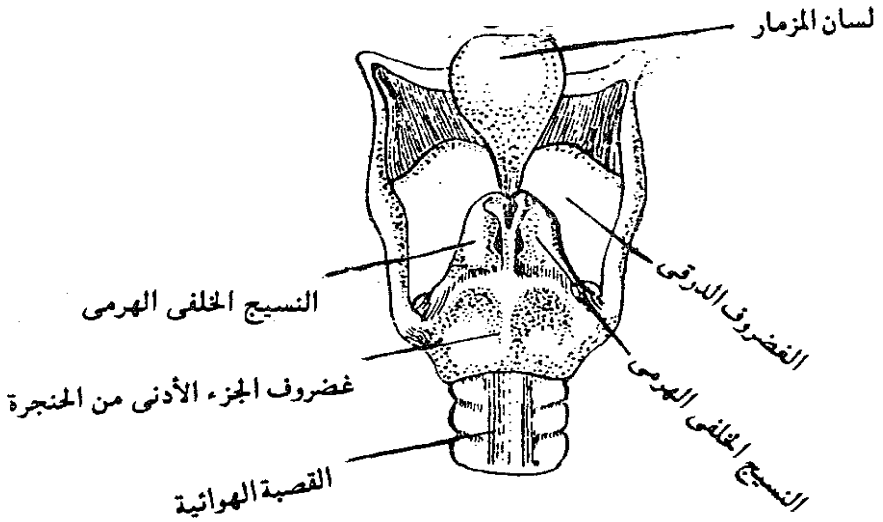
تقع القصبة الهوائية تحت الحنجرة، و هي عبارة عن أنبوبة مرنة أعلاها قاعدة للحنجرة و أسفلها تتفرع إلى أنابيب اقل منها قطرا، تصل إلى الرئتين و تتشعب فيهما لتوصل الهواء إليهما، و دورها في إحداث الصوت أنها توصل الهواء الخارج من الرئة إلى الحنجرة و ما فوقها، كما أنها تعد غرفة رنين للصوت حين يحدث في الأجزاء العليا منها ،لأنها دائما منفتحة لا تنطبق بحكم الحلقات الغضروفية المكونة لها[6][7].



الشكل 3.1: القصبة الهوائية و الحنجرة[5]

3.3. الحنجرة:

الحنجرة عبارة عن صندوق غضروفي يقع على قمة القصبة الهوائية و هي مفتوحة من الأعلى و من الأسفل و هذا يسمح بمرور الهواء من القصبة الهوائية إلى الحلق فالفم أو الأنف و العكس[5].



الشكل 4.1: الحنجرة من الخلف[5]

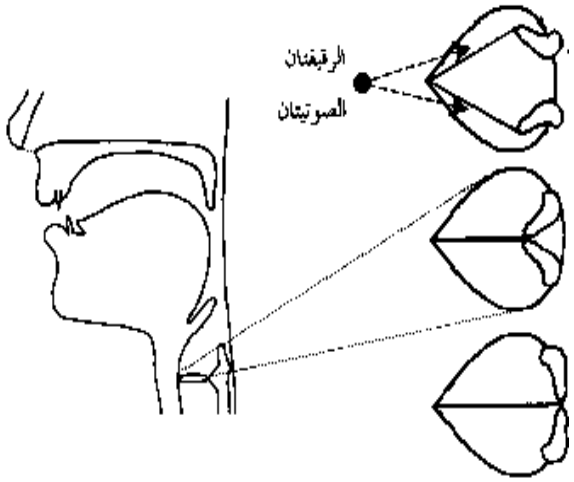
و تتكون الحنجرة من عدة أجزاء و هي:

1.3.3.الغضروف الحلقي: و هو قاعدة الحنجرة و يعد أعلى القصبة الهوائية أيضا[8].

2.3.3. الغضروف الدرقي: هو اكبر غضاريف الحنجرة و هو ما يسمى بالبروز الحنجري (أو تفاحة ادم) [8].

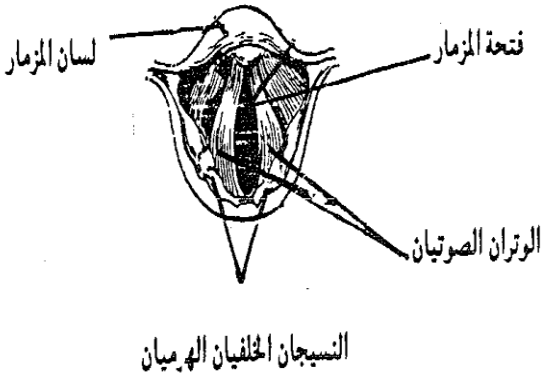
3.3.3. الغضروفان الهرميان: و كل منهما هرمي الشكل له ثلاثة رءوس تشمل قاعدته على نتوئين: احدهما من الأمام و يسمى النتوء الصوتي لاتصال الوترين الصوتيين به، و الثاني خلفي و يسمى النتوء العضلي حيث تنغرز فيه ألياف العضلات المحركة[6].

4.3.3. الوتران الصوتيان: هما رابطان مرنان يشبهان الشفتين يمتدان أفقيا من الخلف إلى الأمام، و يلتقيان عند البروز الذي نسميه تفاحة ادم، و هما قابلان للحركة من الخلف حيث يتصلان بالغضروفين الهرميين[9][10].



الشكل 5.1: الوتران الصوتيان[5]

لوضع الوترين الصوتيين ثلاث حالات كما في الأشكال الثلاثة التي إلى اليمين: فالشكل الأعلى يوضح وضعهما أثناء نطق الأصوات المهموسة و الثاني يبين وضعهما أثناء نطق الأصوات المجهورة و الأخير في حالة نطق الهمزة، حيث ينغلقان تماما[5].



الشكل 6.1: رسم تبسيطي للوترين الصوتيين و هما مفتوحان

5.3.3. المزمار:

هو الفتحة الواقعة بين الوترين الصوتيين، و هي تنقبض و تنبسط بنسب مختلفة مع الأصوات، مما يؤدي إلى اختلاف شد الوترين الصوتيين و استعدادهما للاهتزاز، و هي تكون في وضع انفتاح مع الصوت المهموس، و في وضع فتح و إغلاق متكرر مع الصوت المجهور، و في وضع تضييق مع الصوت الموشوش، و تدعى الأصوات التي ينتجها أصواتا حنجرية، أو مزمارية[11].

6.3.3. لسان المزمار:

للمزمار غطاء نسميه لسان المزمار و يقع في المنطقة التي خلف مستقر اللسان مباشرة و لهذا الغضروف وظائف صوتية منها، انه يعمل على تكييف الرنين بما يحدثه من تغيير في حجم الحنجرة[12].

7.3.3. العظم اللامي:

و هو قوس عظمي قاعدته إلى الجانب الأمامي من الرقبة، و طرفاه إلى الداخل و هو يشكل امتدادا إلى أعلى للجدار الأمامي للحنجرة مع انه لا يرتكز عليه و إنما يربطه به غشاء رقيق[6].

4.3. الحلق:

هو الجزء الذي بين الحنجرة و الفم، و فضلا عن انه مخرج لأصوات لغوية خاصة، يستغل بصفة عامة كفراغ رنان يضخم بعض الأصوات بعد صدورها من الحنجرة و هو ينقسم إلى ثلاثة أقسام و هي[4]:

1.4.3. الحلق الأقصى:

و هو الفراغ الممتد بين سقف الحنجرة، و فم المريء من أسفل، و لسان المزمار و العظم اللامي عند جذع اللسان من أعلى، و هو قابل لان يتسع أو يضيق بتقدم أو تأخر جذع اللسان[6].

2.4.3. الحلق الأدنى:

يبدأ من العظم اللامي حتى مؤخرة الحنك الرخو (الطبق) فيكون أعلاه فتحتان: الخلفية أول تجويف الأنف من الداخل، و الأمامية أول تجويف الفم من الداخل أيضا، و أما الجدار الأمامي للحلق الأدنى فهو أعلى جذع اللسان[6].

3.4.3. فتحة الخيشوم الخلفية:

و هي فتحة في أعلى الحلق الأدنى تبدأ من خلف نهاية سقف الحنك الرخو (الطبق) حيث تتصل بالأنف عن طريق الخياشيم الخلفية.

ملاحظة:

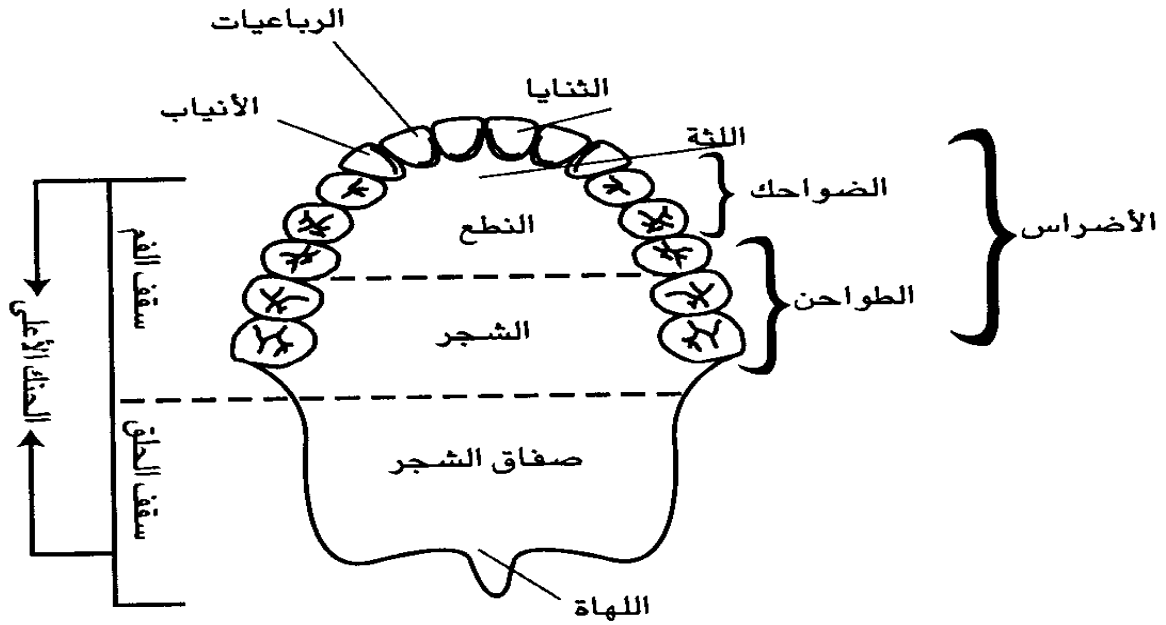
الحلق عند علماء التجويد يختلف عنه عند المحدثين في علم الصوتيات، فهو عند علماء التجويد يمثل أقصى الحلق و وسطه و أدناه فمجموع هذه المخارج الثلاثة يعطي عندهم الحلق، بينما في علم الصوتيات يصفون الحلق بأنه ذلك التجويف الواقع فوق الحنجرة و هو ينتهي عند اللهاة و غشاء الحنك الأعلى من جهة، و أصل اللسان من جهة أخرى، أي هو ما يمثل منطقة أدنى الحلق عند علماء التجويد[13].

5.3 تجويف الفم:

يعتبر أهم التجاويف التي فوق الحنجرة لدوره الكبير في إنتاج الأصوات، لاحتوائه على عدد كبير من أعضاء النطق التي تساهم في تشكيل الأصوات، بالإضافة إلى انه يمكن أن يتغير بشكل كبير في الشكل و الحجم عن طريق تحريك اللسان، و هو يحتوي على الأعضاء التالية[10]:

1.5.3. الحنك الأعلى:

هو الجزء الأعلى من تجويف الفم يبدأ من لثة الأسنان العليا حتى اللهاة وهو يدخل في تكوين مخارج كثير من الأصوات باتصال اللسان في أوضاعه المختلفة مع ما يقابله من أجزاء الحنك الأعلى[13].



الشكل 7.1: الحنك الأعلى[2]

و يتكون الحنك الأعلى من:

2.5.3. اللثة:

و هي الغشاء اللحمي الذي يكسو أصول الأسنان، و هو أملس تقريبا[6].

3.5.3. الحنك الصلب:

يقع تحت الحجرة الأنفية حيث يفصل بينها و بين الحجرة الفموية و له وظائف نطقية كثيرة، فهو يستعمل موضعا لنطق عدد من الأصوات و يساعد أيضا على إحداث عملية الرنين[12].

4.5.3. الحنك اللين:

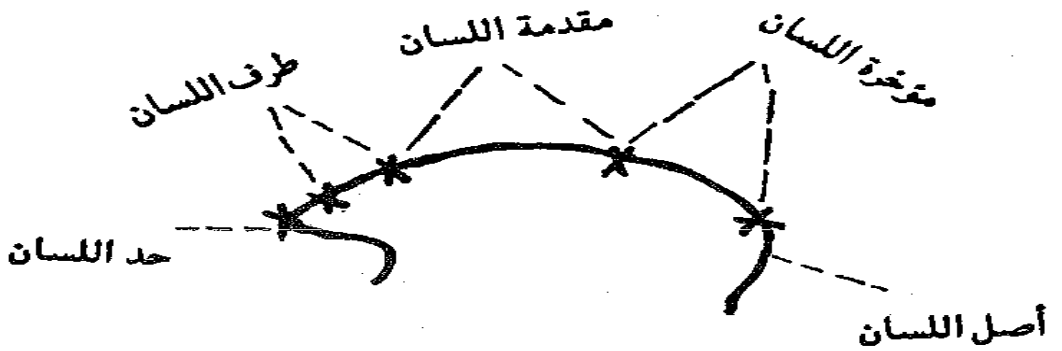
و يسمى أيضا الطبق و هو يستجيب للضغط بالأصبع أو بأسلة اللسان كما يستجيب الإسفنج لذلك، و تسمى الأصوات التي تنتمي إلى هذه المنطقة (الأصوات الطبقية) [12].

5.5.3. اللهاة:

و هي عضلة متحركة مخروطية الشكل متدلية إلى أسفل من الطرف الخلفي للحنك اللين و هي مرنة و قابلة للتحرك الوظيفي، و تؤدي دورا مهما في الغنة و الصوت الأنفي[7].

6.5.3. اللسان:

يعتبر من أهم الأعضاء في عملية النطق و لهذا تسمى به اللغة، لأنه مرن و كثير الحركة في الفم عند النطق، فهو ينتقل من وضع إلى آخر فيكيف الصوت اللغوي حسب أوضاعه المختلفة[4].



الشكل 8.1: أجزاء اللسان[2]

7.5.3. الأسنـان:

الوظائف النطقية للأسنان لا تقل أهمية عن الهضمية فثمة أصوات لا تنطق بصورة صحيحة في حال عدم وجود الأسنان مثلما تنطق مع وجودها، و للإنسان 32 سنا, 16 في الفك العلوي و مثلها في الفك السفلي.

8.5.3. الشفتان:

الشفتان ثنيتان لحميتان عند انطباقهما تغطيان الفم من فوق و من تحت و هما من أعضاء النطق المهمة، لأنهما قابلتان للحركة بداعي الرغبة، فهما ذواتا حركة إرادية، و تتخذان عدة أوضاع عند النطق مما يؤثر في صفات الأصوات و أنواعها، فهما إما تنفرجان، أو تستديران أو تنطبقان.

6.3. التجويف الأنفي (الخيشوم):

و هو تجويف يندفع الهواء من خلاله عندما ينخفض الحنك اللين فيفتح الطريق أمام الهواء الخارج من الرئتين ليمر عن طريق الأنف، و هذه هي الحال عند النطق بالنون و الميم[9].

4. الصوت اللغوي:

هو اثر سمعي يصدر باختيار الإنسان من تلك الأعضاء المسماة أعضاء النطق، و هذا الأثر يظهر في صورة ذبذبات معدلة و موائمة لحركات الفم بأعضائه المختلفة حيث أن كل اللغات تتكون من أصوات تصدرها أعضاء النطق، هذه الأصوات—لتصبح ذات معنى- يجب أن توضع في شكل تتابعي محدد مكونة الكلمات[14].

في اللغة العربية 34 صوتا لغويا متميزة و مستعملة هي الهمزة ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع غ ف ق ك ل م ن هـ و ي ثم الفتحة و الضمة و الكسرة و ألف المد و واو المد و ياء المد[6].

1.4. تصنيف أصوات اللغة العربية:

بعد مرور الصوت من الحنجرة يأخذ طريقه في الفم إلى الخارج, و في الفم تتحقق تفرقة بين نوعين من الأصوات اللغوية, فإذا اعترض طريق الهواء الخارج من الحنجرة جزء من أجزاء الفم نتجت الصوامت مثل الباء و

التاء و الثاء...، و إذا لم يحدث هذا الاعتراض نتجت الحركات و هي الفتحة و الضمة و الكسرة و الفتحة الطويلة أو ألف المد و الضمة الطويلة أو واو المد و الكسرة الطويلة أو ياء المد.

و قد اتفق علماء الأصوات على تقسيم الأصوات اللغوية قسمين رئيسيين، و هما:

1.1.4. الصامت:

هو الصوت المجهور أو المهموس الذي يحدث أثناء نطقه اعتراض في مجرى الهواء في الفم سواء أكان الاعتراض كاملاً كما في نطق صوت الدال، أم كان الاعتراض جزئيا يسمح بمرور الهواء بصورة ينتج عنه احتكاك مسموع كالذال، و يدخل فيها الأصوات التي لا يمر الهواء أثناء النطق بها من الفم إنما يمر من الأنف كالنون و الميم، و الأصوات التي ينحرف هواؤها فلا يخرج من وسط الفم و إنما يخرج من جانبيه أو احدهما كاللام[9].

2.1.4. الصائت (الحركة):

هو الصوت المجهور الذي يحدث أثناء نطقه أن يمر الهواء حرا طليقا خلال الحلق و الفم دون أن يقف في طريقه عائق و دون أن يضيق مجرى الهواء ضيقا من شانه أن يحدث احتكاكا مسموعا.

و الصوامت مستقلة و متمايزة عن بعضها البعض وفق معايير، و هي:

- دراسة موضع النطق، و عمل الأعضاء على إخراج الصوت، و تدرس تحت عنوان مخارج الأصوات.
- دراسة العائق الذي يعترض الهواء، و تأثيره على درجة انفتاح الآلة المصوتة، أو على درجة إقفالها.
- دراسة ما يرافق اجتياز الهواء الحاجز أو العائق، و كيفية التلفظ بالصوامت.

كما استخرج علماء الأصوات العربية صفات كالاستعلاء و الإستفال و الإطباق و الانفتاح، و غيرها.

2.4. مخارج الصوامت(الحروف) العربية:

مخرج الحرف هو الموضع الذي يتولد فيه الحرف و يخرج [6] ويمكن أن يقع مخرج الصوت في أي مكان من الجهاز الصوتي يكون فيه على الأقل عضو صوتي متحرك بدءا بالشفتين و انتهاء بالوترين الصوتيتين، و يوصف

الصوت باسم المكان الذي يخرج منه[5].

من الملاحظ انه يمكن تقسيم الجهاز النطقي إلى قسمين[16]:

أ– القسم المتحرك: ويضم الرئتين، و الوترين الصوتيين، و الحنجرة، و اللهاة، و الطبق، و الفك الأسفل، و اللسان، و الشفتين.

ب– القسم الثابت: ويضم الجدار الخلفي للحلق، و الغار، و اللثة، و الأسنان.

5. تقسيم الصوامت حسب مخارجها:

1.5. الأصوات الحنجرية (المزمارية):

الهمزة: تنتج بالغلق المحكم للوترين الصوتيين، ثم انفراجهما دفعة واحدة.

الهاء: تنتج بانفراج الوترين الصوتيين انفراجا كبيرا بحيث يمر بينهما الهواء الخارج من الرئتين إلى الخارج عن طريق الفم دون أن يعترض طريقه عائق أو تضييق.

2.5. الأصوات الحلقية:

الحاء: ينطق بخروج الهواء من الحنجرة دون اهتزاز الوترين، ثم يخرج محتكا بجدران الحلق، ليخرج من الفم.

العين: تنطق كالحاء غير أن الوترين يهتزان معها و هي أدخل من منطقة احتكاك الحاء فهي النظير المهتز للحاء.

3.5. الأصوات اللهوية:

القاف: عند مرور الهواء من الحنجرة لا يهتز الوتران و يتشكل عائق يمنع هذا الهواء من المرور بالتقاء اللهاة مع ما يقابلها من أقصى اللسان، و بانفراجهما يمر الهواء و ينتج صوت القاف.

4.5. الأصوات الطبقية:

الغين: يمر الهواء الخارج بالحنجرة فيهتز الوتران، حتى إذا وصل إلى أصل اللهاة من الحنك يجد ممرا ضيقا بين أصل اللهاة و أقصى اللسان، فيخرج و ينتج عنه صوت الغين.

الخاء: تنطق مثل الغين إلا أن الوترين لا يهتزان، إذن فالخاء هي النظير غير المهتز للغين.

الكاف: يخرج الهواء مارا بالحنجرة بدون اهتزاز للوترين، و يعترض طريقه عائق ناتج عن التقاء أقصى اللسان و أقصى الحنك الصلب بما يجاور الحنك اللين أو يشترك معه ما يمنع الهواء من المرور، ثم يعقبه تباعد بين أقصى اللسان و أقصى الحنك الصلب ينتج عنه صوت الكاف.

5.5. الأصوات الغارية:

هي الأصوات التي تخرج من بين الغار و اللثة مع مقدم اللسان، و هي:

الشين: تنتج بمرور الهواء من الحنجرة بدون أن يهتز الوتران، ثم يعرض له تضييق بين مقدم اللسان و مقدم الحنك و مؤخر اللثة يخرج الهواء من خلاله فينتج صوت الشين.

الياء: تنطق بخروج الهواء مارا بالحنجرة، فيهتز الوتران، و في الفم يتقلص اللسان إلى الخلف و بارتفاع وسطه نحو ما فوق مقدم الحنك من وسطه حتى يقترب منه مكونا ممرا ضيقا يخرج الهواء منه محدثا صوت الياء.

الجيم: يخرج الهواء مارا بالحنجرة فيهتز الوتران، ثم يعرض له حاجز بالتقاء وسط مقدم اللسان بمقدم الحنك التقاء محكما ثم يفارقه في شيء من البطء، فينتج صوت الجيم.

6.5. الأصوات اللثوية:

و هي التي يشترك فيها مقدم اللسان مع اللثة العليا:

اللام: عند مرور الهواء في الحنجرة يهتز الوتران، ثم يجد عائقا ناتج الالتقاء بين طرف اللسان و أعلى لثة الثنايا العليا، و لكن اللسان يفسح جانبا للهواء فيخرج من جانبي اللسان مع بقاء الغلق عند طرف اللسان، و لهذا وصفت اللام بأنها جانبية.

النون: تخرج بمرور الهواء من الحنجرة فيهتز الوتران، و يمر الهواء عن طريق الأنف بعد أن فتحت له اللهاة الطريق إلى التجويف الأنفي، فيخرج منها محتكا بجدران الأنف و فراغاته، فنسمع صوت النون.

الراء: صوت ينطق بمرور الهواء من الحنجرة، فيهتز الوتران، و في الفم يمتد طرف اللسان إلى لثة الثنايا العليا، و يطرقها عدة طرقات سريعة فيخرج الصوت مكررا.

7.5. الأصوات الأسنانية اللثوية:

و هي الأصوات التي تنطق باشتراك مقدم اللسان بما فيه طرفه مع اللثة و أصول الثنايا العليا:

الضاد: يخرج الهواء مارا بالحنجرة فيهتز الوتران، و في الفم يحدث تضييق بين حافتي اللسان – أو إحداهما– و ما يحاذيهما من الأضراس العليا، كما يحدث غلق محكم بين مقدم اللسان و اللثة و أصول الثنايا العليا، و بخروج الهواء من الشدقين أو احدهما ينتج صوت الضاد.

الدال: يمر الهواء الخارج من الرئتين بالحنجرة فيهتز الوتران، ثم يعترض طريقه تضييق ناتج عن التقاء طرف اللسان بلثة الثنايا العليا التقاء محكما مع لمس أعلى صفحتها، بدون ارتفاع لمؤخر اللسان أو تقعر لوسطه و بانفراجهما ينتج حرف الدال.

الطاء: ينتج مثل الدال مع ارتفاع لمؤخر اللسان و تقعر لوسطه فالفرق بينهما هو الإطباق.

التاء: ينتج مثل الدال و لكن بدون اهتزاز للوترين، إذن فالنظير المهتز للتاء هو الدال.

السين: تنطق عن طريق مرور الهواء بالحنجرة دون أن يهتز الوتران، و بسبب التضييق الذي في الفم بين أسلة اللسان و صفحة الثنايا العليا و بدون ارتفاع لمؤخر اللسان أو طرفه و عدم تقعر لوسطه، يتكون ممر ضيق يخرج منه الهواء محدثا صفيرا نسميه السين.

الزاي: و ينطق مثل السين و لكن مع اهتزاز الوترين، و لهذا فإن الفرق بين الزاي و السين هو اهتزاز الوترين.

الصاد: و ينطق مثل السين، مع ارتفاع لمؤخر اللسان و طرفه و تقعر وسطه و هي صفة الإطباق، و هذا هو الفرق الوحيد بين السين و الصاد، إذن فالصاد هي النظير المطبق للسين.

8.5. الأصوات الأسنانية:

هي أصوات تخرج بامتداد طرف اللسان بين الثنايا العليا و السفلى و هي:

الذال: يمر الهواء الخارج من الرئتين بالحنجرة فيهتز الوتران، ثم يعترض طريقه تضييق ناتج عن امتداد طرف اللسان بين أطراف الثنايا العليا و السفلى، فيخرج الهواء محتكا فينتج صوت الذال.

الظاء: ينطق مثل الذال، و لكن الفرق بين الظاء و الذال هو الإطباق الناتج عن وجود تضييق بين مؤخر اللسان و مؤخر الحنك، إذن فان النظير المطبق للذال هو الظاء.

الثاء: ينطق مثل الذال أيضا، و لكن بدون أن يهتز الوتران، لهذا فإن الذال هي النظير المهتز للثاء.

