جامعة آل البيت

كلية الأمير الحسين بن عبدالله لتكنولوجيا المعلومات

قسم علم الحاسوب

رسالة ماجستير في علم الحاسوب

بعنوان

خوارزمية جديدة لتجزئة ملف الموجة المعتمد على صوت المتكلم للتعرف على تطابق الأصوات

A New Algorithm for Speaker Wave File Segmentation for Voice Sequence Matching Recognition

إعداد الطالب

فيصل كامل علي عوده

٠٠٢٠٩٠١٠٠٥

المشرف **د. إسماعيل عبابنه**

المشرف المشارك **د. بسام حمــّو**

٢٠٠٦ م

خوارزمية جديدة لتجزئة ملف الموجة المعتمد على صوت المتكلم للتعرف على تطابق الأصوات

A New Algorithm for Speaker Wave File Segmentation for Voice Sequence Matching Recognition

إعداد الطالب

فيصل كامل علي عوده

٠٠٢٠٩٠١٠٠٥

المشرف	المشرف المشارك
د. إسماعيل عبابنه	د. بسام حمّو

أعضاء لجنة المناقشة	التوقيع
_____	_____
د . اسماعيل عبابنه	_____
د . بسام حمّــو	_____
د . مأمون ربابـعة	_____
د . أكرم حمارشـة	_____
د . موسى حبـيب	_____

قدمت هذه الرسالة استكمالا لمتطلبات الحصول على درجة الماجستير في علم الحاسوب في كلية تكنولوجيا المعلومات في جامعة آل البيت .

نوقشت وأوصي بإجازتها بتاريخ : ٢٣ / ٣ / ٢٠٠٦ م

الإهـــــداء

أهدي هذا العمل المتواضع إلى الأحبة والداي وزوجتي وأبنائي ، وإلى كل من أعانني على إكمال هذا العمل وإخراجه إلى حيز الوجود .

الشـــكر

بادئ ذي بدء ألهج بالشكر والثناء لله سبحانه وتعالى على أن أعانني ويسر لي سبل التعلم والاستزادة من العلم ، حتى أتقدم بهذا العمل المتواضع .

ثم أتشرف بتقديم جزيل الشكر والامتنان للدكتور الفاضل إسماعيل عبابنه على كل ما قدمه لي طيلة فترة دراستي الجامعية .

و أتشرف بتقديم جزيل الامتنان وحسن الدعاء للدكتور الفاضل بسام حمّو على كل ما قدمه لي من تعاون كامل لاخراج هذا العمل .

و أقف وقفة إجلال وتقدير أمام أعضاء لجنة المناقشة على جهودها ووقتها الثمين الذي منحتني إياه .

و أخيرا أقدم جزيل الشكر وموفور المحبة لكل شخص مد يد العون لي على إخراج هذا العمل ليرى النور ، وتعم به الفائدة إن شاء الله تعالى .

وفق الله الجميع لما يحبه و يرضاه ، إنه سميع مجيب .

قائمة المحتويات

المـــوضـــوع	الصفحة

الإهـــداء ..... ج

الشكر ..... د

قائمة المحتويات ..... هـ

قائمة الأشكال ..... ح

قائمة الجداول ..... ي

قائمة المصطلحات ..... ك

ملخص الدراسة ..... ل

الفصل الأول

الوسائط المتعددة (Multimedia) ١

١-١ مقدمة ١

١-٢ الهدف من استخدام الوسائط المتعددة ٢

١-٣ ما يجب توفره في جهاز الحاسوب لتشغيل الوسائط المتعددة ٣

١-٤ الأصوات جزء من الوسائط المتعددة ٣

الفصل الثاني

الصوت والتسجيل الصوتي (Audio and Voice) ٤

٢-١ استخدام الصوت في الوسائط المتعددة ٤

٢-٢ أهداف استخدام الأصوات والتسجيلات الصوتية ٤

٢-٣ عيوب استخدام الصوت في الحواسيب ٥

٢-٤ الأصوات ٥

٢-٥ ملفات التسجيل الصوتي ٦

٢-٦ مكتبة ملفات التسجيل الصوتي ٦

٢-٧ تنسيق ملف التسجيل الصوتي ٧

٢-٨ تنسيق بيانات ملف التسجيل الصوتي ٨

٢-٩ العوامل المؤثرة في حجم ملف التسجيل الصوتي ٨

٢-١٠ أنواع ملفات التسجيل الصوتي ١٠

٢-١١ الملفات ذات التنسيق (RIFF) أو تنسيق الملفات المتغيرة المصدر (Recourse Interchange File Format) ١٢

٢-١٢ نظرة عامة على ملف الموجة (Wave File) ١٤

٢-١٣ قطع ملف الموجة (Wave File Chunks) ١٦

المـــوضـــوع	الصفحة

الفصل الثالث

تجزئة الملفات (File Segmentation)..... ٢٠

٣-١ الدراسات السابقة في موضوع تجزئة الملفات ٢٠

٣-٢ عملية تجزئة الملفات (File Segmentation Process)..... ٢١

٣-٣ الأهداف من عملية التجزئة..... ٢٢

الفصل الرابع

التعرف الآلي إلى الأصوات (Automatic Speech Recognition) ٢٣

٤-١ الدراسات السابقة في موضوع التعرف إلى الأصوات..... ٢٣

٤-٢ التعرف إلى الأصوات ٢٣

٤-٣ بعض مجالات استخدام أنظمة التعرف إلى الأصوات ٢٤

٤-٤ محددات تعتمد عليها أنظمة التعرف إلى الأصوات ٢٤

٤-٥ أنواع أنظمة التعرف إلى الأصوات..... ٢٥

٤-٦ طرق التعرف إلى الأصوات..... ٢٦

٤-٧ كيف تعمل نظم التعرف إلى الأصوات..... ٢٧

الفصل الخامس

الخوارزمية الجديدة لتجزئة ملف الموجة والتعرف إلى تطابق الصوت

(New Algorithm for Wave File Segmentation and Voice Sequence Matching Recognition)..... ٢٩

٥-١ لغة البيسك المرئية (Visual Basic) و برنامج الدراسة..... ٢٩

٥-٢ الخوارزميات المستخدمة في البرنامج ٣٠

٥-٣ بيئة فحص البرنامج (ملاحظات الفحص)..... ٣٠

٥-٤ نتائج مراحل التنفيذ (المتوسطات) للبرنامج ٣٦

٥-٥ الواجهات الرسومية لبرنامج تجزئة الملفات..... ٣٩

٥-٦ الواجهات الرسومية لبرنامج التعرف إلى تطابق الأصوات..... ٤٩

المـــوضـــوع	الصفحة

الفصل السادس

النتائج والتوصيات ٥٦

٦-١ **نتائج عملية تجزئة ملفات الصوت** ٥٦

٦-٢ **نتائج عملية التعرف إلى تطابق الأصوات** ٥٦

٦-٣ **خوارزمية طبقت باستخدام التطبيق MATLAB** ٥٧

٦-٤ **تطبيق الخوارزميتين على مجموعة من ملفات الصوت** ٥٨

٦-٥ **نتائج المقارنة بين الخوارزمية الجديدة وخوارزمية طبقت باستخدام التطبيقMATLAB.** ٦٣

٦-٦ **النتيجة النهائية** ٦٥

الخلاصة ٦٦

الأعمال المستقبلية ٦٧

قائمة المراجع والمصادر ٦٨

قائمة مراجع الإنترنت ٦٩

الملخص باللغة الإنجليزية ٧٠

قائمة الأشكال

رقم الشكل	الوصـــــف	الصفحة
(٢-١)	الأعداد المختلفة للقنوات في ملف التسجيل الصوتي	١٠
(٢-٢)	أسماء امتدادات الملفات المدعمة من قبل مكتبة التسجيل الصوتي	١١
(٥-١)	خوارزمية قراءة ملف قيد التجزئة	٣١
(٥-٢)	خوارزمية التطبيع	٣٢
(٥-٣)	خوارزمية تجزئة الملفات	٣٤
(٥-٤)	خوارزمية التعرف إلى الأصوات	٣٧
(٥-٥)	واجهة البدء لبرنامج التجزئة	٣٩
(٥-٦)	واجهة البرنامج الرئيسة	٤٠
(٥-٧)	قائمة ملف ومحتوياتها	٤٠
(٥-٨)	درجة قيمة الدقة المرغوب بها	٤١
(٥-٩)	قائمة التحكم	٤١
(٥-١٠)	قائمة العمليات	٤٢
(٥-١١)	القائمة الجانبية تحكمات العملية	٤٣
(٥-١٢)	عملية اختيار محلل الصوت	٤٣
(٥-١٣)	تحكمات محلل الصوت	٤٤
(٥-١٤)	صندوق الحوار لفتح الملف المراد تجزئته	٤٤
(٥-١٥)	موجات الصوت لملف وقد حولت لرسم بياني	٤٥
(٥-١٦)	الواجهة الرسومية لتجزئة الملف	٤٦
(٥-١٧)	الرسم البياني للصوت في ملف	٤٦
(٥-١٨)	صندوق الحوار لاختيار عدد العينات غير ذات المعنى المرغوب باختيارها بين العينات ذات المعنى عند عملية التجزئة	٤٧
(٥-١٩)	ملف الصوت بعد تجزئته	٤٧
(٥-٢٠)	التقرير الذي يظهر في نهاية عملية التجزئة	٤٨
(٥-٢١)	الواجهة لبدء عملية التعرف على تطابق الأصوات	٤٩
(٥-٢٢)	الواجهة الرئيسة لبرنامج التعرف على تطابق الأصوات	٥٠
(٥-٢٣)	صندوق الحوار الخاص باختيار ملف و فتحه	٥١
(٥-٢٤)	عملية فتح الملف	٥٢
(٥-٢٥)	عملية اختيار مشغل الأقراص الذي يحوي النماذج	٥٣

رقم الشكل	الوصـــــف	الصفحة

(٥-٢٦) الواجهة التي تظهر نتائج المقارنات تباعا للملفات ٥٣

(٥-٢٧) الواجهة التي تظهر نتائج المقارنات النهائية ٥٤

(٥-٢٨) الواجهة التي تظهر التقرير في معرفة الصوت ٥٤

(٥-٢٩) الواجهة لتأكيد الخروج من البرنامج ٥٥

(٦-١) الأشكال المختلفة للملفات المخرجة من عملية التجزئة ٥٦

(٦-٢) الواجهة التي تظهر نتائج المقارنات للملفات..... ٥٦

(٦-٣) نقاط البداية والنهاية لموجات الكلمات في ملف معين ٥٩

(٦-٤) شكل الموجات لمجموعة من الكلمات..... ٥٩

(٦-٥) رسم بياني يوضح العلاقة بين أعداد الكلمات الكلية في كل ملف وأعداد الكلمات الكاملةالناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين..... ٦٢

(٦-٦) رسم بياني يوضح العلاقة بين أعداد الكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين ٦٣

(٦-٧) رسم بياني يوضح العلاقة بين أعداد الكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين ٦٤

(٦-٨) رسم بياني يوضح العلاقة بين النسبة المئوية للكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين ٦٥

قائمة الجداول

رقم الجدول	الوصـــــف	الصفحة
(٢-١)	تركيب ملف (RIFF)	١٢
(٢-٢)	تركيب ملف موجة حسب التنسيق (RIFF)	١٣
(٢-٣)	قيم القطع لملف من نوع (RIFF)	١٤
(٢-٤)	تنسيق القطع لملف من نوع (RIFF)	١٥
(٢-٥)	تنسيق قطعة "fmt" لملف من نوع موجة	١٦
(٢-٦)	بعض أنواع شيفرات الضغط الشائعة	١٧
(٢-٧)	تنسيق قطعة "data" لملف من نوع موجة	١٨
(٦-١)	أعداد الكلمات الكاملة والمشوهة في الملفات قبل التجزئة وبعدها من خلال تطبيق الخوارزميتين الجديدة و MATLAB	٦٠
(٦-٢)	أعداد الكلمات الكاملة والنسب المئوية لعدد الكلمات الكاملة الناتجة بعد تطبيق الخوارزميتين على عدد من الملفات لتجزئتها	٦٤

قائمة المصطلحات

المصطلح باللغة العربية	المصطلح باللغة الإنجليزية
أحادي القناة	Mono
تأطير	Framing
تجزئة	Segmentation
تحليل الخصائص	Feature Analysis
ترويسة	Header
تسجيل الأصوات	Audio Recording
التعديل	Offset
تعديل شيفرة النبضة الخطية	Pulse Code Modulation
التعرف الآلي على الأصوات	Automatic Speech Recognition
ثنائي القناة	Stereo
ثنائية	Bit
جمعية الوسائط المتعددة التفاعلية	Interactive Multimedia Association
حجم النافذة	Window_size
خطوة القطع	Cut_ Step
ضغط مخسر	Lossy Compression
الطاقة	Energy
العتبة	Threshold
عملية التطبيع	Normalization
العينة	Sample
قطع	Chunks
قيمة الدقة	Accuracy_Value
لغة البيسك المرئية	Visual Basic
المحاكاة	Simulation
مدى التردد	Bandwidth
مدى قطع الموجة	Cut_Wave_Range
المرحلة الوسط للبيانات	Metadata
مطابقة النماذج	Template Matching
معدل القطع الصفري	Zero Crossing Rate
مكبر الصوت	Microphone
ملف الموجة	Wave File
ملفات التسجيل الصوتي	Audio Files
نظام تشفير التنبؤ الخطي	Linear Predictive Coding
نماذج	Forms
نموذج ماركوف المخفي	Hidden Markov Model
واجهات البرامج التطبيقية	Application Program Interface
وحدات	Modules
وسائط متعددة	Multimedia
وسائل الإعلام	Hypermedia

ملخص الدراسة

ظهرت الوسائط المتعددة لتعلب دورا هاماً و ريادياً في توصيل المعلومات والبيانات لمستخدمي أجهزة الحاسوب على اختلاف أنماطهم وأعمالهم وتطلعاتهم ، وذلك من خلال العديد من الطرق الجذابة والمثيرة ، وباستخدام الكثير من الوسائل والأجهزة التي يسهل استخدامها والتعامل معها وتوفرها على أبسط المستويات .

وقد سعت دراسات كثيرة لتطوير الوسائط المتعددة من كلا الجانبين ، جانب المعدات وجانب البرمجيات التي تطبق بوساطتها ، وذلك لمجابهة الاحتياجات والرغبات المختلفة لدى المستخدمين ، فقد ظهرت أجهزة ومعدات لم تكن لتظهر لولا الحاجة إليها في حقل ما أو لخدمة اتجاه علمي معين . وما لبثت أن ظهرت إثر ذلك برمجيات تمكن المستخدم من الحصول على الاستفادة المثلى من تلك الأجهزة والمعدات .

ومن بين ذلك ظهرت إمكانية تسجيل الأصوات و الصور المتحركة على أجهزة الحاسوب ، واسترجاع ما تم تخزينه على هذه الأجهزة ، دون الحاجة إلى الرجوع إلى ما تم اختراعه في السابق من أجهزة مثل جهاز تسجيل الأصوات و جهاز عرض الأفلام (الفيديو) ، فاستغنى من ملك جهاز حاسوب مزود بالوسائط المتعددة عن تلك الأجهزة وغيرها مما لم يذكر.

وفي خضم هذه الثورة الهائلة لتكنولوجيا الوسائط المتعددة ظهرت الحاجة إلى العمل على النظر في ما تم تسجيله وتخزينه باستخدام هذه الوسائط المتعددة ، وكيفية الاستفادة منه بشكل أو بآخر ، أو استرجاعه بثوب مختلف بحيث يمكن استخدامه استخدامات متنوعة في غير الأغراض التي صنع من أجلها .

وقد قام عدد من الباحثين في مجال الوسائط المتعددة بأبحاث كثيرة ومتنوعة حول موضوع تجزئة الملفات المنتجة بوساطة تلك الوسائط ، بحيث يتم الحصول على ملفات مجزأة يحتوي كل منها على بيانات متنوعة يستفاد منها استفادات مختلفة.

وكانت هذه الدراسة تقدم خوارزمية جديدة لتجزئة ملفات التسجيل الصوتي من نوع ملفات الموجة من أجل التعرف إلى صوت المتحدث في الملف الجزئي الناتج عن عملية التجزئة، حيث استطاعت الدراسة أن تحول جزء من الكم الهائل من الثنائيات المبهمة في ملف موجة معين ، إلى ثنائيات ذات معنى يمكن سماع ما تحويه من بيانات ضمن ملف موجة جزئي يمكن تشغيله على مشغلات متوفرة بشكل واسع ، ثم مقارنة الملف الجزئي الناتج مع عدد من الملفات المخزنة لتحديد صوت المتكلم في هذا الملف من بين تلك الملفات ومعرفة صوته .

وقد احتوت الدراسة على عدة عمليات صغيرة احتيج إليها لاتمام الخوارزمية الجديدة ، مثل عملية التطبيع (Normalization) ، وعملية التصغير (Minimization) .

وقد تم تقسيم هذه الدراسة إلى ستة فصول ، يعرض الفصل الأول مقدمة عامة في مفهوم الوسائط المتعددة و معداتها و برمجياتها ، ثم كان الفصل الثاني الذي تطرق إلى مفهوم الصوت و عملية تسجيل الأصوات في ملفات من نوع الموجة ، وطبيعة تركيب تلك الملفات وكيفية التعامل معها . وبعد ذلك جاء الفصل الثالث ليبين عملية تجزئة الملفات ، وما تم من دراسات سابقة في هذا المجال وتم عرض بعض الفوائد التي ترتجى من عمليات التجزئة تلك .

أما الفصل الرابع فاحتوى على مفهوم عملية التعرف على الأصوات و أهميتها ، وأهم الدراسات التي قام بها الباحثون والدارسون . وبعد ذلك أتى الفصل الخامس وهو الأهم في هذه الدراسة و الذي أوضح أهداف الدراسة و آلية تحقيق تلك الأهداف وعرض الخوارزمية ، ثم كان الفصل السادس والأخير لعرض نتائج هذه الدراسة عن طريق مقارنة نتيجة تطبيق الخوارزمية الجديدة مع خوارزمية طبقت باستخدام التطبيق MATLAB ثم الخلاصة والعمل المستقبلي .

ن

الفصل الأول

الوسائط المتعددة (Multimedia)

١-١ مقدمة

تتنوع البيانات والمعلومات التي يتم تداولها في جهاز الحاسوب تنوعاً كبيراً حيث نجد البيانات النصية التي يمكن معالجتها خلال تطبيقات كثيرة مختلفة مثل معالجات النصوص أو الجداول الإلكترونية وغيرها الكثير . كما نجد البيانات الصوتية المختلفة التي تسجل على الحاسوب أو التي تسترجع منه خلال تطبيقات متعددة مثل برامج مشغلات الأصوات .

ثم ظهرت بيانات الصور الثابتة و الصور المتحركة (الأفلام) ، التي يمكن التعامل معها من خلال الماسحات الضوئية أو الكاميرات الرقمية أو غيرها من الأجهزة الحديثة .

وتعرّف الوسائط المتعددة بأنها الحقل المهتم بتكامل النصوص والرسوم والأشكال والصور الثابتة و المتحركة ، والتسجيل الصوتيّ ، وأية وسائط أخرى ، من خلال جهاز الحاسوب ، حيثما أمكن تمثيل المعلومات وتخزينها ونقلها ومعالجتها رقميّاً .

كما يمكن تعريف الوسائط المتعددة بأنها لفظ عام للأجهزة والبرمجيات والمعلومات والبيانات التي تتمثل بأكثر من تنسيق . وكمثال على ذلك فإن تطبيق العروض التقديمية يُعد شكلاً من أشكال الوسائط المتعددة ، لأنه يستطيع الجمع بين الصوت والنص والرسوم البيانية والصور بالإضافة إلى الحركة .

وقد كانت عملية مشاهدة فيلم متحرك أو سماع قطعة موسيقية على الحواسيب في بدايات تطورها ضربا من ضروب المستحيل . ولكن بعد أن أصبحت البيانات تمثل بشكل رقمي بعد أن كانت تمثل بشكل تناظري ، صار من البديهي القيام بمشاهدة ذاك الفيلم أو سماع تلك القطعة الموسيقية من خلال برمجيات الوسائط المتعددة .

وتتميز أنظمة الوسائط المتعددة بمجموعة من المميزات البسيطة منها أن هذه الوسائط يجب أن تدار بوساطة الحواسيب ، وأنها متكاملة أي تكمل بعضها بعضاً ، وأن البيانات المتداولة من خلال هذه الوسائط يجب أن تمثل بشكل رقمي ، وأن التمثيل النهائي للوسائط المتعددة عادة ما يكون بشكل تفاعلي .

و يُسمىّ التطبيق الذي يستخدم مجموعة من الوسائط المتعددة مثل النصوص و الرسوم البيانية و الصور والأصوات و الحركة و الأفلام ، بتطبيق الوسائط المتعددة ، ويمكن اعتبار وسائل الإعلام (Hypermedia) نوعاً من أنواع تطبيقات الوسائط المتعددة .

وتتنوع عناصر بيانات الوسائط المتعددة كثيراً وتختلف حسب استخداماتها ،ومن هذه الأنواع : النصوص ، والفاكس ، وصور الوثائق ، والصور الفوتوغرافية ، والخرائط الجغرافية ، والأوامر الصوتية ، ورسائل التسجيل الصوتي ، والموسيقى ، والرسوم المتحركة، وأشرطة الفيديو.

وقد استخدمت تقنيات ضغط البيانات بسبب الكم الهائل من البيانات للتسجيل الصوتي أو الأفلام ، إذ تقاس كمية هذه البيانات بالميجابايت (أي مليون بايت) ، لذا فإنّ عمليات التخزين والنقل والمعالجة تسبب كثيراً من ضياع الوقت والجهد ، فكان استخدام تقنيات الضغط للبيانات هاما في هذه الحالات .

١-٢ الهدف من استخدام الوسائط المتعددة

إنّ التأثيرات الخاصة في حواس الإنسان لها سحر كبير في مشاعره واتجاهاته وتفكيره ، وهذا ما تقوم به الوسائط المتعددة سواء في مجال التعليم ، أو مجال الترفيه والتسلية كما في الألعاب أو في مجال إنتاج الأفلام والإبداعات الأخرى لبني الإنسان . ولذلك فهناك أهداف عديدة لاستخدام الوسائط المتعددة منها ما يلي :

١. التعمق في المسائل وحلول المشاكل التي تواجه فئات المستخدمين ، و إعطاء تفاصيل أكبر وأوسع حول الموضوعات المختلفة .
٢. الوصول المباشر إلى المحتوى المعيّنِ ، حيث تتيح الوسائط المتعددة المجال أمام المستخدمين للوصول إلى أدقّ تفاصيل ما يبحثون عنه بطريقة أسرع وأسهل .
٣. التمثيل المتعدّد للمعلوماتِ نفسها ،فنجد المعلومة الواحدة تتمثل بأكثر من صورة حتى يتم ترسيخها لدى المستخدم بأكثر من حاسة فيسهل حفظها أو التعامل معها بيسر وسرعة .
٤. التزويد بتقديم عرض شكليّ متعدد للمعلومات ، فيمكن أن تعرض المعلومة بشكل صوتيّ أو آخر صوريّ أو آخر حركيّ .

كما يفيد استخدام الوسائط المتعددة المستخدمين خلال الاعتبارات التالية التي تنعكس عليهم وعلى سلوكياتهم وهي :

١. الأمثلة : التي تفيد في التعليم عن طريق التقليد والمحاكاة ، فيقوم الطلبة مثلاً بتقليد صوت تهجئة لكلمة معينة بعد سماعها صوتيا .
٢. الإرشاد : يُمكن أَن يُزوّد بتفاعليةَ كاملةَ ، لكن التفاعل يجب أَن يُبدأَ من قبل المستخدم. وهو يفيد لأَتمتة المهام ويخفي التعقيدات لدى المستخدمين الجدد.
٣. التدريب : إن فائدة التدريبات هي التزويد بتفاعليةَ كاملة ، والمساعدة عند الحاجة . حيث أنها تحسّن الكفاءة وتُخفّض نسب الأخطاء .

١-٣ ما يجب توفره في جهاز الحاسوب لتشغيل الوسائط المتعددة

لتشغيل البرامج الخاصة بالوسائط المتعددة على جهاز حاسوب معين ، يجب توفير وحدة معالجة مركزية ذات سرعة عالية نسبيّاً . ووحدة تخزين ثانوية ذات سعة كبيرة ، لتحمل ملايين البيانات الناتجة عن استخدام هذه الوسائط .

بالإضافة إلى ذلك يجب أن تكون وحدة الذاكرة الرئيسة ذات سعة عالية ، مما يعود بالسرعة الملحوظة عند التفاعل مع الوسائط المتعددة . كما أنّ توافر مشغل اسطوانات مدمجة ، وبطاقة تشغيل الفيديو ، وبطاقة الصوت ، بالإضافة إلى الفأرة ، وسماعات صوت مناسبة مع وجود مكبر صوت (Microphone) ، هو من الأمور المهمّة للتعامل مع الوسائط المتعددة .

١-٤ الأصوات جزء من الوسائط المتعددة

و كما أشير سابقاً فإنّ الوسائط المتعددة تتألف من عناصر كثيرة منها ما يعنى بالصورة أو الصوت أو الحركة أو الصور المتحركة وغيرها . وفي مجال هذه الدراسة سوف نركز على الجانب الذي يخدم الصوت وتطبيقاته كواحد من عناصر الوسائط المتعددة ، وسنخصص للدراسة والبحث في مجال الصوت نوعاً من أنواع ملفات التسجيل الصوتيّ وهو ملف الموجة (Wave File) .

الفصل الثاني

التسجيل الصوتي و الصوت (Audio and Voice)

٢-١ استخدام الصوت في الوسائط المتعددة

عند التعامل مع الوسائط المتعددة ، فإن جانب الصوت عادة ما يكون مهملاً وذلك لأنّ الجانب المرئيّ من تلك الوسائط قد غلب على الجانب الصوتيّ مما أدى إلى انحسار مميزات الصوت . أما في الوقت الحاضر فإنّ أغلب الحواسيب الشخصية تحوي بطاقات للصوت ومكبرات الصوت بالإضافة إلى السماعات .

وقد حسّن جانب الصوت كثيراً من تطبيقات الوسائط المتعددة في عدة نواح منها على سبيل المثال : إمكانية سماع محاضرات أو سماع قطع موسيقية عن طريق الإنترنت ، أو التفاعل مع برمجيات العروض التقديمية .