9.5. الأصوات الأسنانية الشفوية:

تخرج من بين الثنايا العليا و الشفة السفلى و هو صوت الفاء:

الفاء: يمر الهواء الخارج من الرئتين بالحنجرة دون أن يهتز الوتران، ثم يعترض طريقه تضييق ناتج عن التقاء أطراف الثنايا العليا بباطن الشفة السفلى، فيخرج الهواء محتكا فينتج صوت الفاء.

10.5. الأصوات الشفوية:

و هي الأصوات التي تخرج من بين الشفتين:

الباء: ينطق بمرور الهواء الخارج من الرئتين بالحنجرة فيهتز الوتران، ويستمر الهواء في الخروج حتى يصل إلى الشفتين، حيث تنطبقان انطباقا محكما في نقطة اقرب إلى باطنهما ثم تنفرجان فيحدث انفجار تسمع معه صوت الباء.

الميم: ينطق كالباء و لكن انطباق الشفتين يكون في نقطة اقرب إلى ظاهرهما و أثناء هذا الانطباق تهبط اللهاة، فيمر الهواء من الأنف وتكتسب صفة الغنة كالنون.

الواو: بمرور الهواء من الحنجرة يهتز الوتران، فإذا وصل اللسان ارتفع أقصاه إلى ما يقابله من أقصى الحنك، فيضيق الممر بينهما، وتستدير الشفتان مكونة فتحة دائرية ضيقة، فيخرج الهواء محدثا صوت الواو.

6. صفات أصوات اللغة العربية وتقسيم الصوامت حسبها :

صفة الحرف حلية صوتية تصحبه عند نطقه، و هي وسيلة للتمييز بين الحروف التي لها نفس المخرج، و الصفات قسمان، قسم يمثل الصفات العامة و هي خمس ثنائيات لابد للحرف أن يتصف بصفة من كل ثنائية منها, و قسم يتمثل في صفات خاصة تتصف بها بعض الحروف بشروط معينة، و ثنائيات الصفات العامة هي: الجهر أو الهمس و الشدة أو الرخاوة،و الاستعلاء أو الإستفال، الإطباق أو الانفتاح و الذلاقة أو الاصمات[6].

1.6. الصفات العامة:

1.1.6. الجهر:

تقسيم الأصوات إلى مجهورة و مهموسة يرتبط باهتزاز الوترين الصوتيين، حيث يقترب الوتران الصوتيان احدهما من الآخر، فيتشكل بينهما فراغ يسمح بمرور الهواء مع إحداث اهتزازات منتظمة لهما.

و الأصوات المجهورة، كما عددها سيبويه، تسعة عشر صوتا، و هي: الهمزة، و الألف، و العين، و الغين، و القاف، و الجيم، و الياء، و الضاد، و اللام، و النون، و الراء، و الطاء، و الدال، و الزاي، و الظاء، و الذال، و الباء، و الميم ، و الواو[16]، و قال عن حروف الجهر أنها:(حروف اشبع الاعتماد في مواضعها فمنع النفس أن يجري معها حتى ينقضي الاعتماد فيجري النفس).

2.1.6. الهمس:

الهمس عكس الجهر، وفيه ينفرج الوتران الصوتيان بعضهما عن بعض أثناء مرور الهواء، و لا يهتزان، و لا يحدثان ذبذبات، لان انفراجهما يسمح للهواء بالخروج دون أن يقابله أي عائق.

و الأصوات المهموسة اثنا عشر صوتا عند بعض المحدثين و هي: ت، ث، ح، خ، س، ش، ص، ط، ف، ق، ك، هـ.

أما سيبويه فقد جعل الأصوات المهموسة عشرة و هي: الهاء و الخاء و الكاف و الشين و السين و التاء و الصاد و الثاء و الفاء و قال: أن الصوت المهموس هو حرف اضعف الاعتماد في موضعه حتى جرى النفس معه.

ويوصف الصوت المهموس بأنه الصوت الخفي، وحروفه مجموعة في: (فحثه شخص سكت) [18].

3.1.6. الشدة:

انحباس الصوت ثم خروجه في قوة و جمعها بن الجزري في قوله (أجد قط بكت) [18] و سميت شديدة لأنها تمنع الصوت من أن يجري فيها لشدتها و صلابتها, يقول سيبويه (و من الحروف – الشديدة – و هو الذي يمنع الصوت أن يجري فيه) و تابعه في ذلك جملة من علماء اللغة و قال الانباري هي: (حروف صلبة لا يجري فيها الصوت) و للمبرد في تفسير الشديدة مذهب يختلف عن سيبويه فعنده أنها حروف تمنع النفس, و أما المجهورة فهي التي يرتعد الصوت فيها،و أما الرماني فيرى أن الشديد هو الذي يشتد الاعتماد فيه بلزوم موضعه لا شدة الوقع،والمجهور يقوى الاعتماد فيه بشدة الوقع وأضاف المحدثون إلى الأصوات الثمانية الشديدة صوت الضاد[19]

4.1.6. الرخاوة:

و هي أن يسمح بمرور النفس و سميت رخوة لأنها لا تمنع الصوت أن يجري فيها لرخاوتها ،فتقول : المس و الرش فتجد الصوت جاريا مع السين و الشين، و الحروف الرخوة ثلاثة عشر حرفا، ما عدا الشديدة ومابين الرخوة والشديدة، وهذه الحروف هي : (ه ح غ خ ش س ص ض ز س ظ ذ ف) و يسمي المحدثون الأصوات الرخوة: احتكاكية، في حين يسمون الشديدة :انفجارية و جعلوا الرخوة خمسة عشر حرفا بإخراج الضاد و إضافة الواو و الياء و العين[19].

5.1.6. بين الشدة والرخاوة(البينية):

أصوات يكون جريان الصوت في مخرجها جزئيا لعدم كمال غلقه، وقد أطلق عليها المحدثون مصطلح (المائعة) و هي: ل،ن، ع،م،ر و جمعت في: (لن عمر) و عند آخرين هي ثمانية تجمع على (لم يروعنا)، لان العين

يجري فيها الصوت كالحاء و لم يمنع امتناع غيره، و اللام فهو مع انحرافه و شدته لم يعترض على الصوت كاعتراض الحروف الشديدة، فلا هو مثل الرخوة لان طرف اللسان لا يتجافى عن موضعه، و لا مثل الشديدة فيمنع الصوت، و الراء يجري فيه الصوت تكريره و انحرافه إلى اللام، و النون و الميم تجري فيها الغنة من الأنف، و الواو والياء يتسع لهما المخرج أكثر من اتساع غيرهما، فيمتد الصوت بهما كونهما لينين،و أما الألف فاشد اتسعا[19].

6.1.6. الاستعلاء:

سميت ذلك لتصعد الصوت و استعلائه بعد اعتمادك على مخارجها، و يجوز أن تكون سميت مستعلية، لخروج صوتها من جهة العلو، و أصوات الاستعلاء سبعة، و هي: خ،ص،ض،ط،ظ،غ،ق و يجمعها قولك (خص ضغط قض)، و بين المستعلية و المنطبقة فارق، هو أن المستعلية نوعان: منها ما يستعلي ثم ينطبق، و هن حروف الإطباق، و منها ما يستعلي و لا ينطبق كالخاء و الغين و القاف و بعضهم أضاف إلى المستعلية غير المنطبقة صوتي الحاء و العين، و أضاف بعض المحدثين إلى حروف الاستعلاء اللام و الراء المفخمتين لارتفاع مؤخرة اللسان بهما.

7.1.6. الإستفال:

الإستفال ضد الاستعلاء و الحروف التي تكتسب صفة الإستفال لا يتصعد فيها اللسان إلى الحنك، و لا يستعلى بعد الاعتماد على المخرج، و حروف الإستفال غير حروف الاستعلاء السبعة (خ غ ق ص ض ط ظ) و هي :ء ، ب، ت، ث، ج، ح، د، ذ، ر، ز، س، ش،ع، ف، ك، ل، م، ن، هـ، و، ي، ا[16].

8.1.6. الإطباق:

و هو ارتفاع مؤخر اللسان حتى يقترب من الحنك أثناء النطق و هو انحصار الصوت الصامت بين اللسان و الحنك الأعلى، لارتفاع ظهر اللسان إلى الحنك الأعلى حتى يلتصق به، و الأصوات المطبقة هي: الصاد و الضاد و الطاء و الظاء، و صفة الإطباق غير الطبقية التي تعبر عن مخرج أصوات معينة منها:ك،غ،خ...

وأضاف المحدثون أصواتا أخرى عدها بعضهم ليست من المطبقة بل مفخمة كالقاف والراء و الغين و اللام[18].

9.1.6. الانفتاح:

الانفتاح هو ضد الإطباق، و يكون بابتعاد اللسان عن الحنك الأعلى و جريان النفس و الصوت عند النطق، دون عائق بين اللسان و الحنك، أي انه يكون نتيجة انفراج ظهر اللسان و عدم إطباقه على الحنك الأعلى عند النطق، و أصوات الانفتاح غير أصوات الإطباق و هي:ء، ب، ت، ث، ج، ح، خ، د، ذ، ر، ز، س، ش، ع، غ، ف، ق، ك، ل، م، ن، هـ، و، ي[16].

10.1.6. الذلاقة:

الذلق لغة هو الطرف و هو سرعة النطق بالحرف في سهولة و يسر بالاعتماد على طرفي اللسان و الشفة عند النطق، و يجب عدم الخلط بين الأصوات الذلقية مخرجا، و المذلقة صفة فالأصوات الذلقية لا تخرج إلا من ذلق اللسان أي من طرفه، و هي اللام ، و النون، و الراء، أما الأصوات المذلقة، فهي التي تخرج من ذلق اللسان ،كالراء ،و اللام ، والنون، أو من ذلق الشفة و هي: الباء، و الفاء، و الميم، و لقد وصفت هذه الأصوات بالمذلقة لاتفاق المعنى اللغوي مع ما لهذه الأصوات من الخفة، و سهولة النطق دون كلفة، سواء أكان ذلك من ذلق اللسان أو من ذلق الشفة[16].

11.1.6. الإصمات:

الأصوات المصمتة ضد المذلقة، و هي أصوات العربية ماعدا الأصوات المذلقة الستة : الراء، اللام، النون، الباء ، الفاء، و الميم و سميت هذه الأصوات مصمتة لأنها أصمتت أي منعت أن تختص ببناء كلمة في لغة العرب حروفها أكثر من ثلاث، و حروف الاصمات هي: أ، ت، ث، ج، ح، خ، د، ذ، ر، ز، س، ش، ص، ض، ط، ظ، ع، غ، ق، ك، هـ، و، ى[18].

2.6. صفات خاصة:

1.2.6. الصفير:

هو صوت سببه انحصار النفس عند خروجه من بين صفحتي الثنايا العليا و طرف اللسان.

2.2.6. الإنحراف:

و هو تجافي جانبي اللسان دون طرفه لترك منفذ يخرج منه الصوت و هذا ينطبق على اللام.

3.2.6. التكرير :

و هو ارتعاد طرف اللسان عند خروج الصوت، و الصوت الذي يتسم بهذه الخاصية هو الراء.

4.2.6. التفشي :

و هو انتشار الهواء بين الفم و الأسنان عند النطق بالحرف، ويحدث عند نطق حرف الشين[18].

5.2.6. الإستطالة:

و هو امتداد الصوت بالضاد من أول حافة اللسان إلى آخرها، أي حتى تتصل بمخرج اللام[16].

قال مكي بن أبي طالب: (لما اجتمع فيها من القوة بالجهر والإطباق والاستعلاء فقويت بذلك واستطالت في الخروج من مخرجها حتى اتصلت باللام لقرب مخرج اللام من مخرجها) [19].

6.2.6. الخفاء:

و هو خفاء الصوت و قد وصفت به الهاء و حروف المد و النون الساكنة قال القرطبي: (أما الخفية فالهاء و الألف و الياء و الواو، و ذلك لاتساع مخرجهن، و مما يشرك هذه الحروف في الخفاء، النون إذا سكنت في غير إظهارها و لا إدغام و لا قلب) [19].

7.2.6. القلقلة:

اضطراب الصوت و تقلقل المخرج عند النطق به و يقول سيبويه عن حروف القلقلة(و اعلم أن من الحروف

حروفا مشربة ضغطت من مواضعها، فإذا وقفت عليها خرج معها من الفم صويت و نبا اللسان عن موضعه، و هي حروف القلقلة)،و أصوات القلقلة خمسة و هي: ب،ج،د،ق،ط و جمعها ابن الجزري في (قطب جد)[16].

8.2.6. الغنة:

الغنة هي جريان الصوت و خروجه من الخيشوم، و أصواتها الميم، و النون[5].

9.2.6. اللين:

هو خروج الصوت في سهولة و امتداد و يحدث عند خروج الواو والياء بشرط أن يكونا ساكنين بعد فتح.

10.2.6. الهاوي:

هو الصوت اللين الذي يتسع فيه تجويف الفم، و يسمى الألف من حروف المد بالهاوي، قال سيبويه (الهاوي هو حرف اتسع لهواء الصوت مخرجه اشد من اتساع مخرج الياء والواو لأنك قد تضم شفتيك في الواو ، و ترفع في الياء لسانك قبل الحنك و هي الألف) و لكن الخليل _شيخه _قد أطلق على الحروف الألف و الواو و الياء و الهمزة اسم الهاوية و سماها أحرفا جوفا[19].

7.الحركات:

الحركات أصوات تحدث من تذبذب الوترين الصوتيين عندما يندفع الهواء من الرئتين ليمر بالحنجرة و بعدها يخرج مارا بالحلق و الفم، دون أن يعترض طريقه حاجز أو عائق أو تضييق من شانه أن يكسبه صفة الصوامت حيث انه ليس للفم دور في تشكيلها سوى انه يعتبر غرفة رنين تأخذ شكلا معينا مع كل نوع من الحركات[4].

تتنوع الحركات في مختلف الأنظمة اللغوية، و اقل عدد يمكن أن يحويه نظام لغوي من الحركات في اللغات المعروفة يتكون من ثلاث حركات هي (a) (u) (i) و هو الشكل الموجود في اللغة العربية الفصحى وتسمى بالفتحة (a) و الضمة (u) و الكسرة (i) [4].

و للحركات مجموعة من الصفات تتميز بها عن بقية الأصوات و هي:

- خروج الهواء من الفم حرا دون أن يعترض مجراه عائق، أو تضييق يحدث احتكاكا مسموعا، و قد ذكر هذا ابن جني عند حديثه عن الألف (الفتحة الطويلة) قد أدرك هذه الخاصية، حيث أشار إلى اتساع مخرجها و خروج الهواء ممتدا حتى ينفذ[9].
- الحركات هي أصوات مجهورة نتيجة تذبذب الوترين الصوتيين عند مرور الهواء من الحنجرة، و هي تختلف عن بعضها البعض بعملية الرنين، لان الممر الذي يسلك الهواء عند خروجه يكون صندوقا رنانا، يغير من طبيعة الصوت الناتج عن ذبذبة الوترين الصوتيين، فتغير شكل هذا الممر، يتسبب في ظهور الصوائت المتمايزة[16].
- الحركات أقوى الأصوات وضوحا في السمع، نتيجة للخاصتين السابقتين، خاصة الأولى منهما[9].

و للحركات عدة خصائص شغلت العلماء و كانت محل اهتمامهم و بحثهم حيث أنهم قاموا بدراسة خصائصها و طريقة نطق كل حركة منها بالإضافة إلى تموضع أعضاء النطق و دور كل منها فوضعوا لذلك مقاييس أشهرها مقياس دانيال جونز الذي عرف الحركة بأنها: صوت مهتز(مجهور)، يخرج الهواء عند النطق به بصفة مستمرة، دون وجود عقبة تعوق خروجه، أو تسبب فيه احتكاكا مسموعا[17].

1.7.الفتحـــة:

الفتحة حركة وسطية قصيرة، نجد أن أعلى نقطة في اللسان أثناء النطق بها هي وسطه فهي حركة متسعة و اللسان معها يكاد يكون مستويا في قاع الفم مع ارتفاع خفيف في وسطه، و يبقى الفم مفتوحا بشكل أوسع، و تكون حجرة الرنين فيه كبيرة أما الشفتان فتكونان أثناء نطق الفتحة مسطحتين منفرجتين، أي أن فراغ الشفتين لا يلعب دورا في إنتاج الفتحة[16].

و لكي تنتج الفتحة فإن خروج الهواء من الرئتين يمر بالمراحل التالية:

- يندفع الهواء من الرئتين و يمر بالحنجرة فيهتز له الوتران الصوتيان.

- يجد الهواء الخارج تضييقا بين مؤخر اللسان و ما يقابله فينتج فراغان في مجرى الهواء، فالأول خلفي محصور بين سطح الحنجرة و مكان التضييق، و الثاني أمامي بين مكان التضييق و الفتحة التي بين الشفتين[17].
- اللسان يكون مستويا في قاع الفم مع ارتفاع بسيط في وسطه.
- تتخذ الشفتان وضع الانبساط[18].
- و بخروج الهواء بعد هذه المراحل ينتج صوت الفتحة القصيرة أو الطويلة.

2.7.الضمـــة:

و هي حركة خلفية قصيرة لان الجزء الخلفي من اللسان يكون اقرب ما يمكن من الحنك اللين و اللهاة أثناء نطقها، و تكون حجرة الرنين الفمية في وضع اللسان هذا ضيقة و الفراغ الفمي اكبر بسبب الانخفاض الكبير للفك السفلي، بالإضافة إلى أن الشفتين تتخذان وضع الاستدارة و تكونان متقدمتين نحو الأمام[16].

و لكي تنتج الضمة فإن خروج الهواء من الرئتين يمر بالمراحل التالية:

- يندفع الهواء من الرئتين و يمر بالحنجرة فيهتز له الوتران الصوتيان.
- يجد الهواء الخارج من الحنجرة تضييقا بسبب تقلص ظهر اللسان إلى الخلف و ارتفاعه إلى الأعلى نحو اللهاة، فيتكون بذلك فراغين في مجرى الهواء الأول خلفي و هو محصور بين سطح الحنجرة و مكان التضييق، و الثاني أمامي فيما بين مكان التضييق، إلى الفتحة التي بين الشفتين.
- تنخفض الحنجرة عند نطق الضمة.
- استدارة الشفتان مع بروزهما إلى الأمام.

3.7.الكسرة:

و هي حركة أمامية قصيرة، لأن الجزء الأمامي من اللسان يكون اقرب ما يكون من الجزء الأمامي من الحنك الصلب أثناء النطق بها، بالإضافة إلى أن حجرة الرنين في داخل الفم تكون في اصغر أحجامها، و يكون

الفم مفتوحا بحده الأدنى، و تكون الشفتان مشدودتين أقصى ما يمكن لهما من الشد[16].

و لكي تنتج الكسرة فإن خروج الهواء من الرئتين يمر بالمراحل التالية[18]:

- يندفع الهواء من الرئتين و يمر بالحنجرة فيهتز له الوتران الصوتيان.
- بارتفاع مقدم اللسان نحو الحنك الصلب ينتج ممر عريض يمر منه الهواء دون أن يحدث احتكاكا مسموعا.
- يتكون فراغان الأول حلقي و الثاني بين اللسان و الحنك الأعلى.
- تكون الشفتان في وضع الانبساط.
- يكون اللسان في أعلى ارتفاع له.

ملاحظة: الحركات تختلف عن بعضها البعض في النطق باختلاف:

1- مستوى ارتفاع اللسان في الفم.

2- المنطقة التي يتم فيها ارتفاع اللسان مقدم الفم، أو وسطه، أو آخره[18].

3- حجم الفراغات الناتجة عن التضييق في مجرى الهواء فوق الحنجرة.

4- وضع الشفتين أثناء النطق.

4.7.الفتحة الطويلة:

و هي امتداد لحركة الفتحة و تنتج باستمرار خروج الهواء عند النطق لفترة أطول من الفتحة القصيرة.

5.7.الضمة الطويلة:

و هي امتداد لحركة الضمة و تنتج باستمرار خروج الهواء عند النطق لفترة أطول من الضمة القصيرة.

6.7.الكسرة الطويلة:

و هي امتداد لحركة الكسرة و تنتج باستمرار خروج الهواء عند النطق لفترة أطول من الكسرة القصيرة.

النظام الفونولوجي العربي حسب تصنيف المحدثين:

– الجدول التصنيفي الوظيفي للأصوات العربية

	مواضع اعتراض مجرى الهواء(المخارج)															مواضع اعتراض مجرى الهواء (المخارج)															
	شفوي					بين أسناني			أسناني لثوي							لثوي مائع			حنكي أمامي				حنكي لهوي				حلقي		حنجري		
المخارج	مزدوج				أسناني				غير صفيري				صفيري										حنكي خلفي	لهوي							
الحروف	ب	م	و	ـُ	ف	ذ	ث	ظ	د	ت	ط	ض	ص	ز	س	ر	ل	ن	ش	ج	ي	ـِ	ك	ق	غ	خ	ع	ح	ء	ه	ـَ
صامت	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	
صائت				+																		+									+
فموي	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
انفي		+																													
شديد	+								+	+	+									+			+	+					+		
رخو	+				+	+	+	+				+	+	+	+				+						+	+	+	+		+	
بيني		+	+													+	+	+			+										
لين			+																												
جانبي																	+														
تكراري																+															
مستفيل	+	+	+	+	+	+	+		+	+				+	+	+	+	+	+	+	+	+	+				+	+	+	+	+
مستعل								+			+	+	+											+	+	+					
مجهور	+	+	+	+		+		+	+			+		+		+	+	+		+	+	+			+		+				+
مفخم								+			+	+	+			+	+							+	+	+					
مرقق	+	+	+	+	+	+	+		+	+				+	+			+	+	+	+	+					+	+	+	+	+
مهموس					+		+			+	+		+		+				+				+	+		+		+	+	+	
منفتح	+	+	+	+	+	+	+		+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
مطبق								+			+	+	+																		

كيفية اعتراض مجرى الهواء (الصفات)

8. الخاتمة:

تحيط بالإنسان الكثير من الظواهر الطبيعية التي قد يكون هو مصدرها تتطلب منه التدبر و إعمال عقله لتحليلها و دراستها وفق ما يتوفر له من معلومات و إمكانيات، و من بين أهم هذه الظواهر الكلام الذي هو في الأصل عبارة عن ظاهرة طبيعية تتمثل في حدوث الصوت، و الذي يعتبر أهم وسائل الإنسان للاتصال لما حبي به من جهاز صوتي يسمح له بإنتاج مجموعة متنوعة من الأصوات و التي كانت محل دراسة العلماء و اللغويين منذ القديم إلى يومنا هذا، حيث أن كل الدراسات و الأبحاث المتعلقة بالكلام لابد أن تمر بدراسة معمقة لمكوناته الأساسية و التي هي الأصوات اللغوية (الفونيمات) لما لها من خصائص و صفات تمكن من فهم الظواهر اللغوية المختلفة و الاستفادة منها، و قد تنوعت أساليب هذه الدراسة بين لغوية تدرس الأصوات من ناحية وظيفتها اللغوية و بين العلمية التي تدرسها كظاهرة طبيعية فتبحث في كيفية إنتاجها و خصائصها الفيزيائية.

قمنا بدراسة الأصوات اللغوية للغة العربية من حيث مصدرها و خصائصها النطقية فتوصلنا إلى النتائج التالية:

- من خلال دراسة البنية التشريحية لأعضاء جهاز النطق و تحديد الأعضاء المسؤولة عن إحداث كل صوت على حدة لاحظنا توزع الأصوات على طول جهاز النطق مما يكسبها صفات تساعد على التمييز بينها بسهولة.
- أصوات اللغة العربية أصوات زفيرية مما يكسبها قوة و وضوحا.
- دراسة الأصوات تمكن نقطها بصورة صحيحة مما يساعد على الاستفادة من خصائصها.
- دراسة مخارج الأصوات وصفاتها تساعد في اختيار المعاملات المناسبة لتحسين أنظمة التعرف على الكلام

الفصـــل الثاني

التعرف الآلي على الكلام

1. مقدمة:

لقد تطورت الدراسات حول الأصوات اللغوية تطورا كبيرا فبعد أن كانت تقتصر على دراسة الأصوات من الجانب النطقي و الجانب البيولوجي و الجانب الفيزيائي ظهر مجال آخر للبحث و هو العلاج الآلي للكلام الذي استفاد من بقية العلوم، فقد أصبحت نظرتنا للأصوات أكثر وضوحا و تعاملنا معها أكثر سهولة لما توصلنا إليه من معلومات حول موجات الصوت و إمكانية استخدامها في مختلف التطبيقات التي يحتاجها الإنسان، و للعلاج الآلي للكلام عدة فروع هي: الفهم الآلي الكلام و التركيب الآلي للكلام، و التعرف الآلي على الكلام و غيرها، حيث يعتبر التعرف من أهمها لما له من تطبيقات و استعمالات.

2. بعض برامج التعرف على الكلام:

لقد مولت وكالة مشروع الأبحاث المتقدمة (ARPA) بوزارة الدفاع في الولايات المتحدة الأمريكية برنامجا استمر لمدة خمس سنوات (1971-1976) للبحث في التعرف على الكلام المتصل، و قد نتج عن هذا المشروع عدة برامج منها Speechlis و HWIM لشركة BBN LBeranek and Bolt (Newman)، و عدة نسخ من HAARPY Hearsay لجامعة كارنيجي ميلون، كما طورتSRI international بالتعاون مع System Development Corporation برنامجSRI SDC و طورت شركة IBM برنامجا باسم Dragon، و قد أدى مجهود مماثل في فرنسا إلى بناء عدة برامج منها Centre de Rechareche en Informatique de (CRIN) in MYRTILLE-I and II Nancy d'Informatique و برنامج ESOP في Laboratoire et de Mechanique pour les Sciences de Centre National d'Etude (CNET) في KEAL و برنامج I'engenieur بجامعة مارسيليا و برنامج Lannion des Telecommunications و برنامج ENSER بجرينبول [21].

و قد توالت الجهود لتطوير برامج التعرف الآلي على الكلام كان منها بعض الدراسات و الأبحاث من طرف العديد من الباحثين متعلقة بالتعرف على الكلام العربي و لكنها كانت قليلة جداً مقارنة باللغات الأخرى كالإنجليزية مثلاً، و من أشهر برامج التعرف الآلي على الكلام العربي نجد برنامجي صخر ASR و IBM Via Voice للإملاء الصوتي العربي.