٢-٢ أهداف استخدام الأصوات والتسجيلات الصوتية

يكمن العديد من الأهداف وراء استخدام الأصوات و التسجيلات الصوتية في الحواسيب، و من أهم تلك الأهداف :

١. التفاعل بين المستخدمين الذين لا يمكن تواصلهم بالصورة .
٢. يحمل الصوت عادة المعاني ويزود بقناة إضافية للمعلومات ، فإن كان المعنى غير واضح عند استخدام الوسائط المرئية ، فإن الصوت قد يوضح ذلك المعنى.
٣. يزود الصوت بمعلومات إضافية لتدعيم أنماط مختلفة من التعليم لدى المتعلمين المختلفين .
٤. يضيف الصوت نوعاً من الواقعية على المعلومات ، فتحمل موسيقى شعب من الشعوب مثلا العاطفة إلى نفس من ينتقل إلى ثقافة ذلك الشعب عبر الأزمنة أو المواقع الجغرافية .
٥. للصوت أهمية بالغة في توجيه الاهتمام بالأحداث الهامة ، وذلك عن طريق الأصوات غير الناتجة عن الأحاديث أو الكلام ، مثل صوت الرعد أو صوت الضحك أو صوت البكاء وغيرها . ويجب هنا عدم الإفراط في استخدام هذه الأصوات لعدم صرف الانتباه عن أجهزة الوسائط المتعددة الأخرى .
٦. سهولة الاتصال حيث يكون التفاعل مع الكلمة المسموعة أكبر من التفاعل مع باقي الوسائط المتعددة .

٢-٣ عيوب استخدام الصوت في الحواسيب

يظهر عدد من العيوب لاستخدام الصوت والتسجيل الصوتيّ مما يؤثر في الإقبال على التعامل مع الصوت والتسجيل الصوتيّ ، ومن هذه العيوب :

١. تكون ملفات الصوت كبيرة الحجم تؤثر في كيفية تخزينها و التعامل معها ، مثلها في ذلك مثل ملفات الوسائط المتعددة الأخرى .
٢. يمكن أن يعتبر المستخدم الصوت غير قابل للتذكر على عكس الأمور المرئية ، مثل مشاهدة قطعة من شريط فيديو مثلا .
٣. صعوبة الحصول على نوعية صوت عالية الجودة ، لوجود الأصوات التجارية وبالأخص الموسيقى التي يمكن الحصول على نسخ غير قانونية منها بشكل كبير.
٤. يجب وجود معدات معينة خاصة للتعامل مع الصوت وملفاته وبرامجه، مثل السماعات ومكبرات الصوت والتطبيقات الخاصة.

٢-٤ الأصوات

ينشأ الصوت عادة عند الاهتزاز السريع للأشياء محدثا تغييرات متتابعة في ضغط الهواء بمقدار يتراوح بين (٢٠) مرة في الثانية و (٢٠٠٠٠) مرة في الثانية ، وعند حدوث هذه التغييرات في ضغط الهواء ، فإنّ ذلك يؤدي إلى اهتزاز الأجزاء المستقبلة في أذن الإنسان أو غيره ممّا يفضي إلى سماع هذه الأصوات ، ويكون الصوت في هذه الحالة على شكله التناظريّ.

وعند تسجيل الأصوات بشكلها التناظريّ كما هو معتاد ، فإنّ التغييرات في ضغط الهواء تحوّل إلى إشارات كهربائية دائمة التغيير يمكن مشاهدتها على شكل موجات ، وتحدد أشكال هذه الموجات عناصر منها سعة الموجة والزمن بالإضافة إلى التردّد .

ويتم تخزين الصوت عادة بتحويل الإشارات الكهربائية إلى تغييرات دقيقة في قطبية المغناطيسات لقطع معدنية مجهرية موجودة على شريط التسجيل الصوتي . وعند إعادة سماع الأصوات من هذا الشريط فإنّ تلك التغييرات المغناطيسية تحوّل مرة أخرى إلى إشارات كهربائية تكبر عبر سماعات ، وهذا يحوّل الإشارات الكهربائية إلى تغييرات فيزيائية في ضغط الهواء لتتمكن الأذن من سماع هذه الأصوات مجددّاً .

وعلى مدى السنوات الماضية قامت التكنولوجيا بتطوير طرق تسجيل الأصوات والموسيقى و كذلك المعدات التي تستخدم لذلك . وتستخدم الحواسيب لتسجيل الأصوات وتخزينها على أقراص مدمجة أو على مسجلات MP3 ، وعلى وسائط أخرى كثيرة التنوع .

ولتحويل هذه الأصوات من صورتها التناظرية (نبضات الصوت) إلى الصورة الرقمية لإتمام عملية تسجيلها على الوسائط التخزينية ، يجب تجزئة هذه الأصوات إلى ما يسمى

(عينات) وتجري هذه العملية أثناء عملية تسجيل الصوت ، ويقوم كرت الصوت الموجود في جهاز الحاسوب بتحويل الإشارات الكهربائية إلى أرقام تمثّل تلك العينات . وتخزن هذه العينات في متتاليات مكونة من مجموعة من الأصفار و الواحدات حسب الصوت المسجل . ويقوم الحاسوب بأخذ لقطات من هذه العينات في فترات زمنية صغيرة ، ويعرّف معدل أخذ العينات بعدد العينات التي تؤخذ في كل ثانية . أو يمكن تعريف السرعة التي يحوَّل فيها الصوت إلى عينات بمعدل أخذ العينات ، ويقاس بالكيلوهيرتز .

ومن الأهمية بمكان أن تكون جودة تسجيل الصوت عالية ، أي أن يقترب التمثيل الرقمي للصوت من تمثيله التناظريّ ، وهذا يتحكم به عن طريق معدل أخذ العينات ، فإذا زاد المعدل فإن نوعية الصوت المسجل تصبح أفضل . فمثلاً يكون الصوت الذي تؤخذ منه عينات عند (٤٤ كيلوهيرتز) أفضل من الصوت الذي تؤخذ منه عينات عند (٢٢ كيلوهيرتز) ، وهذا يعني بالضرورة إيجاد مساحات أكبر للتخزين لنوعيات الأصوات الأفضل .

٢-٥ ملفات التسجيل الصوتي

يوجد عدد كبير من البيانات السمعية ، ومن أنواع الملفات التي تحتويها ، وهذا ما يجعل المشاركة في ملفات البيانات السمعية في البيئات غير المتجانسة أمراً صعباً . وقد قامت بعض المنظمات العالمية مثل جمعية الوسائط المتعددة التفاعلية (IMA) (Interactive Multimedia Association) بمحاولة تطوير مقاييس معينة لتكوين معلومات عن التسجيل الصوتي في تنسيق مقبول بين منتجي المعدات المختلفين ، وتسهيل التعامل مع الوسائط المتعددة الأخرى .

٢-٦ مكتبة ملفات التسجيل الصوتي

هي عبارة عن مجموعة من التطبيقات و البرامج الفرعية التي تستطيع تزويد المستخدم بمجموعة من واجهات البرامج التطبيقية (API) الأنيقة والمنظمة ، وذلك لامكانية التعامل مع التنسيقات المختلفة لملفات التسجيل الصوتي المتنوعة مثل تنسيقات :

Compressed Audio Interchange Format File (AIFC)

، Audio Interchange Format File (AIFF)

، Recourse Interchange Format File (RIFF)

Sun Audio File (.snd/ .au) ، WAVE وغيرها من الأنواع الكثيرة الأخرى .

و يكمن الهدف من وجود هذه المكتبة إظهار شفافية تنسيق كل من الملف والبيانات . فالاستدعاءات لفتح ملف ، والتعامل مع أو معالجة البيانات في المرحلة الوسط (Metadata) ، وقراءة وكتابة عينات البيانات ، سوف تكون قابلة للعمل مع أي تنسيق مدعّم من قبل هذه المكتبة لأي ملف تسجيل صوتيّ .

٢-٧ تنسيق ملف التسجيل الصوتيّ

يوجد العديد من التنسيقات الخاصة بملفات التسجيل الصوتيّ ، والتنسيق هو عبارة عن تركيب يحتوي على معلومات عن البيانات ، ويحتوي على البيانات نفسها . وهذه المعلومات يمكن أن تكون مخزنة فيما يسمى بترويسة (Header) في بداية الملف ، أو أن تكون مبعثرة في قطع معينة (Chunks) خلال الملف ، وفي هذه الحالة تكون هذه القطع ذات صلة وثيقة بالتسجيل الصوتيّ .

ويوجد نوعان من أنواع التنسيقات الخاصة بملفات التسجيل الصوتي هما :

١. تنسيقات لنوع غير مضغوط من الملفات يسمّى تعديل شيفرة النبضة الخطية (Pulse Code Modulation) (PCM) ، وفي هذا النوع من التنسيقات تحول الموجات الصوتية في عملية التسجيل الصوتيّ إلى رقم ، وكل رقم من هذه الأرقام يمثل عينة من عينات ملف الصوت ،ومن أمثلة تنسيقات هذه الملفات : AIFF ، WAV وغيرها.
٢. تنسيقات لنوع مضغوط من الملفات يسمى الضغط المخسر (Lossy Compression)، والملفات التي تتبع هذا النوع من التنسيقات تستخدم خوارزميات معقدة لتمثيل التردّدات المهمة لإشارات الصوت لتأخذ مساحة أقل في التخزين مثل ملفات Media Player 3 (MP3) و ملفات Windows Media Audio (WMA) .

أما المعلومات وثيقة الصلة بالتسجيل الصوتيّ فتحتوي على تنسيق البيانات ، ومعدل أخذ العينات ، وعدد القنوات ، وعدد العينات ويمكن أن تحتوي على معلومات أخرى إضافية.

وتقوم مكتبة التسجيل الصوتيّ بدعم نوعين من أنواع تنسيق الملفات هما :

١. تنسيق شكل الموجة Waveform / RIFF المبنيّ على استخدام القطع (Chunks) .
٢. التنسيق العام المستخدم من قبل SUN , DEC , NeXT ، وهذا التنسيق مبنيّ على استخدام الترويسة المتبوعة بالبيانات .

وينصح هنا باستخدام تنسيق الملف ذي الترويسة حيثما أمكن ذلك ؛ لأنّ المحددات لبيانات التسجيل الصوتيّ لا يمكن تحديدها من العينات أنفسها .

٢-٨ تنسيق بيانات ملف التسجيل الصوتي

وهي طريقة تخزن بها عينات التسجيل الصوتيّ ، وأبسط هذه الطرق طريقة تسمى تعديل شيفرة النبضة الخطية . وهذا التنسيق يمكن التعامل معه بسهولة كبيرة وهو مفضل لكثير من برامج خلط وتحرير ملفات التسجيل الصوتي .

و تقوم مكتبة التسجيل الصوتيّ بدعم تنسيق البيانات المسمى (١٦ بت الخطي) والمستخدم مع الأقراص المدمجة و في محطات العمل الخاصة بيونكس (UNIX) . أما التنسيق المسمى (٨ بت الخطي) فيستخدم من قبل شركة ماكينتوش (Macintosh) وفي الحواسيب الشخصية .

٢-٩ العوامل المؤثرة في حجم ملف التسجيل الصوتيّ

توجد ثلاثة عوامل مشتركة بين جميع ملفات التسجيل الصوتي ، و تؤثر في حجم الملف وهي :

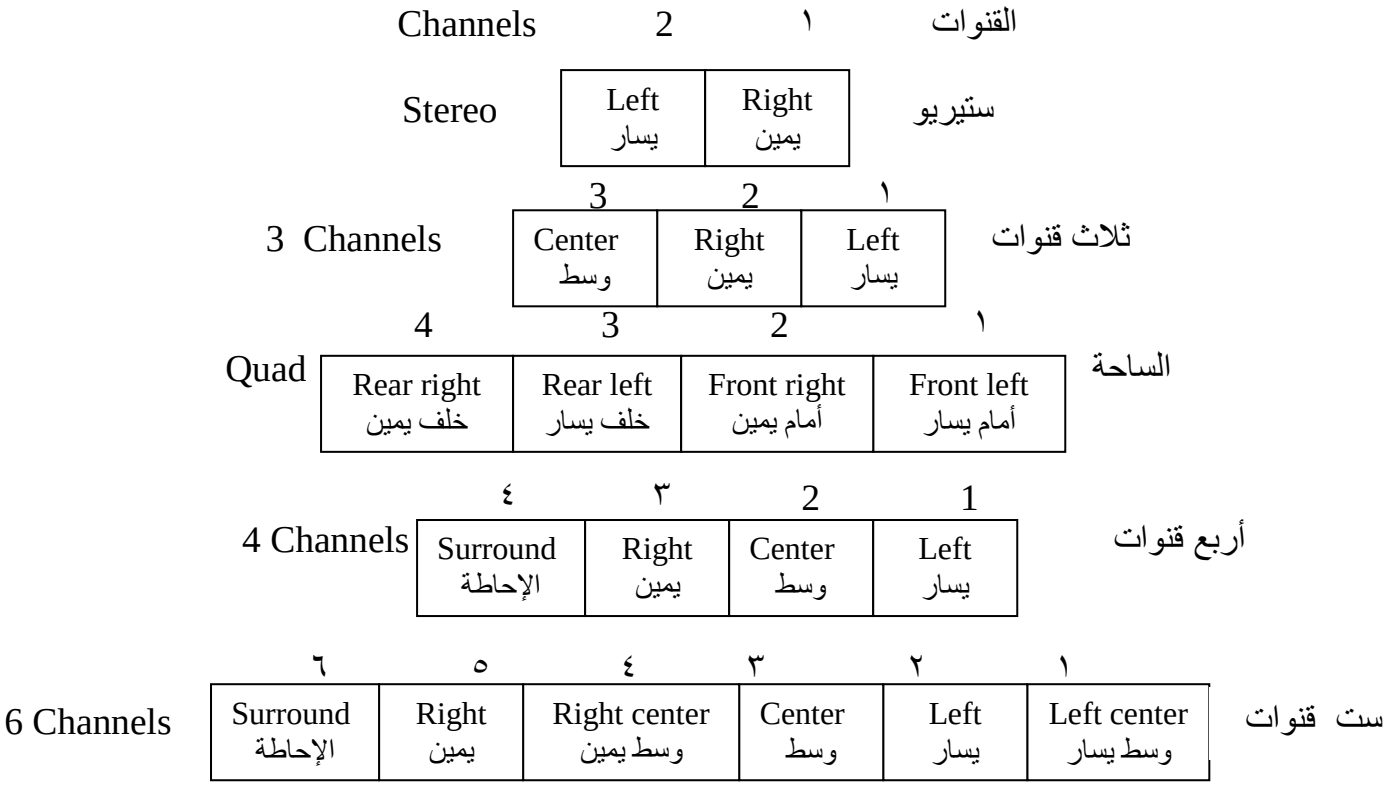
١. معدل أخذ العينات : تخزن ملفات التسجيل الصوتيّ الرقمية كمتتاليات طويلة مكونة من عينات الصوت ، تماما كما تخزن الصور المتحركة في فيلم كصور ثابتة متتالية . والمعدل الذي يحسب عند أيّ من هذه الصور الثابتة يسمى معدل أخذ العينة (كما يسمى التردّد) . فمثلا يمكن أن يكون معدل أخذ العينة هو (١٦ كيلوهيرتز) ، وهذا يعني أنّ (١٦) ألف عينة في الثانية أخذت عندما حوُل التسجيل الصوتي التناظري إلى تسجيل صوتيّ رقميّ . إن نوعية الصوت لملف معين يمكن أن تزداد بأخذ المزيد من العينات في نفس الفترة الزمنية ، مما يؤثر على زيادة حجم الملف .
٢. عدد الثنائيات في العينة : الثنائيات هي الأجزاء لنظام الترقيم المستخدم في الأجهزة الرقمية مثل الحواسيب ، والثنائية الواحدة تكون رقما أكبر يسمى الكلمة (Word) والذي يستخدم لتمثيل مستوى الجهد الكهربائي (السعة) للإشارة التناظرية . وبشكل عام فإنه لتمثيل العينة تستخدم إما (٨) ثنائيات ، أو (١٦) ثنائية . وللحصول على نوعية صوت أفضل ، تستخدم (١٦) ثنائية لتمثيل العينة، وهذا ينتج ملفا بضعف حجم ملف استخدم فيه (٨) ثنائيات لتمثيل العينة .
٣. يمكن أن تكون ملفات التسجيل الصوتيّ منسقة بحسب عدد قنواتها إلى عدد من التنسيقات، مثل ملفات ذات تنسيق أحادي القناة (Mono) أو ملفات ذات تنسيق ثنائي القناة (Stereo) الذي عادة ما يحتاج مثل هذا التنسيق إلى ضعف حجم الملف من نوع التنسيق أحادي القناة السابق لكونه يعمل على قناتين منفصلتين من قنوات التسجيل الصوتيّ . ويوجد تنسيقات أخرى ، مثل تنسيق الثلاث قنوات وتنسيق الساحة (Quad) وتنسيق الأربع قنوات وتنسيق الست قنوات .

وعند استخدام المحددات الثلاثة السابقة في أنظمة التعرف الآلي إلى الأصوات تراعى أن تكون تلك المحددات ضمن ما يلي:

١. بالنسبة لمعدل أخذ العينات : إنّ معدل أخذ العينات المناسب لأغلب نظم التعرف إلى الأصوات يساوي (٨٠٠٠ عينة / ثانية) أي (٨ كيلوهيرتز)، وبعض المستخدمين يفضله مساوياً لـ(١٦٠٠٠ عينة /ثانية) أي (١٦ كيلوهيرتز) ، لأنّ هذا المعدل الأخير يعطي معلومات أكثر دقة بالتردد العالي. ويدعم هذا أن الإنسان له خاصية تتمثل في كون صوته أو حديثه ذا مدى تردد (Bandwidth) منخفض يتراوح بين (١٠٠ هيرتز) و (٨ كيلوهيرتز) . ولو كان الحاسوب مجهزاً تجهيزاً عالياً لكان من الأفضل استخدام معدل أخذ العينات مساويا لـ(١٦ كيلوهيرتز) .

٢. أمّا عن عدد الثنائيات في العينة : فيمكن استخدام (٨ ثنائيات / عينة) وهذا سيتيح تسجيل القيم من (٠) إلى (٢٥٥) قيمة ، وهذا يعني أنّ كل عنصر من عناصر مكبر الصوت ، يمكن أن توضع في واحد من (٢٥٦) موضع . أمّا استخدام (١٦ ثنائية / عينة) فيعطي عدد قيم يصل إلى (٦٥٥٣٦) قيمة .

٣. أمّا بالنسبة لعدد القنوات في العينات : فإنّ التنسيق أحادي القناة يفي بغرض استخدام الملف الصوتيّ في سبيل التعرف إلى الأصوات . (Stephen Cook, 2003)



شكل رقم (٢-١) يبين الأعداد المختلفة للقنوات في ملف التسجيل الصوتي.

٢-١٠ أنواع ملفات التسجيل الصوتي

توجد أنواع كثيرة من ملفات التسجيل الصوتي المختلفة بحسب المعدات والبرامج التي تقوم بعملية تسجيل الأصوات في الملفات . وينتج من خلال تخزين البيانات السمعية ملفات ذات تنسيقات خاصة بالملف نفسه ، وتنسيقات أخرى خاصة بالبيانات التي يحتويها هذا الملف .

ومن الأنواع المختلفة لملفات التسجيل الصوتي ملفات الموجة (Wave Files) ذات الامتداد (wav) ، وملفات مشغلات الوسائط ٣ (Media Player 3) ذات الامتداد (MP3) ، وملفات الصوت البسيطة (Basic Audio Files) ذات الامتداد (au) أو (snd) ، وملفات المشغل الحقيقي (RealPlayer) ذات الامتداد (ra) أو (ram) ، وملفات مشغل الوسائط الخاص بالنوافذ (Windows Media Player) (WMA) ، وملفات (MIDI) ذات الامتداد (midi) أو (mid) ، وغيرها الكثير من أنواع ملفات الصوت والتسجيل الصوتي المختلفة .

ولأسماء الملفات عادة امتدادات مختلفة تتكون حسب نوع الملف وكيفية إنتاجه . وتدعم مكتبة التسجيل الصوتيّ أسماء امتدادات الملفات من قبيل :

- قانون ميو (μ) .u
- قانون أ .al
- الامتداد العام (مستخدم من قبل Next, SUN, DEC, HP) .au , .snd
- مستخدم من قبل مايكروسوفت ويتبع تنسيق شكل الموجه RIFF .wav
- ١٦ بت الخطي .L16
- ٨ بت الخطي .l8
- ٨ بت الخطي المعدل .lo8

شكل رقم (٢-٢) يبين أسماء امتدادات الملفات المدعمة من قبل مكتبة التسجيل الصوتي.

وتستخدم مكتبة التسجيل الصوتيّ عدداً من القوانين لتحديد تنسيق البيانات في ملف تسجيل صوتيّ على النحو التالي :

١. إذا كان ملف التسجيل الصوتيّ يحتوي على ترويسة ، فيتم تجاهل امتداد اسم الملف .
٢. إذا لم يحتوِ الملف التسجيل الصوتي على ترويسة ، فإنه يتم التعامل معه كملف خام للتسجيل الصوتيّ ، وهنا يستخدم امتداد اسم الملف لتحديد المعلومات عن الملف .

و نستطيع دراسة أحد أنواع ملفات التسجيل الصوتي الكثيرة والمتنوعة دراسة معمقة لبحث إمكانية إجراء خوارزمية جديدة على مثل هذا النوع من ملفات التسجيل الصوتي من شأنها تجزئة الملف بطريقة جديدة .

وقد اختير أحد هذه الأنواع وهو الملفات ذات الامتداد (.wav) وذلك بسبب :

١. سهولة التعامل مع مثل هذا النوع من الملفات من ناحية تنسيقه الداخليّ نسبيّاً .
٢. إمكانية تشغيل وتجربة هذا النوع من الملفات على أيِّ حاسوب يعمل بنظام التشغيل النوافذ (Windows) .

٢-١١ الملفات ذات التنسيق (RIFF) أو تنسيق الملفات المتغيرة المصدر

(Recourse Interchange File Format)

يستخدم هذا التنسيق للملفات الخاصة بتخزين العديد من أنواع البيانات و حفظها وبشكل رئيس بيانات الوسائط المتعددة ، مثل : التسجيل الصوتي والفيديو . وتقوم شركة مايكروسوفت بالتعامل مع مجموعة كبيرة من الملفات في برامجها المتنوعة و الكثيرة ، وأحد هذه الأنواع من الملفات يسمى بتنسيق الملفات المتغيرة المصدر (Resource Interchange File Format) ، ويعرف اختصاراً بـ(RIFF) ، ويتكون أي ملف من النوع (RIFF) عادة في بنائه الداخلي من قطع رئيسة وقطع فرعية ولكل منها نوع يتمثل بأربعة حروف .

يمكن اعتبار تركيب الملف ذي التنسيق (RIFF) كله كقطعة كبيرة واحدة تحوي القطع الفرعية الأخرى ، وأول محتويات القطعة الرئيسة ما يسمى (نوع الشكل) (Form Type) الذي يصف النوع العام لمحتويات الملف . ويمكن عرض تركيب ملف (RIFF) بالشكل التالي:

التعديل Offset (بالنظام ١٦) (hex)	Contents المحتويات
0000	"R" , "I" , "F" , "F" الحروف المكونة لنوع الملف
0004	Length of the entire file حجم كل الملف
0008	Form Type نوع الشكل
000C	First chunk type نوع أول قطعة
0010	First chunk size حجم أول قطعة
0014	First chunk data بيانات أول قطعة

جدول رقم (٢-١) يبين تركيب ملف (RIFF) من موقع

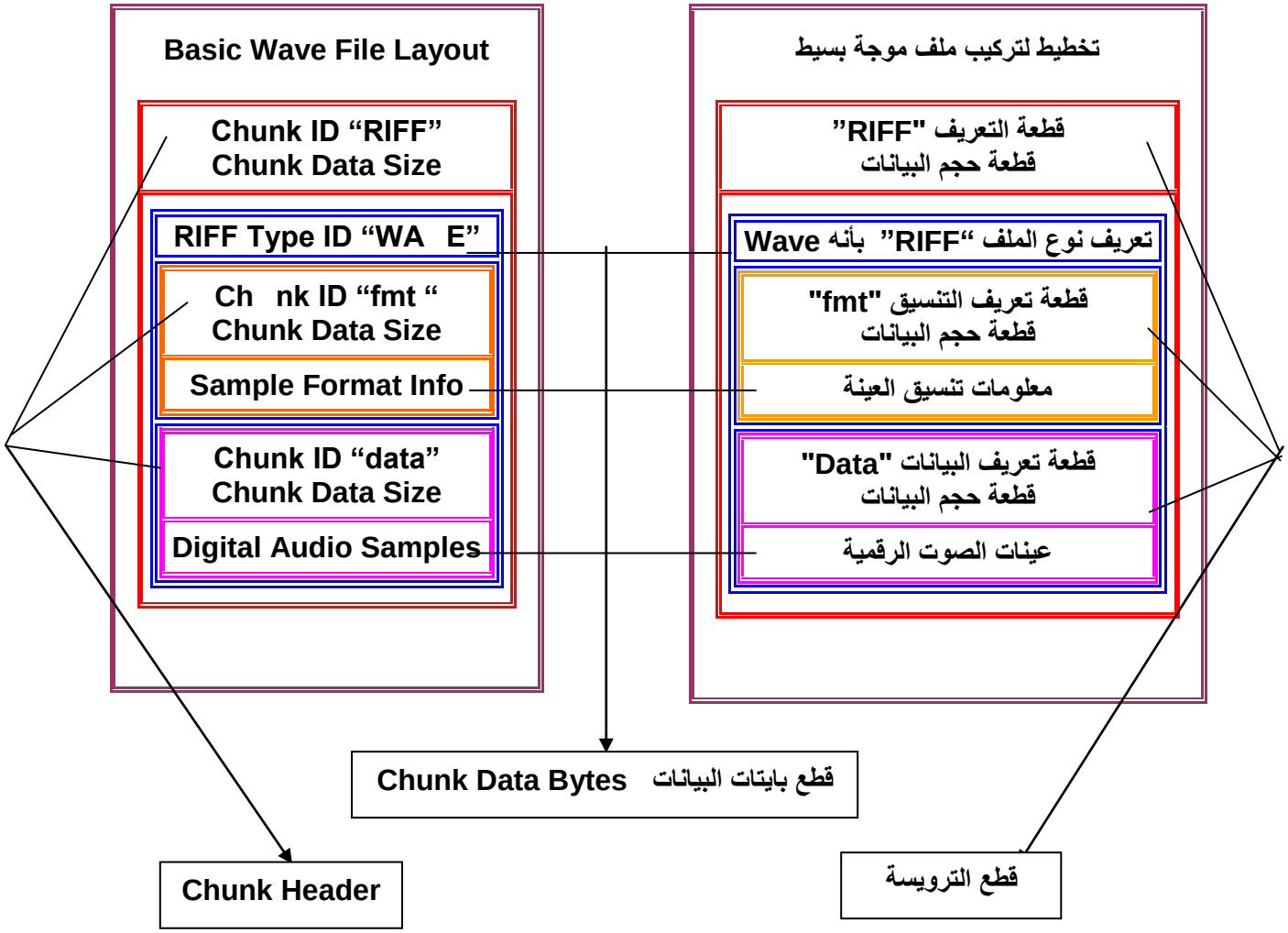
(www.sonicspot.com/guide/wavefiles.html#data**)**

ويستخدم التنسيق القياسي (RIFF) في ملفات الموجة ، حيث يقوم بتجميع محتوى الملف، مثل : (تنسيق العينة ، عينات الصوت الرقمية) في قطع منفصلة كل منها يحتوي ترويسة خاصة وبيانات على شكل ثنائيات . وتقوم قطعة الترويسة بتحديد حجم ونوع قطع بايتات البيانات .