تواريخ مهمة[22]:

نستطيع تلخيص المراحل المهمة التي مر بها التعرف الآلي على الكلام ببعض التواريخ التالية:

1952: التعرف على 10 أرقام بواسطة جهاز الكتروني سلكي من اجل متكلم واحد

1960: استعمال التقنيات الرقمية

1965: التعرف على الحروف في الكلام المستمر

1968: التعرف على الكلمات المعزولة بواسطة أنظمة على الحواسيب الكبيرة (إلى غية 500 كلمة)

1969: استعمال المعطيات و المعارف اللسانية

1972: أول آلة توجه للتسويق للتعرف على الكلمات

1978: تسويق نظام تعرف بمعالج على بطاقة عبارة عن دارة مطبوعة

1985: تسويق أول أنظمة تعرف على الآلاف من الكلمات.

1986:إطلاق المشروع الياباني هاتف ATR مع الترجمة الآلية في الزمن الحقيقي.

1988: ظهور آلات الإملاء الأولى للكلمات المعزولة

1990: أول تطبيق حقيقي للتخاطب بين الإنسان و الآلة

1994: IBM تطلق نظامها الأول للتعرف الآلي على الكلام على أجهزة الحاسوب PC

1997: إطلاق نظام الإملاء للكلام المستمر من IBM.

3. الموجات الصوتية الناتجة عن الجهاز الصوتي:

إن الجهاز الصوتي عند الإنسان قادر على إصدار أصوات متعددة، أهمها الأصوات اللغوية التي يتألف منها الكلام و ترجع هذه القدرة إلى بنية الجهاز الصوتي الذي يتكون من مجموعة من الأعضاء التي لكل منها شكل معين و موضع محدد، كما تعتبر الحنجرة أهم هذه الأعضاء في إصدار اغلب الأصوات لاحتوائها على الوترين الصوتيين اللذان هما مصدر التردد الأساسي للأصوات اللغوية حيث أن اهتزازهما ينتج موجات ذات تردد منتظم، غير أن هذه الموجات لا تبقى على حالها حتى تخرج من الجهاز الصوتي حيث يعترض طريقها الهواء الموجود داخل التجاويف التي فوق الحنجرة (التجويف الحلقي و التجويف الفموي و التجويف الأنفي) و هي تؤثر على التردد الأساسي للموجة الصوتية فتضفي عليه صفات لم تكن فيه من قبل، فدور هذه التجاويف يظهر في عملية الرنين الذي ينتج عنه تجمع الترددات في مجموعات تعرف بالحزم الصوتية، و هذه الحزم هي التي تشكل نوع الصوت الذي نستطيع من خلاله التمييز بين الأصوات ذات الأنواع المختلفة.

4. طرق تمثيل الكلام:

المقياس الزمني للموجات الصوتية هو الملي ثانية، و يمثل الزمن على المحور الأفقي عند تمثيل الكلام بشكل موجة أو الرسم الطيفي الناتج عن التحليل بالسوناجراف، و لمعرفة الزمن أهمية كبيرة في تحديد و وصف الأصوات بشكل عام و بصفة خاصة تلك الأصوات التي يغلق الممر الصوتي عند نطقها ثم ينفتح محدثا ما نسميه الانفجار مثل ما يحدث عند نطق الدال، كما يلعب الزمن دورا هاما في تحديد الفترة اللازمة للانتقال من صوت إلى آخر.

يتم تمثيل الكلام أي الإشارة الصوتية بثلاث طرق هي [23]:

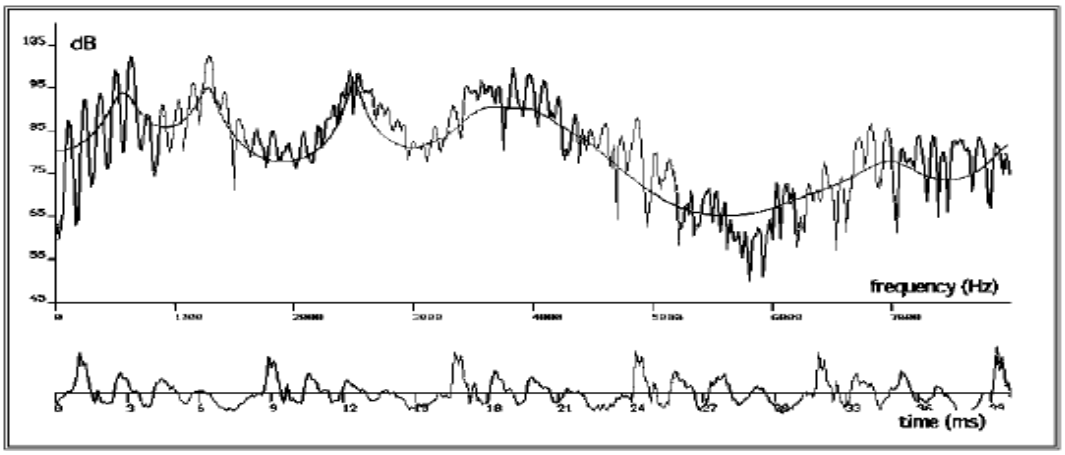
1.4. بشكل موجة:

أي تغيرات الإشارة بدلالة الزمن ، و هذه التغيرات هي إما تغيرات في مطال الإشارة أو طاقتها ، و نسميه عادة (الشكل 2-1 أسفل).

2.4. بشكل طيف الإشارة:

حيث ندرس تغيرات مطال الإشارة أو طاقتها بحسب التردد، و هو ما يمثل عادة (الشكل 2-1 أعلى) و هذه الطريقة أسهل في عملية المعالجة.

تظهر الأشكال التالية كيفية تمثيل الاشارة مع الزمن، و تمثيل طيف هذه الاشارة



الشكل 1.2: تمثيل الاشارة مع الزمن ، و تمثيل طيف الاشارة

3.4. بشكل سوناغرام spectrogram:

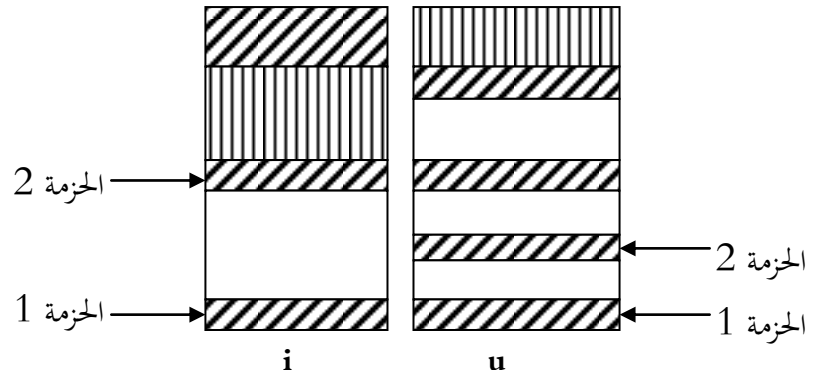
و هو تمثيل ثلاثي الأبعاد يرتبط بالمتغيرات الزمن و التردد و طاقة الحرف حيث نرسم spectrogram كرسم ثنائي البعد (الزمن و التردد) و يتجلى البعد الثالث بدرجات اللون الرمادي التي تعبر عن الطاقة (2-3)

تظهر مجموعة الحزم بشكل شرائط سوداء أفقية في الرسم الطيفي (الحزمة الأولى F1 و الحزمة الثانية F2 و الحزمة الثالثة F3) لها دور كبير في التعرف على الأصوات المنطلقة التي يهتز معها الوتران الصوتيان مثل أصوت الحركات و الأصوات الأنفية كالنون و الميم و الأصوات الجانبية كاللام و الأصوات المستمرة غير الاحتكاكية و يمكن أيضا أن تساعد في دراسة و وصف بعض الأصوات الاحتكاكية مثل الهاء و الخاء.

كما أن درجة سواد شرائط الحزم الصوتية التي تظهر في الرسم الطيفي تعبر عن قيمة الشدة لهذه الحزم حيث انه كلما كان الشريط الممثل للحزمة اشد سوادا دل ذلك على أن هذه الحزمة أكثر طاقة و شدة.

أما بالنسبة لبعض الأصوات التي تتكون من ترددات غير منسجمة، فهي لا تظهر في شكل حزم صوتية و إنما تتوزع الطاقة و الشدة على كل تردد دون حدوث تجميع لهذه الترددات حول تردد واحد منها، و قد سميت من اجل ذلك بالضوضاء أو الضجة، و تظهر الضوضاء في الرسم الطيفي على هيئة نقاط أو خطوط غير مرتبة أو مجمعة في حزام أو مناطق من الترددات.

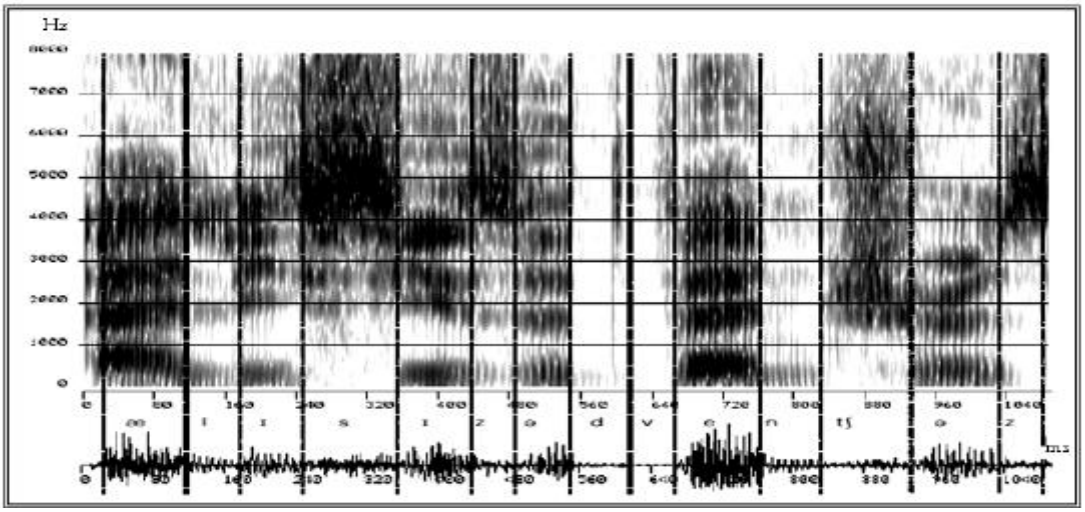
يمثل الشكل التالي رسما طيفيا للحركتين الكسرة (i) و الضمة (u) تظهر فيه الحزم المكونة[10]:



الشكل 2.2: الرسم الطيفي للحركتين i و u [10].

1- الحزمتان (1) و (2) متباعدتان جدا مع الصوت i بينما هما متقاربتان جدا مع الصوت u

2- الحزم العليا التي تظهر في الجزء الأعلى من الصورتين عبارة عن حزم تمييزية ترتبط بالمتكلم الفرد، و ليس لها أي قيمة لغوية[16].



الشكل 3.2: تمثيل الاشارة مع الزمن، يعلوها الرسم الطيفي (السوناغرام).

5. تطبيقات التعرف على الكلام:

تعد وسائل الإدخال جزء من أجزاء تقنية المعلومات التي يتم السعي دائما إلى جعلها أكثر سرعه لنقل البيانات إلى الحاسب لمعالجتها و لذلك تعتبر وسيلة النقل الصوتية أسهل الطرق في الاتصال على الإطلاق. فمن هذا المنطلق نشأت فكرة التعرف على الكلام كوسيلة لإدخال البيانات و التفاعل مع الحاسب في شتى التطبيقات، و هنا سوف يتم ذكر بعض التطبيقات على أنظمة التعرف على الكلام:

1.5. البرامج التعليمية:

للتعرف على الكلام دور كبير في التعليم و التعلم و يتمثل أبرزها في تعلم اللغات وتعليمها، و خاصة تعليم اللغة العربية سواء لأهلها أو لغير الناطقين بها.

2.5. الإملاء الصوتي:

من أهم تطبيقات التعرف على الكلام برامج الإملاء الصوتي التي تقوم بالتعرف على الكلام المنطوق و تحويله إلى كتابة، أو معرفة من هو الشخص الذي يتكلم في مقطع صوتي من بين مجموعة من الأشخاص الذين يحتفظ النظام بمقاطع صوتية لكلامهم، و من هذه التطبيقات القراءات القرآنية أو علم التجويد.

3.5. الأوامر و التحكم:

البرامج التي تتقبل بعض الأوامر المحددة بواسطة الصوت و هي تلك الأنظمة المصممة لأداء الوظائف و الأعمال في نظام يكون التعرف على الكلام إضافة ثانوية مثل "افتح الملف".

4.5. الإرسال الهاتفي:

تسمح بعض أنظمة البريد الصوتي للأشخاص المتصلين بإصدار الأوامر بدل كبس الأزرار، دليل الهاتف الآلي، الهواتف للصم و التي تقوم بتحويل حديث المتصل إلى كتابة مقروءة على شاشة الهاتف، التعرف على الأسماء في الهواتف المحمولة. إضافة إلى أن هناك العديد من التطبيقات لأنظمة التعرف الآلي على الكلام.

6. صعوبات تعرّف الكلام:

تكمن صعوبة مسألة تعرف الكلام في تنوع شكل الإشارة الكلامية الموافقة للكلمة نفسها، و يمكن تصنيف هذه التغيرات كما يأتي:

1.6. تغيرات تتعلق بالمتحدث نفسه:

يختلف نطق المتحدث الواحد لنفس العبارة من وقت لآخر، إذ يمكن أن تنشأ تغيرات صوتية بحسب تغيرات الحالة الفيزيولوجية و النفسية للشخص (هل هو مريض أو صحيح، حزين أو فرح أو غاضب...)،

2.6. تغيرات بين المتحدثين:

تتبع الخلفية الاجتماعية اللغوية لهم مثل، اللهجة و أبعاد و شكل جهازهم الصوتي الفيزيولوجي.

3.6. اختلاف محيط المتحدث:

و هذا ناتج عن وجود إشارات صوتية غير مرغوب فيها (متحدثين آخرين، ضجيج، إغلاق باب...)، أو عن تنوع الميكرفونات المستخدمة و مكان وضعها.

4.6. تغيير طريقة الكلام:

نطق الكلام نادرا ما يكون مضبوطا، و يختلف نطق نفس العبارة من شخص إلى آخر، و بحسب معدل كلامه (سريع أو بطيء)، وبحسب جودة صوته (يقصد الإفهام أو يتكلم على نحو عارض).

5.6. اختلاف نطق الصوت الواحد:

يمكن أن يختلف نطق الصوت الواحد تبعا لما إذا كان ينطق منفردا أو مع كلمات أخرى، عندما لا يكون هناك حدود واضحة في الإشارة الصوتية بين الكلمات المتتالية، و يمكن أن تكون هناك فترات صمت في منتصف الكلمة، أو غياب أي توقف بين الكلمات ا لمتتالية.

7. أنواع أنظمة التعرف على الكلام:

توجد عدة أنواع لأنظمة التعرف على الكلام و قبل التطرق إليها، لابد من ذكر المعايير التي على أساسها يتم تقسيم هذه الأنظمة، و بيئاتها التي سوف تعمل فيها اعتمادا على طريقة النطق حيث انه يمكن أن يكون الكلام متصلاً أو منفصلاً إضافة إلى مدى قدرة النظام على تحديد بداية و نهاية النطق ، لذلك يمكن تقسيم نظم التعرف على الكلام إلى ثلاثة أقسام على حسب الطريقة المستخدمة في التعرف و الاستخدام وهي[52]:

1.7. أنظمة التعرف على الكلمات المعزولة:

أنظمة التعرف على الكلام المنفصل تتعرف على الكلمات المنطوقة بصورة منفصلة، و هنا يجب على المستخدم أن يتوقف بين كلمة و أخرى، و بهذا يستقبل النظام كلمة واحده و يحللها في وقت واحد، أي ما يعني أن النظام يفرق بين الكلمات بواسطة سكوت المستخدم بين كل كلمة و كلمة و لهذا النظام حالتين إما استماع أو صمت، و يعتبر هذا النظام من أسهل الأنظمة في التعرف على الكلام لعدة أسباب[52] [53]:

- إناطة مهمة تفصيل الكلمات إلى المستخدم و ليس إلى النظام.
- عند نطق الكلمات كلمة كلمة تكون طريقة النطق أوضح و أدق للنظام.
- لا نواجه مشكلة co-articulation و هي إلتقاء الحرف في نهاية الكلمة الأولى مع الحرف في بداية الكلمة الثانية مما يسبب صعوبة في التعرف.

2.7. أنظمة التعرف على الكلام المتصل غير المستمر:

تعد كحالة وسطية بين التعرف على الكلمات المنفصلة و التعرف على الكلام المستمر ففي هذا النظام يتم دخول الكلمات إليه بنفس طريقة الكلمات المنفصلة فهي تفترض وجود وقفات زمنية واضحة تفصل بين الكلمات لكن هنا يتم تقليل فتره الصمت بين الكلمات بحيث تبدو كأنها متصلة أو كأنها جملة متقطعة بحيث تسمح للمستخدم بان يتكلم بعدد من الكلمات في وقت واحد، و في هذا النوع من التعرف تكون عملية

التقطيع سهلة فهو يشبه النوع السابق لكنه أصعب في التعرف، و لكن يعاب على هذا النوع عدم مشابهة الواقع الإنساني الذي هو الغرض من هذه التقنية[53].

3.7. أنظمة التعرف على الكلام المستمر (المتصل):

في هذا النظام يتم إدخال جمل كاملة بنطق طبيعي بدون أية شروط على طريقة النطق للمستخدم، و هذا النظام من أصعب الأنظمة تطبيقا في التعرف على الكلام لقلة الدقة في نطق الكلمات عندما تكون في داخل الجملة و ما يطرأ عليها من تشويه نتيجة لانسياب النطق و تدفقه، و صعوبة تحديد حدود كل كلمة على حدة فهو يدرب على التعرف على وحدات الكلام مثل الفونيمات و المقاطع و أن التعرف في الأنظمة المستمرة يتطلب تسلسلاً سمعيا ليتم تقطيعه إلى وحدات صوتية اصغر من الكلمة إلى مقاطع أو إلى أنصاف المقاطع أو إلى فونيمات و بعدها يتم تمييزها، و هذه النوعية من التعرف تستخدم الفونيمات وحدة أساسية للتعرف و تحتاج فضلاً عن الخصائص المميزة للفونيمات إلى معرفة بإشارات اللغة و قواعدها، و هذه المعرفة ضرورية لحل بعض المشاكل، في هذا النوع من التعرف يكون حجم المصادر كبيراً و هذا يعتمد بصورة مباشرة على اللغة المستخدمة.

8. مراحل عملية التعرف على الكلام:

أغلب الأنظمة اليوم تعتمد على مبدأ التعرف على الأنماط لأنه الحاصد على قدر أعلى من نتائج التعرف الحالية، و تمر اغلب أنظمة التعرف على الكلام بالخطوات التالية[50]:

1.8. مرحلة تسجيل و قراءة الإشارة الصوتية:

و هي المرحلة المسؤولة عن تسجيل صوت الكلمة باستعمال تقنيات التسجيل المتوفرة، لتحضير الإشارة الصوتية لمجموعة من المعالجات التي سيتم عليها فيما بعد، حيث تتم عملية اخذ سلسلة من العينات تمثل مطال الإشارة الكلامية ضمن بطاقة تحصيل الصوت، لذا علينا الانتباه إلى خصائص تسجيل هذه الإشارة و التي تتمثل ثلاث محددات هي: تردد التقطيع، عدد بتات الترميز، إضافة للتسجيل بأحد الخيارين Mono أو Stereo .

2.8. مرحلة التقسيم إلى إطارات و استخلاص المعاملات:

تقسيم الصوت إلى قطع متتالية ذات أحجام محددة و استخلاص معاملاتها يعتبر من أصعب العمليات لاختلاف مواصفات الصوت من متكلم لآخر و اختلاف نطق المتكلم نفسه وسرعة كلامه و دقة نطقه و كيفية ربطه للكلمات بعضها ببعض.

3.8. تقنيات المقارنة و الربط للتعرف على الكلام:

إن أي نظام للتعرف على الكلام لابد و أن يمر بعدة مراحل، و تعتبر مرحلة المقارنة و الربط هي المرحلة الرئيسية التي يتحقق فيها معنى التعرف على النطق المراد للآلة التعرف عليه حيث يستخدم أي نظام للتعرف على الكلام تقنية ما في هذه المرحلة، فيما يلي بعض هذه التقنيات المستخدمة حاليا، و قد قمنا بتقسيم تلك التقنيات إلى فئات و هي:

1.3.8. تقنيات إحصائية:

حققت نماذج ماركوف المخفية رواجا في الفترة السابقة ، نتيجة للنجاح الكبير الذي حققته في التعرف على الكلام، يرجع ذلك النجاح إلى الأساس الاحتمالي الذي بنيت عليه ، فالتعامل مع الكلام تعتريه مشاكل النقص و عدم المصداقية ، و تعمل هذه التقنية عن طريق دالة احتمالية لسلسلة ماركوف المخفية، التي تتكون من عدد محدد من الحالات و مجموعة من الدوال العشوائية المرتبطة بكل حالة، تتغير الحالات نسبة لمصفوفة احتمالات انتقالية، حتى الوصول للتوقع النهائي للنطق.

2.3.8. تقنيات المقارنة:

التقنية الأكثر شيوعا في هذا المجال هي تقنية تطويع الزمن الديناميكي (DTW)، تعمل هذه التقنية على مبدأ معالجة التغيرات في محور الزمن المسببة باختلاف معدل النطق، لأن الكلمات التي تمثل بقيم صوتية متشابهة تقارن على أساس أنها متساوية الطول و أزمنتها متطابقة، و تستعمل (DTW) الانحراف المتري لمقارنة

متجهات الخصائص في كل من النطق و النطق المقارن و تحذف الفروق بين العينتين بتطويع محور الزمن بالنسبة للعينة التي تحقق اكبر توافق مع الأخرى.

3.3.8. تقنيات الربط:

أثناء السنوات الأخيرة، ازدهر اتجاه حاسوبي جديد مستند على العلوم العصبية سموه الشبكات العصبية قياسا على العمل الداخلي للإنسان، إن الفكرة الرئيسية وراء الشبكات العصبية هو محاكاة سلوك الربط العصبي في الدماغ، و هي ممتازة في التعرف على الكلام لقدرتها على التكيف مع البيانات الناقصة أو المبهمة، فالكلام قلما يسلم من هذه العيوب، كما تتميز بالسرعة بالمقارنة مع التقنيات الأخرى.

9. تسجيل الأصوات:

الكلام ذو طبيعة معقدة للغاية، لذا يتطلب تمثيلا و علاجا محددا من أجل تخفيف التعقيد من جهة، و ليتناسب مع الآلات من جهة أخرى، لهذا نختار مع هذا التمثيل ثلاث محددات[23]:

1- تمتد حدود السمع ضمن المجال [20 – 20000Hz] و يتراوح أعلى مجال حساسية الأذن من [80–8000Hz]، فإذا ما قارنا مع طيف الإشارة الهاتفية نجد أن طيفها يمتد في المجال [300– 3400Hz] و مع هذا تعطي صوتا مميزا للشخص المتكلم بكلمات مفهومة، لذا يمكن اعتبار أعلى تردد للإشارة الكلامية 4000Hz.

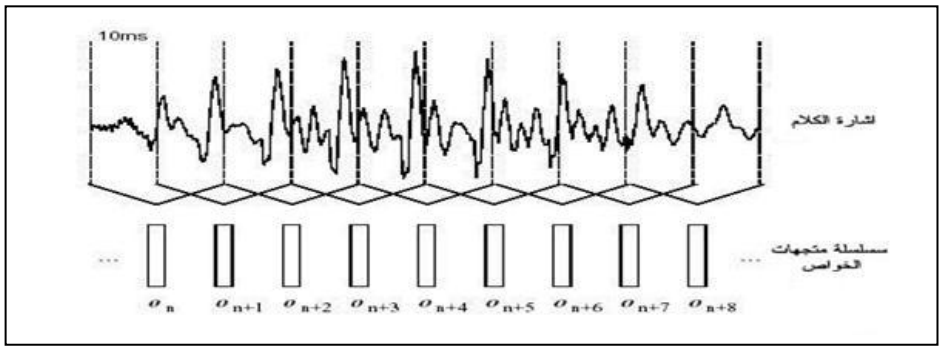
و بتطبيق قانون شانون في تحصيل الإشارة الكلامية الذي يشترط أن يكون تردد التقطيع $Fs \geq 2Fmax$ حيث Fmax التردد الأعظم للإشارة الكلامية. يكفي أن نأخذ تردد التقطيع Fs=8000Hz ، ليعطي إشارة كلام مقبولة جدا و يجب الانتباه عند زيادة هذا التردد ترتفع نوعية الإشارة المسجلة و لكنها تؤدي لزيادة عدد العينات المأخوذة في الثانية مسببة زيادة في زمن المعالجة الرقمية.

2- عدد بتات الترميز و يقصد بها عدد البتات اللازمة لترميز قيمة مطال العينة أو عدد المستويات المأخوذة من الإشارة، و يمكن اختيار طول العينة بإحدى القيم الثلاث (8 bit,16 bit, 32 bit)، مع العلم بان زيادة هذه القيمة لا يؤثر على جودة الإشارة الكلامية بل سيعقد عمليات المعالجة مؤديا إلى زيادة زمنها.

3- التسجيل باستخدام قناة أو قناتين أي التسجيل بأحد الخيارين Mono و Stereo مما يؤثر على حجم ملف التسجيل، كون الخيار Stereo يعطي ضعف حجم الملف الصوتي، الذي تم تسجيله باستخدام الخيار Mono.