وتعني قيمة التعديل (Offset) المذكورة في الجدول السابق ، قيمة العنوان أو الموقع الأساسي في الذاكرة لتعبير معين مضافاً إليه رقم محدد لانتاج قيمة ثانية للعنوان أو الموقع لنفس

التعبير . وتحديد العناوين باستخدام طريقة التعديل تسمى العنونة النسبية ، ويعبر عن التعديل بقيم ممثلة بالنظام السادس عشر . هذا ويمكن استخدام كلمة إزاحة بمعنى كلمة التعديل .

وقد تحتوي بعض أنواع هذه القطع على قطع فرعية فيها ،و الجدول رقم (٢-٢) يوضح القطع الخاصة بالتنسيق "fmt" والأخرى الخاصة بالبيانات "Data" وهما عبارة عن قطعتان فرعيتان من القطعة الرئيسة لـ "RIFF" .



جدول رقم (٢-٢) يبين تركيب ملف موجة حسب التنسيق (RIFF) من موقع

(www.sonicspot.com/guide/wavefiles.html#data**)**

٢-١٢ نظرة عامة على ملف الموجة

يعتبر ملف الموجة من الملفات واسعة الانتشار والاستخدام في الحواسيب الشخصية بسبب انتشار برنامج النوافذ (Windows) ، ولكثرة البرامج المكتوبة لهذا النوع من تخطيطات الملفات ، كما أن هذا النوع من الملفات يستخدم لتخزين البيانات السمعية الرقمية ؛ لذا فإن أغلب البرامج الحديثة التي تخزن وتحفظ الأصوات الرقمية تدعم هذا النوع من التنسيقات .

تنسيق ملف الموجة :

إنّ تنسيق ملف الموجة هو جزء من التنسيق القياسي المسمى بتنسيق الملفات المتغيرة المصدر (RIFF) ، فالترويسات الخاصة بملف الموجة تتبع تركيب تنسيق ملف (RIFF) ، حيث إن أول ثمانية بايتات في ملف الموجة هي عبارة عن قطعة الترويسة لملف (RIFF) القياسي والتي تحتوي بدورها على قطعة التعريف لملف (RIFF) والمحتوية على الحروف (RIFF) ، وقطعة حجم البيانات تساوي حجم الملف مطروحا منه ثمانية بايتات وهي البيتات المستخدمة للترويسة السابقة . وتعد أول أربعة بايتات للبيانات في قطعة (RIFF) محددة لنوع المصدر الموجود في قطعة (RIFF) ، وفي حالة ملف الموجة سنجد كلمة (WAVE) في هذه القطعة . وبعدها تظهر جميع القطع الخاصة بملف الموجة التي تعرف شكل موجة التسجيل الصوتي .

Offset التعديل	Size الحجم	Description الوصف	Value القيمة
0x00	4	Chunk ID قطعة التعريف	"RIFF" (0x52494646)
0x04	4	Chunk Data Size قطعة حجم البيانات	(File Size) minus (8) (حجم الملف) مطروحاً منه (٨)
0x08	4	RIFF Type نوع ملف من التنسيق السابق	مثلا "WAVE" (0x57415645)
0x10	Wave chunks قطع ملف الموجة		

جدول رقم (٢-٣) **يبين** قيم القطع لملف من نوع RIFF. من موقع

(www.sonicspot.com/guide/wavefiles.html#data**)**

ملف الموجة

يتركب ملف الموجة من قطع خاصة لتعريفه ، وهذه القطع تكون غالباً قطعتين : إحداهما للتنسيق والأخرى للبيانات ، ووظيفتهما وصف تنسيق عينات الصوت الرقمية ، ووصف العينات أنفسها .

ومن الطبيعي هنا أن تكون قطعة التنسيق موجودة في مكان ما قبل قطعة البيانات ، خاصة إذا سجل الصوت الرقمي من خلال مصدر بطيء ويقوم على طريقة التسجيل الخطية مثل

الإنترنت . وإذا حدث أن كانت قطعة البيانات قبل قطعة التنسيق ، فإنه عند إعادة سماع الصوت سوف يكون من الواجب إجراء تغيير الترتيب للقطع السابقة مرة أخرى قبل التشغيل .

إنّ القطع في أي ملف من تنسيق (RIFF) أو في ملف الموجة تخزن حسب الجدول التالي:

Offset التعديل	Size الحجم	Description الوصف
0x00	4	Chunk ID قطعة التعريف
0x04	4	Chunk Data Size قطعة حجم البيانات
0x08	Chunk Data Bytes قطعة بايتات البيانات	

جدول رقم (٢-٤) **يبين** تنسيق القطع لملف من نوع RIFF من موقع

(www.sonicspot.com/guide/wavefiles.html#data**)**

يتكون ملف الموجة من عدد من القطع المختلفة داخل القطع المذكورة سابقاً التي يكون لكل منها عمل خاص بها ، ولها مميزات محددة تساعد في أداء العمل بشكل تام .

٢-١٣ قطع ملف الموجة (Wave File Chunks)

أ. قطعة التنسيق (Chunk "fmt") يوضح الشكل التالي ما يتعلق بكل قطعة فرعية من قطعة التنسيق "fmt" من ملف الموجة من حيث الحجم و الوصف و القيمة .

Offset التعديل	Size الحجم	Description الوصف	Value القيمة
0x00	4	Chunk ID قطعة التعريف	"fmt " (0x666D7420)
0x04	4	Chunk Data Size قطعة حجم البيانات	16 + extra format bytes ١٦+ بايتات تنسيق إضافية
0x08	2	Compression code شيفرة الضغط	From 1 to 65,535
0x0a	2	Number of channels عدد القنوات	From 1 to 65,535
0x0c	4	Sample rate معدل أخذ العينة	From 1 to 0xFFFFFFFF
0x10	4	Average bytes per second معدل البايتات في كل ثانية	From 1 to 0xFFFFFFFF
0x14	2	Block align محاذاة الكتلة	From 1 to 65,535
0x16	2	Significant bits per sample الثنائيات المهمة في كل عينة	From 2 to 65,535
0x18	2	Extra format bytes بايتات التنسيق الإضافية	From 0 to 65,535
0x1a	Extra format bytes بايتات التنسيق الإضافية		

جدول رقم (٢-٥) يبين تنسيق قطعة "fmt" لملف من نوع موجة . من موقع (www.sonicspot.com/guide/wavefiles.html#data)

وتفصيل الجدول السابق فيما يلي يتضمن شرحاً موجزاً عن كل قطعة من القطع السابقة من حيث أهميتها وعملها ، حيث إن قطعة التنسيق "fmt" تحتوي على معلومات عن كيفية تخزين البيانات واسترجاعها ، ونوع شيفرة الضغط المستخدمة ، وعدد القنوات ، ومعدل أخذ العينات ، و عدد الثنائيات في كل عينة ، ومحددات أخرى غيرها .

١. قطعة التعريف (Chunk ID) : وهي دائما تحتوي على كلمة "fmt" أي القيمة (0x666D7420) .

٢. قطعة حجم البيانات (Chunk Data Size) : وهي تمثل الحجم القياسي لبيانات تنسيق الموجة وهي تساوي (١٦) بايت ، مضافاً إليها حجم بايتات التنسيق الإضافية التي يُحتاج

إليها لتنسيق موجة محددة ، وهذه الحالة تكون عندما لا تحتوي على بيانات غير مضغوطة من نوع تعديل شيفرة النبضة الخطية (PCM) . يلاحظ هنا أن المتتالية تنتهي بفراغ (0x20) .

٣. شيفرة الضغط (Compression Code) : تمثل أول كلمة من تنسيق البيانات نوع الضغط المستخدم على البيانات ، والمحتوى في القطعة "RIFF" التي ذكرت سابقاً . وهذه بعض أمثلة أنواع الضغط الممكن توافرها في هذه القطعة :

Code الشيفرة	Description الوصف
(1) means in Hex (0x0001)	PCM / uncompressed
(6) means in Hex (0x0006)	ITU G.711 a-law
(7) means in Hex (0x0007)	ITU G.711 μ-law
(17) means in Hex (0x0011)	IMA ADPCM
(80) means in Hex (0x0050)	MPEG

جدول رقم (٢-٦) يبين بعض أنواع شيفرات الضغط الشائعة من موقع

(www.sonicspot.com/guide/wavefiles.html#data**)**

٤. عدد القنوات (Number of channels) : ويحدد هذا المتغير عدد الإشارات الصوتية المنفصلة المشفرة في قطعة بيانات الموجة . وتعني القيمة (١) هنا أن الإشارة هي إشارة أحادية (Mono) ، أما الرقم (٢) فيعني أن الإشارة ثنائية (Stereo) .

٥. معدل أخذ العينات (Sample Rate) : وهو عدد العينات المأخوذة في الثانية الواحدة .

٦. معدل البايتات في الثانية (Average bytes per second): يمثل هذا المتغير عدد البايتات في الثانية لبيانات الموجة الواجب إدخالها كمتتابعة إلى محوِّل بيانات لامكانية تشغيل ملف الموجة. وهذه المعلومات مفيدة عند تحديد فيما إذا كانت البيانات ستتابع من المصدر بسرعة كافية للحفاظ على تتابع إعادة تشغيل التسجيل . ويتأثر هذا المعدل بمعدل أخذ العينات و محاذاة الكتلة .

٧. محاذاة الكتلة (Block Align) : يمثل هذا المتغير عدد البايتات في شريحة من العينات ، ويتأثر بعدد القنوات للصوت ، وبالثنائية المهمة لكل عينة .

٨. الثنائية المهمة لكل عينة (Significant Bits Per Sample) : ويحدد عدد الثنائيات التي تعرف كل عينة ، وهذه القيمة عادة ما تكون (٨ ،١٦ ، ٢٤ ،٣٢) وإذا لم يكن عدد الثنائيات من مضاعفات الرقم (٨) – أي أن البايت لم يحاذى - فإن عدد البايتات

المستخدمة في كل عينة يتم تقريبه لأقرب حجم بايت ، وتملأ البايتات الفارغة بالأصفار ثم تهمل .

٩. بايتات تنسيق إضافية (ُExtra Format Bytes) : وقيمة هذا المتغير تحدد عدد بايتات التنسيق الإضافية اللاحقة ، وتعتمد هذه القيمة على نوع شيفرة الضغط المستخدمة في ملف الموجة ، أي معلومات الضغط المطلوبة لتشفير بيانات الموجة .

ب. قطعة البيانات (Chunk "data")

تحتوي هذه القطعة من الملف على عينات البيانات الرقمية السمعية ، والتي ستشفر اعتمادا على التنسيق وطريقة الضغط المحددة من قبل قطعة التنسيق "fmt" . وإذا كانت شيفرة الضغط = (١) – أي تعديل شيفرة النبضة الخطية (PCM) غير المضغوطة – فإنّ بيانات الموجة تحتوي قيم عينات خام .

وتحتوي ملفات الموجة قطعة بيانات واحدة ، ولكنها قد تحتوي عدداً أكثر إذا كانت محتواة في قطعة تسمى قطعة قائمة الموجة (Wave List Chunk) .

يوضح الشكل التالي ما يتعلق بكل قطعة فرعية من قطعة البيانات " data" من ملف الموجة ، من حيث الحجم و الوصف والقيمة :

Offset	Length	Description	Value
0x00	4	chunk ID قطعة التعريف	"data" (0x64617461) كلمة "بيانات"
0x04	4	chunk size قطعة الحجم	Depends on sample length and compression يعتمد على طول العينة ونوع الضغط
0x08	sample data بيانات العينات		

جدول رقم (٢-٧) يبين تنسيق قطعة "data" لملف من نوع موجة. من موقع (www.sonicspot.com/guide/wavefiles.html#data**)**

وتفصيل الجدول السابق فيما يلي يتضمن شرحاً موجزاً عن كل قطعة من القطع السابقة من حيث أهميتها وعملها ، حيث إنّ قطعة البيانات " data" تحتوي على معلومات عن قطعة التعريف ، و قطعة الحجم .

١. قطعة التعريف (Chunk ID) : وتحتوي هذه القطعة على كلمة بيانات (data) ممثلة بالرقم (0x64617461) .

٢. قطعة الحجم (Chunk Size) : وهي تمثل حجم للبيانات وهي تعتمد على طول العينة وما فيها من بايتات ، وعلى نوع شيفرة الضغط المستخدمة .

٣. قطعة بيانات العينات (Sample Data Chunk) : وتحتوي على جميع عينات الصوت السمعية ، ويلاحظ على العينات عند تمثيلها بثمانية ثنائيات ، فهي تحدد تركيباً بقيم عديمة الإشارة (Unsigned values) . أما إذا مثلت العينات بأي حجم آخر من الثنائيات ، فتحدد بقيم ذات إشارة (Signed values) . فمثلاً عينة ممثلة بـ (١٦) ثنائية يمكن أن تعطى مدى من قيمة (-٣٢٧٦٨) وحتى قيمة (٣٢٧٦٧) مع قيمة وسط تمثل السكون (صفر) .

هذا ويحتوي ملف الموجة على كثير من القطع الأخرى غير التي تم ذكرها سابقاً ، نوردها بمُسمياتها هنا فقط لاهتمام هذه الدراسة بما تم ذكره تفصيلاً فيما سبق ، ومن أسماء القطع الأخرى : قطعة الحقيقة (Fact Chunk - fact) ، قطعة الوقت الحقيقي (Cue Chunk - cue) ، قطعة لائحة التشغيل (Playlist Chunk - plst) ، القطعة الخاصة بالبيانات المتعلقة باللائحة (Associated Data List Chunk - list) ، قطعة الملصق (Label Chunk - labl) ، قطعة النص المعنون (Labeled Text Chunk - ltxt) ، قطعة الملاحظات (Note Chunk - note) ، وقطعة العينة (Sample Chunk - smpl) ، و قطعة الأداة (Instrument Chunk - inst) . وكل من هذه القطع له عمله الخاص به ومميزاته التي يعمل بها ولكن هدف الدراسة بعيد عن التعمق في مثل هذه التفاصيل ، ويمكن التوسع في دراسة هذه القطع في أبحاث أخرى غير هذه الدراسة .

الفصل الثالث

تجزئة الملفات (File Segmentation)

٣-١ الدراسات السابقة في موضوع تجزئة الملفات

تمت دراسات كثيرة ومتعددة حول موضوع تجزئة ملفات الصوت والتسجيل الصوتي للتعرف على الأصوات ، وظهرت أوراق بحثية كثيرة تدور حول هذا الموضوع بحيث أصبح كثيرٌ من ملفات التسجيل الصوتي - بعد هذه الدراسات والأبحاث- تقطع وتجزأ في برامج (أو بشكل نظريّ) بحيث يستفاد منها في نواح مختلفة عدة .

ومن هذه الدراسات ما قام به كل من جورج تزانتيك و بيري كوك (George Tzanetakis, Perry Cook) في دراسة تدعى (تجزئة ملفات التسجيل الصوتي المتعدد المميزات للتصفح والتذييل) (Multifeature Audio Segmentation For Browsing And Annotation) ، التي قدمت خوارزمية عامة لتجزئة ملفات الصوت البسيطة ، و المعتمدة على خصائص متعددة . وقد استخدم الباحثان نوع من المتصفحات المعتمدة على التجزئة والتصنيف ، الذي طبق على موضوعات التجزئة المختلفة لتقييمها .

ونفذ الباحثان تجارب متعددة للمقارنة بين التجزئة التي تقام يدوياً ، والتجزئة الآلية . وبينت النتائج أن المواضيع البشرية ثابتة عند تجزئة ملفات الصوت ، كما يمكن أتمتة هذه العملية.

وبعد ظهور النتائج عقد الباحثان مقارنات بين عملية التجزئة التي تنفذ بشكل يدويّ والتي تنفذ بشكل آلي ، على ملفات الصوت . وخلص الباحثان في نهاية البحث إلى أنّ هذا البحث يبين القوة الكامنة لعملية تجزئة ملفات الصوت بشكل آلي .

أما ما قام به الباحثون (ألفر بيتكيون ، و لورنت كوفرر ، و بيير كوفرر) (Olivier Pietquin, Laurent Couvreur, Pierre Couvreur) من تقديم ورقة بحثية بعنوان (القطاعات التطبيقية للتجزئة الآلية المعتمدة على صوت المتحدث للمواد السمعية) (Applied Clustering for Automatic Speaker-Based Segmentation of Audio Material) تبحث في خوارزمية تتعلق بالتجزئة لصوت متحدث في ملف صوت ، للاستفادة من ذلك في عمليات استرجاع ملف الصوت لأحد المتكلمين في الملف .

وتحتوي هذه الخوارزمية على عمليتين منفصلتين : إحداهما تدعى عملية الفصل (splitting) ، والأخرى عملية الدمج (Merging) ، واعتمد الباحثون في هذه الخوارزمية على تسجيلات صوتية إخبارية من هيئة الإذاعة البريطانية (BBC) .

وأظهرت النتائج بأنّ عملية الفصل نفذت بدقة عالية ، ولكن رافقها فقدان قليل من معدل التنبؤ ، بينما كانت نتائج عملية الدمج مُرْضية .

واعتمد الباحثون في خوارزميتهم على معادلات رياضية كثيرة ومتنوعة ، و قاموا بإجراء التجارب الكثيرة ، حتى ظهرت نتائج تلك العمليتين ، وعقدوا مقارنات متعددة على النتائج التي توصلوا إليها .تمت دراسات كثيرة ومتعددة حول موضوع تقسيم ملفات الصوت والتسجيل الصوتي ، مما أنتج برمجيات وبرامج متعددة تخدم عمليات التقسيم تلك ، حيث أصبح كثير من ملفات التسجيل الصوتي تقطع إلى قطع وأجزاء يستفاد منها في نواح عدة مختلفة .

و قدم الثنائي (توماس سوونج و تل فان) (Thomas Soong, Til T. Phan) ورقة بحثية عن التجزئة بعنوان (تعريف حديث المتكلم الغير معتمد على نص ، المستند على فكرةِ انتزاع الميزةِ) (Text-Independent Speaker Identification based on the Idea of Feature Extracting) ، وهي عبارة عن منهجية تستخدم المعادلات الرياضية لتقليل عدد العينات المتعامل بها في كل لحظة .

هذا ويجب أن تحتوي عملية التجزئة على منهجية للتقليل أو التصغير يمكن أن تكون انتزاع ميزة أو أية منهجية أخرى ، لذا فإنّ العينات يمكن أن تعالج بسهولة دون الحاجة لزيادة تعقيد عمليات معالجة عدد هائل من العينات .

٢-٣ عملية تجزئة الملفات (File Segmentation Process)

تعرف عملية التجزئة لملفات الصوت بأنها العملية التي يتم فيها تقطيع سيل الثنائيات الهائلة العدد الممثلة للبيانات الصوتية إلى قطع صوتية متجانسة ، ويعتمد هذا التجانس على طبيعة العمل المنفذ على هذا الملف .(J. Ajmera, I. McCowan, and H. Bourlard, 2004)

كما يقصد بعملية تجزئة الملف تقسيمه إلى أجزاء أصغر حجماً ، تحتوي مقاطع محددة من كلمات أو أصوات يستفاد منها في نواح مختلفة ، بحيث لا يمكن الاستفادة من هذا الملف إذا بقي على حالته الأصلية بغير تجزئة .

وعمليات التجزئة كثيرة ومتعددة ، منها ما يتم على ملفات النصوص (Text Files) ، و منها ما يتم على ملفات الفيديو (Video Files) ، أو على ملفات التسجيل الصوتيّ (Audio Files) .

وجميع عمليات التجزئة ترنو لتحقيق أهداف مشتركة فيما بينها ، منها تصغير حجم الملف ليصار إلى التعامل معه بشكل معين ، أو تسهيل وتبسيط عملية تحميله أو تنزيله من الإنترنت مثلاً ، أو تخزين الملف بشكله المجزأ على وسائط التخزين المختلفة ، أو القيام بالتعرف على صوت المتحدث من خلال أجزائه .

٣-٣ الأهداف من عملية التجزئة

يمكن تصنيف العمليات التي قد يحتاج فيها إلى عملية تجزئة التسجيل الصوتيّ إلى الفروع التالية :

١. عمليات التعرف الآلي على الأصوات (Automatic Speech Recognition) (ASR) وهو من أهم الحقول في هذا المجال ، و مازال العمل جارياً والأبحاث تصدر في هذا الجانب المهم من جوانب الذكاء الاصطناعي .
٢. عمليات التجزئة (Segmentation) : لاستخدام الصوت عبر الإنترنت ، وتقليل الوقت المطلوب لذلك الاستخدام ، وتقليل الوقت لعمليات تحميل الملفات والبرامج و تنزيلها .
٣. عمليات المحاكاة (Simulation) : عند التعامل مع ملفات الصوت أو معالجتها بسهولة و بوضوح وانسيابية ، فإنّ ذلك يؤدي إلى فهم جيد لتركيب ملفات الصوت ، و إمكانية تطويع ملفات الصوت لاستخدامها في أي نوع من أنواع الدراسات أو الأبحاث أو الألعاب أو تطبيقات الوسائط المتعددة ، أو تطوير صفحات الإنترنت ، أو الأمن ، وكثير من التطبيقات الأخرى .

و قد كان الاهتمام الرئيس لعمليات التجزئة للملفات منصبّاً على عمليات التعرّف الآليّ على الأصوات ، حيث كانت عمليات التجزئة لملفات الأصوات المساعد الرئيس لتسهيل وتيسير عمليات التعرف الآلي لتلك الأصوات ، بالإضافة إلى استخدام عمليات التجزئة في عمليات الفهرسة والتذييل أيضاً .

وفي هذه الدراسة ، تم التركيز على نوع من أنواع عمليات التجزئة ، وهو الذي يهتم بتجزئة ملفات التسجيل الصوتي (Audio Files) ، وتحديدا ملف الموجة ، وذلك من أجل استخدام الملفات المجزأة في عملية التعرف الآلي على الأصوات .

الفصل الرابع

التعرف الآلي إلى الأصوات (Automatic Speech Recognition)

٤-١ الدراسات السابقة في موضوع التعرف إلى الأصوات

أما في مجال التعرف إلى الأصوات فقد وجدت دراسات عدة في هذا المجال مثل الدراسة التي قام بها كل من كيث بين (Keith Bain) و د. ساره باسون (Sara Basson) و د. مايك والد (Mike Wald) ، وكانت حول موضوع التعرف إلى الخطاب في قاعات المحاضرات الجامعية : مشروع التعليم المحرر (Speech Recognition in University Classrooms : liberated Learning Project) .

وكان هدف الدراسة الإجابة عن استفسارات منها : هل تستطيع تكنولوجيا التعرف إلى الأصوات تحويل المحاضرات من صورتها الصوتية إلى محاضرات نصية في قاعات المحاضرات . وهل تستطيع تلك التكنولوجيا تغيير طرق أخذ الملاحظات في غرف المحاضرات من شكلها التقليدي إلى طرق أخرى يستطيع المعاقون استخدامها ؟

٤-٢ التعرف إلى الأصوات

يعد موضوع التعرف على الأصوات فرعاً من فروع علم الذكاء الاصطناعي ، حيث طرقت كثير من الدراسات أبواب هذا المنحى و أجادت في تحليلاتها واختباراتها ، وتعددت طرق بحثها من جميع الجهات ، حتى وجدت العديد من النتائج على هيئة تطبيقات مختلفة ، تسعى لكي يصل المستخدم إلى مبتغاه في الإجابة عما يدور من أسئلة حول ماهية الأصوات التي يحصل عليها من ملفات الأصوات أو التسجيلات الصوتية .

ويمكن تعريف مفهوم التعرف إلى الأصوات بأنه العملية التي يقوم فيها الحاسوب بالتعرف إلى كلمات منطوقة ، أو بمعنى آخر التحدث إلى جهاز الحاسوب و تعرفه بشكل صحيح إلى ما تم قوله . (Stephen Cook , 2003)

كما يمكن تعريفه بأنه تقنية تُحوِّل الكلمات و الأصوات التي تصدر عن الإنسان إلى إشارات كهربائية ، تنقل بدورها إلى نموذج تشفير يقوم بتوضيح ما قصد من تلك الكلمات أو الأصوات .

٤-٣ بعض مجالات استخدام أنظمة التعرف إلى الأصوات

بالاعتماد على تعريف نظام التعرف إلى الأصوات بأنه برنامج يسمح للمستخدمين بالتواصل من خلال الحاسوب باستخدام الحديث بدلاً من استخدام لوحة المفاتيح ، فإن هذا المجال قد فتح أبواباً كثيرة تستفيد منها فئات غير قليلة من المستخدمين ، ومن أهمها:

١. فئة المعاقين حركيّاً : بعض الطلاب و كبار السن لديهم إعاقات حركية تمنعهم من استخدام لوحات المفاتيح العادية أو حتى الفأرة . لذا فإن أنظمة التعرف على الأصوات تتيح المجال أمامهم لاستبدال طرق الإدخال التقليدية بتلك الأنظمة .
٢. فئة الذين يعانون من صعوبة النطق : تستطيع تقنية التعرف إلى الأصوات مساعدة الطلاب الذين يعانون من صعوبة النطق وذلك بمساعدتهم في إمكانية تهجئة الكلمات أو حتى كتابتها . (M.Hoffman , L.Coleman,1999)

بالإضافة إلى المجالات السابقة يوجد العديد من المجالات الأخرى التي تُستخدم فيها أنظمة التعرف إلى الأصوات مثل المجالات التالية:

١. إصدار الأوامر وهو الاستخدام الأكثر شيوعاً لدى أنظمة التعرف إلى الأصوات حيث تستخدم في المجال الطبي ، والأوامر القانونية والتجارية كما في معالجات النصوص العامة .
٢. الأنظمة التي تدار بشكل أوامر وتعليمات ، مثل الأمر (Open) ، (Save) .
٣. بعض أنظمة الهاتف ، حيث يطلب من المتصل إعطاء أوامر للجهاز بدلاً من الضغط على مفاتيح الهاتف .
٤. بعض الهواتف النقالة التي يسمح فيها بطلب رقم معين صوتيّاً مثل (العمل) ، (المستشفى) وغيرهما . (Stephen Cook , 2003)

٤-٤ محددات تعتمد عليها أنظمة التعرف إلى الأصوات

توجد بعض المحددات التي لابد من معرفتها عند التعامل مع أنظمة التعرف إلى الأصوات حتى تتم هذه العملية بأعلى معدل للدقة ، ومن هذه المحددات :

١. النطق : وهو كيفية لفظ الكلمة أو الكلمات والتي تحدد معنى واحد في الحاسوب.
٢. الاعتماد على المتكلم : وتصمم أنظمة التعرف إلى الأصوات المعتمدة على المتكلم ، على متكلم محدد ، بافتراض ثبات حديثه وحركاته الصوتية . أما الأنظمة غير المعتمدة على المتكلم فتصمم لنوعيات مختلفة من أصوات المتكلمين.
٣. الكلمات : وهي قائمة الكلمات التي يستطيع النظام أن يتعرف عليها ، وهنا تكون الكلمات الصغيرة الحجم ايسر على النظام للتعرف عليها من الكلمات الكبيرة الحجم .