10. استخلاص المعاملات:

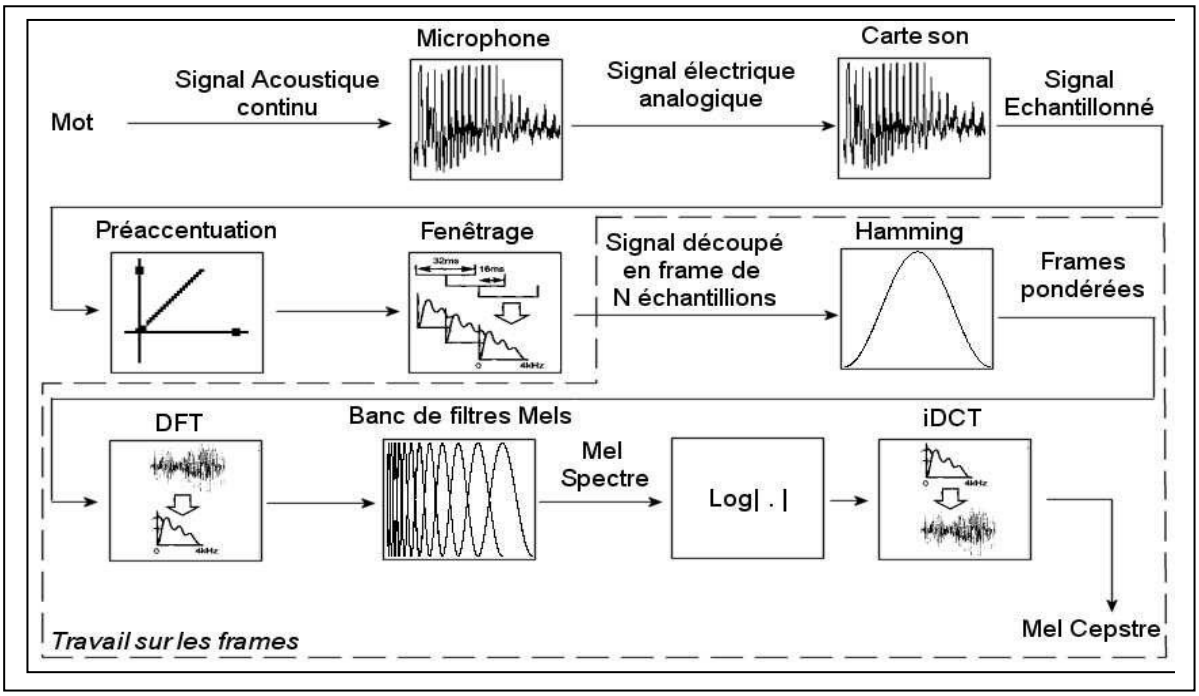
تتمثل في تحليل الإشارة لاستخلاص المعاملات الصوتية بدقة حيث يتم معالجة الإشارة بتقسيمها إلى إطارات متساوية في الطول ثم يتم استخلاص المعاملات لكل إطار لتكون النتيجة النهائية متجه من المعاملات يمثل ذلك الإطار، هناك عدة طرق تستخدم لاستخلاص المعاملات لأية إشارة صوتية فيمكن استخدام الطاقة (Energy) و معدل قطع الصفر (Zero crossing) و معاملات (Mel – Frequency cepstrum coefficients (MFCC) و معاملات التنبؤ الخطي ((Linear predictive coefficients (LPC) و غيرها، و الشكل (2) يوضح ما سبق ذكره على جزء مقتطع من إشارة كلام تم تقسيمه إلى عدة مقاطع (Frames) و إيجاد متجه المعاملات لكل مقطع على حدة:



الشكل 4.2: تقسيم إشارة الكلام و استخراج متجه المعاملات منها[24].

1.10. تحليل الإشارة الصوتية بـ (MFCC) Mel– Frequency Cepstrum Coefficients:

تعد MFCC طريقة لتحليل الإشارة الصوتية لتثمين مواصفات الكلام مثل طبقة الصوت و نغمته و حيث تعتمد بالأساس على ترشيح الترددات المنخفضة خطيا، و العالية لوغارتيما لاستخلاص المعاملات المهمة للصوت و في ما يأتي شرح مفصل للمراحل التي يمر بها الصوت لاستخلاص المعاملات منه، الشكل (5) يوضح المخطط الكتلي لمعالجة (MFCC) و تشمل المراحل الآتية [24] [25]:



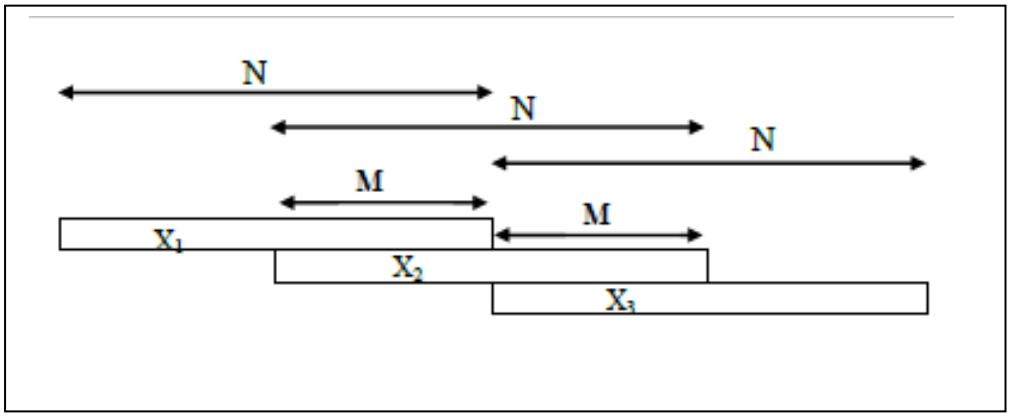
الشكل 5.2: مخطط كتلي يوضح مراحل حساب معاملات MFCC [25].

2.10. تقسيم الإشارة إلى نوافذ زمنية:

يستخدم التحليل الطيفي لجزء قليل من الإشارة لان مواصفاتها ثابتة، حيث انه يمكننا اعتبار إشارة الكلام إشارة مستقرة لا تتغير طبيعتها مع الزمن خلال نوافذ زمنية قصيرة نسبيا، و يتم تقسيم الإشارة إلى عدة نوافذ متداخلة بحيث تحوي كل نافذة N عينة، و نأخذ تداخلا بين النوافذ بإزاحة مقدارها M عينة للتأكد من عدم ضياع معلومات بين النوافذ فإذا اعتبرنا أن القيمة النموذجية لطول النافذة هي 30 ms و مقدار الإزاحة بين

نافذة و أخرى هي 15 ms عندها من اجل تردد تقطيع Fs=8000Hz فان N=240 و M=120 مع الانتباه إلى أن عدد النوافذ سيختلف من كلمة لأخرى كون الكلمات مختلفة الأطوال، و بالتالي النافذة [23][25]:

$$x_i(m) = s(m+iM) \qquad m= 0, 1 , \dots ,N-1 \qquad \dots\dots\dots\dots\dots\dots(2.1)$$



الشكل 6.2: تهيئة الإطارات بصورة المتداخلة [24].

3.10. المعالجة بالنافذة:

يتم في هذه الخطوة تمرير كل إطار من خلال نافذة بفترة معينة عبارة عن دالة رياضية تشبه في وظيفتها عمل النافذة الاعتيادية بإدخال كمية الضوء و ذلك لتقليل عدم استمرارية الإشارة و إزالة التعرجات الحادة منها في بداية و نهاية كل إطار كما هو موضح في الشكل تقوم هذه النافذة بتنعيم الإشارة و إزالة التعرجات الحادة منها ولكل قالب في الإشارة، و تكون نتيجة النافذة ممثلة بإشارة المعادلة الآتية [24]:

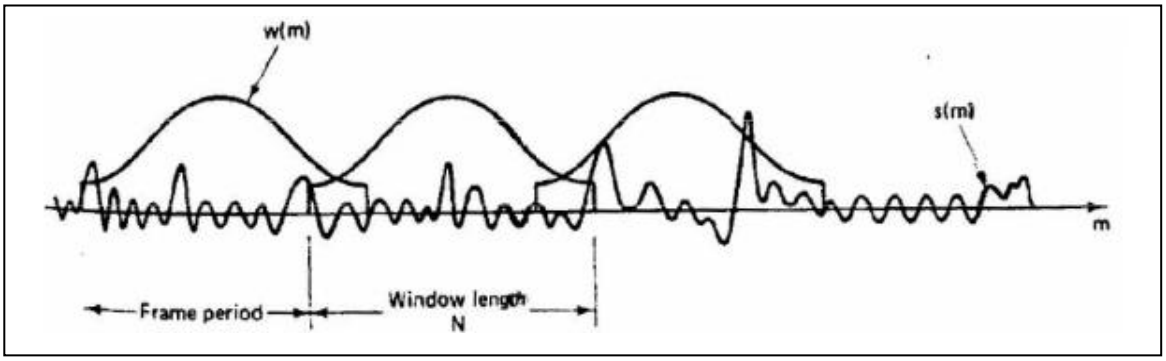
$$x_i(n) = x_i(n).w(n) \qquad , 0 \leq n \leq N-1 \qquad \dots\dots\dots\dots\dots\dots (2.2)$$

حيث:

x_i : عينات القالب الواحد

N : حجم القالب أي عدد عينات القالب

و هناك أنواع عديدة منها النافذة الأكثر شيوعا في أنظمة التعرف على الكلام هي نافذة هامينغ (Hamming window)، و السبب في استخدامها هو خواصها الطيفية الجيدة حيث يتم تثقيل العينة باتجاه مركز النافذة و نخفف من اثر التشوه عند الحواف، و بالتالي نتأكد من عدم تشوه الإشارة و الحصول على تعرف دقيق بنسبة عالية للكلمة، و يوضح الشكل التالي هذه الخطوة.



الشكل 7.2: الضرب بنوافذ هامينغ

تعطى علاقة نافذة هانينغ بالمعادلة التالية:

$$\text{Hanning}(n) = \begin{cases} \alpha - (1-\alpha)\cos\left(\frac{2\pi n}{N-1}\right) & 0 \le n \le N-1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

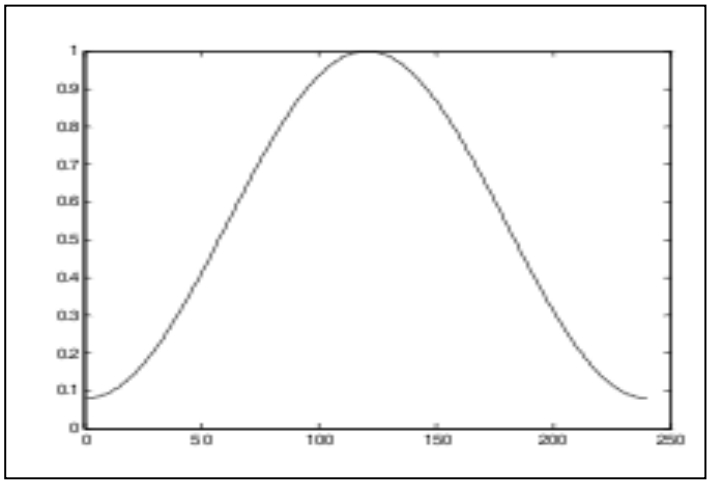
حيث N هو طول النافذة و $0 \le n \le N-1$ و تعتبر نافذة هامينغ حالة خاصة من نافذة هانينغ باعتبار $\alpha=0.54$:

و بالتالي تكون المعادلة لحساب النافذة

$$h(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) & 0 \le n \le N-1 \\ 0 & \text{otherwise} \end{cases} \quad (2.4)$$

$x_i(m) = x_i(m).h(m)$, $0 \leq m \leq N-1$ ..... (2.5)

يبين الشكل التالي شكل نافذة هامينغ [23].



الشكل 8.2: نافذة هامينغ

4.10. تحويل فوريه السريع FFT (Fast Fourier Transform):

تحويل فوريه السريع و هو تطبيق سريع لتحويل فوريه المنفصل (DFT) يتم فيه تحليل الطاقة التي تمتلكها الإشارة عند ترددات معينة، حيث انه من التحليل الطيفي نلاحظ أن المنطلق الترددي يعطينا معلومات مهمة لا نستطيع أن نحصل عليها في المنطلق الزمني، لذلك يطبق هذا التحليل على كل قالب كما في المعادلة الآتية [24][25]:

$$X_n = \frac{1}{N} \sum_{k=1}^{N-1} X_k^* e^{\frac{-2\pi jkn}{N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.6)$$

حيث أن:

N: تمثل حجم القالب

X_k: قيمة العينة قبل FFT

X_n: قيمة العينة بعد FFT

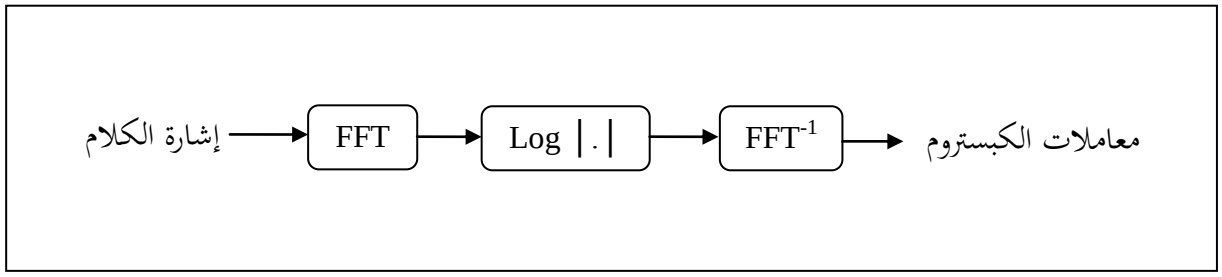
J: تمثل جذرا خياليا.

5.10. التحليل باستخدام الكبستروم (Cepstrum):

الاسم Cepstrum هو كلمة بدايتها (الحروف الأربعة) معكوس بداية كلمة الطيف Spectrum، و هو يعرف رياضيا بأنه تحويل فورييه العكسي للوغاريتم مطال الطيف، و يعبر عنه بالمعادلة:

$$Cepstrum = FFT^{-1}[\log|FFT[s(n)]|] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

و باعتبار إشارة الكلام s(n) إشارة مقطعة فإننا نستخدم تحويل فورييه المتقطع ثم نحسب لوغاريتم المطال ثم نحسب تحويل فورييه العكسي كما يوضحه الشكل التالي:



الشكل 9.2: حساب معاملات الكبستروم.

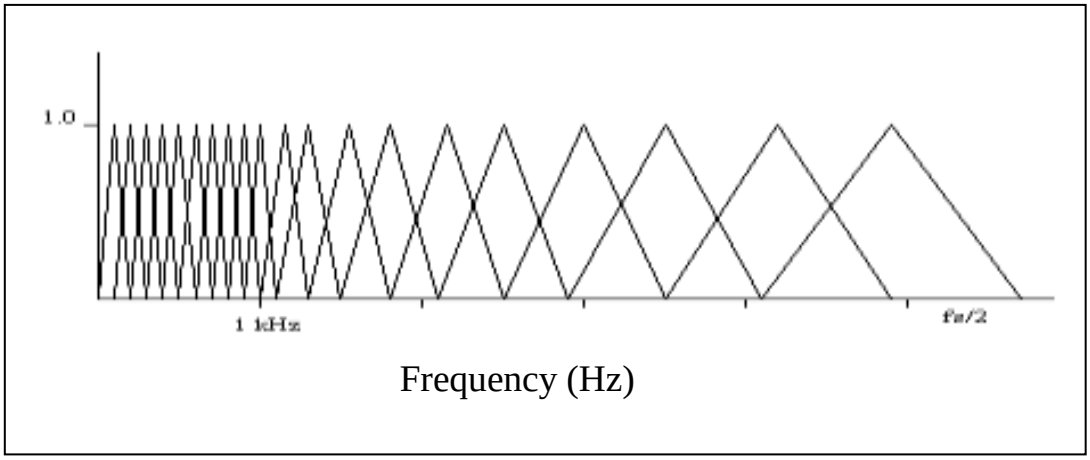
6.10. مقياس Mel:

إن محتويات الترددات لأصوات الإشارات لا تتبع مقياسا خطيا، لهذا فكل نغمة في الترددات الطبيعية (f) تقاس بالهرتز (Hz) أما طبقة الصوت، فتقاس بمقياس Mel الذي يعطينا ترددات خطية بالمسافات التي يكون ترددها أقل من 1000Hz و يعطي ترددات لوغارتمية في المسافات الأعلى من 1000Hz لهذا يتم استخدام الصيغة التقريبية لحساب Mel لتردد معين إذا أعطينا f ب (Hz) :

$$Mel(f) = 2595^{*}\log10(1+f/700) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$

إن الهدف من مقياس الترددات Mel هو عمل بنك الترشيح (Filter Bank) الذي يحوي مرشحا واحدا لكل مكون من مكونات Mel المطلوبة و يتم بموجبه تجميع مرشحات تمرير الحزمة مثلثة الشكل (Triangular Bandpass Filter) يحدد عرض المرشح بوساطة ثابت التردد Mel التي في النهاية تمثل

بنك الترشيح، و يتم بوساطة هذه الطريقة تغطية كل الترددات بدون أي فقد للمعلومات بحيث أنه عند إدخال الإشارة الناتجة عن FFT إلى المرشح تظهر مواقع القوة لتلك الإشارة، و الحالة المثالية لعدد المرشحات في هذه الطريقة تساوي 20 و التي سوف تعتبر عدد معاملات معادلة MFCC، و يمثل الشكل (2) بنك الترشيح الذي يطبق على الإشارة [24][25].



الشكل 10.2: بنك الترشيح Mel

7.10.معاملات MFCC:

في الخطوة الأخيرة يتم تحويل نتيجة بنك الترشيح إلى المنطلق الزمني و هي MFCC و التي تجهزنا بتمثيل جيد لمواصفات الطيف لجزء من إشارة الرقم الصوتي للإطار المحلل و تستخدم المعادلة الآتية لحساب MFCC [25][24]:

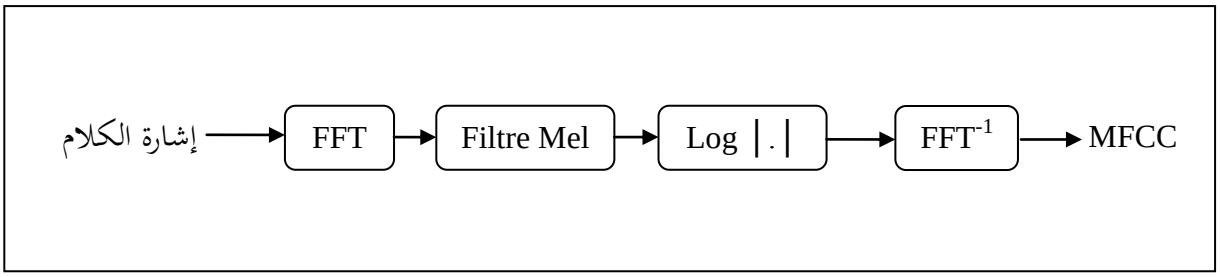
$$MFCCi = X_k \cos[\pi i (k - 0.5) / M] \qquad i = 1, 2, \ldots, p \quad \ldots\ldots\ldots (2.9)$$

حيث أن:

M : عدد مرشحات في بنك الترشيح

X_k: هي إشارة الإطار الواحد Mel cepstrum .

P: عدد معاملات المتجه الواحد للإشارة المحللة.



الشكل 11.2: حساب معاملات MFCC

8.10. المشتقات الأولى و الثانية لمعاملات الكبستروم:

و التي تسمى Delta and Delta-Delta Cepstrum و تعطى بالعلاقات التالية:

$$\Delta c_i(n) = \sum_{k=-2}^{k=+2} k c_i(m+k) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.10)$$

$$\Delta\Delta c_i(n) = c_i(m+1) - c_i(m-1) = \sum_{k=-1}^{k=+1} k\, \Delta c_i(m+k) \quad \ldots\ldots\ldots\ldots (2.11)$$

إن كل قيمة من شعاع المشتقات $\Delta c_i(n)$ هي الفرق بين القيمة الموافقة في شعاع معاملات الكبستروم و القيمة ذات الترتيب نفسه في النوافذ المجاورة للنافذة الحالية، و لا نأخذ عددا كبيرا من النوافذ لان هذا يتنافى مع شرط استقرار إشارة الكلام الذي اعتبرناه في بداية هذه المرحلة من المعالجة، لهذا في حالة المشتقات من المرتبة الأولى نأخذ أربع نوافذ مجاورة للنافذة الحالية (اثنين قبلها و اثنين بعدها)، أما في حالة في حالة المشتقات من المرتبة الثانية فان قيم الشعاع $\Delta\Delta c_i(n)$ هي قيمة الفروق لمشتقات الدرجة الأولى $\Delta c_i(n)$ لنافذتين مجاورتين للنافذة الحالية.

إن اعتماد هذه المشتقات في نظم التعرف يؤدي إلى تحسين كبير في أدائها، نلاحظ أن طول أشعة المشتقات المعرفة أعلاه تساوي طول شعاع معاملات الكبستروم، و يتطلب لحساب هذه الأشعة حساب أشعة

معاملات الكبستروم لنوافذ الكلمة كلها، ينتج لدينا من الحسابات السابقة تسلسل من الأشعة $X=\{x_1,x_2,\ldots,x_n,\ldots,x_s\}$ بحيث كل شعاع x_n هو ناتج ضم شعاع الكبستروم $c_i(n)$ و شعاع مشتقاته من المرتبة الأولى $\Delta c_i(n)$ شعاع مشتقاته من المرتبة الثانية $\Delta\Delta c_i(n)$ للنافذة n أي:

$$x_n = (c_1(n), c_2(n), \ldots c_D(n),$$

$$\Delta c_1(n), \Delta c_2(n), \ldots \Delta c_D(n),$$

$$\Delta\Delta c_1(n), \Delta\Delta c_2(n), \ldots \Delta\Delta c_D(n)) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.12)$$

و إذا حددنا عدد معاملات الكبستروم $D = 12$ فان معاملات كل شعاع x_n هو 36 [23].

11. الخاتمة:

لأنظمة التعرف الآلي على الكلام مكانة مهمة ظهرت من خلال التطبيقات الكثيرة لهذه التقنية مثل: إعطاء الأوامر و التحكم، و إدخال المعطيات الصوتية، ... و غيرها، و انجاز مثل هذه الأنظمة يمر بعدة مراحل تتم فيها معالجة إشارة الكلام بشكل متعاقب للوصول إلى النتائج المرجوة، إلا أن هذه المعالجة تعترضها مجموعة من المشاكل و الصعوبات يتطلب حلها استخراج أقصى قدر من المعلومات المفيدة من الإشارة الصوتية المدخلة لتسهيل التعرف عليها، حيث يتم تحويل إشارة الكلام إلى مجموعة من المعاملات تحسب في فترات زمنية منتظمة، لتمثيل الإشارة في أفضل نموذج، و من بين المعاملات الأكثر استخداما نجد معاملات MFCC، حيث أظهرت أنظمة التعرف الآلي على الكلام التي تستعمل هذه المعاملات كفاءة عالية و تحسنا في دقة التعرف.

الفصل الثالث

نماذج ماركوف المخفية

1. مقدمة:

إن مفهوم نماذج ماركوف المخفية (Hidden Markov Model) و خوارزمياته مستلهمة أساسًا من نماذج رياضية معروفة باسم العالم الذي اكتشفها و هو (Andrei Markov) التي ظهرت في بداية القرن العشرين و التي تدعى بنماذج ماركوف و نظرا للاهتمامات المتزايدة بالسلاسل الماركوفية فقد ظهر في ستينيات القرن العشرين أسلوب رياضي خاص، مبني أساسا على فكرة الخاصية الماركوفية هو أسلوب فضاء الحالة، و في سبعينيات القرن العشرين تم إجراء تطوير آخر على السلاسل الماركوفية هو نموذج ماركوف المخفي و هذا يدل على أن نماذج ماركوف المخفية (HMMs) ما هي إلا امتداد لنماذج ماركوف الاعتيادية (MMs) [26][27].

كما يعود الفضل في اكتشاف هذا النموذج إلى الباحث (Leonard E. Baum) و آخرون عندما نشروا مجموعة من المقالات الإحصائية بهذا الخصوص في النصف الثاني من ستينيات القرن العشرين، أما أول تطبيق لنموذج ماركوف المخفي فكان في مجال التعرف على الصوت و ذلك في منتصف سبعينيات القرن العشرين، و في النصف الثاني من ثمانينيات القرن العشرين بدأ استخدام نموذج ماركوف المخفي في تحليل المتتابعات الحياتية، خصوصاً DNA، و منذ ذلك الوقت فرض أنموذج ماركوف المخفي وجوده في مجال المعلوماتية الحياتية، و يستخدم نموذج ماركوف المخفي بشكل خاص في تمييز الأنماط مثل الكلام و الكتابة و غيرها، كما تم تطبيق هذه النماذج في مجالات كثيرة من مجالات الحياة، و قد بدت المسالة صعبة جدا في بادئ الأمر ولكن بمرور الزمن و نتيجة للبحوث و الدراسات التي جرت حول الموضوع فقد تم تسهيله نظريا و تطبيقيا[28].

و في عام (1989) قدم الباحث (Rabiner) بحثا موسعا حول (HMMs) حاول من خلاله و بشكل دقيق و منظم استعراض الجوانب النظرية لهذا النوع من الصياغة الإحصائية و إظهار كيفية تطبيقها في مسائل التعرف على الكلام، و قد عد هذا البحث المصدر الأساسي للعديد من البحوث الأخرى المختصة بنماذج ماركوف المخفية.

2. تعاريف أساسية:

2. 1. العملية التصادفية:

إن أي ظاهرة حقيقية تجري في حيز المعلمة المدروسة (كالزمن مثلا) هي عملية تصادفية إذا كانت حالات تلك الظاهرة في أي جزء من حيزها (في أي وقت مثلا) تمثل نتائج تجربة عشوائية تخضع لقوانين الاحتمالات، هذا و تعرف العملية التصادفية أيضا بالعملية العشوائية أو عملية الفرصة.

تعريف: العملية التصادفية $\{ X(t) , t \in T \}$ هي العملية المكونة من خلال قيم مختلفة تأخذها متغيرات عشوائية معينة، معرفة على تجربة ما، عند قيم مختلفة من فضاء المعلمة المدروسة (عند أزمان مختلفة مثلا) حيث أن (T) تمثل مجموعة من المعلمات و تشير إلى فضاء المعلمة [28][29].

2. 2. فضاء الحالة:

تعريف: إن مجموعة القيم الممكنة للعملية التصادفية المتعلقة بـ $X(t)$ تسمى فضاء الحالة و يرمز لها بالرمز S.

و يكون فضاء الحالة متقطعا إذا احتوى على نقاط منتهية أو نقاط غير منتهية قابلة للعد، أما في الحالات الأخرى فهو مستمر[28][29].

2.3. سلاسل ماركوف (الخاصية الماركوفية):

إن الظواهر التي تتميز بخاصية انه إذا علمت حالتها في الماضي ، فانه سوف لن يكون هناك إي تأثير لحالات هذه الظاهرة في الزمن الماضي على ما سيجري لها في المستقبل، تعرف بسلاسل ماركوف، و أن الخاصية نفسها تدعى بالخاصية الماركوفية. و يقال للعملية التصادفية بأنها تحمل الخاصية الماركوفية إذا كانت تحقق ما يأتي :

$$P(X_n = j / X_{n-1} = i, X_{n-2} = a, X_{n-3} = b, \ldots, X_0 = c) = P(X_n = j / X_{n-1} = i) = p_{ij}$$

لجميع قيم n, i, j, a, b, c [28][30] .

4.2. المصفوفة الاحتمالية الانتقالية:

تسمى احتمالات الانتقال من الحالة (i) في الخطوة (n) إلى الحالة (j) في الخطوة (n+1) بالاحتمالات الانتقالية ويرمز لها p_{ij}، و يمكن تمثيل التوزيع الاحتمالي للانتقالات بين الحالات بمصفوفة تسمى المصفوفة الانتقالية حيث يتم وضع العناصر $p(i,j)$ (لكل $i,j \in E$) على شكل مصفوفة مربعة، ويرمز لها بـ P إذ أن العنصر (i,j) من هذه المصفوفة يمثل الاحتمالية الانتقالية $p_{ij} = p(X_n = j/X_{n-1}=i)$

و بافتراض أن $E = \{1,2,\ldots,N\}$ ، فانه يمكن كتابة المصفوفة الانتقالية ذات الخطوة الواحدة (P) بالشكل التالي[26][30]:

$$P = \begin{array}{c} X_n = \\ X_{n-1} = \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \ldots & N \end{array} \\ \begin{pmatrix} P_{11} & P_{12} & \ldots & P_{1N} \\ P_{21} & P_{22} & \ldots & P_{2N} \\ \vdots & \vdots & \ldots & \vdots \\ P_{N1} & P_{N2} & \ldots & P_{NN} \end{pmatrix} \end{array} \begin{array}{c} 1 \\ 2 \\ \vdots \\ N \end{array}$$

يقال للمصفوفة المربعة (P) (المعرفة أعلاه) بأنها مصفوفة ماركوف (أو مصفوفة الانتقال) الممثلة للعملية $\{X_n , n \in N\}$ ، على فضاء الحالة (E) ، إذا تحقق الشرطان الآتيان:

1 -جميع عناصرها غير سالبة (لكونها احتمالات). لكل قيم $(i, j \in E)$ تكون $p(i,j) \geq 0$

2 -مجموع كل صف من صفوفها يساوي 1 (لان مجموع الاحتمالات الكلية يساوي 1).

لكل قيم $(i \in E)$ تكون $\sum_{j \in E} p(i,j) = 1$

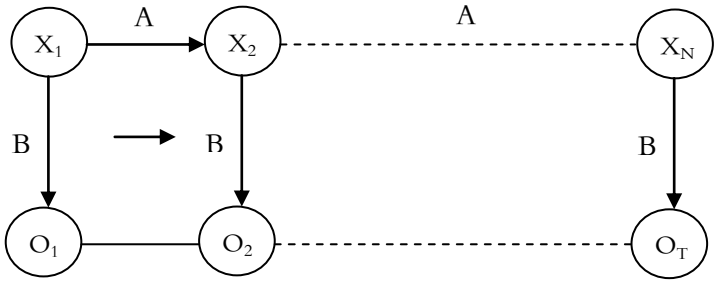
ملاحظة: إذا كانت المصفوفتان A وB تصادفيتين فان حاصل ضربهما هي أيضا مصفوفة تصادفية[30].

3. نماذج ماركوف المخفية:

تعد نماذج ماركوف المخفية أداة إحصائية قوية تستخدم لنمذجة العمليات التي تتغير مع مرور الزمن, فهي تساعد على ملاحظة مجموعة من الأحداث المخفية و إيجاد احتمال وقوعها اعتمادا على مجموعة من الأحداث تمت ملاحظتها مسبقا[31][32].

تعرف نماذج ماركوف المخفية بأنها عبارة عن عملية تصادفية مضاعفة إذ أن الحالات لا يمكن ملاحظتها مباشرة فيجب البحث عن طريقة غير مباشرة لملاحظتها (أي إضافة طبقة تصادفية ثانية للعملية الماركوفية)، لذا فإن معلمات النموذج الماركوفي المخفي هي $\lambda=(A, B, \pi)$ [33]، و لأن نموذج ماركوف المخفي هو امتداد لنموذج ماركوف الاعتيادي(MM) فتكون معلمة نموذج ماركوف المخفي $\lambda=(A, B, \pi)$ ما هي أيضا إلا امتداد لمعلمة نموذج ماركوف الاعتيادي $\lambda=(A, \pi)$ [34][35].

و يمكن أن نرى ذلك بوضوح من خلال الشكل(2) أدناه:



الشكل 1.3:يوضح النموذج التخطيطي لنموذج ماركوف المخفي

في الشكل أعلا يمكن تمثيل المعلمات بالشكل الآتي[36]:

x : تمثل الحالات (states)

o : تمثل المشاهدات الممكنة.

A : احتمالات انتقال الحالة.

B : احتمالات الإخراج.

1.3. عناصر نماذج ماركوف المخفية:

يتم التعبير عن نماذج ماركوف المخفية باستخدام العناصر التالية [33][37]:

- T = طول متسلسلة المشاهدات
- $Q=\{q_1,q_1,\ldots,q_N\}$ هي مجموعة الحالات لعملية ماركوف (أي الحالات المخفية)
- N: هي عدد الحالات المخفية في النموذج و أن $S=\{s_1,s_2,\ldots,s_N\}$ هي مجموعة من الحالات
- $V = \{v_1,v_2,v_3,\ldots,v_M\}$ مجموعة رموز المشاهدات.
- M : عدد رموز الملاحظات (المشاهدات) (V) و أن $V= \{v_1,v_2,v_3,\ldots,v_M\}$ هي مجموعة من المشاهدات.
- A : مصفوفة احتمالية انتقال الحالة $A=\{a_{ij}\}$ و هي تمثل احتمال الانتقال من الحالة i إلى الحالة j خلال وحدة زمنية واحدة (خطوة واحدة).

و تعطى بـ:

$$A=\{a_{ij}\} \ , \ a_{ij}=P[q_{t+1}=S_j \mid q_t=S_i] \ ; \ 1\leq i,j \leq N$$

إذ أن $q_t=i$ تمثل الحالة i خلال الزمن t.

- B : مصفوفة التوزيع الاحتمالي لرمز المشاهدة في الحالة j $B=\{b_j\ (k)\}$، و هي تمثل احتمال مشاهدة الرمز v_K في الحالة j، مصفوفة احتمالية رابطة بين الحالات المخفية و المشاهدات إذ أن:

$$B=\{b_j\ (k)\} \ , \ b_j\ (k) = P\ (O_t = V_k \mid \ q_t = S_j) \ ; \ 1 \leq j \leq N$$

$$1 \leq k \leq M$$

إذ أن O_t تمثل ملاحظة الرمز i خلال الزمن t.

- π : مصفوفة توزيع الحالة الابتدائية، $\{\pi_i\} = \pi$ إذ أن:

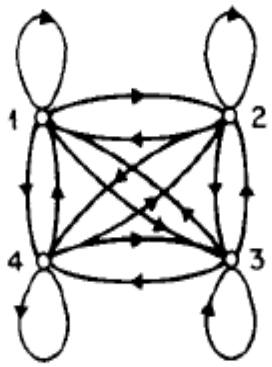
$\pi = \{\pi_i\}, \pi_i = P[q_1 = S_i] ; 1 \leq i \leq N$

- $O=\{O_1,O_1, \ldots,O_T\}$ متسلسلة المشاهدات

2.3. انواع نماذج ماركوف المخفية:

يمكن تصنيف نماذج ماركوف المخفية إلى نوعين و ذلك حسب الانتقالات بين حالاتها [28][38]:

النموذج الثبوتي: و هو النموذج الذي تكون فيه كل الحالات انتقالية.



الشكل 2.3: النموذج الثبوتي

نموذج الأيسر– الأيمن : و هو الذي تكون فيه بعض الحالات انتقالية ، بحيث أن:

$$a_{ij} = 0 \qquad , \forall \quad j < i$$

الشكل 3.3: نموذج الأيسر– الأيمن

و هذا النوع من النماذج يستخدم و بشكل واسع في نمذجة متسلسلات الإشارات.

3.3. المسائل الأساسية لنماذج ماركوف المخفية:

لكي تكون نماذج ماركوف المخفية مفيدة في كثير من التطبيقات في العالم الحقيقي هناك ثلاث مسائل أساسية يجب أن يتم حلها و هذه المسائل هي[35][38]:

المسالة (1):

تعمل مسألة التقييم على حساب احتمالية متسلسلة المشاهدات $p(O|\lambda)$ للنموذج عندما يكون المعطى هو النموذج $\lambda= (A, B, \pi)$ و متسلسلة المشاهدات $O =\{ O_1 O_2 \ldots O_T \}$، أي تحديد الإمكان لمتسلسلة المشاهدات (O) عندما يكون المعطى هو النموذج، و تحل عن طريق الخوارزمية الأمامية –الخلفية Forward- Backward Algorithm [34] [37].

المسالة (2):

إذا كان المعطى هو النموذج $\lambda= (A, B, \pi)$ و متسلسلة المشاهدات $O = \{O_1 O_2 \ldots O_T\}$، تتناول هذه المسألة كيفية تعديل معلمات النموذج $\lambda= (A, B, \lambda)$ لتعظيم $P(O|\lambda)$، و إيجاد متسلسلة الحالة المثلى لعملية ماركوف المخفية ، و التي تمثل أفضل توضيح للمشاهدات[28] [34].

المسالة (3):

إذا كان المعطى هو متسلسلة المشاهدات $O = \{O_1 O_2 \ldots O_T\}$ و الأبعاد (N) و (M) ، تعمل مسألة التدريب على إعادة تقدير معلمات النموذج $\lambda=(A,B,\pi)$ الذي تكون فيه $P(O|\ \lambda\ )$ عظمى، و تحل هذه المسألة عن طريق Baum-Welch [28][37].

4.3. حلول المسائل الأساسية لنماذج ماركوف المخفية:

1.4.3. حل المسالة (1) (مسألة التقييم):

إن الطريقة المباشرة لحسباها تكون من خلال سرد متسلسلة حالة ذات طول (T) (حيث أن (T) تمثل عدد المشاهدات) و لنتأمل المتسلسلة الثابتة للحالة:

$Q = \{q_1 q_2 \ldots q_T\}$

حيث أن (q_1) تمثل الحالة الابتدائية، أما الاحتمالية لمتسلسلة المشاهدات (O) بالنسبة لمتسلسلة الحالة في المعادلة (1) فهي:

$$P(O|\ Q\ ,\ \lambda) = \prod_{t=1}^{T} P(O_t | q_t\ ,\ \lambda) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (3.1)$$

و ذلك بافتراض الإستقلال الإحصائي للمشاهدات، أي أن:

$$P(O|\ Q\ ,\ \lambda) = P(O_1 | q_1\ ,\ \lambda) * P(O_2 | q_2\ ,\ \lambda) * \ldots * P(O_T | q_T\ ,\ \lambda)$$

$$= \prod_{t=1}^{T} P(O_t | q_t\ ,\ \lambda)$$

$$= \prod_{t=1}^{T} b\ q_t\ (O_t) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (3.2)$$

و أن الإحتمالية لمتسلسلة الحالة (Q) هي:

$$P(\ Q\ |\ \lambda) = P(q_1\ ,\ q_2\ ,\ \ldots\ q_T |\ \lambda)$$

$$= P(\ q_1\ ;\ \lambda) \prod_{t=2}^{T} P(q_t |\ q_{t+1}\ ;\ \lambda)$$

$$= \pi\ q_1 \prod_{t=2}^{T} a_{\ qt-1\ qt} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (3.3)$$

و ذلك حسب الخاصية الماركوفية حيث أن (π) تشير إلى احتمالية الحالة الابتدائية للحالة ($q1$).

أما الاحتمالية المشتركة لـ (O) و (Q)، أي ورود بالوقت نفسه، فهي تمثل النتيجة للمعادلتين أي أن:

$$P(O , Q \mid \lambda) = P(O \mid Q , \lambda) * P(Q \mid \lambda) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.4)$$

و من هنا يمكن الحصول على الاحتمالية لـ (O) عندما يكون المعطى هو النموذج (λ) و ذلك عن طريق جمع هذه الاحتمالية المشتركة لكل المتسلسلات الحالة (q) ، حيث أن :

$$P(O \mid \lambda) = \sum_{\text{all q}} P (O \mid Q , \lambda) * P(Q \mid \lambda)$$

$$= \sum_{q1,q2,\ldots,qT} \pi_{q1} * b_{q1} (O_1) * a_{q1\,q2} * b_{q2} (O_2) * \ldots * a_{qT-1\,qT} * b_{qT} (O_T)$$

إن هذه الحسابات تكون غير عملية و إن كانت قيم (N) و (M) صغيرة، فمثلا إذا كانت (N = 5) (و هي تمثل عدد الحالات المخفية) و كانت (T=100) (و هي تمثل عدد المشاهدات)، فيجب القيام بـ $2(100) * 5^{100} \approx 10^{72}$ عملية حسابية، و من ذلك يتضح بان هذه الحسابات تتطلب جهدا كبيرا لحل المسالة (1)، و لذلك استخدم الباحثون الخوارزميات الأمامية–الخلفية لحل تلك المسالة، و فيما يأتي توضيح لهذه الخوارزميات [28][38]:

أولا الخوارزميات الأمامية [33][39]:

بإمكاننا أن نعرف المتغير الأمامي $\alpha_t(i)$ بأنه احتمالية متتابعة المشاهدات الجزئية , O1, O2, … Ot و الحالة S_i عند الزمن t عندما يكون المعطى هو النموذج λ، هذا و يمكن التعبير عنه رياضيا بالصيغة الآتية:

$$\alpha_t(i) = P(O_1, O_2, \ldots , O_t , q_t = S_i \mid \lambda) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

و إن المتغيرات الأمامية يمكن حسابها من خلال الخطوات الآتية:

البداية:

$$\alpha_t(i) = \pi_i \, b_i \, (O_1) \, , \, i = 1, 2, \dots , N \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.7)$$

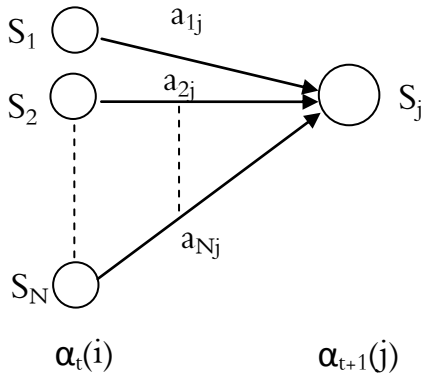
التعاقب:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t \, (i) \, a_{ij}\right] b_j \, (O_{t+1}) \qquad , t = 1, 2, \dots , T-1 \qquad \dots\dots\dots\dots\dots\dots (3.8)$$
$$j = 1, 2, \dots , N$$

الإنهاء:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.9)$$

و يمكن توضيح الخوارزمية الأمامية من خلال الشكل (3) بالآتي:



الشكل 4.3: متتابعة العمليات المطلوبة لحساب المتغير الأمامي.

إن حسابات ال **(α)** تتطلب $N(N+1)(T-1)+N$ عملية ضرب و $N(N-1)(T-1)$ عملية جمع و هي أفضل من الحسابات المباشرة و التي تتطلب $(2T^* N^T)$ عملية حسابية[28][38].

الخوارزمية الخلفية [33][40]:

بأسلوب الخوارزمية الأمامية نفسه يمكن أن نعرف رياضيا المتغير الخلفي $\beta_t(i)$ كما يأتي:

$$\beta_t(i) = P \, (O_{t+1} \, O_{t+2} \dots O_T \, | \, q_t=S_i, \lambda) \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.10)$$

أي انه يمثل احتمالية متتابعة المشاهدات الجزئية من الزمن t+1 إلى T عندما يكون المعطى هو Si عند الزمن t و النموذج λ .

و يمكن حساب المتغيرات الخلفية من خلال الخطوات الآتية [40]:

البداية:

$$\beta_t(i) = 1 \; ; \; i = 1, 2, \dots , N \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.11)$$

التعاقب:

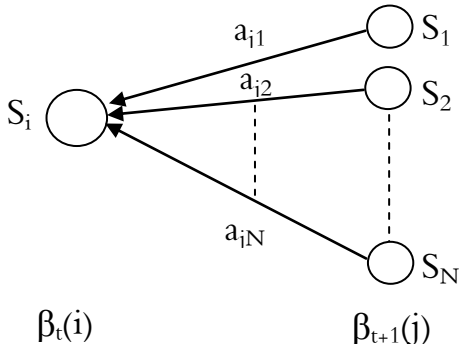
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} \, b_j \, (O_{t+1}) \, \beta_{t+1}(j) \qquad \begin{matrix} , \; t = T-1, \; T-2, \dots , \; 2,1 \\ i = 1, 2, \dots , N \end{matrix} \qquad \dots\dots\dots\dots (3.12)$$

الإنهاء:

يمكن حساب الإحتمالية $P(Q|\lambda)$ باستخدام المتغيرات الأمامية و الخلفية كما يأتي:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \, \beta_t(i) \qquad , \text{for any t} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.13)$$

و الشكل (4) يوضح العمليات المطلوبة لحساب الإحتمالية الخلفية:



الشكل 5.3: متتابعة العمليات المطلوبة لحساب المتغير الخلفي $\beta_t(i)$

2.4.3. حل المسالة (2):

هناك بعض الطرائق الممكنة لحل المسالة (2)، أي لإيجاد متسلسلة الحالة المثلى المقترنة بمتسلسلة المشاهدات المعطاة منها :

خوارزمية (Viterbi):

هي طريقة البرمجة الحركية المستخدمة لحساب مسار الحالة الانتقالية الأكثر احتمالا عندما يكون المعطى هو متسلسلة المشاهدات للرموز، و هي مشابهة للخوارزمية الأمامية ما عدا كون استخدام الـ "اكبر" " Max " بدلا من الـ "مجموع" " Σ ".

و لإيجاد أفضل متسلسلة حالة وحيدة $Q = \{q_1\ q_2\ \dots\ q_T\}$ بالنسبة لمتسلسلة المشاهدات المعطاة $O=\{O_1\ O_2\ \dots\ O_T\}$ يجب تعريف المقدار:

$$\delta_t(i) = \max_{q1\ ,\ q2\ ,\ \dots\ qt-1} P(q_1\ ,\ q_2\ ,\ \dots\ q_t = i\ ,\ O_1\ O_2\ \dots\ O_t \mid \lambda) \quad \dots. (3.14)$$

إذ أن $\delta_t(i)$ تمثل أعلى احتمالية على طول المسار الوحيد خلال الزمن (t) و المحسوبة بالنسبة لمشاهدات (t) الأولى و المنتهية في الحالة (S_i) ، فضلا عن استخدام $\psi_t(t)$ و التي تساعد في حفظ تتابع الأثر (Keep Track) للمسار الفعلي، أي أن $\psi_t(t)$ ستوضح أي حالة خلال الزمن (t−1) تقود لاحتمالية $\delta_t(i)$ الأعلى خلال الزمن (t).

أما خطوات خوارزمية (Viterbi) فهي كالآتي:

البداية:

$$\delta_1(i) = \pi_i * b_i(O_1)\ ,\ i = 1, \dots, N \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.15)$$

$$\psi_1(i) = 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.16)$$

التعاقب:

$$\delta_t(i) = \max_{1\le i\le N} [\delta_{t-1}(i) * a_{ij}] * b_j(O_t) \quad , t = 2, \dots, T \quad \dots\dots\dots (3.17)$$
$$J = 1, \dots, N$$

$$\psi_t(j) = \operatorname{argmax}_{1\le i\le N} [\delta_{t-1}(i) * a_{ij}] \quad , t = 2, \dots, T \quad \dots\dots\dots (3.18)$$
$$J = 1, \dots, N$$

النهاية:

$$P^* = \max_{1\le i\le N} [\delta_T(i)] \quad \dots\dots\dots (3.19)$$

$$q^*_T = \operatorname{argmax}_{1\le i\le N} [\delta_T(i)] \quad \dots\dots\dots (3.20)$$

التتابع المعاكس

$$Q^* = \{ q^*_1 \dots q^*_T \} \quad \dots\dots\dots (3.21)$$

$$q_t^* = \psi_{t+1}(q^*_{t+1}) \quad , t = T-1 , T-3 , \dots ,1 \quad \dots\dots\dots (3.22)$$

ثم تقرا المتسلسلة الأفضل للحالات من متجهات الـ (ψ_t) [28] [41].

3.4.3. حل المسالة (3):

المطلوب في المسالة (3) هو إعادة تقدير معلمات النموذج (A, B, π) و ذلك لتعظيم الاحتمالية لمتسلسلة المشاهدات عندما يكون المعطى هو النموذج و يمكن حل هذه المسالة باستخدام خوارزمية (Baum–Welch) [31][33].

خوارزمية باوم ويلش(Baum–Welch) :

هذه الخوارزمية تناقش خطوات التكرار بالاعتماد و بشكل ابتدائي على العمل الكلاسيكي للعالم (L.E.Baum) و زملائه و ذلك لاختيار معلمات النموذج، و لوصف خطوات إعادة تقدير معلمات (HMM) ، يتم تعريف المتغيرين الآتيين:

أما المتغير الأول هو:

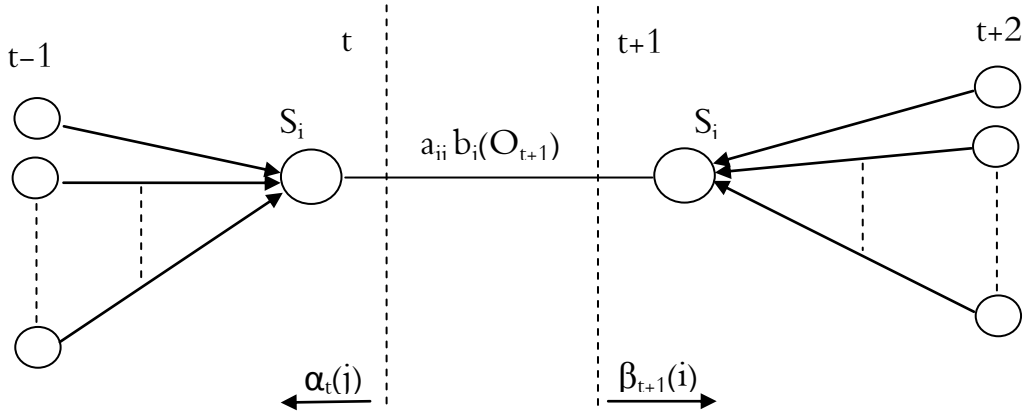
$$\zeta t(i,j) = P(q_t = S_1 , q_{t+1} = S_j \,|O, \lambda)$$

$$= \frac{P(O , q_t = S_1 , q_{t+1} = S_j \,|\lambda)}{P(O|\lambda)}$$

$$= \frac{\alpha_t (i)\, a_{ij}\; b_i(O_{t+1})\;\; \beta_{t+1}(j)}{P(O|\lambda)}$$

$$\zeta t(i,j) = \frac{\alpha_t (i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t (i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)} \quad , t = 1, \dots , T-1 \qquad \text{.....(3.23)}$$

و الشكل (5) يوضح هذه العملية :



الشكل 6.3: إعادة تقدير نموذج ماركوف المخفي باستخدام $\zeta t(i,j)$

المتغير الثاني هو:

$$\gamma_t(i) = \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\,\beta_t(i)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.24)$$

كما يمكننا حساب المتغير $\gamma_t(i)$ باستخدام المتغير $\zeta_t(i, j)$)

$$\gamma_t(i) = \sum_{j=1}^{N} \zeta_t(i, j) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.25)$$

و باستخدام الصيغ المذكورة أنفا لإعادة تقدير نموذج جديد من النموذج القديم نقوم بالإجراءات الآتية:

العدد المتوقع للانتقالات في الحالة i عند الزمن t

$$\pi_i = \gamma_1(i) \quad (t=1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.26)$$

$$a_{ij} = \frac{\text{العدد المتوقع للانتقالات من الحالة } S_i \text{ إلى الحالة } S_j}{\text{العدد المتوقع للانتقال من الحالة } S_i}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \zeta_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.27)$$

$$b_j(k) = \frac{\text{العدد المتوقع للأزمنة في الحالة } j \text{ و رموز المشاهد } V_K}{\text{العدد المتوقع للأزمنة في الحالة } j}$$

$$b_j(k) = \frac{\sum_{t=1, O_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(t)} \quad ; j = 1, 2, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.28)$$

من المعادلات أعلاه نحصل على نموذج جديد يرمز له ب $\overline{\lambda}$ الذي يحقق [33]:

$$P(O|\overline{\lambda}) > P(O|\lambda) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.29)$$

و باستخدام الصيغ أعلاه يمكن إدراج الصيغ الآتية و ذلك لإعادة تقدير معلمات(HMM) (A,B,π)

بالاعتماد على الخطوات السابقة و باستخدام (و بشكل متكرر) $\overline{\lambda}=(\overline{A}, \overline{B}, \overline{\pi})$ بدلا من $\lambda=(A,B,\pi)$ و تكرار حسابات إعادة التقدير فسوف تعدل الاحتمالية لـ (O) المشاهدة من النموذج إلى أن يتم التوصل إلى بعض النقاط المحددة، و النتيجة النهائية لخطوات إعادة التقدير هذه تعرف بـ (تقدير الإمكان الأعظم لـ (HMM) [28].