٤. مقدار الدقة : وهي قدرة النظام على التعرف إلى كيفية نطق الكلمات ، ويُعدُّ النظام نظاماً جيداً إذا كانت قيمة الدقة فيه تقارب الـ (٩٨%) أو أكثر.

(Stephen Cook, 2003)

٤-٥ أنواع أنظمة التعرف إلى الأصوات

يمكن تصنيف أنواع الأنظمة الخاصة بالتعرف إلى الأصوات تصنيفات كثيرة ومتنوعة حسب الحاجة من تلك الأنظمة والمقدرة على إمكانية التعرف على نطق للكلمات أو وجود الكلمات في الحاسوب أو غيرها من المحددات .

و أحد هذه التصنيفات يتضمن تقسيم أنواع أنظمة التعرف إلى الأصوات إلى صنفين رئيسين أوّلهما مطابقة النماذج (Template Matching) ، والثاني ويعرف بتحليل الخصائص (Feature Analysis) .

١. فالنوع الأول وهو مطابقة النماذج وهو الأسلوب الأكثر سهولة وبساطة ودقة عند استخدامه بشكل صحيح ، و يعتمد هذا النظام على حساب المعدل الإحصائيّ لعدد من العينات لكلمة ما ، ومقارنته بالمعدل الإحصائي لنفس الكلمة المخزنة كنموذج في جهاز الحاسوب .

ويكون مفهوم التعرف إلى الأصوات بهذه الطريقة منحصراً بعدد من كلمات النماذج المخزنة في ذاكرة الحاسوب وتعد بالمئات ، ويعتمد كذلك على المتكلمين الذين صدرت منهم تلك الكلمات النماذج .

ويتطلب تطبيق مثل هذا النظام عدداً من الخطوات تبدأ عند النطق بالكلمة أو التعبير من خلال مكبر للصوت ، ثم تحويل هذه الكلمة إلى إشارات كهربائية تحول بعدها باستخدام محول خاص لتحويل الإشارات التناظرية إلى الصورة الثنائية لتلك الكلمة ، ويتم تخزين تلك الصورة الثنائية للكلمة في الذاكرة .

عند إدخال تلك الكلمة مرة أخرى من خلال الحاسوب للتعرف عليها ، يقوم الحاسوب بربط الكلمة المدخلة مع الكلمة المخزنة كنموذج في الذاكرة ، وذلك باستخدام جمل شرطية بسيطة.

ولكنه ليس من الأمر المنطقي تخزين أصوات لجميع المتكلمين حتى يتمكن النظام من التعرف إلى كلمة من تلك الكلمات المخزنة لهؤلاء المتكلمين . لذا فإن الأمر الطبيعي في مثل هذه الحالات أن يدرب النظام على صوت متكلم محدد بتكرار هذا المتكلم لكلمة ما عدة مرات قبل أن يبدأ النظام في التعرف على ذلك الصوت أو تلك الكلمة .

مثل هذا النظام الخاص بالتعرف على الأصوات يسمى بالنظام المعتمد على المتكلم .

٢. أما النوع الثاني و هو تحليل الخصائص ، فهو نظام لا يعتمد على المتكلم ، ويقوم على أساس استخدام نظام تشفير التنبؤ الخطي ((Linear Predictive Coding (LPC) من خلال إيجاد الخصائص المتشابهة بين المدخلات المتوقعة ، والمدخلات الصوتية المحولة إلى صورة رقمية ثنائية . ثم تقارن هذه التشابهات مع مجموعة واسعة من المتكلمين . لذا لا يحتاج مثل هذا النظام إلى تدريب كل متكلم على كلمات محددة للمقارنة معها عند التعرف على تلك الكلمات .

وقد أثبتت الأنظمة غير المعتمدة على المتكلمين صعوبتها عند التطبيق لتنوع اللهجات ، واختلاف تصاريف الكلمات الكثيرة ، لذا كانت درجة الدقة لمثل هذه الأنظمة تنحصر بين (٩٠%) و (٩٥%) .(Jim Baumann, 2000)

٦-٤ طرق التعرف إلى الأصوات

تستخدم الأنظمة السابقة طرقاً مختلفة في التعرف إلى الأصوات ، فمنها ما يستخدم كلمات مفردة فقط في عملية التعرف إلى الأصوات ، ومنها ما يستخدم كلمات متتابعة ، ومنها ما يعتمد على الحديث المتصل ، ومنها ما يستطيع التعرف إلى مستخدم بعينه .

إنّ أغلب أنظمة التعرف إلى الأصوات تعتمد على استخدام الكلمات المفردة ، حيث يطلب من المتكلم إعطاء كلمة واحدة ثم يتوقف ثم يعطي كلمة أخرى ثم يتوقف وهكذا ... ، وهذا جيد في الحالات التي يعطي فيها المتكلم أوامر أو كلمات فقط . أما في حالة إعطاء المتكلم لحديث فيبدو هذا الأمر غير منطقي إذا تم بالصورة السابقة نفسها .

وفي حالة استخدام النظام لطريقة إعطاء سلسلة من الكلمات المتلاحقة ، فيجب الحرص هنا على أن تكون الكلمات دقيقة النطق ، ولا تدغم حروف نهايات الكلمات بأوائل الكلمات التي تليها .

أما في حالة استخدام الحديث المتصل في أنظمة التعرف إلى الأصوات ، فإنّ الأمر يبدو معقداً بدرجة كبيرة عند التعامل به ، و تبرز صعوبة بالغة في عملية التطبيق ، بسبب الصعوبة التي تواجه النظام في تحديد حدود الكلمات المنطوق بها ، واعتبار النظام مدققاً إملائياً للحديث فضلا عن التعامل مع ألفاظ لا تعد كلمات تصدر أثناء الحديث مثل (ايه ، ووو، ....الخ).

أما الأنظمة التي تستطيع التعرف إلى مستخدم بعينه من خلال حديثه ، فتدخل ضمن أنظمة الحماية والأمن الشديدة التعقيد .

٤-٧ كيف تعمل نظم التعرف إلى الأصوات

إنَّ معظم أنظمة التعرف إلى الأصوات الحديثة ترتكز على مفهوم مطابقة النماذج لأن هذا المفهوم يجمع بشكل جيد بين تقنيات الحوسبة الحديثة و الرغبة في الوصول إلى أعلى درجة دقة منشودة .

وبناء على هذا المفهوم كان لزاماً أن تُتّبع الخطوات التالية لتطبيق نظام التعرف إلى الأصوات حتى يتم الحصول على أفضل النتائج .

١. تسجيل الأصوات (Audio Recording) يتم تسجيل الأصوات بعدة طرق مختلفة حسب الإمكانات المتاحة في جهاز الحاسوب ، ولابد هنا من اتباع الشروط الصحيحة في عملية التسجيل من عدم وجود مصدر إزعاج أثناء عملية التسجيل ، والتلفظ بالكلمات واحدة بعد الأخرى مثلا مع ترك سكتات بعد الكلمات ، وطريقة حفظ البيانات في ملفات ذات محددات خاصة .

٢. التصفية القبلية (Pre-Filtering) : وتجرى بعدة طرق مختلفة بالاعتماد على الخصائص المتعلقة بنظام التعرف إلى الأصوات ، ومن هذه الطرق طريقة تسمى بنك المصفيات (Bank-of-Filters) والتي تستخدم عدد من مصفيات الأصوات لتجهيز العينة ، ويستخدم نظام تشفير التنبؤ الخطي لحساب الفروقات (الأخطاء) ، كما تستخدم النماذج المختلفة للتحليل .

٣. التأطير(Framing) : ويقصد بهذه العملية وضع عينات الصوت في نوافذ أو إطارات معينة ذات أحجام محددة ، وتهدف هذه العملية لتجهيز حدود العينة لعملية التحليل والتي تشمل إزالة حواف العينة مثل الأصوات الباهتة غير المرغوب فيها .

٤. عملية التصفية الإضافية (Additional Filtering) : وهي لا تستخدم دائماً ، وتهدف إلى تجهيز كل إطار من العينات تجهيزاً نهائياً قبل عملية الربط و المقارنة ، وتحتوي على عملية تسمى التطبيع (Normalization) .

٥. عملية الربط والمقارنة (Comparison and Matching) : وهناك تقنيات كثيرة متنوعة لتنفيذ هذه العملية ، وتهدف لمقارنة وربط عينة أو إطار معين بعينة معلومة مخزنة في جهاز الحاسوب .

ومن التقنيات المستخدمة في هذه العملية ، نموذج ماركوف المخفي (Hidden Markov Model (HMM)) ، تحليل الترددات (Frequency Analysis) ، التحليل التفاضلي (Differential Analysis) ، تقنيات الاختصار في الجبر الخطي .(Linear Algebra Tech

(Shortcuts/ ،التشويه الطيفي (Spectral Distortion) ، وغيرها الكثير. وكل هذه الطرق تهدف لإعطاء الاحتمالية والدقة في التطابق . (Stephen Cook , 2003)

الفصل الخامس

الخوارزمية الجديدة لتجزئة ملف الموجة والتعرف إلى تطابق الصوت

(New Algorithm for Wave File Segmentation and Voice Sequence Matching Recognition)

٥-١ لغة البيسك المرئية (Visual Basic) و برنامج الدراسة

لقد تم استخدام لغة بيسك المرئية (Visual Basic) في عمل برنامج الدراسة و إعدادها، لما لهذه اللغة من أهمية في عصر ثورة المعلومات ، و لما لها من قدرة عالية في توفير الوقت والجهد مقارنة بلغات البرمجة الأخرى ، و حيث إنه يمكن استخدام الواجهات الرسومية (Graphical User Interface) التي تزود بها هذه اللغة المستخدمين لها ، فتسهل عليهم استخدام الرسومات المرئية مع التطبيقات المختلفة بطريقة تفاعلية سهلة وشيقة .

و قد تطورت لغة بيسك المرئية في السنوات الأخيرة فأصبحت تقدم للمبرمجين أفضل الأدوات المساعدة والمفيدة التي تؤدي إلى تحسين التطبيقات ، وتسهل عمليات البرمجة ، وتوفر الوقت ضمن بيئة تطوير متكاملة ((Integrated Development Environment (IDE)) .

و بالإضافة إلى ما سبق ذكره من مميزات لهذه اللغة ، وجدت أن التعامل مع ملفات الأصوات ببياناتها التي لا تعدو أن تكون إما (٠) أو (١) ، سهلٌ جدّاً باستخدام لغة بيسك المرئية ، والتي تتعامل مع تلك الملفات التي تتركب من تلك الثنائيات مباشرة دون الحاجة إلى تحويلها إلى شكل آخر من أشكال البيانات . و مما زاد في سهولة التعامل أيضاً ميزة في لغة بيسك المرئية وهي إمكانية التعامل مع محتويات الملفات بشكل تتابعيّ ، الأمر الذي نحتاجه في التعامل مع ملفات الصوت .

ولقد كان لإمكانية تقسيم العمل في البرامج المكتوبة بلغة بيسك المرئية إلى وحدات (Modules) و نماذج (Forms) الأثر الكبير في تسهيل العمل و إتمامه بشكل ميسر وسريع .

و هذه الأسباب مجتمعة كانت وراء اختيار لغة بيسك المرئية لتكون لغة برمجة النظام قيد الدراسة الذي احتوى على تطبيق لخوارزمية جديدة لتجزئة أحد أنواع ملفات الصوت وهو ملف الموجة ، ثم إدخال مخرجات هذا البرنامج في نظام آخر للتعرف على الأصوات .

٥-٢ الخوارزميات المستخدمة في البرنامج

تبحث هذه الدراسة في خوارزمية جديدة لتجزئة أحد أنواع ملفات الصوت وهي الملفات ذات الامتداد (wav) أي ما يسمى بملفات الموجة ، وقد اعتمد في إخراج هذا البرنامج وتنفيذه على عدد من الخوارزميات المختلفة مثل خوارزمية التطبيع ، وخوارزمية قراءة ملف الصوت ، وخوارزمية الرسم البياني للموجات ، وهذه الخوارزميات ساعدت على تنفيذ الخوارزمية الرئيسة وهي خوارزمية التجزئة .

٥-٣ بيئة فحص البرنامج (ملاحظات الفحص)

وعند تطبيق هذا البرنامج لابد من مراعاة بعض الشروط القبلية التي يجب أن تتوافر في ملف الصوت قيد التجربة ، وهذه الشروط هي :

١. أن يتم تسجيل الصوت باستخدام جهاز الحاسوب دون ضوضاء تؤثر على البيانات المسجلة ، حيث يتم تسجيل الصوت في بيئة هادئة تماماً بعيدة عن أية ازعاجات صوتية حتى لا تؤثر على الكلمات المسجلة في الملف .
٢. أن تسجل الكلمات بوضوح تام ، و أن تنطق الكلمات كلمة كلمة ، وذلك بالتوقف بين تلك الكلمات ، أي وضع سكتات بينها ، أي أن تكون الكلمات مفصولة عن بعضها البعض دون أن تدغم نهاية كلمة ما ببداية كلمة أخرى بعدها .
٣. أن يحفظ الملف - بعد الانتهاء من التسجيل - بنوع معين وهو (PCM) الذي تم ذكره سابقاً ،وهو التنسيق الخاص بملفات الموجة التي تدور الدراسة حوله .
٤. أن تكون محددات الملف كما يلي : (٨ كيلوهيرتز) من حيث معدل أخذ العينات ، وهو المعدل المناسب لصوت الإنسان ، و(١٦ بت) من حيث عدد الثنائيات في العينة ، و أن يكون عدد القنوات أحادي أي أن يكون التنسيق (Mono) .

وبعد إجراء عملية التسجيل السابقة ، يكون ملف الموجة قد أعد لإجراء عملية تجزئته ، أي تطبيق البرنامج على هذا الملف .

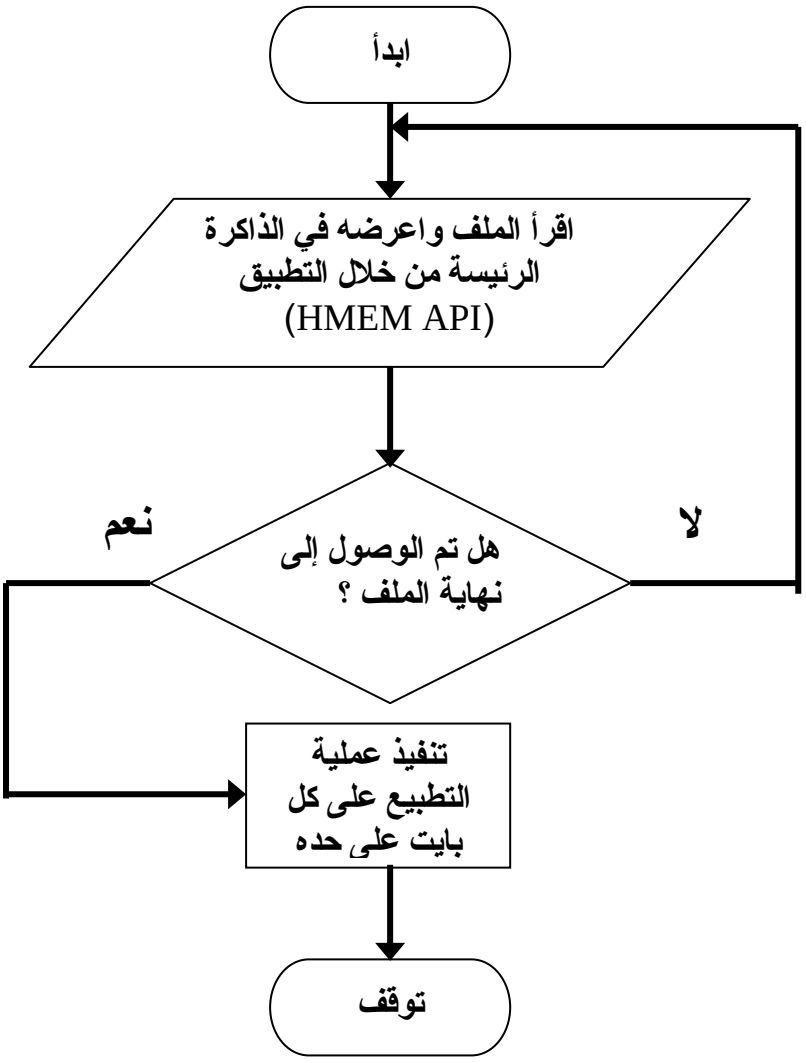
يتم فتح ملف الموجة قيد الدراسة وذلك باختياره من مكان حفظه ، ويتم تحويل بيانات الملف الصوتية والمتوفرة على شكل سيل من الثنائيات المكونة من (٠) و(١) ، لتصبح على شكل تمثيل بياني مرسوم وذلك من خلال تطبيق الخوارزمية التالية على تلك البيانات ، وهنا ستصبح دراسة هذا التمثيل أيسر وأسهل :

١. يتم قراءة ملف الموجة باستخدام التطبيق (HMEM API) وهو تطبيق من مكتبة التسجيل الصوتي في لغة بيسك المرئية يستخدم لعرض ملف الموجة على الذاكرة الرئيسة

للحاسوب (RAM) ، من أجل قراءة هذا الملف بصورة أسهل وأسرع مما لو استخدم تطبيق فتح الملف العادي (Open File Function) .

٢. يجري دوران داخل البرنامج ينتهي عند آخر بايت من بايتات ملف الموجة ، مبتدئاً بالقيمة (٤٤) وهي قيمة البايت اللاحق لبايتات الترويسة ، وذلك للتأكد من تمثيل بايتات البيانات فقط ولا يشمل بايتات رأس الملف (Header) ، ويتم الحصول هنا على القيم بشكل بايتات .

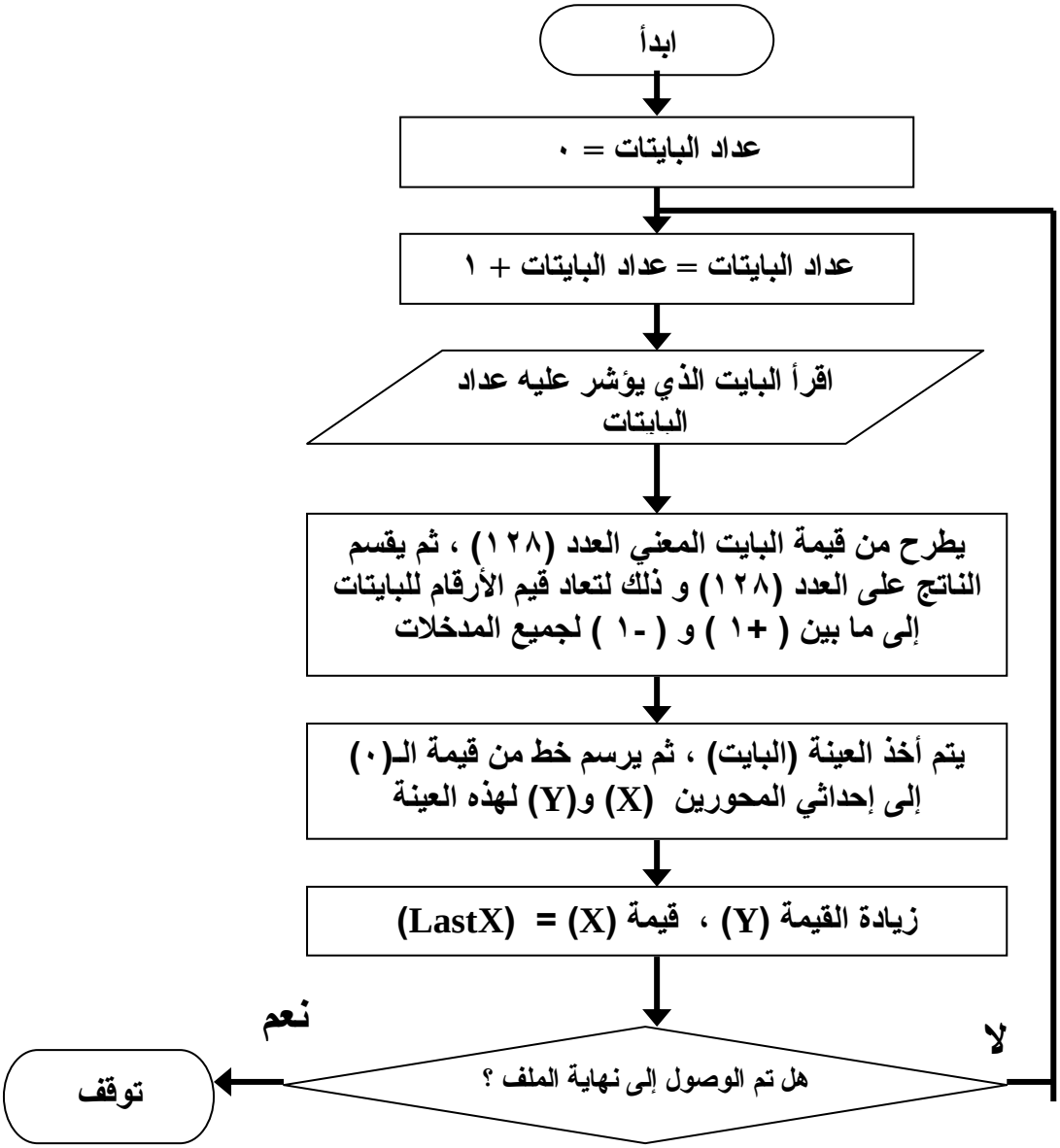
٣. يتم تنفيذ عملية التطبيع (Normalization) على كل بايت على حده .



شكل رقم (٥-١) يبين خوارزمية قراءة ملف قيد التجزئة .

أما عملية التطبيع المذكورة سابقا فهي عملية مهمة تهدف إلى وضع حدود للقيم التي تمثل البايتات من (–١) وحتى (+١) و تتم من خلال الخوارزمية التالية :

١. يطرح من قيمة البايت المعني العدد (١٢٨) ، ثم يقسم الناتج على العدد (١٢٨) وذلك لتعاد قيم الأرقام للبايتات إلى ما بين (+١) و (-١) لجميع المدخلات .
٢. يتم أخذ العينة ، ثم يرسم خط من قيمة الـ(صفر) إلى إحداثي المحورين (X) و (Y) لهذه العينة.
٣. يتم زيادة القيمة (Y) و الحصول على عينة أخرى و توضع قيمة (X) لتصبح مساوية للقيمة (LastX) .
٤. تجرى عملية الدوران في البرنامج حتى إتمام قراءة جميع العينات .



شكل رقم (٥-٢) يبين خوارزمية التطبيع

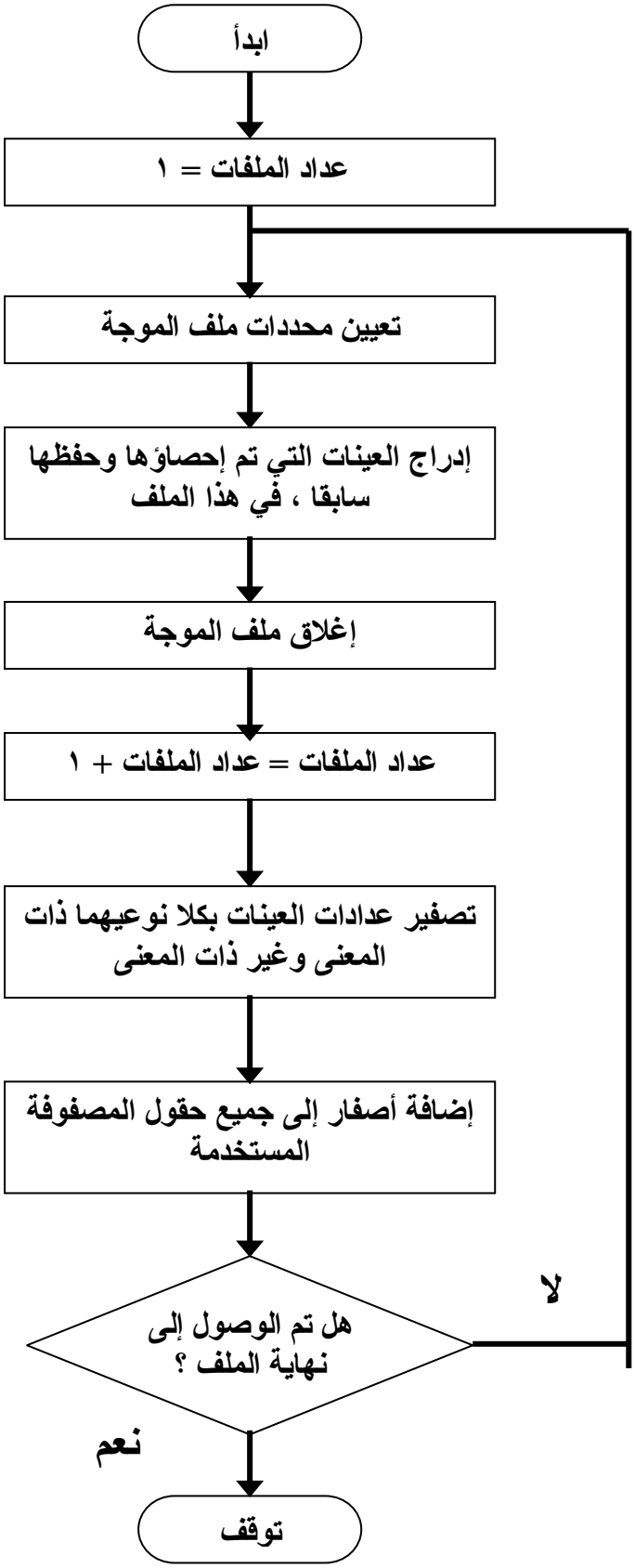
إنّ البايت يمكن أن يخزن القيم من ($٢^{٠}$ =١) وحتى ($٢^{٨}$=٢٥٦) ، وكذلك يمكن أن تمثل العينات بقيم سالبة وهذا غير مطبق في عملية التطبيع ، لأنّ الاختلافات في القيم السالبة والموجبة للعينات يمكن أن يؤدي إلى :

١. تمثيل غير صحيح لملف الموجة .
٢. زمن أطول من المطلوب للحسابات التي تجرى على العينات .

ويمكن تجنب هذه المشاكل إذا طبقت عملية التطبيع بطرح الرقم (١٢٨) من قيمة العينة ثم تقسيم الناتج على (١٢٨) حتى يصبح الناتج الجديد عددا بين العدد (-١) و العدد (+١) .

وبعد تنفيذ خوارزمية التطبيع ، و بعد أن تكون المتغيرات التالية قد أدخلت من قبل المستخدم وهي :قيمة الدقة (Accuracy_Value) ، حجم النافذة (Window_size) ، مدى قطع الموجة (Cut_Wave_Range) ، خطوة القطع (Cut_ Step) ، فيتم إحصاء عدد العينات غير ذات المعنى والمتراصة بعضها وراء بعض بحيث لا يفصل بينها بأي عينة ذات معنى . فإذا كان عدد العينات غير ذات المعنى يزيد على قيمة محددة تسمى حجم النافذة فهذا يعني وجود فترة صمت ، وهنا يتم تنفيذ الخوارزمية التالية :

١. إنشاء ملف الموجة .
٢. تعيين محددات ملف الموجة.
٣. إدراج العينات التي تم إحصاؤها وحفظها سابقاً ، في هذا الملف.
٤. إغلاق ملف الموجة .
٥. زيادة عدد الملفات بواحد .
٦. تصفير عدادات العينات بكلا نوعيها ذات المعنى وغير ذات المعنى .
٧. إضافة أصفار إلى جميع حقول المصفوفة المستخدمة .
٨. الاستمرار بالعمل وإعادة الدوران لجميع الوظائف السابقة حتى نهاية الملف الرئيسي .



شكل رقم (٥-٣) يبين خوارزمية تجزئة الملفات

و بتنفيذ الخطوات السابقة فإن ذلك سينتج ما يلي :

١. سيتكون ملف جديد بالمعلومات المضافة لوصف محتويات الملف .
٢. كل ملف جديد سيتكون له اسم فريد يتكون من (DIV) متبوعا برقم (٠٠٠) ، وامتداد (.WAV)
٣. إن عدد الملفات المتكونة مساوية لعدد مرات إيجاد العينات غير ذات المعنى (طول حجم النافذة) .

و كان الهدف من هذه الدراسة بعد تجزئة ملفات الصوت ، التعرف إلى تلك الأصوات التي تم الحصول عليها من خلال الملفات الجزئية التي أُنتجت بعد إجراء الخوارزمية الجديدة .

وتمت عملية التعرف إلى تلك الأصوات بعقد مقارنات بين الملف المراد معرفة صوت المتكلم فيه ، وعدد من الملفات المجزأة المخزنة سابقاً في مجلد معين ، ثم تنفيذ خطوات الخوارزمية التالية :

١. يتم قراءة الملف المراد معرفة صوت المتكلم فيه .
٢. يقسم الملف السابق إلى عدد من الأجزاء المحتوية على العينات ، حسب رغبة المستخدم
٣. يؤخذ معدل العينات لكل جزء من الأجزاء السابقة .
٤. يكرر ما تم في الخطوة السابقة حتى الانتهاء من جميع أجزاء الملف .
٥. يحسب المعدل الكلي للعينات في الملف من خلال معدلات الأجزاء الخاصة بالملف .
٦. تؤخذ الملفات المخزنة تباعاً ، فتجزأ إلى أجزاء كما فعل بالملف المراد معرفة صوت المتكلم فيه .
٧. يستخرج المعدل لكل معدلات الأجزاء في كل ملف من الملفات المخزنة .
٨. يقسم معدل الأجزاء للملف المراد معرفة الصوت فيه على معدل الأجزاء لكل ملف من الملفات المخزنة حسب الشروط التالية :
   أ- في حالة كون معدل الأجزاء للملف المراد معرفة الصوت فيه أصغر من المعدل للملف النموذج .
   ب- أما في حالة كون معدل الأجزاء للملف المخزن أصغر من معدل الأجزاء للملف المراد معرفة الصوت فيه ، فتعكس عملية القسمة .
٩. تؤخذ القيمة المطلقة لكل القيم بعد استخراج النتائج ، وتحول إلى نسبة مئوية .
١٠. تقرأ قيم النسب المئوية لكل مقارنة من المقارنات التي تمت بين الملف المراد معرفة الصوت فيه ، والملفات المخزنة في قاعدة الكلمات .
١١. كلما كانت نسبة المقارنة أكبر تكون درجة تشابه الصوت أكبر إلى أن تصل إلى ١٠٠ % ، وهنا تصبح المقارنة بين الملف ونفسه .

وقد استخدم النموذج الرياضي للنوع الأول وهو مطابقة النماذج و الذي يعتمد على حساب المعدل الإحصائيّ لعدد من العينات لكلمة ما ، ومقارنته بالمعدل الإحصائي لنفس الكلمة المخزنة كنموذج في جهاز الحاسوب .

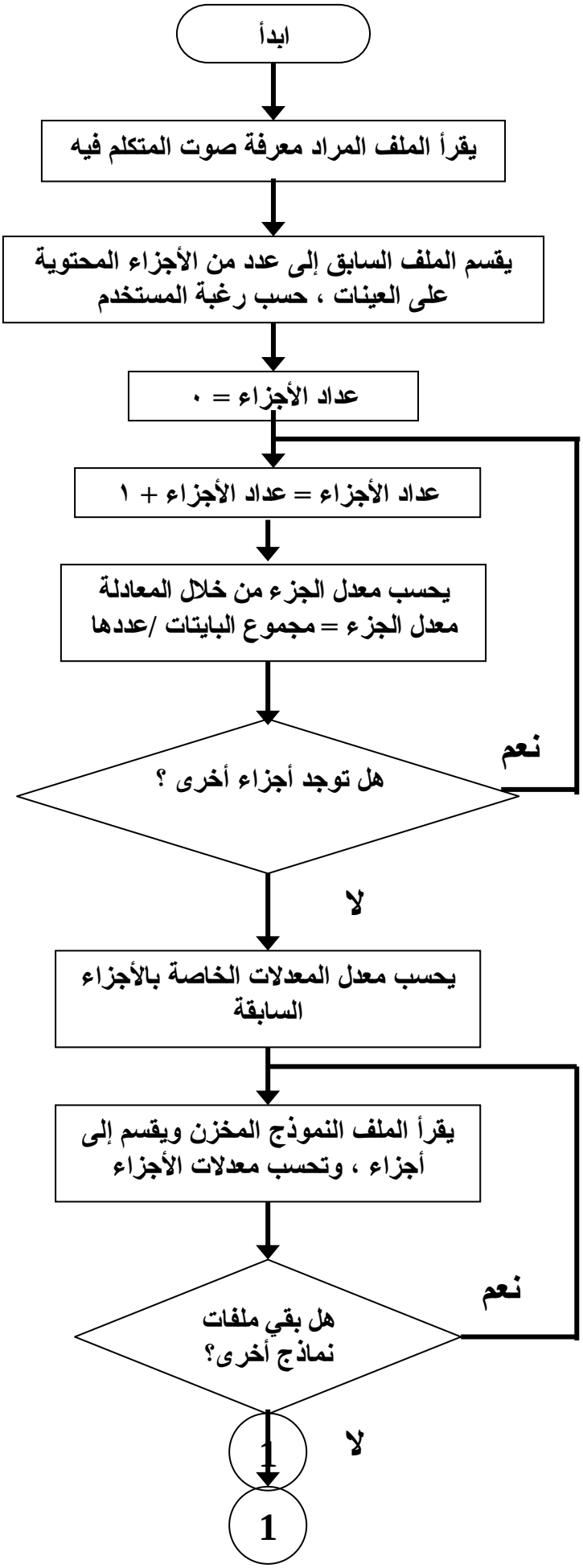
٤-٥ نتائج مراحل التنفيذ (المتوسطات) للبرنامج

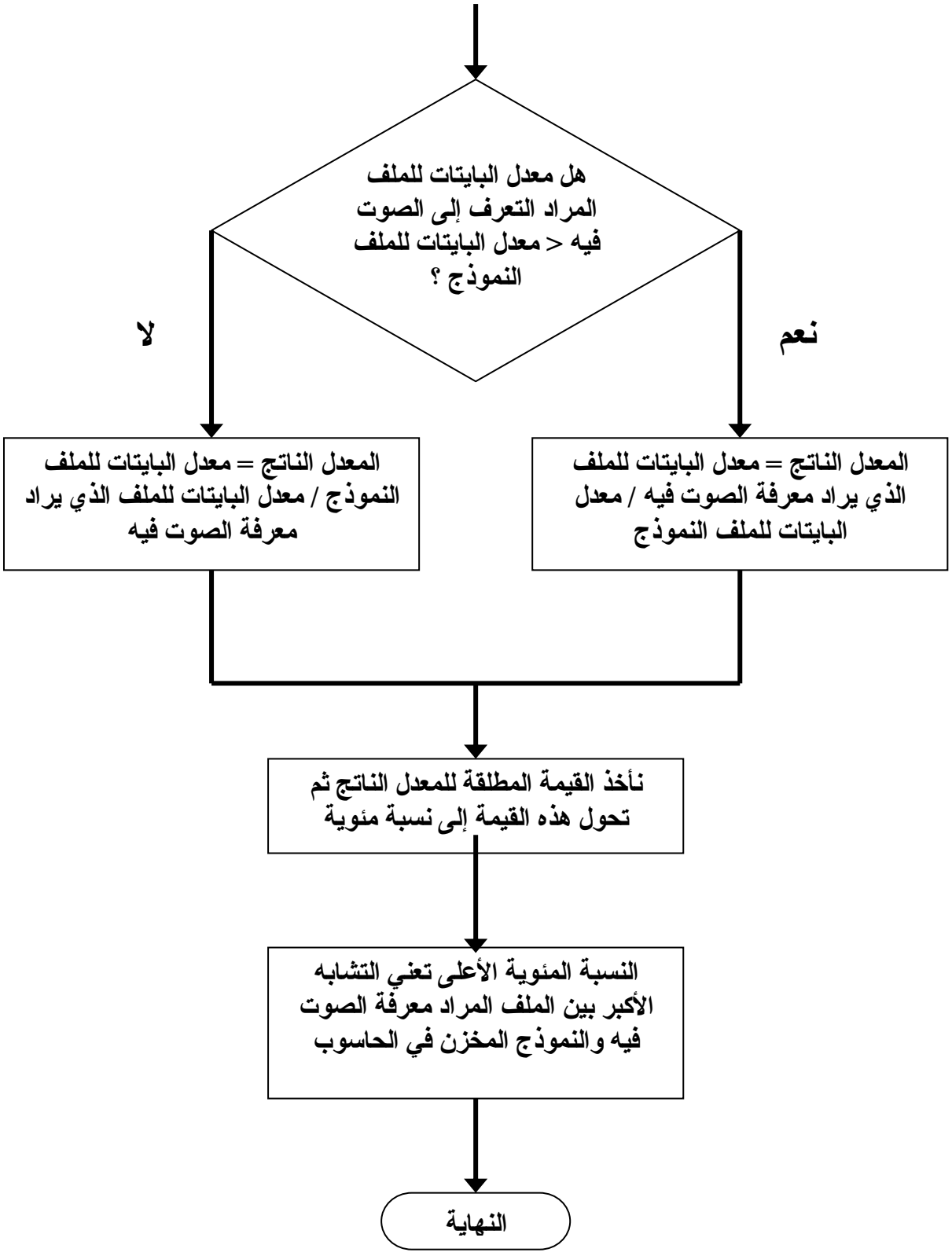
ولاستخراج المعدل الإحصائي الذي يبين مدى التوافق بين الملف المراد معرفة الصوت فيه والملف النموذج المراد المقارنة به نفذت الخطوات التالية :

١. يتم قراءة الملف الذي يحتوي على العينات في الذاكرة .
٢. يقسم الملف الى عدد من الأجزاء كل منها يحتوي على عدد متساو من العينات.
٣. تجمع قيم العينات في كل جزء من الاجزاء السابقة ثم يقسم المجموع على عدد العينات في ذلك الجزء لاستخراج المعدل الإحصائي للبايتات في كل جزء .
٤. تجمع المعدلات الإحصائية الناتجة من كل جزء من الأجزاء ثم يقسم المجموع على عدد الأجزاء التي قُسِّم الملف اليها .وتستخدم المعادلة التالية للحصول على ذلك المعدل :

$$\text{المعدل الإحصائي الناتج} = \frac{\text{معدل البايتات للملف الذي يراد معرفة الصوت فيه}}{\text{معدل البايتات للملف النموذج}}$$

٥. يُجرى العمل السابق على الملف النموذج المخزن في ذاكرة الجهاز الثانوية ، والملف المراد معرفة صوت المتكلم فيه.
٦. بعد الإنتهاء من العمل السابق ، يقارن ما تم التوصل اليه من معدلات احصائية لكلا الملفين ، وذلك من خلال قسمة المعدل الإحصائي الأصغر على المعدل الإحصائي الأكبر ثم ضرب الناتج في ١٠٠% لتنتج نسبة المقارنة بين المعدلين الإحصائيين .
٧. كلما كانت نسبة المقارنة أكبر كانت درجة تشابه الصوت أكبر .


ابدأ
يقرأ الملف المراد معرفة صوت المتكلم فيه
يقسم الملف السابق إلى عدد من الأجزاء المحتوية على العينات ، حسب رغبة المستخدم
عداد الأجزاء = ٠
عداد الأجزاء = عداد الأجزاء + ١
يحسب معدل الجزء من خلال المعادلة
معدل الجزء = مجموع البايتات /عددها
هل توجد أجزاء أخرى ؟
نعم
لا
يحسب معدل المعدلات الخاصة بالأجزاء السابقة
يقرأ الملف النموذج المخزن ويقسم إلى أجزاء ، وتحسب معدلات الأجزاء
هل بقي ملفات نماذج أخرى؟
نعم
لا
1
1



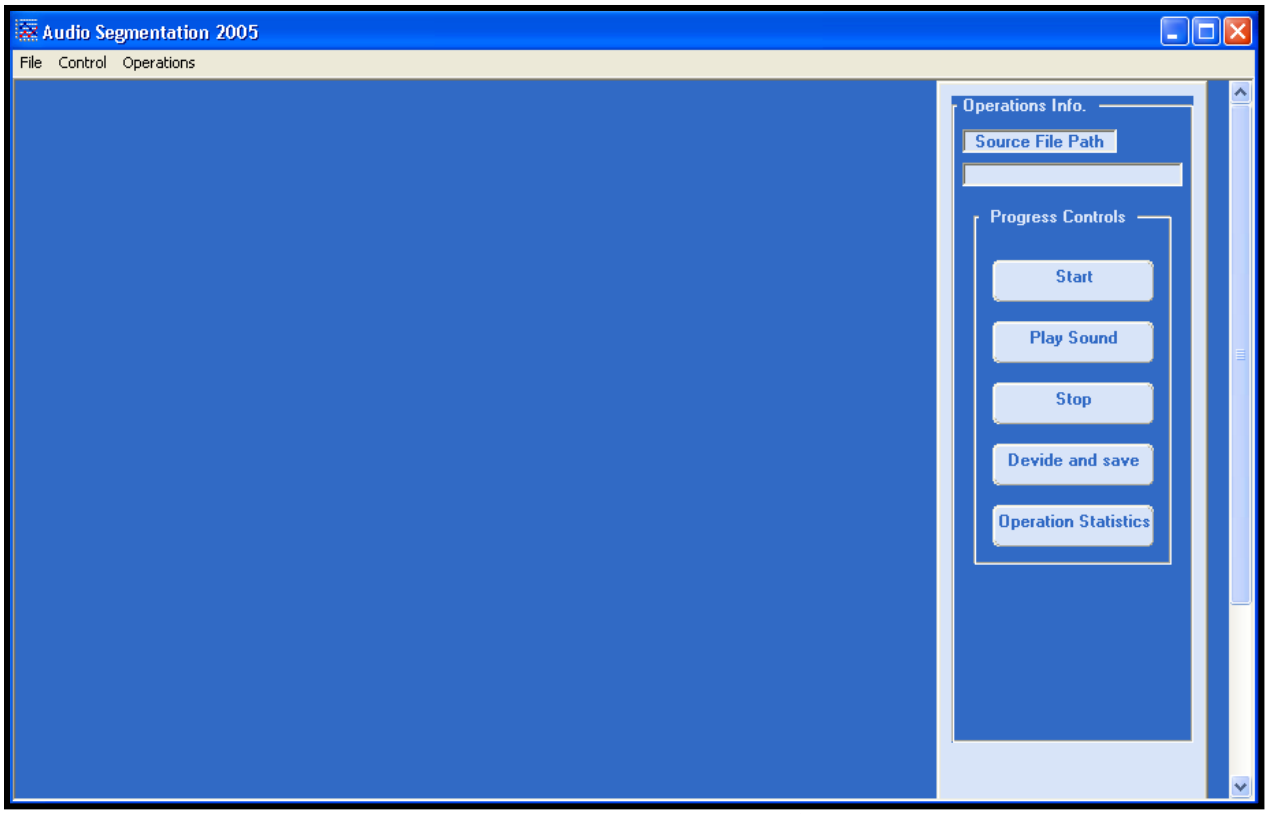
شكل رقم (٥-٤) يبين خوارزمية التعرف إلى الأصوات

٥-٥ الواجهات الرسومية لبرنامج تجزئة الملفات

بعد أن يتم تسجيل الملف الصوتي وتراعى في تسجيله الشروط التي تم ذكرها فيما سبق يُبدأ بتنفيذ البرنامج ، حيث تظهر الواجهة التالية التي تبين اسم البرنامج "تجزئة ٢٠٠٥" "SEGMENTATION 2005" ، وفي أسفل الواجهة يظهر شريط ينبئ بمقدار تحميل البرنامج على الذاكرة الرئيسة .

شكل رقم (٥-٥) يبين واجهة البدء لبرنامج التجزئة

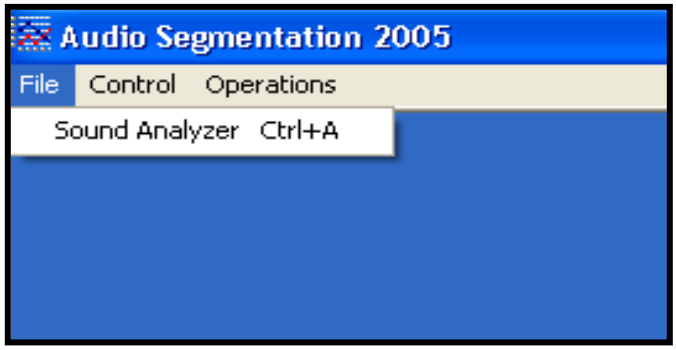
فإذا تم التحميل ظهرت الواجهة الرسومية التالية التي تحتوي ثلاثة قوائم رئيسة هي : ملف (File) ، تحكم (Control) ، عمليات (Operation) . وكل قائمة من هذه القوائم تحتوي مجموعة من الأوامر .



شكل رقم (٥-٦) يبين واجهة البرنامج الرئيسة

أما الأوامر والتعليمات التي تحتويها تلك القوائم فهي:

القائمة : ملف(File) تحتوي الأمر : محلل الصوت (Sound Analyzer) ، كما يمكن الوصول إلى هذا الأمر عن طريق استخدام مفتاح التحكم (Ctrl) بالإضافة إلى المفتاح (A) .



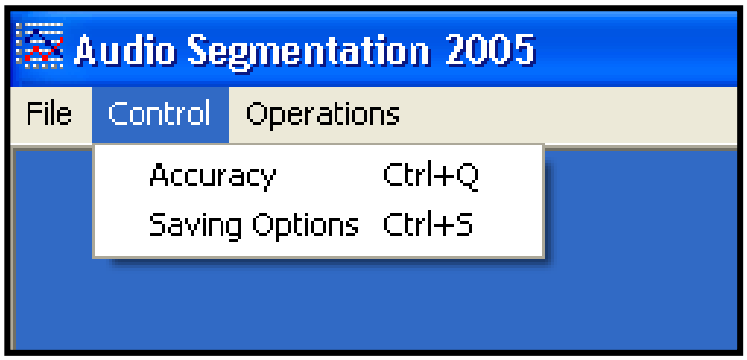
شكل رقم (٥-٧) يبين قائمة ملف ومحتوياتها

القائمة : تحكم (Control) وتحتوي مجموعة الأوامر التالية :

١. مقدار الدقة (Accuracy) ويمكن اختيار هذا الأمر من خلال الضغط على المفتاح تحكم (Ctrl) بالإضافة إلى مفتاح (Q) . وعند اختيار هذا الأمر تظهر واجهة تحمل ثلاثة خيارات للدقة ، وهي تمثل ثلاثة قيم للدقة وهي (١٦ ، ٣٢ ، ٦٤) :

شكل (٥-٨) يبين درجة قيمة الدقة المرغوب بها

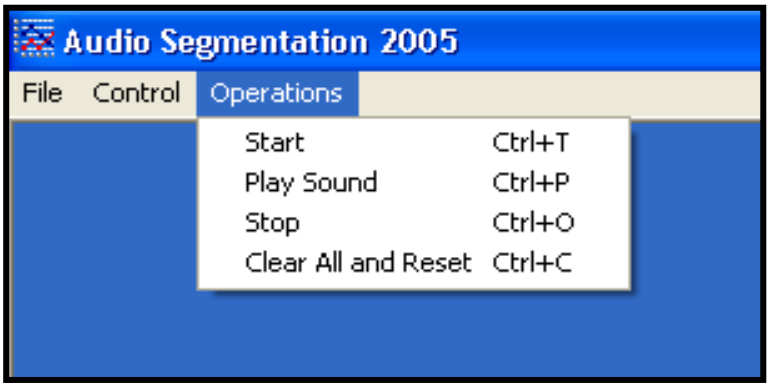
٢. حفظ الخيارات (Saving Options) ويمكن اختيار هذا الأمر أيضاً من خلال الضغط على المفتاح (Ctrl) بالإضافة إلى مفتاح (S).



شكل رقم (٥-٩) يبين قائمة التحكم

القائمة : عمليات (Operations) وتحتوي مجموعة الأوامر التالية :

١. ابدأ (Start) : ويمكن اختيار هذا الأمر أيضاً من خلال الضغط على المفتاح (Ctrl) بالإضافة إلى مفتاح (T).
٢. شغل الصوت (Play Sound) : ويمكن اختيار هذا الأمر أيضا من خلال الضغط على المفتاح (Ctrl) بالإضافة إلى مفتاح (P).
٣. توقف (Stop) : ويمكن اختيار هذا الأمر أيضاً من خلال الضغط على المفتاح (Ctrl) بالإضافة إلى مفتاح (O).
٤. مسح الكل والبدء ثانية (Clear All And Reset) : ويمكن اختيار هذا الأمر أيضاً من خلال الضغط على المفتاح (Ctrl) بالإضافة إلى مفتاح (C).



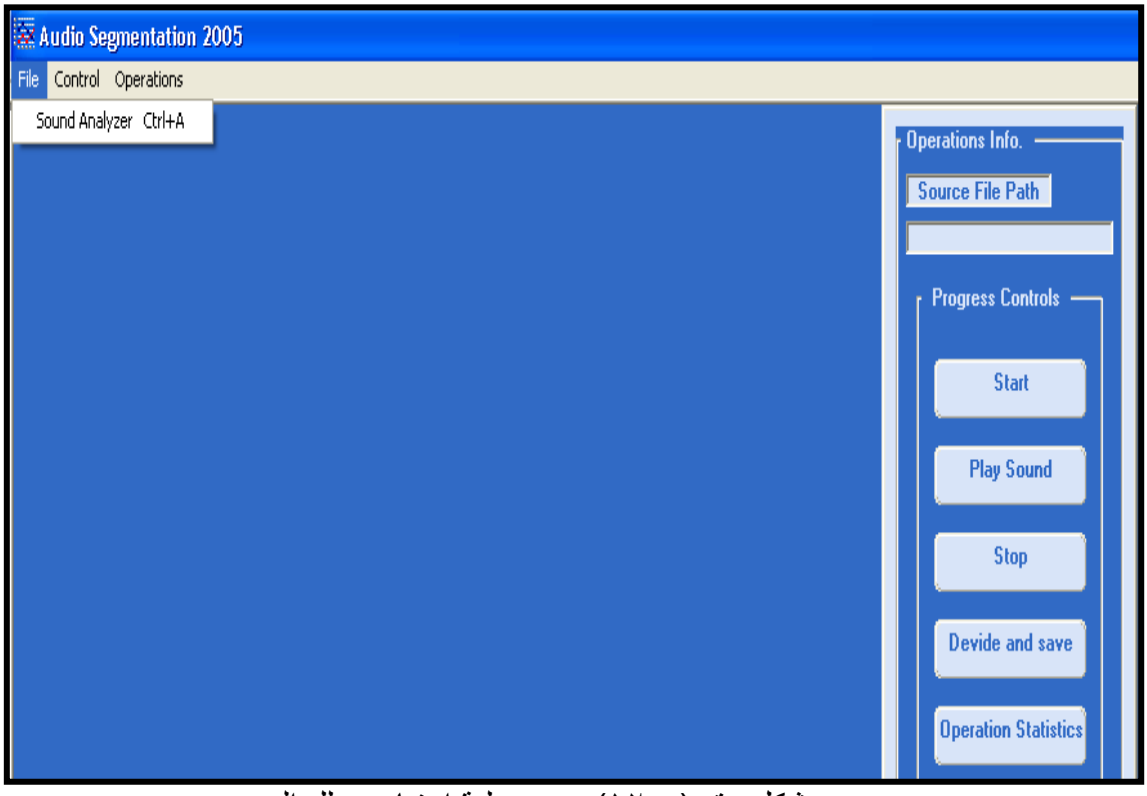
شكل رقم (٥-١٠) يبين قائمة العمليات

وتوجد قائمة جانبية تسمى تحكمات العملية (Progress Controls) لتسهيل العمل تحتوي عدداً من المفاتيح التي يحمل عدد منها نفس العمليات المدرجة في لائحة عمليات السابقة، مثل : ابدأ ، شغّل الصوت ، توقف .