كما يجب ملاحظة قيود النموذج $\overline{\lambda}=(\overline{A}, \overline{B}, \overline{\pi})$ أن تكون مقيدة بالقيود الآتية [38]:

$$\sum_{i=1}^{N} \pi_i = 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.30)$$

$$\sum_{i=1}^{N} a_{ij} = 1 \quad ; i = 1, 2, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.31)$$

$$\sum_{i=1}^{N} b_j(k) = 1 \; ; i = 1, 2, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.32)$$

إن الجانب المهم من خطوات إعادة التقدير هو شرط (أو قيد) التصادفية لمعلمات (HMM)، و بالذات المتحققة في كل تكرار أي أن :

إن عملية إعادة التقدير التي تمثل عملية التكرار بدأت من اختيار أفضل قيم تخمينية (أو تقديرية) لـ $\lambda=(A, B, \pi)$ ، و إذا لم تتوافر هذه القيم فسوف يتم اختيار قيم عشوائية بحيث أن :

$$\pi_i \approx 1/N \qquad a_{ij} \approx 1/N \qquad b_j(k) \approx 1/M$$

مع التأكيد على أن تكون (A, B, π) دائما ذات صف تصادفي، و أن لا تكون قيم عناصر المصفوفة متساوية [28].

مع ملاحظة أن تكون مقيدة بالقيود المذكورة سابقا [33].

يتم تدريب نماذج ماركوف المخفية باستخدام خوارزمية باوم ويلش و التي تعطى بالخطوات التالية [28][31][37]:

1 -تهيئة النموذج $\lambda= (A, B, \pi)$

2 -حساب $\alpha_t (i)$

3 -حساب $\beta_t(i)$

4 -حساب $\zeta_t (i, j)$ و التي تمثل احتمال الوجود في الحالة i خلال الزمن t و الانتقال من الحالة j خلال الزمن $t+1$ لمتتالية الملاحظات $O = \{O_1,O_2,O_3, ..O_T\}$ و لنموذج $\lambda= (A, B, \pi)$

5 -حساب $\gamma_t (i,j)$ و التي تمثل احتمال الوجود في الحالة i خلال الزمن t لمتتالية الملاحظات $O = \{O_1,O_2,O_3, ..O_T\}$ و لنموذج $\lambda= (A, B, \pi)$

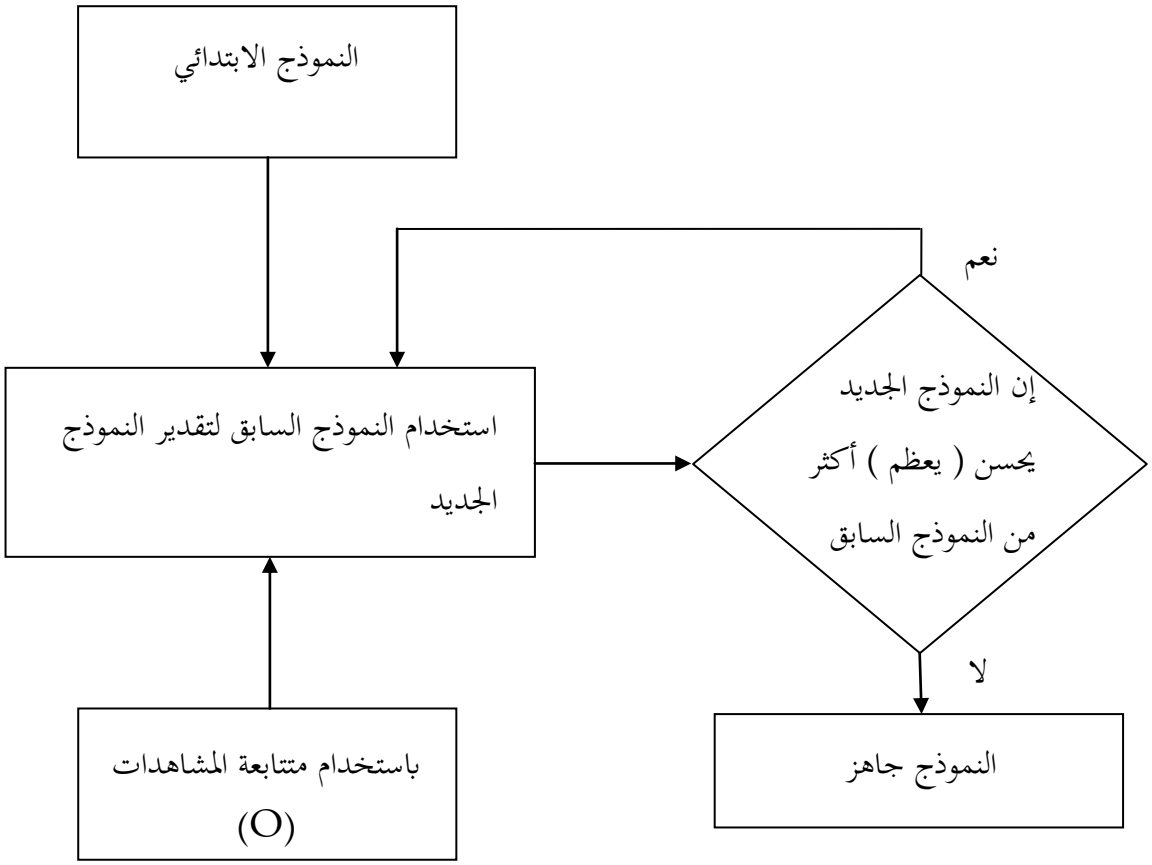
6 -إعادة تقدير النموذج $\lambda= (A, B, \pi)$

7 -إذا ازدادت قيمة $P(O|\lambda)$ و $\lambda = \lambda$ عندئذ اذهب إلى الخطوة 9.

8 -عندما يكون $P(O|\lambda) > P(O|\lambda)$ يدل على انه تم تحسين النموذج و في الحالة المغايرة يتم تكرار الخطوات من 1 إلى 7

9 -النهاية يجب التوقف اذا لم تكن $P(O|\lambda)$ تزايدية.

يمكن توضيح مجمل خطوات خوارزمية (BW) من خلال الشكل (6) [34]:



الشكل 7.3: كيفية تعظيم الاحتمالية لنماذج ماركوف المخفية باستخدام خوارزمية باوم ويلش[33].

5.3. تقييس نماذج ماركوف المخفية:

إن تطبيق الخوارزميات الأساسية لنماذج ماركوف المخفية على المتسلسلات قصيرة الطول يعطي نتائج جيدة. و لكن في حالة المتسلسلات الطويلة فإن الخوارزميات لا تعطي نتائج مقبولة حتى لو تم استخدام أفضل الأجهزة و اللغات الحاسوبية المتطورة و يعود السبب في ذلك إلى كون أن قيم المتغيرين (الأمامي والخلفي) عندما تكون المتسلسلة طويلة تتناقص تدريجيًا نحو الصفر أو القيم القريبة منه وبذلك يصعب حساب خوارزمية (BW) لأن المتغير $\gamma_t(i)$ ينتج من حاصل ضرب كل منهما مما يجعل قيمة المقام كمية غير معرفة [35][38]، لذلك تم اللجوء إلى استخدام أسلوب التقييس الذي يعمل على تقييس كل من $\alpha_t(i)$ و $\beta_t(i)$ فنحصل على α_t

(i) و $\beta_t(i)$ على التوالي على أن تكون متناسبة مع α_t (i) و α_t (i) يكون المجموع يساوي (1) لكل الحالات الممكنة (أي لجميع قيم i) و كالآتي:

هذا و أن:

$$\alpha_t(i) = \prod_{t=1}^{T} \eta_k \, \alpha_t(i) \quad , \; i = 1, \ldots, N \quad ; t = 1, \ldots, T \qquad \ldots\ldots\ldots\ldots\ldots (3.33)$$

و كذلك فان:

$$\prod_{t=1}^{T} \eta_k = \frac{1}{\sum_{i=1}^{N} \alpha_{T-1}(i)} = \frac{1}{P(O|\lambda)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.34)$$

أو:

$$\text{Log}\,[P(O\,|\,\lambda)\,] = -\sum_{k=1}^{T-1} \log(\eta_k) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.35)$$

حيث يمكن حساب لوغاريتم (p) و ليس (p)، هذا و أن التقييس سيمر بعدة خطوات تتضمن إعادة التقدير لكل من a_{ij} و $b_j(k)$ و π_i و يتم الحصول عليها من خلال تطبيق الخطوات الآتية [31] [37]:

يتم حساب (α) بنفس خوارزمية (α) إجراء بعض التغييرات عند كل خطوة و كالآتي[28]:

$$\alpha_0(i) = \frac{\pi_i \, b_i(O_1)}{\sum_{i=1}^{N} \pi_k \, b_i(O_1)} \quad , \; i = 1, 2, \ldots, N \qquad \ldots\ldots\ldots (3.36)$$

$$\alpha_t(i) = \frac{b_i(O_1) \sum_{j=1}^{N} \alpha_{t-1}(i) \, a_{ij}}{\sum_{i=1}^{N} b_k(O_1) \sum_{j=1}^{N} \alpha_{t-1}(i) \, a_{ik}} \quad , \; i = 1, 2, \ldots, N \qquad \ldots. (3.37)$$

$$, \; t = 1, 2, \ldots, T-1$$

ثانيا خوارزمية تقييس الاحتمالية الخلفية:

يمكن تقييس الاحتمالية الخلفية (β) بالطريقة المستخدمة نفسها لتقييس الاحتمالية الأمامية (α) ذلك أن:

$$\beta_t(i) = \beta_t(i) \prod_{t=1}^{T} \eta_k \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.38)$$

حيث أن $\beta_T(i) = \beta_T(i)$ ، و أن:

$$\alpha_t(i)\ \beta_t(i) = \prod_{t=1}^{T} \eta_k\ \alpha_t(i)\ \beta_t(i) = \frac{\alpha_t(i)\ \beta_{t+1}(j)}{P(O|\lambda)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3.39)$$

حيث يتم حساب الاحتمالي الخلفية (β) بطريقة حساب الاحتمالية الخلفية (β) نفسها مع إدخال (η) (المحسوبة سابقا من خوارزمية تقييس الاحتمالية الأمامية (α)) في القانون و كالاتي:

$$\beta_T(i) = \beta_T(i) = 1 \qquad , i = 1, \ldots, N \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3.40)$$

$$\beta_t(i) = \eta_{t+1} \sum_{i=1}^{N} \beta_{t+1}(j)\ a_{ij}\ b_j(O_{t+1}) \qquad , t = 1, \ldots, T \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3.41)$$

ثالثا صيغ التحديث باستخدام α و β :

يمكن حساب الـ (γ) باستخدام الـ (α) و الـ (β) و كالآتي:

$$\gamma_t(i) = \frac{\alpha_t(i)\ \beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\ \beta_t(j)} \qquad , i = 1, \ldots, N \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3.42)$$

$$; t = 1, \ldots, T-1$$

و حساب (ζ) بصيغ مختلفة و كالآتي:

$$\zeta_t(i,j) = \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\ \eta_{t+1}\, \beta_{t+1}(j)}{\sum_{j=1}^{N} \alpha_t(j)\, \beta_t(j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.43)$$

أو:

$$\zeta_t(i,j) = \frac{\gamma_t(i)\, a_{ij}\, b_j(O_{t+1})\ \eta_{t+1}\, \beta_{t+1}(j)}{\beta_t(i)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.44)$$

و باستخدام (γ) و (ζ) المحسوبتين أعلاه، يمكن حساب (a, b, π) المعدلة (و التي يرمز لها بـ (a, b, π)) و كالآتي [28]:

$$\pi_i = \gamma_1(i) \qquad , i = 1, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.45)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \zeta_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad , i,j = 1, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.46)$$

$$b_j(k) = \frac{\sum_{t=1,\ O_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(t)} \qquad ; j = 1, 2, \ldots, N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.47)$$

رابعا خوارزمية تقييس الاحتمالية(δ) [37][38]:

أما عند استخدام خوارزمية (Viterbi) لإعطاء الإمكان لمتسلسلة الحالة، فليس من الضروري عمل تقييس إذا استخدمت اللوغارتمات بالطريقة الآتية، حيث يعرف المقدار $\delta_t(i)$ كالآتي:

$$\delta_t(i) = \max_{q1, q2, \dots qt} \log\{ P(q_1 q_2, \dots q_t, O_1 O_2 \dots O_t | \lambda) \} \quad \dots..(3.48)$$

و هذه الاحتمالية يمكن إيجادها بشكل تعاقبي و كالآتي:

البداية:

$$\delta_t(i) = \log(\pi_i) + \log[b_i(O_1)] \quad , i = 1, \dots, N \quad \dots\dots\dots\dots\dots(3.49)$$

$$\psi_1(i) = 0 \quad \dots\dots\dots\dots\dots(3.50)$$

التعاقب:

$$\delta_t(i) = \max_{1 \le i \le N} [\delta_{t-1}(i) + \log(a_{ij})] + \log[b_j(O_t)] \quad , t = 1, \dots, T \dots (3.51)$$

$$J = 1, \dots, N$$

$$\psi_t(j) = \operatorname*{argmax}_{1 \le i \le N} [\delta_{t-1}(i) + \log(a_{ij})] \quad , t = 1, \dots, T \quad \dots\dots\dots\dots(3.52)$$

$$J = 1, \dots, N$$

النهاية:

$$\text{Log } P^* = \max_{1 \le i \le N} [\delta_T(i)] \quad \dots\dots\dots\dots (3.53)$$

$$q^*_T = \operatorname*{argmax}_{1 \le i \le N} [\delta_T(i)] \quad \dots\dots\dots\dots (3.54)$$

التتابع المعاكس

$$Q^* = \{ q^*_1 \dots q^*_T \} \quad \dots\dots\dots\dots(3.55)$$

$$q_t^* = \psi_{t+1}(q^*_{t+1}) \quad , t = T-1, T-2, \dots, 1 \quad \dots\dots\dots\dots(3.56)$$

4. الخاتمة:

إن تطور تقنيات نماذج ماركوف المخفية جعلها أداة إحصائية فعالة لمعالجة العمليات التي تتغير مع تغير الزمن، فهي تساعد على ملاحظة مجموعة من الحالات و إيجاد احتمال وقوعها بعد تدريبها على حالات أخرى تمت ملاحظتها سابقا، و من المميزات الخاصة التي تتمتع بها نماذج ماركوف المخفية هي الخاصية الماركوفية و هي أن الحالة التي تحدث في زمن ما تعتمد في حدوثها فقط على الحالة في الزمن الذي يسبقها مباشرة، كما أن لنماذج ماركوف المخفية فضاء حالة يضم مجموعة منتهية من الحالات التي يمكن أن تكون رمزا أو حرفا أو كلمة أو أي شيء آخر، مما جعلها قابلة للاستخدام في العديد من المجالات كالتعرف على الكلام.

الفصل الرابع

تجارب و نتائج

1. مقدمة:

لقد قمنا بانجاز نظام التعرف الآلي على الكلام العربي المتصل القائم على نماذج ماركوف المخفية(HMM) بالاعتماد على مجموعة أدوات ماركوف المخفية(HTK) من اجل التقليل من خطوات البرمجة للأجزاء المختلفة من نظام التعرف، و مجموعة أدوات HTK عبارة عن برنامج مفتوح المصدر لانجاز العمليات المختلفة، تتميز بأنها فعالة و مرنة من ناحية حرية اختيار الأوامر و إمكانية إضافة نماذج أخرى.

و الجدول 1.4 يوضح قائمة من البرامج المجانية التي تعتمد نماذج ماركوف المخفية.

	HTK	CSLU (ASR)	ISIP (ASR)	SPHINX
المؤسسة	Microsoft et Combridge University	Oregon graduate institut	Mississipi University	Camegie Mellon University
URL	htk.eng.cam.ac.uk	www.cslu.ogi.edu	www.isip.msstate.edu	fife.speeshcs.cmu.edu
لغة البرمجة	C	C, Tel/Tk	C++	C, perl, Java
بيئة العمل	Unix, Linux, windows	Windows	Unix	Unix, Linux, windows
تاريخ أول إصدار	1993	1992	1997	1987

الجدول 1.4: خصائص بعض البرامج المستعملة في انجاز أنظمة التعرف الآلي على الكلام التي تعتمد على نماذج ماركوف المخفية[42].

و قد قمنا بتحقيق النظام بواسطة:

1- مجموعة أدوات HTK

2- استخدام برنامج متخصص في تحليل الإشارة الصوتية PRAAT.

3- لغة البرمجة Perl.

2. انجاز المدونة:

الخطوة الأولى التي يتعين القيام بها قبل البدء بانجاز نظام التعرف هو انجاز مدونتي التدريب و الاختبار، ففي عملنا هذا اتخذنا مدونة تتكون من مجموعة من الجمل باللغة العربية تحتوي على الفونيمات في مواضع مختلفة (بداية و وسط و نهائية الكلمة) حتى يكون التعرف على الفونيمات في سياقات حقيقية مختلفة، و لهذا الغرض اخترنا مجموعة من المتكلمين لانجاز المدونة و تسجيل الجمل في شكل ملفات صوتية هم ثلاثة أشخاص واحد من ذكر و اثنان من الإناث تتراوح أعمارهم بين (24-31عاما)، قام كل واحد منهم بتسجيل الجمل خمس مرات.

الفونيم	الكتابة الصوتية	الفونيم	الكتابة الصوتية
الفتحة	AE	س	S
الألف الممدودة	AE:	ش	SH
الضمة	UH	ص	SS
الضمة الممدودة	UW	ض	DD
الكسرة	IH	ط	TT
الكسرة الممدودة	IY	ظ	DH2
همزة (ء)	E	ع	AI
ب	B	غ	GH
ت	T	ف	F
ث	TH	ق	Q
ج	JH	ك	K
ح	HH	ل	L
خ	KH	م	M
د	D	ن	N
ذ	DH	ه	H
ر	R	و	W
ز	Z	ي	Y

الشكل 1.4: قائمة الفونيمات الأساسية للغة العربية.

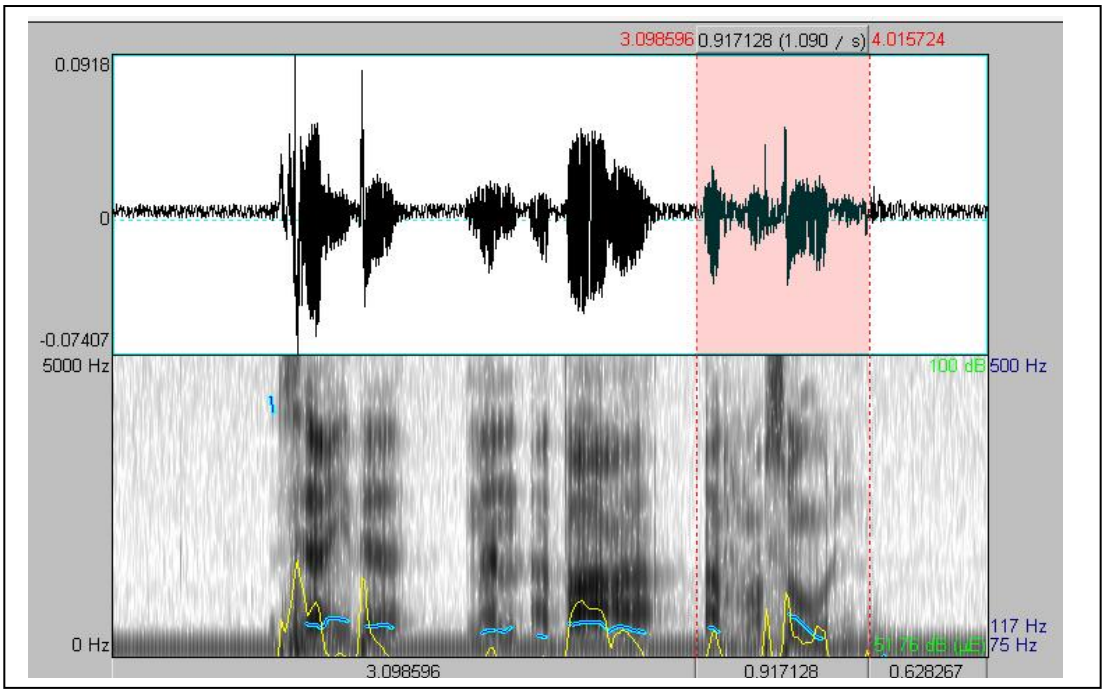
3. مرحلة تسجيل الأصوات:

عند إصدار الكلام أمام لاقط الصوت (الميكروفون) تتغير الموجات الصوتية إلى موجات تماثلية تقوم بطاقة الصوت بتحويلها إلى عينات رقمية، لذلك فانه عند تسجيل إشارة الكلام يتم تقطيع هذه الإشارة ضمن بطاقة تحصيل الصوت بتردد تقطيع مناسب، لذا علينا الانتباه إلى خصائص تسجيل هذه الإشارة و التي تتمثل في تحديد طريقة الترميز.

وفقا لما سبق فان تسجيل الكلمات ضمن هذا البحث تم بتحديد خصائص التسجيل كما يلي:

الخاصية	القيمة
التقطيع	11025 kHz, 16 bit
الصيغة المستخدمة	Mono, WAV
المدونة	مجموعة من الجمل باللغة العربية
المتكلمون	3 (واحد ذكر و اثنان إناث)

جدول 3.4: الخصائص المختارة من اجل انجاز المدونة



الشكل 1.4: تسجيل و معاينة الأصوات باستعمال برنامج PRAAT

4. مراحل انجاز النظام:

1.4. تحضير الملفات الضرورية التي نحتاجها في إنجاز النظام:

تبدأ مرحلة انجاز نظام التعرف بتحديد الموارد الضرورية التي نحتاجها في وقت لاحق، إذ نقوم بتعريف المدونة في ملف (corpus)، كما نقوم بتعريف listeMots و هي قائمة الكلمات التي تشكل المدونة corpus و نضيف إليها (SENT-END وSENT-START) و كلمات هذه القائمة مرتبة حسب الترتيب الأبجدي.

```
آلَامُ
آمِنٍ
آيَاتُ
.
.
يُوقِعُ
يُوَلِّدُ
SENT-END
SENT-START
```

الشكل 2.4: listeMots قائمة الكلمات التي تشكل قاعدة البيانات corpus

حيث سيتم كتابة فونيمات كلمات listeMots كتابة صوتية و تحفظ في القاموس Dictionnaire.

```
آلَامُ E AE: L AE: M UH
آمِنٍ E AE: M IH N IH N
آيَاتُ E AE: Y AE: T UH
.
.
يُوقِعُ Y UW Q IH AI UH
يُوَلِّدُ Y UH W AE L IH D UH
SENT-END [] sil
SENT-START [] sil
```

الشكل 3.4: القاموس Dictionnaire

بعدها سيتم تعريف قائمة الفونيمات (ListePhones) التي يتم استنتاجها بواسطة HDMan لتحديد معجم لقاعدة البيانات corpus بشكل قائمة و هي على النحو التالي[43]:

E ; AE:; L; M ; UH ; IH ; N ; Y ; T ; AE ; B ; IY ; TH ; JH ; H ; GH ; AA ; HH ; D ; Z ; F ; DH2 ; Q ; R ; KH ; DD ; AH ; UX ; S ; DH ; SH ; SS ; AH:; AA:; TT ; W ; K ; AI ; UW ; IX ; sil

```
HDMan -A -D -T 1 -m -w ListeMots\\listeMots.txt -n Dictionnaire\\ListePhones -i -l Dictionnaire\\dlog Dictionnaire\\dict_htk  Dictionnaire\\Dictionnaire.txt
```

بعدها نقوم ببناء برنامج باستعمال Perl يقوم بإنشاء ملف Transcription_mots يحتوي اسم الملف الصوتي المخصص للتدريب و الجملة التي تقابله بشكل قائمة من الكلمات.

2.4. تحليل إشارة الكلام و بناء قاعدة البيانات:

نلاحظ وجود عدد من المشاكل المتعلقة بتعقيد الكلام،و العدد الكبير من المعلومات غير المهمة في عملية التعرف الذي تحتوي عليه الإشارة الصوتية فضلاً عن أنها تحتاج إلى مساحة خزن كبيرة، فمن أجل حل المشاكل و تسهيل خطوات المعالجة و تقليلها يتم استخلاص الصفات الأساسية للإشارة الصوتية و بناء قاعدة البيانات من معاملات (MFCC) مما يؤدي إلى تقليص حجم البيانات، حيث يتم أولا تحديد الجملة و قراءة الملف الخاص بها ثم يحلل هذا الملف بصورة منفصلة لاستخلاص متجهات الصفات لغرض خزنها في قاعدة البيانات. و تكرر هذه العملية مع جميع الجمل لغرض استكمال قاعدة البيانات، علماً أن عدد معاملات المتجه الواحد لكل إطار من إطارات الإشارة يساوي 12 معاملا من معاملات MFCC الأولى باستثناء معامل C0. من اجل كل معامل يتم اختيار المشتق الأول (12 المشتقات الأولى في المجموع) و المشتق الثاني (12 المشتقات الثانية)، بهذا يتم الحصول على متجه من 36 من المعاملات الصوتية المقابلة لكل إطار في الإشارة.

يتم استخراج هذه المعاملات لملفات WAV على نوافذ قدرها 25 ms بفضل الأداة HCopy باستخدام ملف التكوين كمعلمة إدخال حسب الأمر[43].

HCopy -D -C config -S liste_fich

ملف التكوين config يسمح بتحديد المعايير الضرورية من اجل مرحلة التحليل الصوتي.

```
#Example of an acoustical analysis configuration file
SOURCEFORMAT = WAV
WINDOWSIZE = 250000.0 # = 25 ms = length of a time frame
TARGETRATE = 100000.0 # = 10 ms = frame periodicity
ENORMALISE=F
USEHAMMING = T # Use of Hamming function for windowing frames
ZMEANSOURCE=T
PREEMCOEF = 0.97 # Pre-emphasis coefficient
NUMCHANS = 26 # Number of filterbank channels
CEPLIFTER = 22 # Length of cepstral liftering
TARGETKIND = MFCC_D_A
NUMCEPS = 12
```

الشكل 4.4: ملف التكوين config[43].