وتحتوي هذه القائمة الجانبية على مفتاح التجزئة والحفظ (Divide and Save) للدخول إلى عملية التجزئة وهو المفتاح الأساسي لتطبيق الخوارزمية الجديدة ، ومفتاح إحصائيات العمليات (Operation Statistics) لإعطاء تقرير نهائي عن معلومات حول عملية التجزئة .

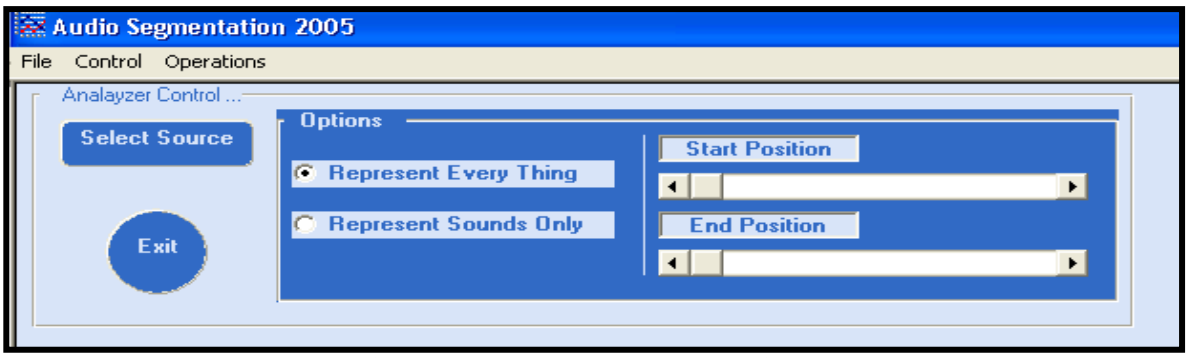
شكل رقم (٥-١١) يبين القائمة الجانبية تحكمات العملية

عند بداية تشغيل البرنامج نقوم باختيار لائحة ملف (File) ثم نختار محلل الصوت (Sound Analyzer) الذي يتم من خلاله اختيار الملف المراد تجزئته .



شكل رقم (٥-١٢) يبين عملية اختيار محلل الصوت

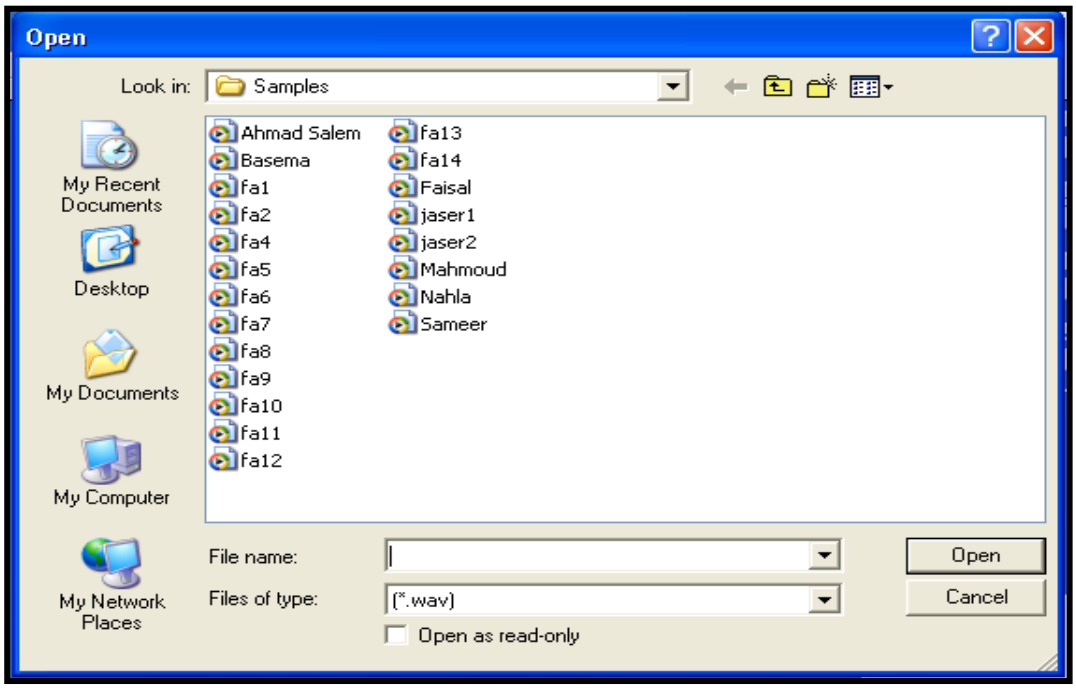
لتظهر الواجهة الرسومية التالية تحت اسم تحكمات المحلل (Analyzer Control) :



شكل رقم (٥-١٣) يبين تحكمات محلل الصوت

وتحتوي هذه الواجهة على الخيارات التالية :

١. اختيار المصدر (Source Select) : وعند الضغط على هذا المفتاح تظهر شاشة اختيار الملفات ومنها يتم اختيار الملف المراد تجزئته .



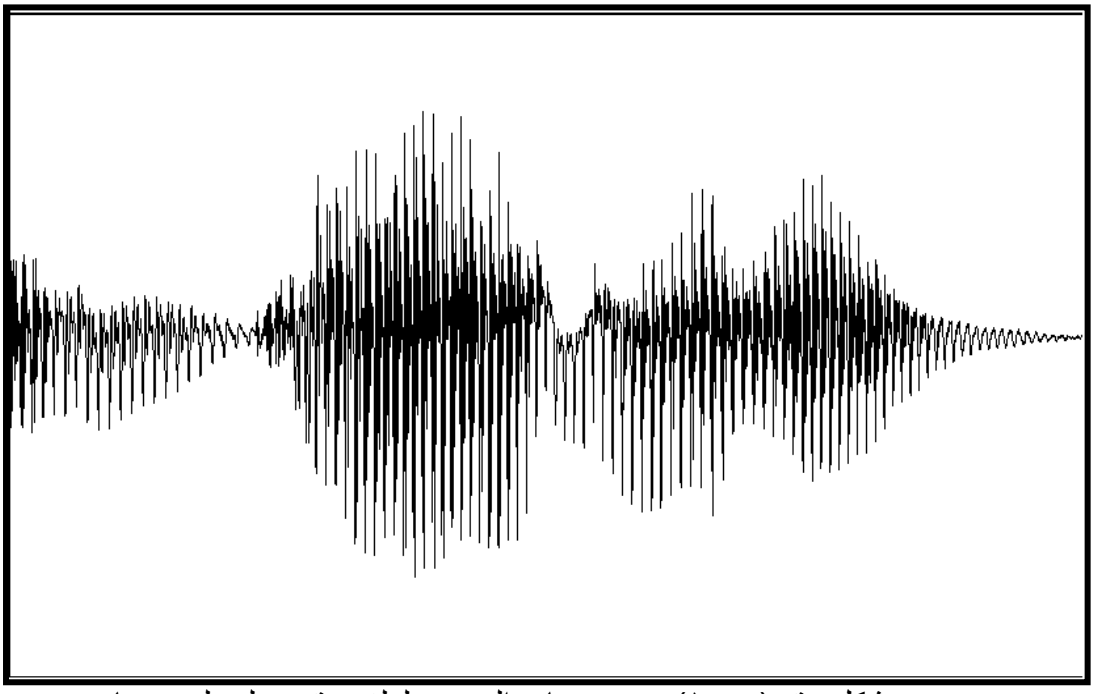
شكل رقم (٥-١٤) يبين صندوق الحوار لفتح الملف المراد تجزئته

٢. خيارات (Options): وفي هذا الخيار يمكن أن يتم رسم جميع محتويات ملف الصوت (Represent Everything) أو رسم الصوت فقط (Represent Sound Only) . ويمكن أيضاً اختيار نقط بداية المقطع المرغوب فيه (Start Position) ونهايته (End Position) .

٣. مفتاح الخروج من البرنامج (Exit) .

وعند اختيار أحد الملفات من شاشة اختيار المصدر ، نقوم بالضغط على مفتاح ابدأ من القائمة الجانبية (تحكمات العملية) ، أو باختيار قائمة خيارات ثم اختيار الأمر ابدأ ، أو بالضغط على مفتاح تحكم والمفتاح (T) سوية .

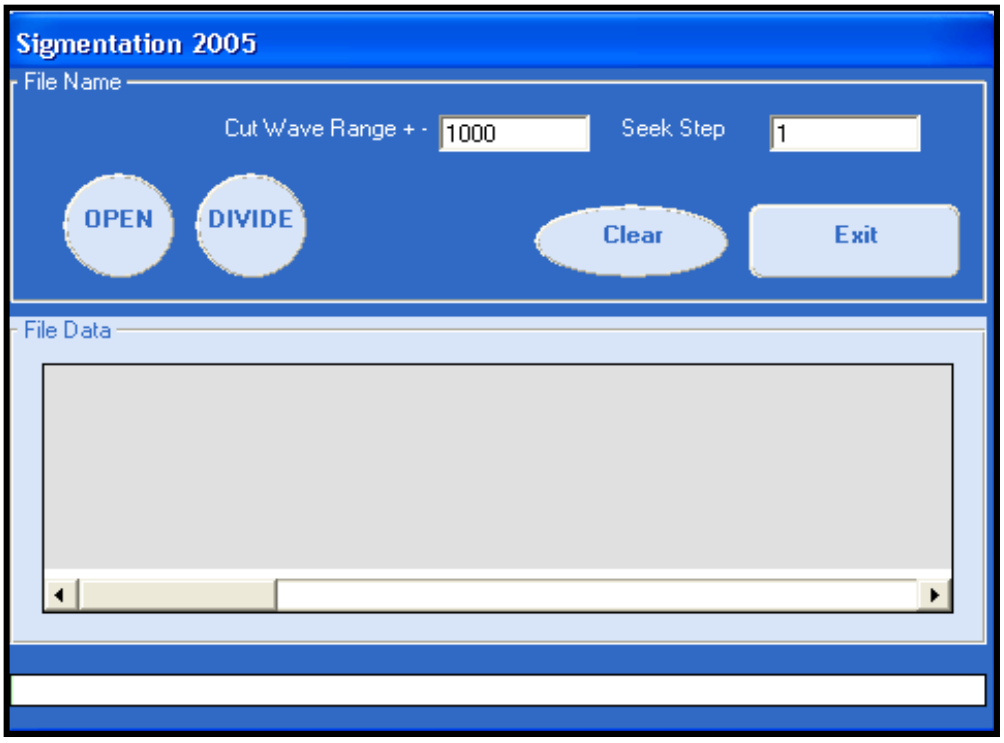
يتم تحويل محتوى الملف الذي تم اختياره إلى رسم بياني يظهرعلى الشاشة كما يلي :



شكل رقم (٥-١٥) يبين موجات الصوت لملف وقد حولت لرسم بياني

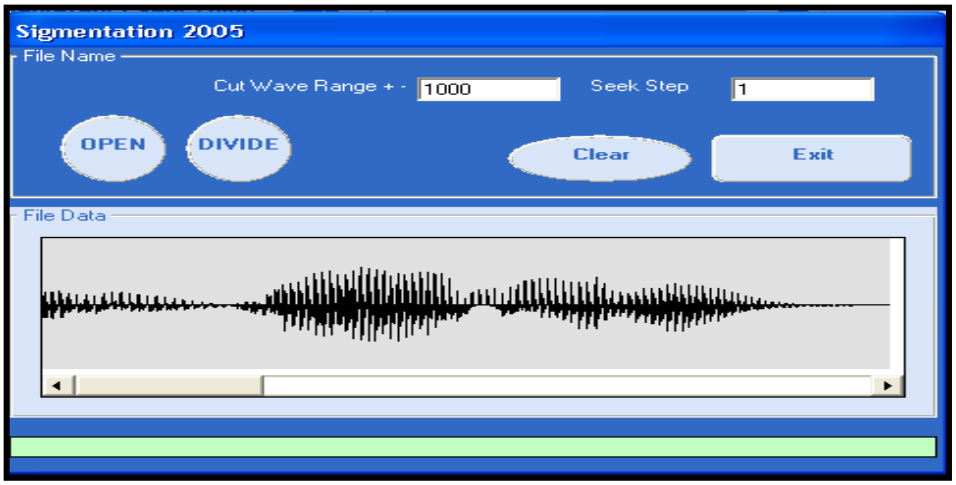
و إذا أراد المستخدم سماع الصوت المسجل في هذا الملف فإنه يقوم بالضغط على مفتاح تشغيل الصوت من القائمة الجانبية ، أو أنه يقوم باختيار نفس الأمر من قائمة عمليات ، أو بالضغط على مفتاح تحكم والمفتاح (P) سوية . أما إذا أراد المستخدم توقيف تشغيل الصوت فبإمكانه فعل ذلك بالضغط على مفتاح توقيف ، أو اختيار نفس الأمر من قائمة عمليات ، أو بالضغط على مفتاح تحكم والمفتاح (O) سوية .

وعند تمثيل بيانات الملف الصوتي على شكل موجات برسم بياني ، يمكن البدء بعملية التجزئة من خلال النقر على مفتاح التجزئة والحفظ (Divide and Save) لتظهر الواجهة التالية :



شكل رقم (٥-١٦) يبين الواجهة الرسومية لتجزئة الملف

وتحتوي هذه الواجهة على مفتاح فتح (Open) الذي يؤدي لفتح الملف استعداداً لتجزئته، كما في الواجهة التالية :

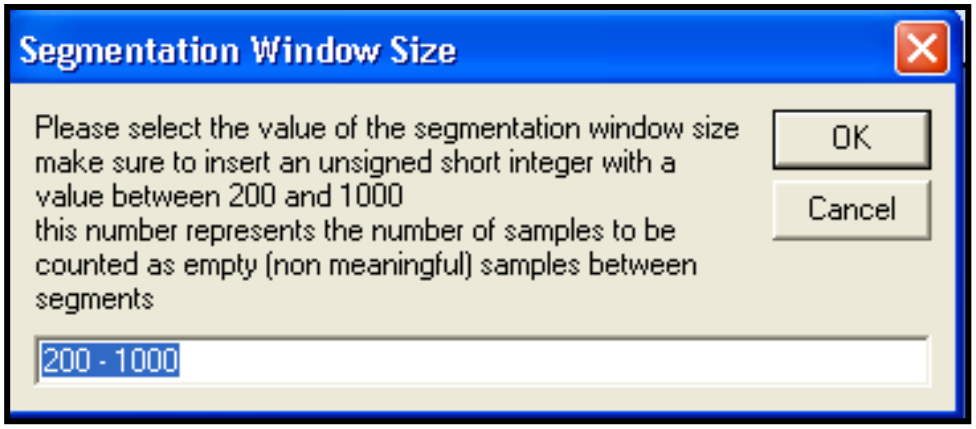


شكل رقم (٥-١٧) يبين الرسم البياني للصوت في ملف

وتحتوي أيضاً على مفتاح تجزئة (Divide) ، والذي يجب القيام بعملين قبل الضغط على هذا المفتاح ، وهما :

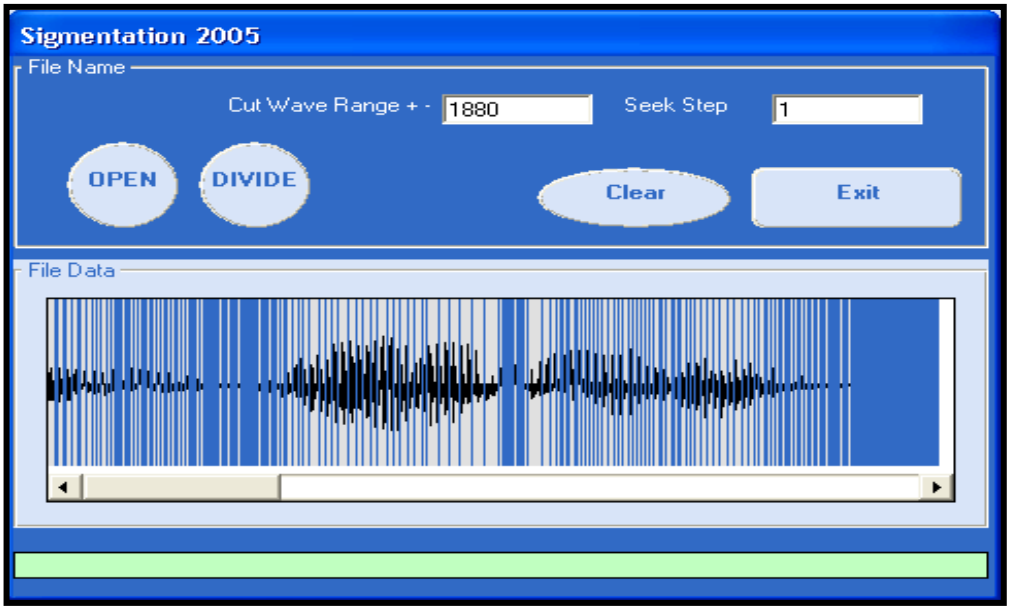
تحديد مدى قطع الموجة (Cut Wave Range) حيث نحدد هنا الرقم الذي يحدد فيما إذا كان الصوت مرتفعاً فيعطى رقماً صغيراً ، وإذا كان الصوت منخفضاً كان الرقم كبيراً .

أما العمل الآخر فهو تحديد عدد العينات التي يتم البحث عنها وتكون عينات غير ذات معنى وتكون متراصة وراء بعضها وراء بعض لا يفصل بينها بأي عينة ذات معنى . ويتأتى ذلك من خلال الضغط على مفتاحي التحكم مع (S) .



شكل رقم (٥-١٨) يبين صندوق الحوار لاختيار عدد العينات غير ذات المعنى المرغوب باختيارها بين العينات ذات المعنى عند عملية التجزئة .

وبعد ذلك يتم الضغط على مفتاح التجزئة (Divide) ، وتظهر الواجهة التالية :

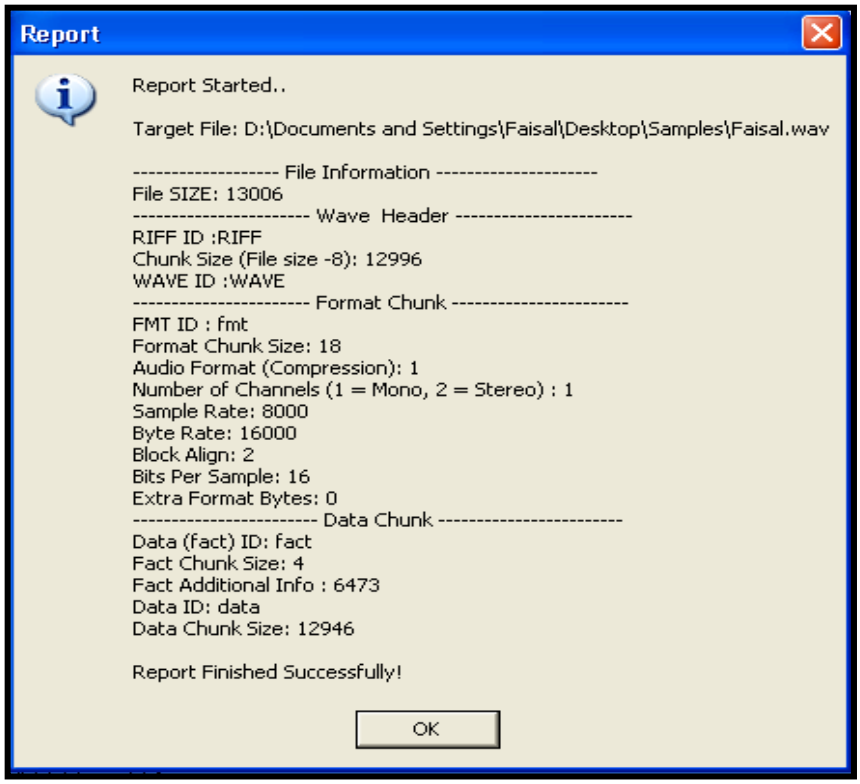


شكل رقم (٥-١٩) يبين ملف الصوت بعد تجزئته

وتظهر الخطوط الرأسية لتعبر عن مناطق التجزئة التي حدثت على الملف ، وفي هذه الأثناء تظهر نتائج عملية التجزئة في مكان وجود البرنامج في الذاكرة .

ويوجد مفتاحين آخرين هما مفتاح المسح (Clear) الذي يسمح الملف المجزأ في الخطوات السابقة من الذاكرة ، ومفتاح الخروج (Exit) للخروج من عملية التجزئة .

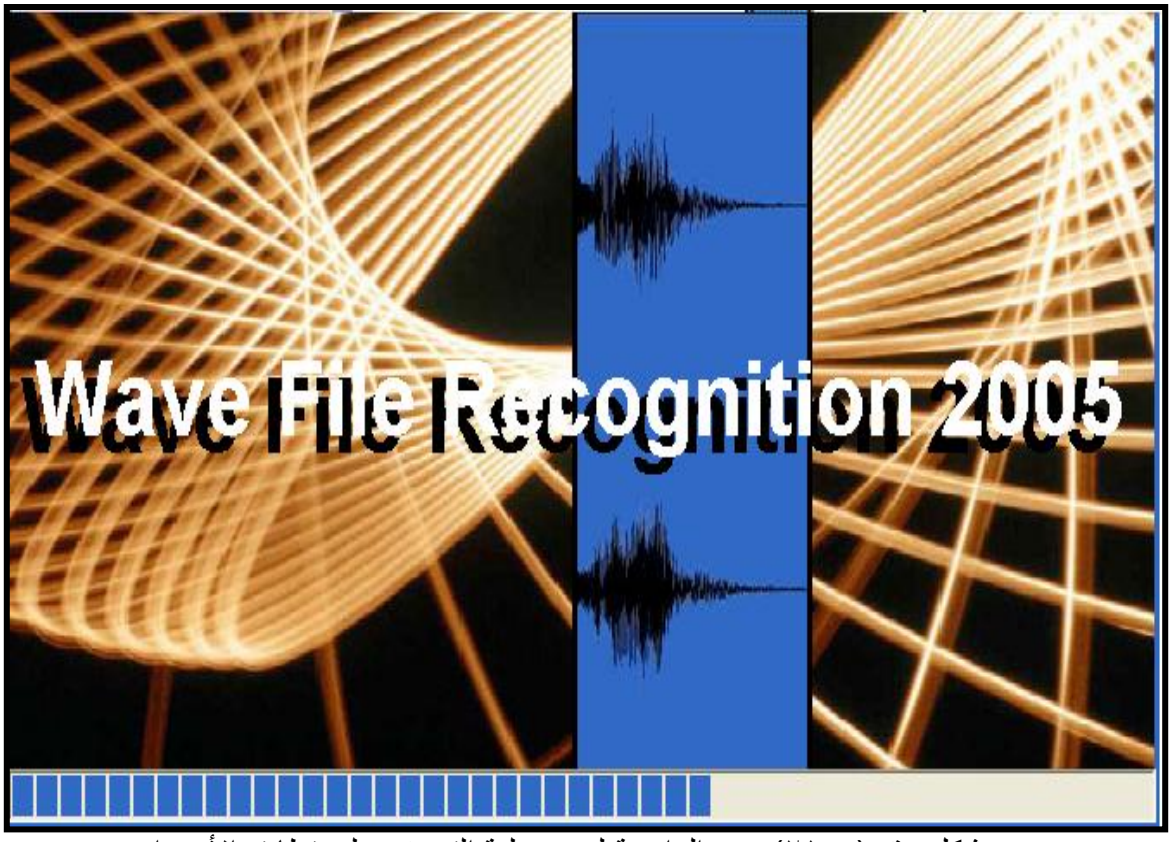
وتحتوي القائمة الجانبية تحكمات العملية (Progress Controls) على مفتاح إحصائيات العمليات (Operation Statistics) لإعطاء تقرير نهائي عن معلومات حول عملية التجزئة ، وتظهر الواجهة الرسومية التالية :



شكل رقم (٥-٢٠) يبين التقرير الذي يظهر في نهاية عملية التجزئة

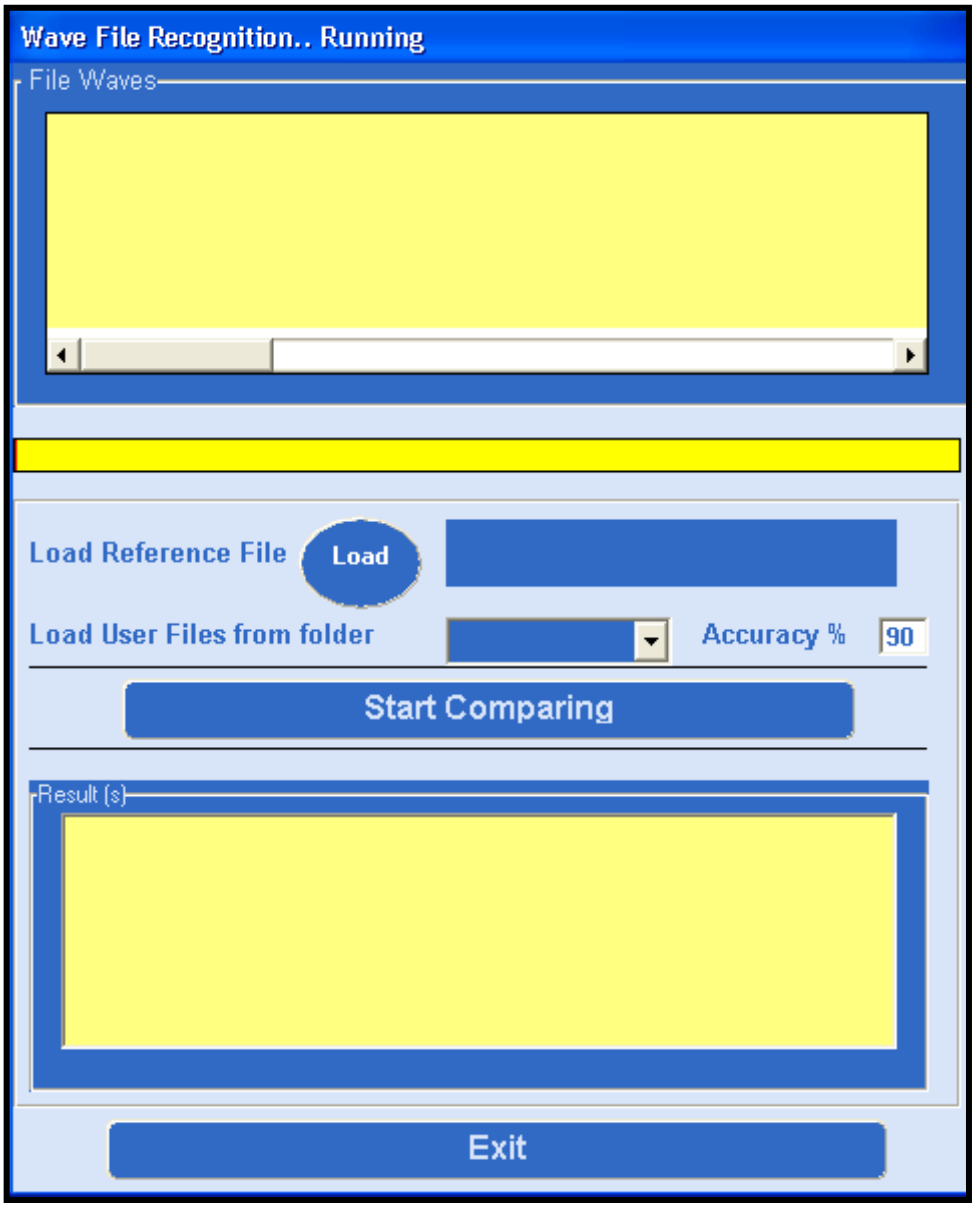
٥-٦ الواجهات الرسومية لبرنامج التعرف إلى تطابق الأصوات

عند تشغيل برنامج التعرف إلى الأصوات ، تظهر الواجهة الرسومية التالية (Wave File Recognition 2005) التي تنبئ بمدى تحميل البرنامج على ذاكرة الحاسوب الرئيسة:



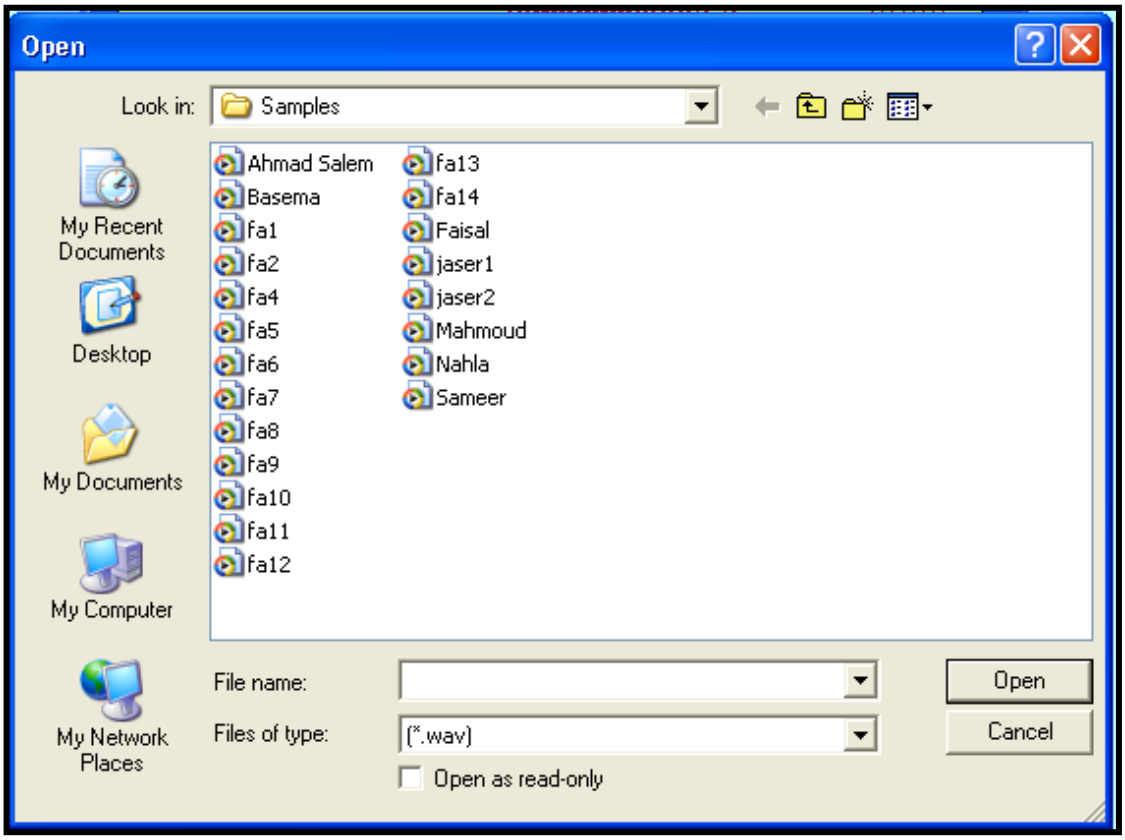
شكل رقم (٥-٢١) يبين الواجهة لبدء عملية التعرف على تطابق الأصوات

ثم تظهر الواجهة الرئيسة للبرنامج ، و هي على الصورة التالية :



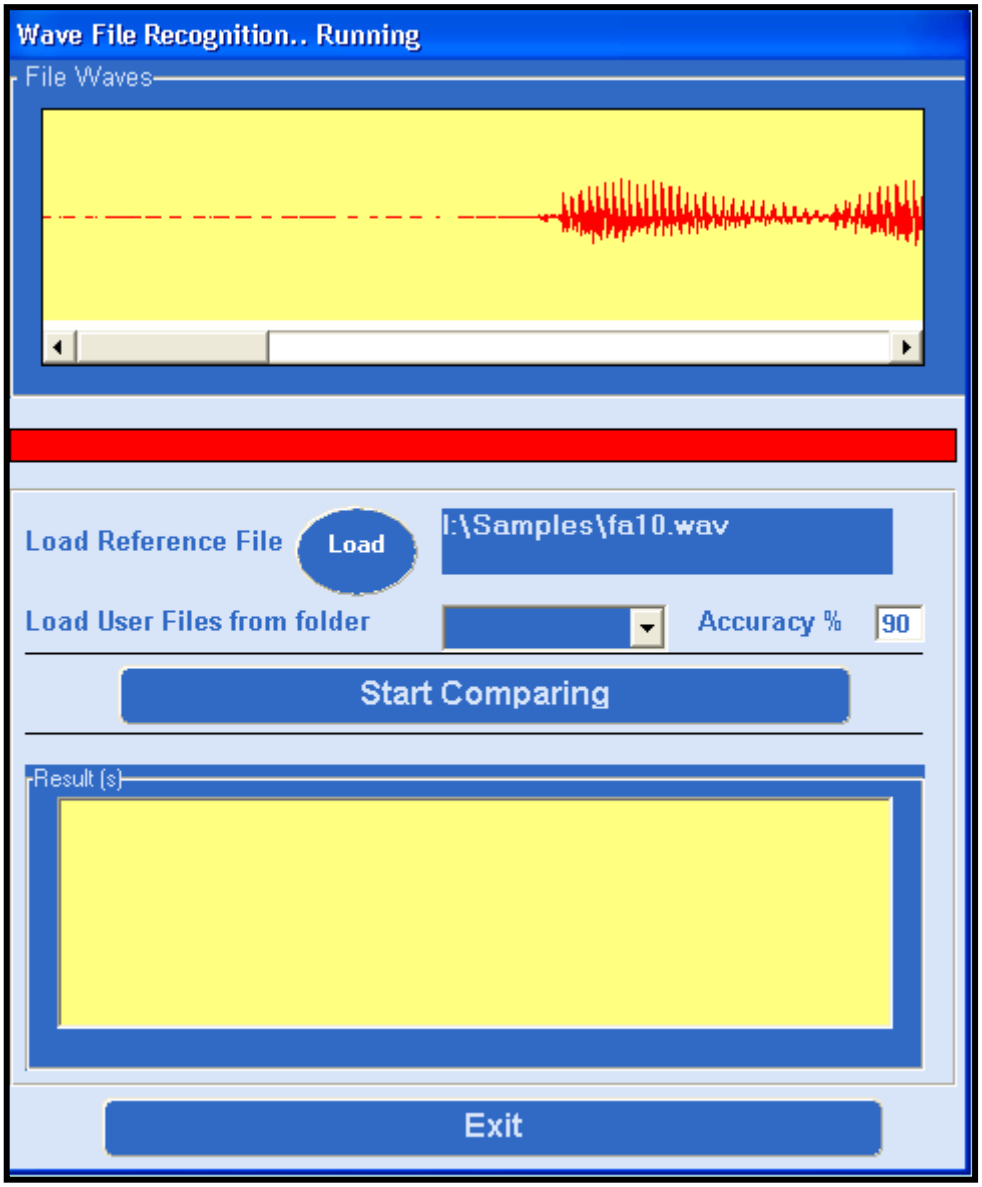
شكل رقم (٥-٢٢) يبين الواجهة الرئيسة لبرنامج التعرف على تطابق الأصوات

تحتوي الواجهة الرسومية السابقة على مجموعة من المفاتيح والخيارات و النوافذ البسيطة ، حيث تبدأ الواجهة بنافذة الهدف منها إظهار التمثيل البياني لبيانات الملف المجزأ المختار من خلال المفتاح تحميل (Load) ، والذي يتم اختيار أحد الملفات المجزأة ، والمرغوب في التعرف إلى الصوت الذي يحتويه هذا الملف من خلال البيانات الصوتية فيه ، وذلك من خلال صندوق الحوار التالي :



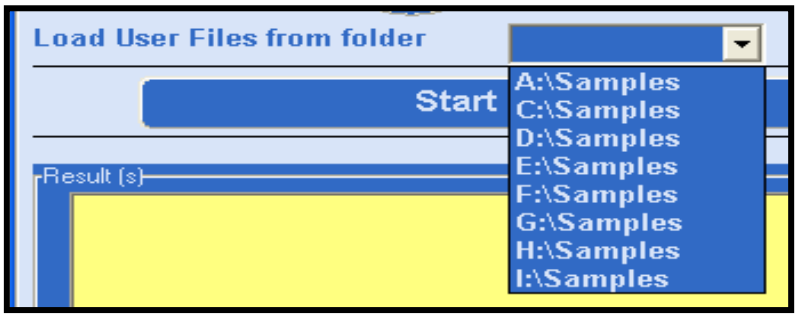
شكل رقم (٥-٢٣) يبين صندوق الحوار الخاص باختيار ملف و فتحه

بعد اختيار الملف المرغوب في التعرف إلى الصوت فيه ، تظهر الواجهة الرسومية التالية لتوضح عملية فتح الملف :



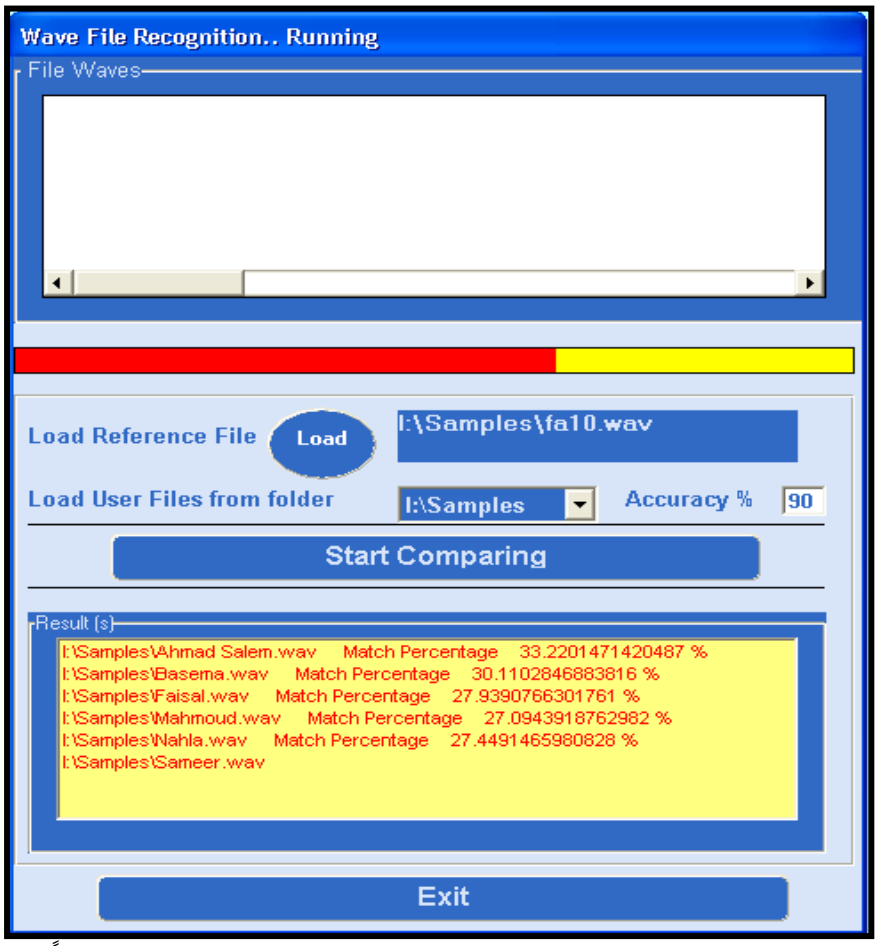
شكل رقم (٥-٢٤) يبين عملية فتح الملف

تأتي بعد ذلك الخطوة الخاصة بتحديد الملفات المراد مقارنة الملف السابق بها ، وذلك بتحديد المجلد الموجودة به الملفات الخاصة بالمقارنة والمجهزة في السابق . وقد تم تسمية هذا المجلد باسم(Samples) ، ويتم ذلك من خلال صندوق الاختيار المسمى تحميل ملفات المستخدم من مجلد (Load User Files from Folder) .



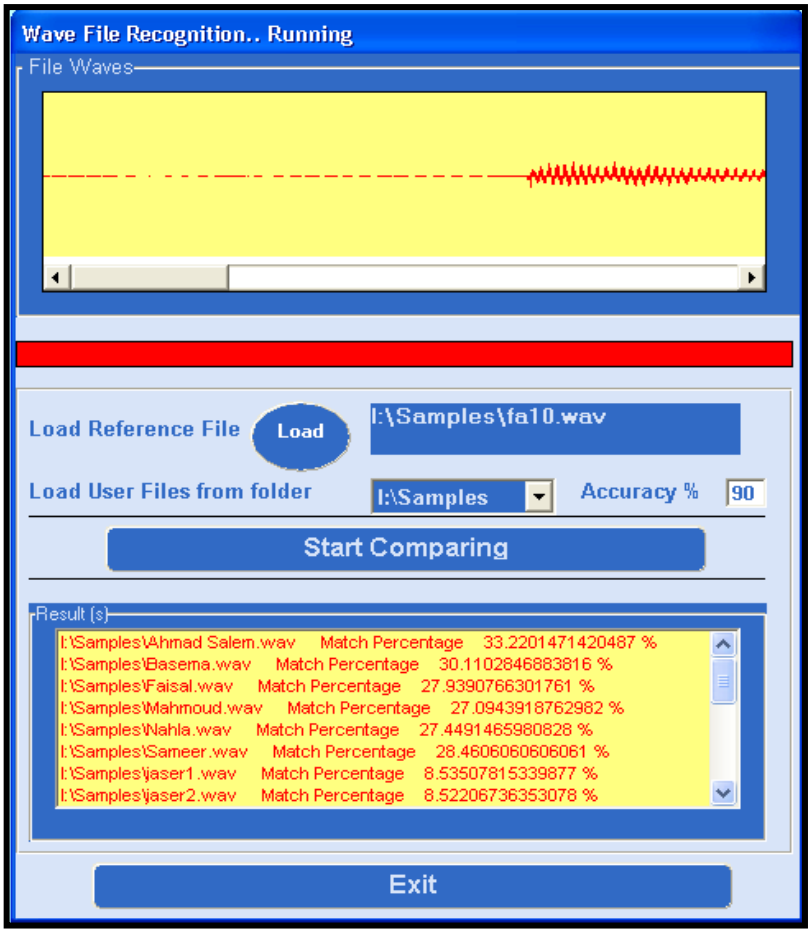
شكل رقم (٥-٢٥) يبين عملية اختيار مشغل الأقراص الذي يحوي النماذج.

بعد تحديد المجلد المناسب من القرص المناسب ، نضغط على مفتاح ابدأ المقارنة (Start Comparing) ، فتبدأ المقارنة وتظهر الواجهة الرسومية التالية :



شكل رقم (٥-٢٦) يبين الواجهة التي تظهر نتائج المقارنات تباعاً للملفات.

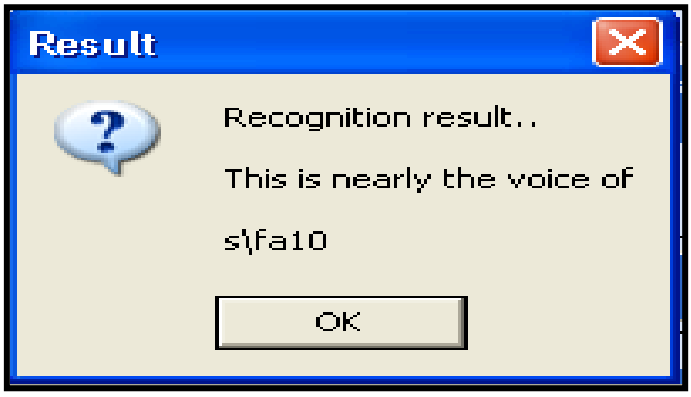
تظهر في نافذة النتائج (Results) نسب المقارنة بين الملف المرغوب في التعرف على الصوت فيه ، والملفات النماذج حتى تظهر الواجهة الرسومية التالية:



شكل رقم (٥-٢٧) يبين الواجهة التي تظهر نتائج المقارنات النهائية

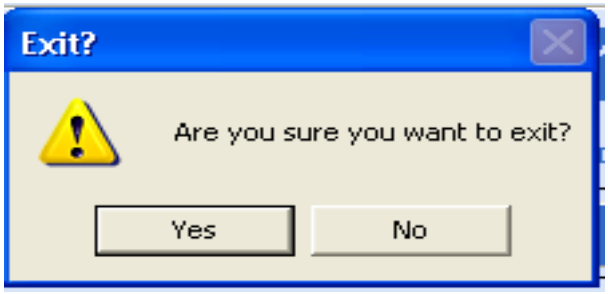
أما اختيار الدقة (Accuracy %) فيؤثر في عدد الأجزاء المراد تقسيم الملف إليها لغاية استخراج المعدلات للبايتات .

وعند انتهاء عمليات المقارنات تظهر الواجهة الرسومية التالية لتخبر عن تشابه الصوت في الملف المراد التعرف إلى الصوت فيه مع الصوت المخزن في أحد الملفات النموذج المخزنة.



شكل رقم (٥-٢٨) يبين الواجهة التي تظهر التقرير في معرفة الصوت

وعند الرغبة في الخروج من البرنامج يظهر صندوق الحوار التالي ليؤكد عملية الخروج نهائياً من البرنامج .



شكل رقم (٥-٢٩) يبين الواجهة لتأكيد الخروج من البرنامج

الفصل السادس

النتائج والتوصيات

٦-١ نتائج عملية تجزئة ملفات الصوت

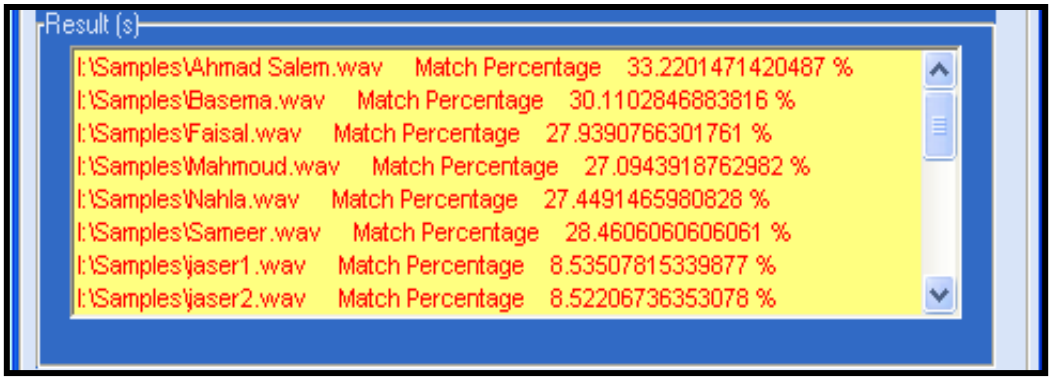
بعد إجراء عملية التجزئة على ملفات الصوت سوف تظهر في مكان وجود برنامج التجزئة على الذاكرة ، الملفات المجزأة كنتيجة لتنفيذ البرنامج ، وتتميز هذه الملفات بحملها أيقونة مشغل الأصوات الذي يمكن تشغيلها باستخدامه ، وتحمل اسماً يتكون من اختصار كلمة قسم (DIVISION) وهو (DIV) ، متبوعاً برقم يدل على ترقيم الملفات حسب ترتيب خروجها من عملية التجزئة مثل (٠٠٥) ، ويكون امتداد الملف (WAV) و يكون نوع الملف هنا ملف موجة كما هو الحال في الملف الذي تم تجزئته .

شكل رقم (٦-١) يبين الأشكال المختلفة للملفات المخرجة من عملية التجزئة .

٦-٢ نتائج عملية التعرف إلى تطابق الأصوات

بعد إجراء عملية التجزئة ، والحصول على الملفات المجزأة ، التي قد تحتوي على حروف لكلمات ، أو تحتوي على كلمات من جمل ، نأخذ أحد تلك الملفات التي تحتوي الكلمة أو الحرف و نقارنها بملفات نماذج مخزنة .

تظهر النتيجة للمقارنات بين الملفات على هيئة نسب مئوية مثل النسب التالية:



شكل رقم (٦-٢) يبين الواجهة التي تظهر نتائج المقارنات للملفات

٣-٦ خوارزمية طبقت باستخدام التطبيق MATLAB

يوجد الكثير من البرامج والتطبيقات المتوفرة بين أيدي المستخدمين التي تهدف إلى تجزئة ملفات الصوت ، ولكنها تختلف عن بعضها البعض في أسلوب التجزئة ، ونوعية الملفات التي تقوم بتجزئتها وهذا ما يعطي المستخدم حرية في اختيار البرامج والتطبيقات المناسبة للوصول إلى أهدافه بطريقة تلبي حاجاته .

ومن بين هذه التطبيقات والبرامج برنامج يدعى MATLAB الذي يقوم بالكثير من التطبيقات في المجالات المختلفة ، مثل معالجة الأصوات و الصور و غيرها من المعالجات المختلفة ، وقد تم اختيار خوارزمية طبقت باستخدام التطبيق MATLAB الخاصة بتجزئة ملفات الصوت لتطبيقها على الملفات من نوع موجة ، بسبب سهولة استخدام هذا التطبيق ، وسهولة عقد مقارنة بين تلك الخوارزمية ، والخوارزمية الجديدة قيد الدراسة .

أمّا بالنسبة للخوارزمية التي طبقت باستخدام التطبيق MATLAB ، فهي تقوم على أساس تحديد نقاط البدء ونقاط الانتهاء للكلمات في حديث المتكلم في الملف . وهذا التحديد يعتمد على مقياسين هامين هما:

١- الطاقة المخزنة في الصوت .

٢- معدل القطع الصفري .

وهذان المقياسين يعتمدان بشكل كبير على مجموعة عوامل مختلفة مثل :

١- الكلمات التي تبدأ أو تنتهي بطاقة قليلة .

٢- الكلمات التي تنتهي بحروف مهموسة .

٣- الكلمات التي تنتهي بحروف تخرج من الأنف .

٤- الكلمات النهائية للمتحدث التي تنتهي بحروف خفيفة أو النفس بين الكلمات .

وعند تطبيق الخوارزمية تحسب العتبات (Thresholds) الثلاث التالية :

١- عتبة الطاقة العليا (Upper Energy Threshold) .

٢- عتبة الطاقة الدنيا (Lower Energy Threshold) .

٣- عتبة معدل القطع الصفري (Zero Crossing Rate Threshold) .

ويحدث القطع الصفري عندما يكون لنقطتين متجاورتين في الزمن لشكل الموجة إشارتين مختلفتين موجبة وسالبة . و يحسب معدل القطع الصفري عن طريق حساب عدد مرات القطع الصفري للإشارة مع المحور السيني لفترة زمنية محددة . ولعله من المنطقي حساب القطع الصفري لاشارة لها قيمة متوسطة تساوي الصفر .

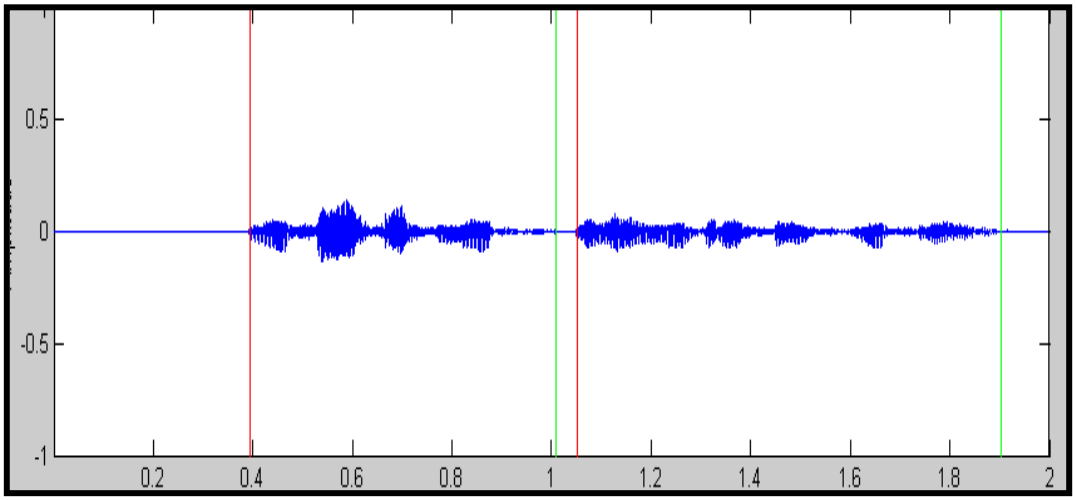
وتتكون الخوارزمية الخاصية بتحديد نقاط البداية والنهاية من الخطوات التالية :

١- يتم التعديل الصفري للإشارة .
٢- يتم أخذ الوحدات من الإشارة .
٣- يُحسب المتوسط الحسابي للطاقة في الإشارة .
٤- يُحسب معدل القطع الصفري .
٥- يتم إيجاد نقاط البداية والنهاية بالرجوع إلى عتبة الطاقة العليا .
٦- تلغى المقاطع الصوتية القصيرة .
٧- يتم إيجاد فهرس للقيمة الصغرى المحلية ، وتمدد نقاط النهاية والبداية حتى عتبة الطاقة الدنيا .
٨- تمدد نقاط البداية والنهاية لتشمل منطقة عتبة القطع الصفري .
٩- يتم إلغاء أجزاء الصوت المكررة .
١٠- يتم تحويل الفهرس المعتمد على نقاط العينات .