3.4. التدريب:

قبل بداية مرحلة التدريب، ننشئ لكل وحدة صوتية نموذج HMM ذو 3 حالات من الانتقالات المسموح بها من اليسار إلى اليمين يتم تهيئتها في المصفوفة الانتقالية، و يتم تهيئة المتوسط بالقيمة 0 و التباين بالقيمة 1 (انظر الجدول ملف التهيئة)، و هذه القيم لنموذج HMM سوف يعاد قياسها لاحقا خلال مرحلة التدريب.

```
 ~o <VecSize> 36 <MFCC_D_A> <StreamInfo> 1 36
 ~h "proto"
<BeginHMM>
 <NumStates> 5
 <State> 2 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 36
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 36
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 3 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 36
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 36
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 4 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 36
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 36
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <TransP> 5
  0.000e+0  1.000e+0  0.000e+0  0.000e+0  0.000e+0
  0.000e+0  6.000e-1  4.000e-1  0.000e+0  0.000e+0
  0.000e+0  0.000e+0  6.000e-1  4.000e-1  0.000e+0
  0.000e+0  0.000e+0  0.000e+0  6.000e-1  4.000e-1
  0.000e+0  0.000e+0  0.000e+0  0.000e+0  0.000e+0
<EndHMM>
```

الشكل 5.4: ملف التهيئة[43].

تسمح مرحلة التدريب بتشكيل النماذج المرجعية للنظام حيث أن نوعية هذه النماذج تؤثر بشكل كبير على نتائج نظام التعرف على الكلام، و يتم تنفيذ التدريب باستعمال أدوات HTK من خلال خطوتين

رئيسيتين هما: التهيئة و إعادة التقييم، و لهذا الغرض فان أكثر الأدوات المستعملة هما الأداتين: HCompV و HERest[43].

4.4. مرحلة تهيئة نماذج HMM بواسطة الأداة HCompV :

تسمح مرحلة التهيئة بتحديث قيم المتوسط و التباين اللذان تساوي قيمهما قبل هذه الخطوة على التوالي 0 و 1، و يتم تنفيذ هذا التحديث على مجموعة بيانات التدريب حتى يسمح في النهاية بالوصول إلى القيم العامة المراد إيجادها لكل حالة HMM باستخدام الأمر:

```
HCompV -C hinit.conf -f 0.01 -m -M HMM0 -S data_train  proto
```

و بعد هذا الأمر، يتم الحصول علي ملف proto جديد في المجلد HMM0 يحتوي على قيم عامة للمتوسط و التباين.

نقوم بإنشاء ملف عام fich_mmf يحتوي نموذجا لكل صوت نقوم بالخطوات التالية:

- وضع الصوت بين علامتي تنصيص(")؛
- إضافة ' h~' قبل الصوت (لاحظ المسافة بعد ' h~')؛
- نسخ محتوى ملف hmm0/proto من السطر 5 فصاعدا (أي بدءا من "<BEGINHMM>" إلى "<ENDHMM>") و لصقه بعد كل صوت.
- ترك سطر فارغ واحد في نهاية الملف.

ننسخ محتويات الملف عدة مرات مثل عدد الفونيمات التي لدينا و نخزن نتيجة النسخ في ملف كلي اسمه fich_mmf، و بالتالي سوف تتم تهيئة جميع الفونيمات بنفس قيمة المتوسط و التباين.

إضافة إلى ذلك، فإن الخيار -f في الأمر HCompV يسمح بإنشاء ملف vFloor يحتوي على عتبة التباين الذي هو جزء من التباين الكلي المقدر، و الفائدة من هذه العتبة هي وضع حد للتباين خلال خطوات التقدير لتجنب القيم المتطرفة.

من ناحية أخرى نقوم بإنشاء الملف macros هو نسخة من vFloor و التي أضفنا إليها رأس الملف،

~o

<STREAMINFO> 1 36

<VECSIZE> 36<NULLD><MFCC_D_A><DIAGC>

الذي يحدد نوع المعاملات و حجم المتجه MFCC، فينتج لنا ملف كالتالي:

```
~o
<STREAMINFO> 1 36
<VECSIZE> 36<NULLD><MFCC_D_A><DIAGC>
~v "varFloor1"
<VARIANCE> 36
 4.513063e-001 5.118352e-001 7.908874e-001 3.573665e-001 6.221036e-001
5.114855e-001 5.549700e-001 7.331246e-001 5.200121e-001 3.645825e-001
5.647354e-001 3.038846e-001 2.246906e-002 2.816695e-002 2.729449e-002
2.325620e-002 2.999017e-002 2.625237e-002 3.628151e-002 3.560720e-002
3.053252e-002 2.753117e-002 3.004724e-002 2.275587e-002 3.346150e-003
4.417596e-003 4.207981e-003 4.052571e-003 5.097604e-003 4.646412e-003
6.421683e-003 6.189519e-003 5.444875e-003 5.192263e-003 5.462960e-003
4.199102e-003
```

الشكل 6.4: إنشاء الملف macros.

5.4. إعادة تقييم نماذج HMM:

تحسين نماذج HMM يتوقف على إعادة تقييم معاملاتها (المتوسط و التباين) باستعمال خوارزمية بوم ولش بفضل الأداة HERest و ذلك من خلال تكرارها ثلاث مرات.

النماذج الناتجة سيتم حفظها في الملف hmm3 و هذا بإعادة تقييم نماذج HMM الواردة في الملف HMMn و حفظها في الملف ذو الترتيب الموالي أي HMMn+1.

HERest – C config – I phones0.mlf – t 250.0 150.0 1000.0 – S liste train –H hmm0/macros – H hmm0/modeles0.mmf – M hmm1 monophones0

6.4. إدراج نموذج التوقف "sp"و إعادة تقييم نماذج HMM:

إلى غاية هذا المستوى نحن لم نعتمد بعد نموذج التوقف "sp" حيث أن الملف ListePhones لا يحتوي على الصوت "sp"، لهذا فإننا في هذه المرحلة سنقوم بإدراج نموذج التوقف "sp" و إعادة تقييم نماذج HMM ، و لهذا الغرض نقوم بإنشاء ملف ListePhones_sp و ننسخ فيه محتوى الملف ListePhones ثم نضيف إليه الصوت "sp".

أما بخصوص إضافة نموذج التوقف "sp" إلى نماذج HMM فانه ينجز أولا بنسخ محتويات المجلد hmm3 إلى hmm4 ثم باستخدام أوامر برنامج Perl ننشئ نموذجا جديدا "sp" في الملف hmm4/fich_mmf بنسخ النموذج "sil" و لصقه و إعادة تسميته بـ "sp" (لا نحذف النموذج "sil" القديم الخاص بنا، فنحن في حاجة إليه و لكن مجرد اخذ نسخة منه)

حيث يتم تهيئة مصفوفة انتقال هذا النموذج بقيم عشوائية و سيتم في وقت لاحق إعادة تقديرها.

خلال الخطوة التالية و بفضل أداة HHEd التي تسمح بإرفاق نموذج التوقف "sp" بنموذج الصمت "sil" (انظر ملف sil.hed).

محتوى الملف sil.hed[43]:

AT 2 4 0.2 sil.transP

AT 4 2 0.2 sil.transP

AT 1 3 0.3 sp.transP

TI silst sil.state[3],sp.state[2]

HHEd – H hmm4/macros – H hmm4/fich_mmf – M hmm5 sil.hed monophones1

و بعد هذا الأمر، ينتج لدينا ملف آخر fich_mmf في المجلد hmm5 .

سيتم إعادة تقييم النماذج الواردة في هذا الملف بعد التكرار لخوارزمية باوم ولش ممثلة في الأداة HERest تماما كما في الخطوة السابقة، و يتم تخزين تقديرات المعاملات الأخيرة في هذه المرحلة في المجلد hmm7.

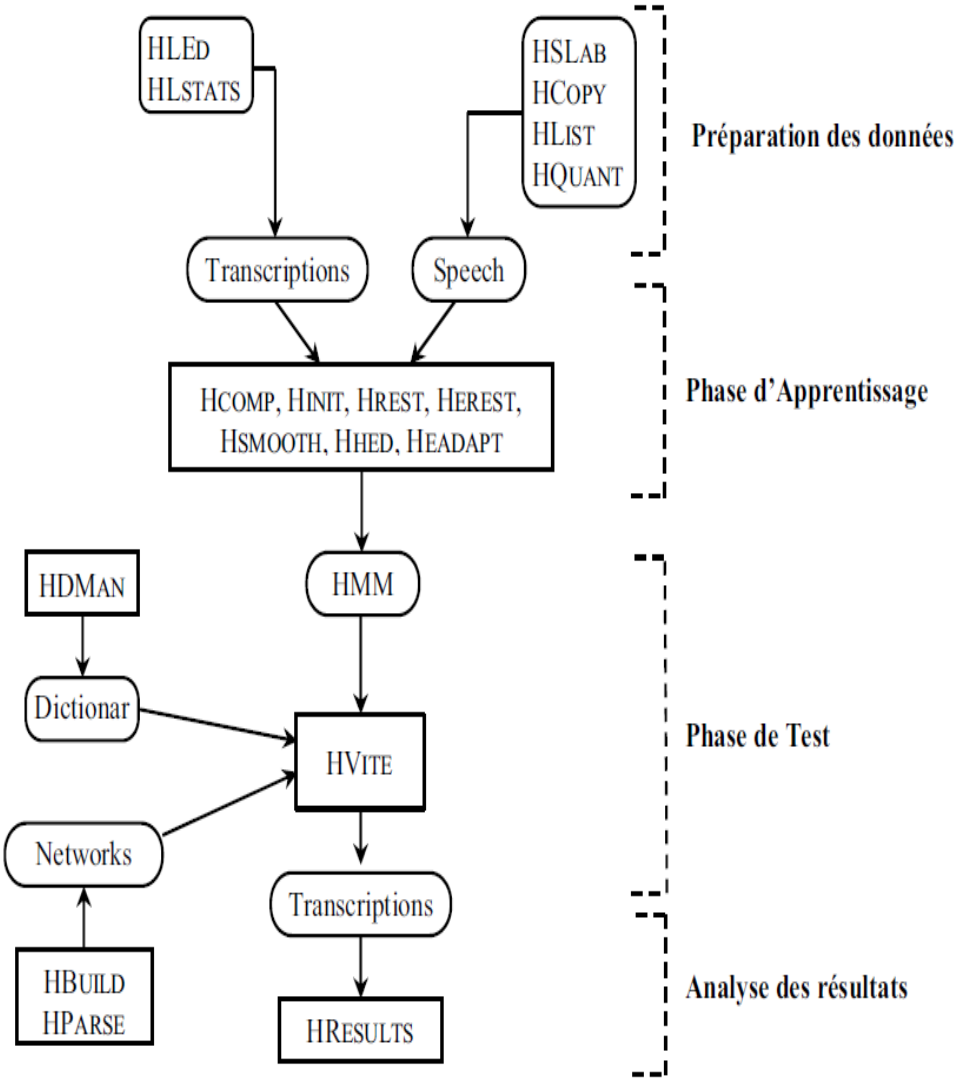
التعرف بواسطة الأداة HVite[43]:

إن نطق الكلام ليس عملية مضبوطة لأن نطق الكلمة نفسها قد يتم بطرق مختلفة من شخص إلى آخر أو بالنسبة للشخص نفسه، مما يؤدي إلى تنوع شكل الإشارة الصوتية الموافقة لها، و ينتج هذا الاختلاف في النطق بحسب سرعة كلام الشخص و جودة صوته إن كان يقصد الإفهام أو يتكلم على نحو عارض، كما يمكن أن تكون التغيرات الفيزيولوجية و النفسية للمتكلم سببا في تغير النطق كالفرح و الحزن و الغضب و المرض...و غيرها، و لجعل نظام التعرف على الكلام فعالا في التعامل مع هذا التنوع في النطق يجب النظر في جميع تغيرات النطق الممكنة بتنظيم قاعدة التدريب و تحليل كل كلماتها المذكورة في القاموس(dictionnaire) و اختيار الأقرب و الأنسب منها بالمقارنة مع الإشارة الصوتية المدخلة و يتم إنجاز هذا العمل بواسطة الأداة HVite:

HVite -A -D -T 1 -l * -o SW -b SENT-END -H sDr\\macros -H sDr\\fich_mmf -i ListeMots\\aligned.mlf -m -t 250.0 150.0 1000.0 -y lab -a -I transcriptionMots -S dataFiles Dictionnaire\\Dictionnaire.txt hmmListr>HViteLog

يسمح تكرار خوارزمية باوم وولش المستخدمة بإعادة تقييم النماذج مع مراعاة التحسينات الأخيرة، ثم يتم تخزين المعاملات الأخيرة الناتجة في مجلد جديد.

بهذا تختتم مرحلة التدريب لنماذج HMM.



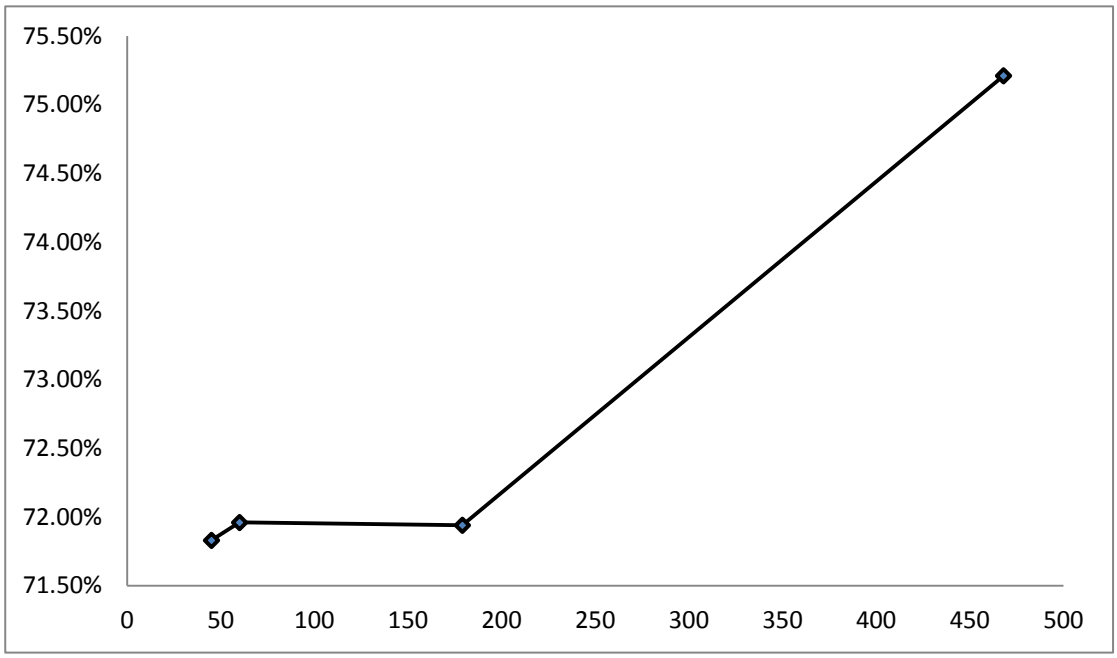
الشكل 7.4: بنية نظام للتعرف بواسطة HTK[43].

5. الاختبارات و التجارب:

نلاحظ أنه خلال انجازنا لنظام التعرف على الكلام كان هناك العديد من المتغيرات التي يمكن أن تؤثر على نسبة التعرف و جودتها، و للوصول إلى أحسن نتائج ممكنة بأقل قدر من العمليات سنقوم بعدة تجارب نلاحظ من خلالها تغيرات نتائج التعرف بدلالة هذه المتغيرات و التي هي: **تغير حجم المدونة و تكرار تسجيل جمل المدونة و نوع المعاملات المستخدمة بالإضافة إلى تكرار التهيئة بـ** HERest.

1.5. تغير نسبة التعرف بدلالة تغير حجم المدونة:

يظهر المنحنى التالي تغير نسبة التعرف عند تغيير حجم المدونة، حيث إن الاختبار تم على مجموعة مؤلفة من 117 جملة تم تكرار تسجيلها 5 مرات من طرف 3 متكلمين واحد من الذكور و اثنان من الإناث، أي لدينا 585 تسجيل، كنا نستخدم في كل مرة جزء منها للتدريب و جزء آخر للاختبار.



الشكل 8.4: منحنى تغير نسبة التعرف بدلالة حجم المدونة

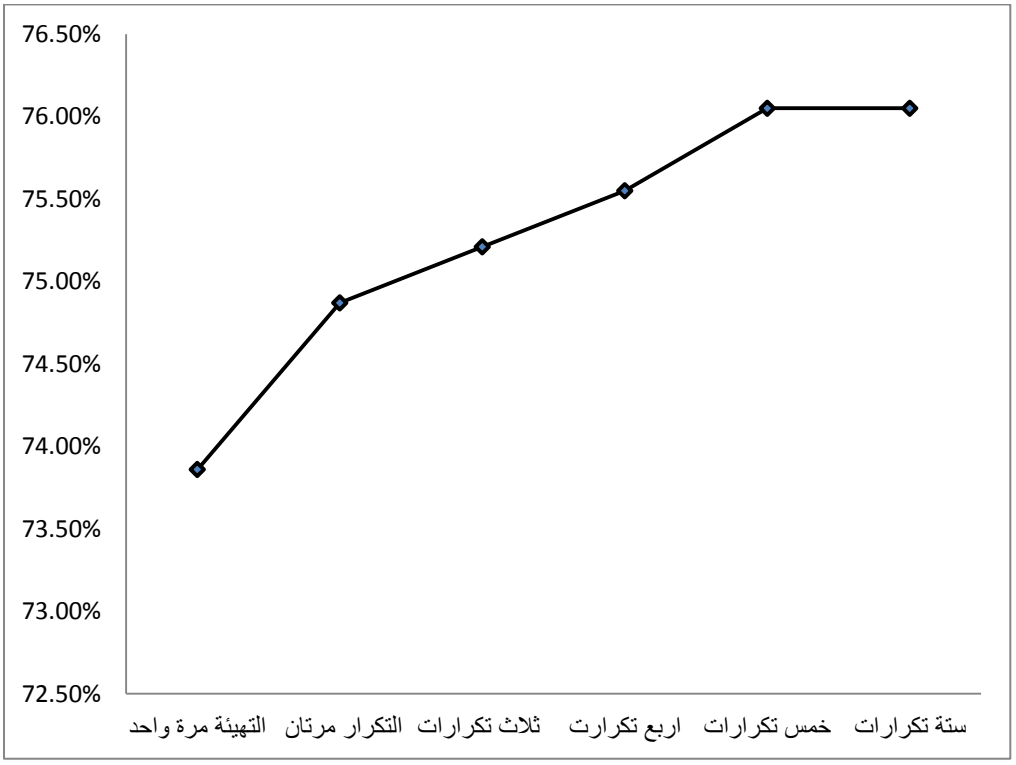
نلاحظ انه كلما زاد حجم المدونة في النظام زادت نسبة التعرف مما يعني أن للمدونة و حجمها اثر ظاهر على النتائج المتحصل عليها من النظام و هذا يفرض علينا إعطاء أهمية كبيرة للمدونة من حيث الحجم و نوعية التسجيل.

و رغم أن زيادة حجم المدونة يؤدي إلى زيادة في نسبة التعرف إلا انه يؤدي أيضا إلى زيادة في العمليات الحسابية، و بالتالي زيادة في وقت المعالجة خاصة مرحلة التدريب، إلا أن هذا ضروري لزيادة فاعلية نظام التعرف بصورة عامة، لأنه بعد عملية التدريب سنتحصل على نماذج سهلة الاستخدام عند الاختبار حيث لا يظهر لحجم مدونة التدريب اثر في مدة التعرف، و لهذا يقترح أن يكون هناك مكتبة من التسجيلات لأحد معاجم اللغة

العربية تساعد على تحسين نسبة التعرف للأنظمة التي هي من هذا النوع و للتسهيل على الباحثين القيام بأبحاثهم و دراساتهم.

2.5. تغير نسبة التعرف بدلالة تكرار التهيئة بـ HERest:

إن التعرف على الكلام باستخدام HMM يتم باستخدام خوارزمية فيتربي، و تستند هذه الخوارزمية على الاحتمال الذي يحسب كل مرة بين الكلمة المراد اختبارها و نموذجها، مما يعطي أهمية كبيرة لحساب النماذج و إعادة تهيئتها و الذي يتم تنفيذه باستخدام الأمر HERest في مجموعة الأدوات HTK، و تظهر نتائج تقييم التعرف على الكلام المتحصل عليها بدلالة تغير تكرار التهيئة بـ HERest في المنحنى :



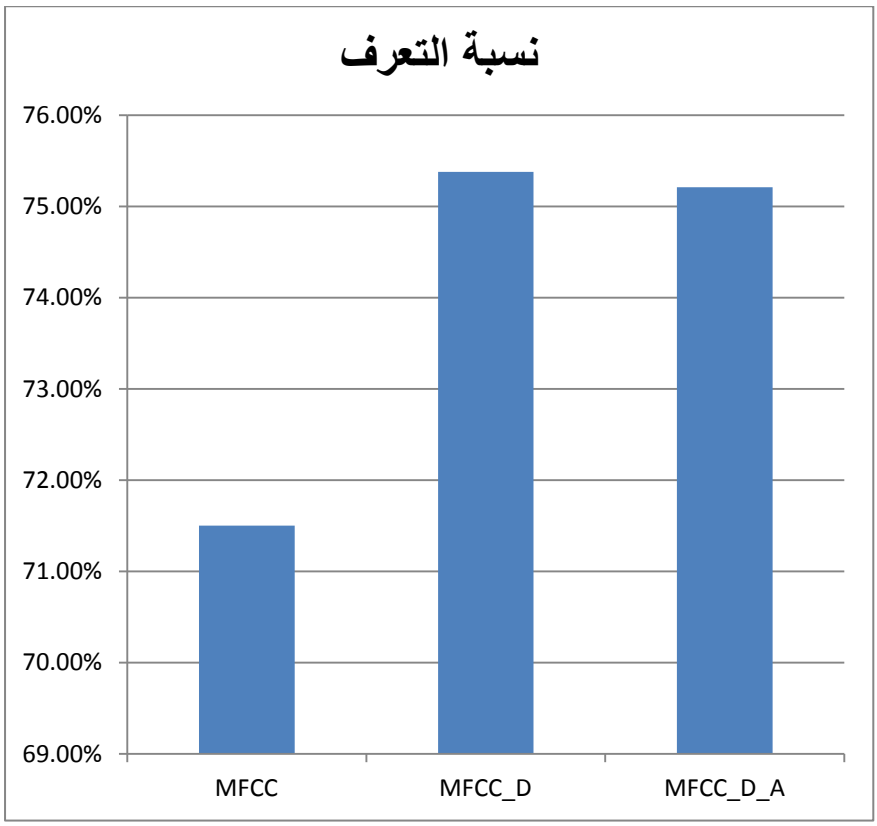
الشكل 9.4: منحنى تغير نسبة التعرف بدلالة تغير عدد تكرارت إعادة التقييم

و بما أن تكرار عملية إعادة تقييم النماذج تأخذ زمنا لا يستهان به في مرحلة التدريب، و نظرا لان نسبة التعرف ابتداء من التكرار الثالث متقاربة، إذا يمكن الاكتفاء بتكرار عملية إعادة تقييم النماذج 3 مرات فقط.

و تبين النتائج الموضحة أعلاه أن نظام التعرف المقترح الذي يعتمد على خوارزمية بوم ولش في إعادة تقييم النماذج ذو كفاءة عالية من حيث عدد الكلمات التي تم تقييمها بشكل صحيح و هذا يترجم قدرة و فعالية نماذج ماركوف المخفية، الذي يكمن أساسا في أنه يقدم خوارزميات قوية للتدريب.

3.5. تغير نسبة التعرف حسب نوع المعاملات المستخدمة:

تم استخدام نظام التعرف بالاعتماد على معاملات من نوع MFCC، باستخدام مجموعات مختلفة من المعاملات الصوتية و هذه المعاملات هي: MFCC، ΔMFCC، ΔΔMFCC، حيث أننا سوف نركز اهتمامنا على هذه المعاملات للقيام بالمقارنة و الجمع بين نتائجها من أجل تحسين دقة التعرف.



الشكل 10.4: تغير نسبة التعرف بدلالة نوع المعاملات المستخدمة

يبين الشكل أن قيم نسبة التعرف التي تم الحصول عليها مع المعاملات الصوتية تتغير حسب نوع المعاملات المختارة من المعاملات الصوتية الثلاثة: MFCC، ΔMFCC، ΔΔMFCC، حيث نلاحظ ارتفاع دقة

التعرف بالنسبة لمعاملات ΔΔMFCC+ΔMFCC+MFCC التي تمثل 36 معاملا حيث تعطي نسبة تعرف 75.21% و نتيجة معاملات ΔMFCC+MFCC (24 معاملا) بنسبة تعرف 75.38% و أما معاملات MFCC (12 معاملا) فهي تعطي دقة التعرف 71.5 % و هي الأقل، و تظهر هذه النتائج أن معاملات ΔΔMFCC+ΔMFCC+MFCC و ΔMFCC+MFCC هي الأكثر ملاءمة لنظام التعرف المقترح، و أن النظام يعطي دقة تعرف اكبر مع عدد معاملات اكبر من 24 معاملا.

4.5. تغير نسبة التعرف بدلالة تغير المتكلم:

بعد تعرضنا لعدد من المتغيرات التي تؤثر على أداء نظام التعرف قمنا بتصميم برنامج للتعرف على الكلام العربي المتصل، باختيار مدونة مؤلفة من 117 جملة تم تكرار تسجيلها 5 مرات من طرف 3 متكلمين واحد ذكر و اثنان من الإناث، أي لدينا 585 تسجيل استخدمنا منها للتدريب و للاختبار، تم تحليلها باستخدام معاملات ΔΔMFCC+ΔMFCC+MFCC ، و تم تكرار حساب النماذج و إعادة تقييمها ثلاث مرات بواسطة خوارزمية بوم ولش الذي يتم تنفيذه باستخدام الأمر HERest في مجموعة الأدوات HTK:

المتكلم	نسبة التعرف
متكلم ضمن مدونة التدريب	% 75.21
متكلم خارج مدونة التدريب	% 62.26

الجدول 5.4: تغير نسبة التعرف بتغير المتكلم

نلاحظ أن نسبة التعرف على تسجيلات مجموعة التدريب كانت 75.21 % أما نسبة التعرف على مجموعة تسجيلات خارج مجموعة التدريب كانت 62.26 % حيث أظهرت النتائج أن نسبة التعرف الأكبر كانت لتسجيلات المتكلم الذي كان ضمن مدونة التدريب، و هذا يعني أن النظام يعمل بشكل جيد لمتحدث ضمن مدونة التدريب كما أظهرت النتائج في الجدول.

الجمل الأصلية المراد التعرف عليها	الجمل الناتجة عن نظام التعرف
أَحَطْتُ بِأَضْدَادِ هَذِهِ الْكَلِمَةِ	الْحَقِّ بِأَضْدَادِ هَذِهِ الْكَلِمَةِ
يَعِيْشُ كَثِيْرٌ مِنَ ٱلْأَشْخَاصِ فِيْ أَحْزَانٍ	يَعْرِفُ كَثِيْرٌ مِنَ ٱلْأَشْخَاصِ فِيْ أَحْزَانٍ
حِيْنَمَا تَثُوْرُ الْأَشْجَانُ تَتَغَيَّرُ الْأَمْزِجَةُ	حِيْنَمَا تَثُوْرُ الْأَشْجَانُ تَتَغَيَّرُ الْأَمْزِجَةُ
أَحْضَرْتُ مِنْ الشَّارَقَةِ أَشْيَاءَ لِلْبَيْعِ	لَحْظَةِ مِنْ الشَّارَقَةِ أَشْيَاءَ لِلْبَيْعِ
فِيْ هَذِهِ الْأَكْوَابِ أَمْشَاجٌ مِنَ الشَّرَابِ	فِيهَا لِي الْأَكْوَابِ أَمْشَاجٌ مِنَ الشَّرَابِ
أَمَّا فِعْلُ الْخَيرِ فَحَسَنٌ	فَمَا فِعْلُ الْخَيرِ فَحَسَنٌ
نَسْمَعُ ٱلْأَنْغَامَ عِنْدَ عَزْفِ الْمَعَازِف	يَسْعَى ٱلْأَنْغَامَ عِنْدَ عَزْفِ الْمَعَازِف
أَبْجَلَنِي هَذَا الطَّعَامُ	الرِّجَالِ فِيهَا الطَّعْنِ
أَنْذَرَ الرَّجُلُ صَدِيقَهُ بِوَقْفِ مُسَاعَدَتِهِ	أَنْ الرَّجُلُ صَدِيقَهُ بِوَقْفِ مُسَاعَدَتِهِ
أَقْتَأَتِ الْمَزْرَعَةُ بِالْقِثَّاءِ فَكَثُرَ نِتَاجُهَا	فَقْسِ النَّاسِ الْمَزْرَعَةُ بِالْقِثَّاءِ فَكَثُرَ نِتَاجُهَا
الْخِتِلافُ أَرِيجِ ٱلْأَزْهَارِ أُعْجُوبَةٌ	تَخْتَلِفُ أَرِيجِ ٱلْأَزْهَارِ أُعْجُوبَةٌ
أَرْذَالُ النّاسِ ذَوُو أَفْئِدَةٍ مَيّتةٍ	الرِّجَالِ النَّاسِ ذَوُو أَفْئِدَةٍ عِزَّةَ
ٱلْفَرْعُ شِبْهُ ٱلْأَصْلِ	الْفَرَسُ شِبْهُ ٱلْأَصْلِ
إِذَا رَأَيْتُ ٱلْأَسْيَافَ تَذَكَّرتُ أَطْيَافَ ٱلْمَاضِي	إِلَّا رَأَيْتُ ٱلْأَسْيَافَ تَذَكَّرتُ أَطْيَافَ ٱلْمَاضِي

الجدول 5.4:نموذج حول النتيجة النهائية لنظام التعرف الآلي على الكلام العربي المتصل

6. الخاتمة:

لقد قمنا في عملنا هذا بالاعتماد على مجموعة أدوات نماذج ماركوف المخفية HTK لانجاز نظام للتعرف الآلي على الكلام العربي المتصل وفق مراحل من المعالجة لأجل الوصول إلى نموذج يمكننا من التعرف بشكل جيد. و قد كانت مجموعة الأدوات HTK تمتاز بالسهولة و المرونة لما تحتويه من أدوات و توفيرها لحرية الاختيار أو إدراج برامج جديدة.

لتقييم أداء النظام أجرينا العديد من التجارب و الاختبارات لتحديد كل مرة عدد الكلمات التي تقييمها بشكل صحيح، فكانت النتيجة أن لنماذج ماركوف المخفية القدرة على التعرف على الكلام العربي المتصل بشكل جيد كما تبين لنا من خلال متابعة أداء النظام أن لحجم المدونة و نوع المعاملات الصوتية تأثير كبير على عملية التعرف، فزيادة حجم المدونة إضافة إلى استعمالنا تحليل الإشارة باستخدام طريقة (MFCC) و مشتقاتها الأولى و الثانية أدى إلى زيادة كفاءة النظام و الحصول على دقة أعلى في التعرف.

و قد قمنا بتطبيق هذا النظام على مجموعة مختلفة من المتكلمين، فوجدنا أن النظام يعمل بكفاءة اكبر مع متكلم معتمد ضمن مدونة التدريب، أكثر من متكلم خارج مدونة التدريب، و قد كانت النتائج التي تم الحصول عليها مشجعة، بالنظر إلى العقبات و المشاكل الكبيرة المتعلقة باستخدام أنظمة التعرف على الكلام المتصل.

الخاتمــة

الخاتمـــــــــــــة

إن التعرف الآلي على الكلام يسمح للمستخدم بتطبيق العمليات المختلفة على الأجهزة و البرامج و التحكم فيها بسهولة بواسطة الكلام الذي يعتبر أسهل وسيلة للتواصل عند الإنسان، لهذا فقد حضي ميدان التعرف الآلي على الكلام عموما بالكثير من الاهتمام ليكون التعامل بين الإنسان و الآلة أكثر مرونة و فاعلية، حيث نلاحظ أن هناك العديد من الطرق المستخدمة في انجاز الكثير من أنظمة التعرف، و لكل طريقة منها محاسن و مساوئ، و من أهم هذه الطرق المستخدمة حاليا نجد نماذج ماركوف المخفية، حيث وجد نتيجة للأبحاث المتعددة أن هذه النماذج تعطي نتائج جيدة، و لذا فإننا في عملنا هذا اعتمدنا على نماذج ماركوف المخفية لانجاز نظام للتعرف الآلي على الكلام العربي المتصل يعتمد على الفونيمات ، وفق مراحل متعددة تمكننا من التعرف بشكل ناجح و فعال.

و للحصول على أفضل النماذج التي تساعد في تحسين أداء النظام المقترح قمنا بإجراء عدة تجارب على مجموعة من المتغيرات تتمثل في: تغير حجم المدونة و نوع المعاملات المستخدمة بالإضافة إلى عدد مرات تكرار تهيئة النماذج.

يمكننا من خلال العمل الحالي استنتاج ما يلي:

1. نماذج ماركوف المخفية قادرة على التعرف بسرعة محسوسة و أداء عاليين كما يبين من خلال متابعة أداء النظام في التعرف.

2. لحجم المدونة تأثير كبير على عملية التعرف، فزيادة حجم المدونة له تأثير فعال بزيادة دقة التعرف على الرغم من أن تأثيرها يكون واضحاً بزيادة الزمن المطلوب لإتمام عملية التمييز.

3. تأثير المعاملات الصوتية على أداء نظام التعرف المقترح حيث إن معالجة الإشارة باستخدام طريقة (MFCC) تزودنا بنموذج جيد لإشارة الرقمية و ذلك بإعطائها أفضل تمثيل ممكن للإشارة وهذا يقود إلى دقة أعلى في أداء التعرف.

4. الاختلاف بين المتكلمين يؤثر على دقة التعرف على الكلام، حيث أن النظام يعمل بشكل جيد مع متكلم معتمد ضمن مدونة التدريب، أكثر من متكلم خارج مدونة التدريب.

و قد أظهرت نتائج التعرف المختلفة التي تم التوصل إليها أن النظام كان قادراً على التعرف على الكلمات و ذالك بنسبة 75.21%, في حالة المتكلمين المعتمدين. و من ناحية أخرى في حالة المتكلمين المستقلين, فإن النظام كان قادراً على التعرف على الكلمات بنسبة 62.26%.

و تعتبر هذه النتائج التي تم الحصول عليها مشجعة و مع ذلك فإن تنفيذ نظام للتعرف الآلي على الكلام العربي المتصل في ظل ظروف واقعية يتأثر بالعديد من المشاكل مثل الضوضاء، و نوعية التسجيل، وصعوبة المفردات و حالة المتكلم، حيث يبقى مهمة صعبة لا تزال تحتاج إلى العديد من البحوث للتعامل مع جميع جوانب التخاطب بين الإنسان و الآلة.

التوصيات:

- الدمج بين نموذج ماركوف المخفي و الطرق الأخرى للتخلص من المساوئ، و تحسين الأداء و النتائج.
- انجاز مدونة (قاعدة بيانات) تحتوي مجموعة كلمات احد معاجم اللغة العربية و تسجيلها بواسطة متكلمين مختلفين للمساعدة على تحسين أداء أنظمة التعرف الآلي على الكلام و مساعدة الباحثين في القيام بدراساتهم و أبحاثهم.
- استخدام التعرف الآلي على الكلام على نطاق واسع جدا يتطلب المزيد من الجهود و الأبحاث لأنه يعتبر مرحلة أساسية لابد منها.

قائمـــــة الجداول

قائمـــــة الجداول:

الجدول 1.1: جدول التصنيف الوظيفي للأصوات العربية 31

الجدول 1.4: خصائص بعض البرامج المستعملة في انجاز أنظمة التعرف الآلي على الكلام بالاعتماد على نماذج ماركوف المخفية 78

الجدول 2.4: قائمة الفونيمات الأساسية للغة العربية 79

الجدول 3.4: الخصائص المختارة من اجل انجاز المدونة 80

الجدول 4.4: تغير نسبة التعرف بتغير المتكلم 93

الجدول 5.4:نموذج حول النتيجة النهائية لنظام التعرف الآلي على الكلام العربي المتصل 94

قائمة الأشكال

قائمة الأشكال:

الشكل 1.1 : الجهاز الصوتي و الجهاز التنفسي 9

الشكل 2.1: أعضاء النطق في الجهاز الصوتي 9

الشكل 3.1: القصبة الهوائية و الحنجرة 10

الشكل 4.1: الحنجرة من الخلف 11

الشكل 5.1: الوتران الصوتيان 12

الشكل 6.1: رسم تبسيطي للوترين الصوتيين و هما مفتوحان 12

الشكل 7.1: الحنك الأعلى 14

الشكل 8.1: أجزاء اللسان 15

الشكل 1.2: تمثيل الاشارة مع الزمن ، و تمثيل طيف الاشارة 36

الشكل 2.2: الرسم الطيفي للحركتين i و u. 37

الشكل 3.2: تمثيل الاشارة مع الزمن، يعلوها الرسم الطيفي (السوناغرام). 37

الشكل 4.2: تقسيم إشارة الكلام و استخراج متجه المعاملات منها. 44

الشكل 5.2: مخطط كتلي يوضح مراحل حساب معاملات MFCC 45

الشكل 6.2: تهيئة الإطارات بصورة المتداخلة 46

الشكل 7.2: الضرب بنوافذ هامينغ 47

الشكل 8.2: نافذة هامينغ 48

الشكل 9.2: حساب معاملات الكبستروم 49

الشكل 10.2: بنك الترشيح Mel 50

الشكل 11.2: حساب معاملاتMFCC 51

الشكل 1.3:يوضح النموذج التخطيطي لنموذج ماركوف المخفي 57

الشكل 2.3: النموذج الثبوتي 59

الشكل 3.3: نموذج الأيسر– الأيمن 59

الشكل 4.3: متتابعة العمليات المطلوبة لحساب المتغير الأمامي 63

الشكل 5.3: متتابعة العمليات المطلوبة لحساب المتغير الخلفي $\beta_t(i)$ 64
الشكل 6.3: إعادة تقدير نموذج ماركوف المخفي باستخدام $\zeta t(i,j)$ 67
الشكل 7.3: كيفية تعظيم الاحتمالية لنماذج ماركوف المخفية باستخدام خوارزمية باوم ويلش 71
الشكل 1.4: تسجيل ومعاينة الأصوات باستعمال برنامج PRAAT 80
الشكل 2.4: listeMots قائمة الكلمات التي تشكل قاعدة البيانات corpus 81
الشكل 3.4: القاموس Dictionnaire 81
الشكل 4.4: ملف التكوين config 83
الشكل 5.4: ملف التهيئة 84
الشكل 6.4: إنشاء الملف macros. 86
الشكل 7.4: بنية نظام للتعرف بواسطة HTK 89
الشكل 8.4: منحنى تغير نسبة التعرف بدلالة حجم المدونة 90
الشكل 9.4: منحنى تغير نسبة التعرف بدلالة تغير عدد تكرارت إعادة التقييم 91
الشكل 10.4: تغير نسبة التعرف بدلالة نوع المعاملات المستخدمة 92

قائمة المصادر و المراجع

قائمة المراجع و المصادر:

[1] ابن منظور، لسان العرب

[2] خولة طالب الإبراهيمي، مبادئ في اللسانيات،دار القصبة للنشر، الجزائر، 2006 م.

[3] بن سينا، أسباب حدوث الحروف ، مطبوعات مجمع اللغة العربية بدمشق.

[4] إبراهيم أنيس، الأصوات اللغوية، مكتبة الانجلو المصرية، مصر، 1999 م.

[5] منصور بن مُحَمَّد الغامدي، الصوتيات العربية ، مكتبة التوبة، السعودية، 2001 م.

[6] مُحَمَّد حسن حسن جبل، المختصر في أصوات اللغة العربية، مكتبة الآداب، مصر، 2012 م.

[7] خالد إسماعيل حسان، في اللسانيات العربية المعاصرة، مكتبة الآداب، مصر، 2008 م.

[8] سعد عبد العزيز مصلوح، دراسة السمع و الكلام، عالم الكتب، مصر، 2005 م.

[9] كمال بشر، علم الأصوات، دار غريب، مصر، 2000 م.

[10] احمد مختار عمر، دراسة الصوت اللغوي، عالم الكتب، مصر، 1997 م.

[11] مُحَمَّد علي الخولي، معجم علم الأصوات ، مطابع الفرزدق التجارية، السعودية، 1982 م.

[12] سمير شريف استيتية, الأصوات اللغوية رؤية عضوية و نطقية و فيزيائية، دار وائل للنشر، الأردن 2002 م

[13] مصطفى أكرور، مخارج وصفات الحروف العربية عند جمهور علماء التجويد، دار الخلدونية، الجزائر، 2013 م.

[14] احمد مختار عمر، أسس علم اللغة، عالم الكتب، مصر، 1998 م.

[15] احمد زرفة، أسرار الحروف، دار الحصاد للنشر، سوريا، 1993 م.

[16] علم الأصوات اللغوية الفونيتيكا، د.عصام نور الدين، دار الفكر اللبناني، لبنان، 1992م.

[17] عبد العزيز علام و عبد الله محمود، علم الصوتيات، مكتبة الرشد، السعودية، 2009م.

[18] فخري مُحَمَّد صالح، اللغة العربية أداء و نطقا و إملاء و كتابة، دار الوفاء، مصر 1994 م.

[19] رشيد عبد الرحمن العبيدي، معجم الصوتيات ، مكتبة .د مروان العطية، العراق، 2007 م.

[20] مسعود بو دوخة، محاضرات في الصوتيات، بيت الحكمة، الجزائر، 2013 م.

[21] الان بونيه، ترجمة: علي صبري فرغلي، الذكاء الاصطناعي واقعه و مستقبله، عالم المعرفة، الكويت، 1993 م.

[22] J-P HATON, Reconnaissance automatique de la parole et dialogue oral homme-machine , techniques de l'ingénieur, Référence H1940 , 10 mars 1992.

[23] ميس عبد القادر الكزبري، تعرف الكلمات المعزولة باللغة العربية، ماجستير في المعلوماتية، جامعة دمشق، سوريا، 2009 م.

[24] يسرى فيصل الارحيم، عمي انصاف جاسم، تمييز الأرقام العربية المفردة النطق باستخدام الخوارزمية الجينية، مجلة الرافدين لعلوم الحاسبات و الرياضيات المجلد (11) العدد (1) 2014 م.

[25] شهلة عبد الوهاب عبد القادر، تمييز أصوات الأرقام العربية، قسم أنظمة الحاسبات، المعهد التقني الموصل، الموصل، العراق.

[26] باسل يونس ذنون الخياط، المتتابعات الماركوفية مع التطبيق في المجال اللغوي، كلية علوم الحاسبات و الرياضيات، جامعة الموصل، المؤتمر العلمي الثاني للرياضيات-الإحصاء و المعلوماتية 6-2009/Dec./7 م.

[27] بان أحمد حسن متراس، رشا رعد هادي المولى، استخدام نماذج ماركوف المخفية في التعرف على صور الوجه المشوه و المركب من صورتين مشوهتين، المجلة العراقية للعلوم الإحصائية (20) 2006 م.

[28] حسن مُحَمَّد الياس، رنا بشار حسين، استخدام سلاسل ماركوف المخفية في تمييز حروف العلة في اللغة الانكليزية، المجلة العراقية للعلوم الإحصائية (9) 2006 م.

[29] باسل يونس ذنون الخياط شعاع محمود عزيز زينة فالح صالح العجوز، خوارزمية حاسوبية لتقدير مرتبة سلسلة ماركوف مع التطبيق، المجلة العراقية للعلوم الإحصائية (19) 2011 م.

[30] باسل يونس ذنون الخياط هنادي داؤد سليم مازن مُحَمَّد غانم العناز، خوارزمية مقترحة للتحقق من مراوحة مشاهدات تنمذج وفق النموذج الماركوفي مع التطبيق، المجلة العراقية للعلوم الإحصائية (18) 2010 م.

[31] مصون نبهان حمصي جربيني، نظام تفاعلي ذكي من اجل التعليم على الشبكة العنكبوتية، رسالة دكتراه، جامعة حلب سوريا 2010 م.

[32] J Bilmes, What HMMs Can Do, UWEE Technical Report Number UWEETR -2002-0003.

[33] باسل يونس ذنون الخياط زينة فالح صالح العجوز، تمييز السلالم الموسيقية باستخدام نموذج ماركوف المخفي، المجلة العراقية للعلوم الإحصائية (15) 2012 م.

[34] T. Aarnio, "Speech recognition with hidden Markov models in visual communication", M.Sc. Thesis, College Computer Sciences, Turku-University, (1999).

[35] بان احمد حسن، رشا رعد هادي، استخدام نماذج ماركوف المخفية في التعرف على صور الوجه الاعتيادي و المحددة حافاته، المجلة العراقية للعلوم الإحصائية (21) 2012 م.

[36] عمر صابر قاسم، تهجين أنموذج ماركوف المخفي باستخدام شبكة ايلمان العصبية الاصطناعية مع التطبيق، مجلة الرافدين لعلوم الحاسبات و الرياضيات المجلد (11) العدد (1) 2014 م.

[37] M. Stamp, " A Revealing Introduction to Hidden, Markov Models", IEEE , Vol. 51 , No. 7 , P. 347 – 356, (2003).

[38] L.R. Rabiner, “ A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition ” , Proceedings of IEEE , Vol. 77 , No. 2 , P. 257 – 286, (1989).

[39] T.Pham, “Alignment-free sequence comparison with vector quantization and hidden Markov models”, proceeding of IEEE of the Computational Systems Bioinformatics, (2003).

[40] S.Duran, “Keyword spotting using hidden markov models”, M.Sc. Thesis, Bogazici University, (2001).

[41] J.Parker, “ Profile Hidden Markov Models ” , Biophysical Journal , Vol. 17 , P. 1335 – 1348, (1999).

[42] A.K. Mohamed Cherif, Reconnaissance Automatique de la Parole par les HMM en Milieu Bruité, Doctorat, Université Annaba, Algerie, 2009.

[43] S. Young, G.Evermann, M. Gales, T.Hain, D. Kershaw, G. Moore, J.Odell, D.Ollason, D. Povey, V.Valtchev, P.Woodland, The HTK Book (for HTK Version 3.3), Cambridge University (2005).

[44]L.R.Rabiner, and B.Juang, “An introduction to hidden Markov models”, IEEE ASSP MAGAZINE, (1986).

[45]M.Hayashi, “Hidden Markov models to identify pilot instrument scanning and attention patterns”, proceeding of IEEE International Conference on Systems, Man and Cybernetics, 2889-2896, Washington D.C.,Oct.5-8, (2003).

[46]W.H.Abdulla, and N.K.Kasabov, " The Concept of Hidden Markov Model in Speech Recognition", Dept. of Knowledge Engineering Lab. Dept.

Information Science, College of Engineering ,University of Otage, New Zealand, (1999).

[47] DOUIB Oualid, reconnaissance automatique de la parole arabe par cmu sphinx 4, magister, université FERHAT Abbas, Algerie, 2013.

[48] Asmaa Amehraye, Débruitage perceptuel de la parole, l'Ecole Nationale Supérieure des Télécommunications de Bretagne, Docteur, 2009.

[49] خليل إبراهيم السيف، سجى جاسم مُحَمَّد، كبس ملفات الكلام باعتماد خواص الإشارة، مجلة الرافدين لعلوم الحاسبات و الرياضيات المجلد (4) العدد (1) 2007 م.

[50] بركات ابراهيم، مصطفى الزين، طلال بهاء الدين، تمييز الكلام العربي باستخدام الشبكات العصوبية الاصطناعية، بكالوريوس الشرف في علوم الحاسوب جامعة السودان للعلوم و التكنولوجيا، 2006 م.

[51] غيداء عبد العزيز الطالب، ارمانيسة نعمان حسون، التعرف على النص العربي المطبوع باستخدام نموذج ماركوف الخفي، مجلة الرافدين لعلوم الحاسبات والرياضيات المجلد (7) العدد(2) 2010م.

[52] فاتن بشير عبد الأحد ،إنعام غانم سعيد، تقطيع الكلمة العربية إلى أحرف و تمييزها، مجلة الرافدين لعلوم الحاسبات و الرياضيات المجلد (2) العدد (2) 2005 م.

[53] رماح معتصم ، رواء عوض ،سوسن عبد اللطيف أحمد، استخدام نماذج ماركوف الخفية في التعرف على حكم الإقلاب،، مجلة اتصالات جمعية الحاسبات العربية المجلد (2) العدد (2) 2005 م.

الفهـــــــرس

الفهـــــــرس

المقدمة 5

الفصـــــــــل الأول: أصوات اللغة العربية و خصائصها

1. مقدمة: 8

2. الصوت: 8

3. الصوت الإنساني و جهاز النطق 8

1.3. الرئتان 10

2.3. القصبة الهوائية 10

3.3. الحنجرة 11

4.3. الحلق 13

5.3 تجويف الفم 14

6.3. التجويف الأنفي (الخيشوم) 16

4. الصوت اللغوي 16

5. تقسيم الصوامت حسب مخارجها 18

6. صفات أصوات اللغة العربية وتقسيم الصوامت حسبها 22

7.الحركات 27

8. الخاتمة 32

الفصـــــل الثاني التعـــرف الآلي عـــلى الكلام

1. مقدمة 33

2. بعض برامج التعرف على الكلام 33

3. الموجات الصوتية الناتجة عن الجهاز الصوتي 34

4. طرق تمثيل الكلام 35

5. تطبيقات التعرف على الكلام 38

6. صعوبات تعرّف الكلام 39
7. أنواع أنظمة التعرف على الكلام 40
8. مراحل عملية التعرف على الكلام 41
9. تسجيل الأصوات 43
10. استخلاص المعاملات 44
1.10. تحليل الإشارة الصوتية بـ (MFCC) 45
2.10. تقسيم الإشارة إلى نوافذ زمنية 45
3.10. المعالجة بالنافذة 46
4.10. تحويل فوريه السريع FFT (Fast Fourier Transform) 48
5.10. التحليل باستخدام الكبستروم 49
6.10. مقياس Mel 49
7.10.معاملات MFCC 50
8.10. المشتقات الاولى و الثانية لمعاملات الكبستروم 51
11. الخاتمة 52
الفصـــــل الثالث نماذج ماركـــوف المخفية
1. مقدمة 54
2. تعاريف أساسية 55
3. نماذج ماركوف المخفية 57
1.3. عناصر نماذج ماركوف المخفية 58
2.3. انواع نماذج ماركوف المخفية 59
3.3. المسائل الأساسية لنماذج ماركوف المخفية 59
4.3. حلول المسائل الأساسية لنماذج ماركوف المخفية 60
5.3. تقييس نماذج ماركوف المخفية 71
4. الخاتمة 76

الفصـــــل الرابع تجـــارب و نتائـــج

1. مقدمة 78

2. انجاز المدونة 79

3. مرحلة تسجيل الأصوات 80

4. مراحل انجاز النظام 81

1.4. تحضير الملفات الضرورية التي نحتاجها في إنجاز النظام 81

2.4. تحليل إشارة الكلام و بناء قاعدة البيانات 82

3.4. التدريب 83

4.4. مرحلة تهيئة نماذج HMM بواسطة الأداة HCompV 85

5.4. إعادة تقييم نماذج HMM 86

6.4. إدراج نموذج التوقف "sp"و إعادة تقييم نماذج HMM 87

5. الاختبارات و التجارب 89

1.5. تغير نسبة التعرف بدلالة تغير حجم المدونة 90

2.5. تغير نسبة التعرف بدلالة تكرار التهيئة بـ HERest 91

3.5. تغير نسبة التعرف حسب نوع المعاملات المستخدمة 92

4.5. تغير نسبة التعرف بدلالة تغير المتكلم 93

6. الخاتمة 94

الخاتمة العامة

قائمة المصادر و المراجع

الفهرس