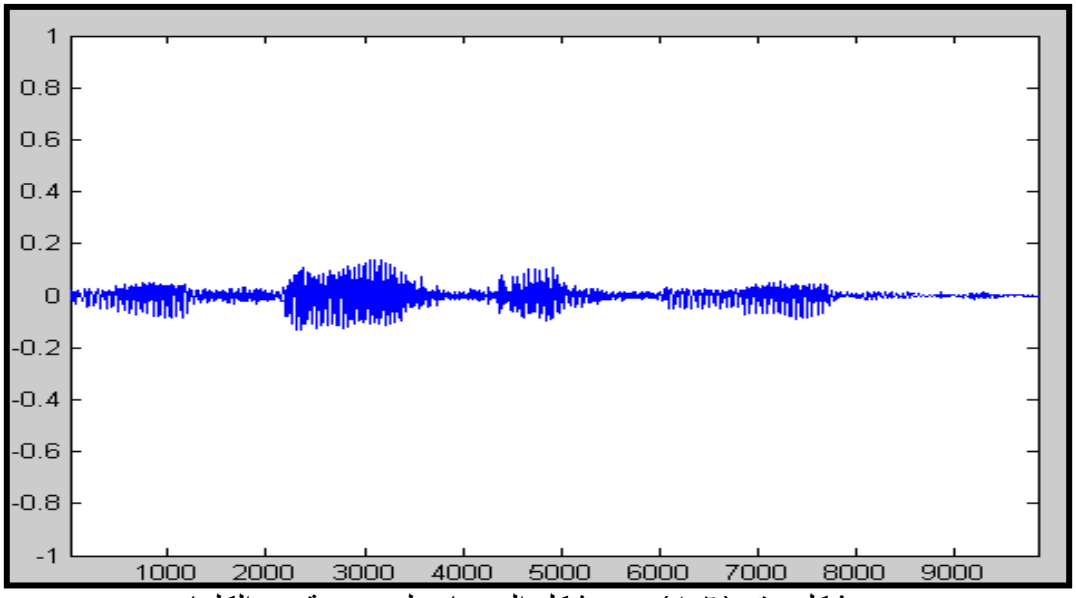
٤-٦ تطبيق الخوارزميتين على مجموعة من ملفات الصوت

تم التعامل مع عدد من ملفات الصوت من نوع الموجة يقدر بخمسين ملفاً ، تحتوي على عدد من الكلمات المختلفة لعدد من الأشخاص . وبعد تجهيز تلك الملفات لتصبح ذات محددات موحدة و ثابتة ، كما تم إحصاء عدد الكلمات التي نُطقت في كل ملف وسجلت تلك الأرقام في سجل خاص لسهولة الرجوع إليها وقت الحاجة ، وقد روعيت الشروط الخاصة بتسجيل الأصوات في هذه الدراسة حيث تم تسجيل الأصوات في بيئة هادئة بعيدة عن أية ضوضاء ، ثم تم تحديد المحددات التي ذكرت سابقا في تسجيل هذا النوع من الملفات .

وبعد ذلك طبقت الخوارزمية باستخدام التطبيق MATLAB على الملفات الواحد تلو الآخر ، ليتم تجزئة تلك الملفات . وكان شكل تجزئة الملفات يأخذ رسما بيانيا لموجات الكلمات ، تُحدد فيه الكلمات المجزأة بخطين رأسيين في بداية الكلمة ونهايتها ثم تُسمع الكلمة المجزأة وتظهر برسم بياني لموجات الكلمة المفردة كما يلي :



شكل رقم (٦-٣) يبين نقاط البداية والنهاية لموجات الكلمات في ملف معين



شكل رقم (٦-٤) يبين شكل الموجات لمجموعة من الكلمات

ثم طبقت الخوارزمية الجديدة على نفس الملفات السابقة ، بعد إدخال و تثبيت حجم النافذة ليساوي ٨٠٠ عينة ، ويكون مدى قطع الموجة مساويا لـ ٢٠٠٠ ، ليتم أيضا تجزئة تلك الملفات باستخدام الخوارزمية الجديدة ، فكانت النتائج من كلا العمليتين كما يلي :

اسم الملف	عدد الكلمات التي يحتويها الملف	عدد الكلمات التي أنتجت من الملف بعد تطبيق الخوارزمية الجديدة عليه		عدد الكلمات التي أنتجت من الملف بعد تطبيق خوارزمية MATLAB عليه	
		الكاملة	المشوهة	الكاملة	المشوهة
م١	٣	٣	٠	٠	٣
م٢	٣	٣	٠	٣	٠
م٣	٤	٢	٢	٢	٢
م٤	٤	٤	٠	١	٣
م٥	٤	٣	١	٣	١
م٦	٤	٣	١	٢	٢
م٧	٤	٣	١	٣	١
م٨	٥	٤	١	٢	٣
م٩	٥	٣	٢	٣	٢
م١٠	٥	٥	٠	٣	٢
م١١	٦	٤	٢	١	٥
م١٢	٥	٥	٠	٣	٢
م١٣	٥	٣	٢	٢	٣
م١٤	٥	٥	٠	٥	٠
م١٥	٥	٣	٢	١	٤
م١٦	٥	٥	٠	٢	٣
م١٧	٦	٥	١	٥	١
م١٨	٦	٤	٢	٢	٤
م١٩	٦	٥	١	٣	٣
م٢٠	٦	٥	١	٢	٤
م٢١	٦	٦	٠	٤	٢
م٢٢	٦	١	٥	١	٥
م٢٣	٦	٥	١	٥	١
م٢٤	٦	٤	٢	٢	٤
م٢٥	٧	٤	٣	١	٦
م٢٦	٧	٦	١	٤	٣

اسم الملف	عدد الكلمات التي يحتويها الملف	عدد الكلمات التي أنتجت من الملف بعد تطبيق الخوارزمية الجديدة عليه		عدد الكلمات التي أنتجت من الملف بعد تطبيق خوارزمية MATLAB عليه	
		الكاملة	المشوهة	الكاملة	المشوهة
م٢٧	٧	٧	٠	٥	٢
م٢٨	٧	٥	٢	٣	٤
م٢٩	٧	٦	١	٥	٢
م٣٠	٧	٧	٠	٣	٤
م٣١	٧	٥	٢	٤	٣
م٣٢	٧	٤	٣	٢	٥
م٣٣	٧	٤	٣	٦	١
م٣٤	٧	٧	٠	٥	٢
م٣٥	٨	٦	٢	٦	٢
م٣٦	٨	٨	٠	٦	٢
م٣٧	٨	٨	٠	٣	٥
م٣٨	٨	٦	٢	٥	٣
م٣٩	٨	٧	١	٣	٥
م٤٠	٨	٨	٠	٨	٠
م٤١	٨	٨	٠	٧	١
م٤٢	٨	٥	٣	٥	٣
م٤٣	٩	٨	١	٤	٥
م٤٤	٩	٧	٢	٧	٢
م٤٥	٩	٨	١	٧	٢
م٤٦	٩	٨	١	٦	٣
م٤٧	٩	٨	١	٥	٤
م٤٨	٩	٥	٤	٢	٧
م٤٩	٩	٨	١	٥	٤
م٥٠	١٢	٧	٥	٧	٥
المجموع	٣٢٩	٢٦٣	٦٦	١٨٤	١٤٥

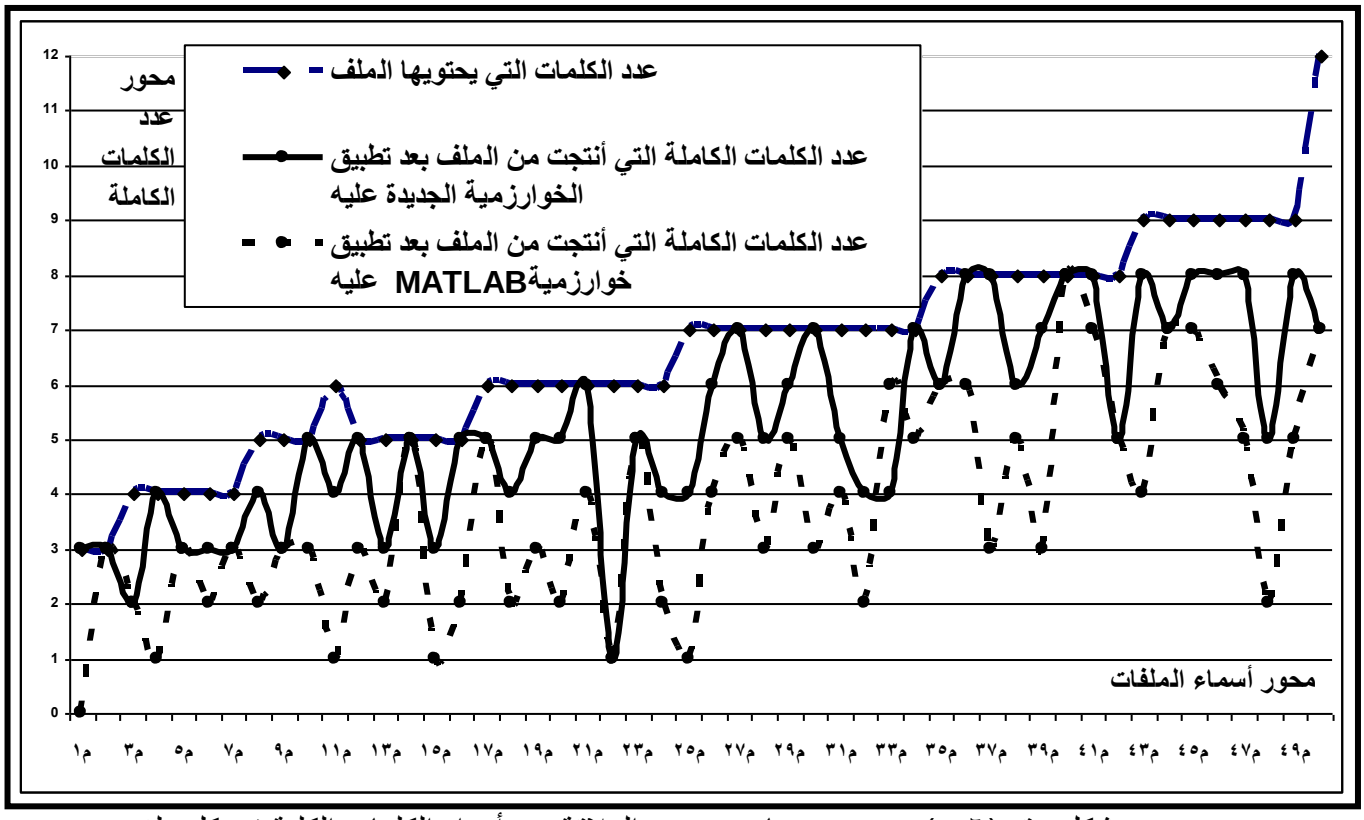
جدول رقم (٦-١) يبين أعداد الكلمات الكاملة والمشوهة في الملفات قبل التجزئة وبعدها من خلال تطبيق الخوارزميتين الجديدة و MATLAB .

يتضح من خلال الجدول السابق أن أعداد الكلمات الكاملة الناتجة من تطبيق الخوارزمية الجديدة على الملفات قيد التجربة كانت في معظم التجارب أكبر من أعداد الكلمات الكاملة الناتجة من تطبيق خوارزمية طبقت باستخدام التطبيق MATLAB ، وذلك عند تجربتها على خمسين ملف صوت من نوع موجة .

وقد اعتمد في خوارزمية التطبيق MATLAB على مبدأ الطاقة ومبدأ القطع الصفري وهما من المبادئ التي تتأثر بشدة في نوع الصوت وحِدَّته و قوة النطق وما إلى ذلك من عوامل تؤثر في نقاط البداية والنهاية للكلمات المنطوقة في الملفات .

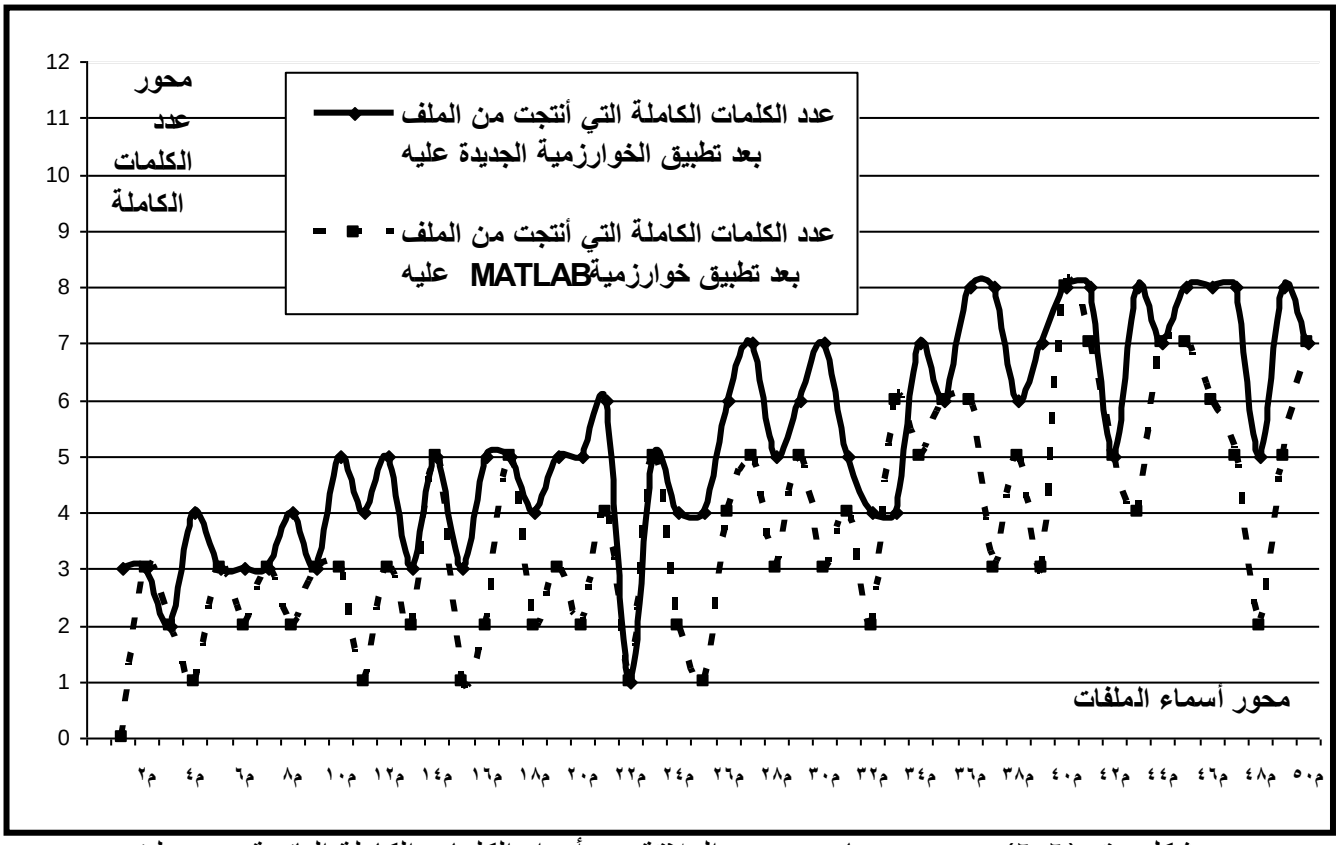
بينما اعتمد في الخوارزمية الجديدة على عدد العينات غير ذات المعنى أي التي تحوى أصفاراً ثنائية ، وهو ما سُمي بحجم النافذة ، كما اعتمد على مدى قطع الموجة ، و هذه المحددات تتأثر إلى حدٍّ ما بشكل أكبر بالعوامل المتعلقة بالكلمات المنطوقة في الملف و المذكورة فيما سبق .

ويبين الرسم البياني التالي العلاقة بين أعداد الكلمات الكاملة الناتجة بعد إجراء عمليتي التجزئة بكلا الخوارزميتين بالإضافة إلى الأعداد الكلية للكلمات في كل ملف .



شكل رقم (٦-٥) يبين رسم بياني يوضح العلاقة بين أعداد الكلمات الكلية في كل ملف وأعداد الكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين .

ويبين الرسم البياني التالي العلاقة بين أعداد الكلمات الكاملة الناتجة بعد إجراء عمليتي التجزئة بكلا الخوارزميتين .



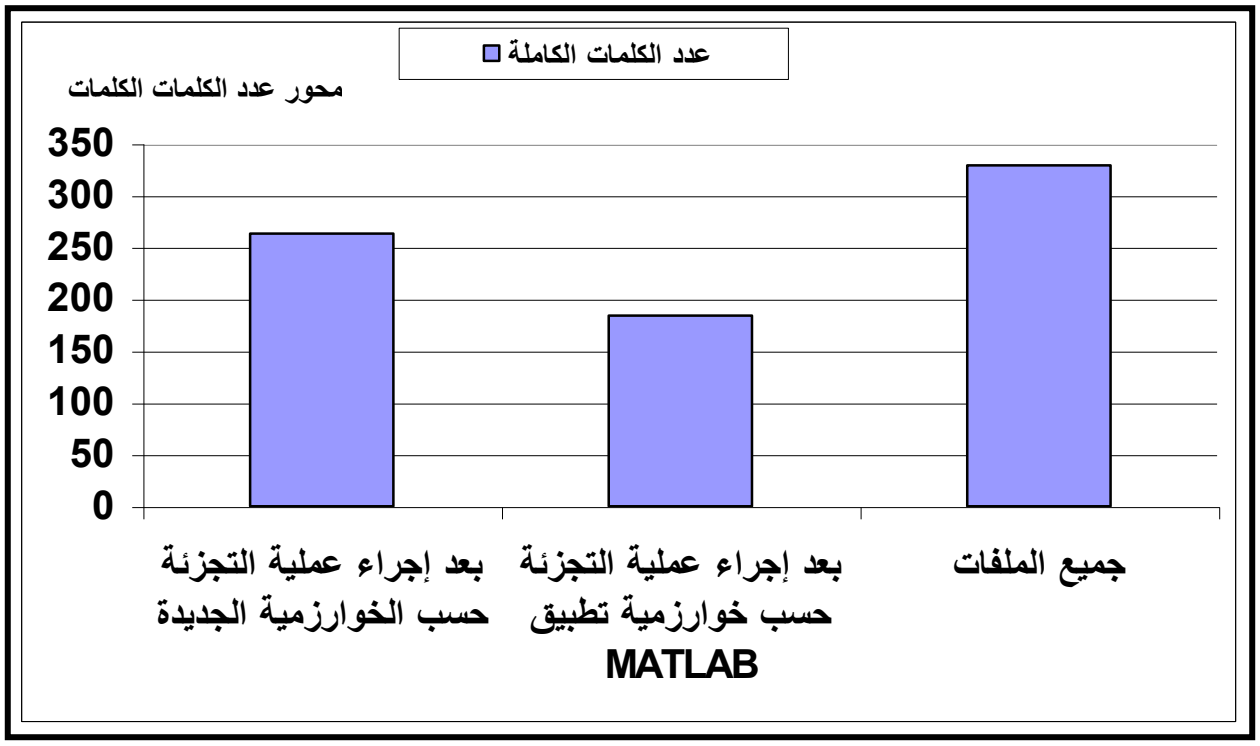
شكل رقم (٦-٦) يبين رسم بياني يوضح العلاقة بين أعداد الكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين .

٦-٥ نتائج المقارنة بين الخوارزمية الجديدة وخوارزمية طبقت باستخدام التطبيق MATLAB

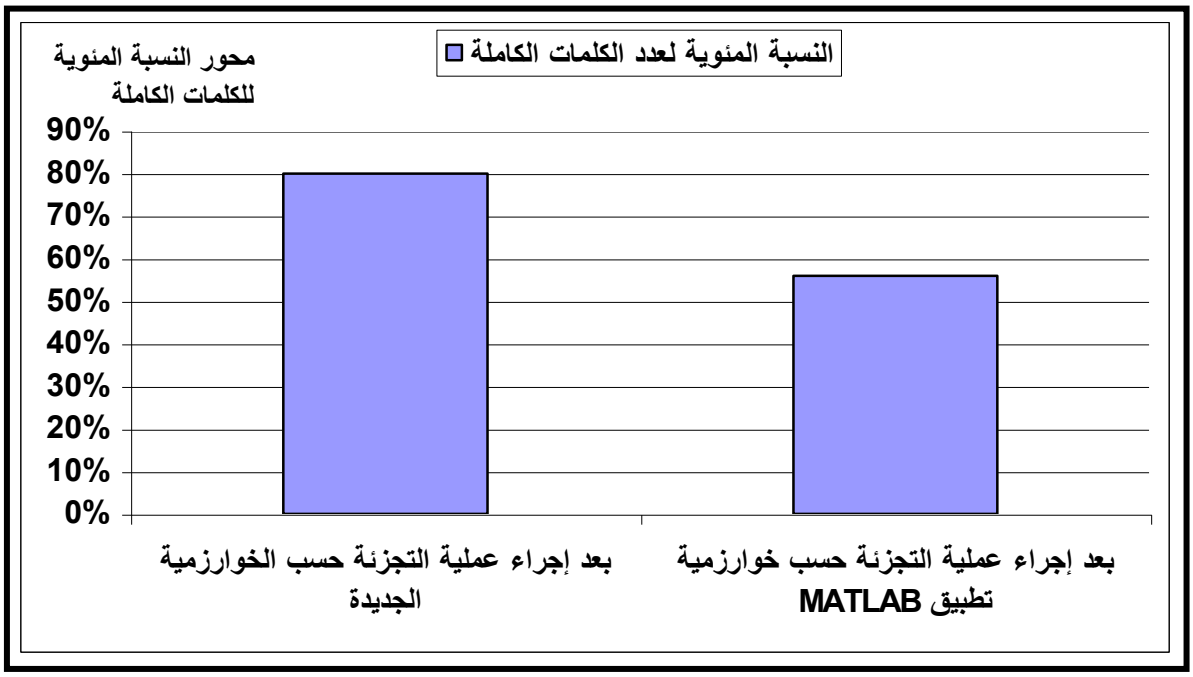
عند استخراج النسبة المئوية لعدد الكلمات الكاملة الناتجة من عملية التجزئة لكل الملفات باستخدام الخوارزمية الجديدة إلى عدد الكلمات الكاملة في جميع الملفات قيد التجربة ، والنسبة المئوية لعدد الكلمات الكاملة الناتجة من عملية التجزئة لكل الملفات باستخدام خوارزمية طبقت باستخدام التطبيق MATLAB إلى عدد الكلمات الكاملة في جميع الملفات قيد التجربة ينتج الجدول البسيط التالي والرسومات البيانية التالية :

نوع الخوارزمية المتبعة	عدد الكلمات الكاملة	النسبة المئوية لعدد الكلمات الكاملة
بعد إجراء عملية التجزئة حسب الخوارزمية الجديدة	٢٦٣	% ٧٩.٩٣٩
بعد إجراء عملية التجزئة حسب خوارزمية طبقت باستخدام تطبيق MATLAB	١٨٤	% ٥٥.٩٢٧
جميع الملفات	٣٢٩	

جدول رقم (٦-٢) يبين عدد الكلمات الكاملة والنسب المئوية لعدد الكلمات الكاملة الناتجة بعد تطبيق الخوارزميتين على عدد من الملفات لتجزئتها .



شكل رقم (٦-٧) يبين رسم بياني يوضح العلاقة بين أعداد الكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين .



شكل رقم (٦-٨) يبين رسم بياني يوضح العلاقة بين النسبة المئوية للكلمات الكاملة الناتجة من عمليتي التجزئة بعد تطبيق الخوارزميتين .

٦-٦ النتيجة النهائية

من خلال مقارنة القيم الناتجة من التجارب السابقة ، ومن خلال التدقيق في الرسومات البيانية المترتبة على تلك التجارب يتبين أن تجزئة ملفات الصوت من خلال الخوارزمية الجديدة يحقق نسبة أفضل بعدد الكلمات الكاملة عن مثيلتها التي تنتج بعد عملية التجزئة من خلال خوارزمية طبقت باستخدام التطبيق MATLAB .

الخلاصة

إنّ هذه الدراسة اقترحت خوارزمية جديدة لتجزئة ملفات الصوت من نوع موجة ، وذلك من أجل استخدام تلك الأجزاء في التعرف إلى الأصوات في الملفات من نوع موجة أيضاً .

وقد بنيت هذه الخوارزمية على التعرف إلى العينات ذات المعنى في ملفات الأصوات والعينات غير ذات المعنى ، والتمييز بينهما ، وعلى دراسة مميزات و محددات ملفات الموجة والتعامل معها وتطويعها حسب ما يرغب المستخدم عمله في هذه الملفات ، ثم نسخ تلك المحددات في تراكيب بيانات مخصصة لإنتاج ملفات موجة جديدة جزئية من تلك التي أجريت عليها عملية التجزئة الجديدة .

ثم صمم برنامج بسيط آخر لإمكانية التعرف إلى الأصوات ، حيث كانت مدخلاته هي مخرجات البرنامج السابق الخاص بعملية التجزئة ، وكانت مخرجات هذا البرنامج هي مقارنات بين ملفات موجة مخزنة كنماذج في الحاسوب ، وملف موجة يراد التعرف إلى الصوت المخزن به على شكل بيانات ثنائية .

و تم عقد مقارنة بين الخوارزمية الجديدة و خوارزمية طبقت باستخدام التطبيق MATLAB من خلال أعداد الكلمات التي تستطيع كل خوارزمية استخراجها من ملفات الصوت من نوع موجة ، وتبين من المقارنة أن الخوارزمية الجديدة أثبتت تفوقاً مُلاحظاً عن خوارزمية تطبيق MATLAB .

الأعمال المستقبلية

لقد جاءت هذه الدراسة خطوة على الطريق لإعطاء الباحثين مجالاً لبناء أنظمة متطورة للتعرف إلى الأصوات بطرق أخرى ، وباستخدام الخوارزمية الجديدة والتي يمكن أن تدار بشكل آلي في المستقبل .

كما يمكن استخدام مخرجات عملية التجزئة قيد الدراسة في التعرف إلى الأصوات والتمييز بينها من حيث الجنس مثلاً ، أو التمييز بين الأصوات من حيث سن المتحدث ، أو حتى في التعرف إلى حالة المتحدث من حيث كونه مدخناً مثلاً أو لا .

ويمكن التعديل على متغيرات برنامج التجزئة ، وجعله ينفذ على ملفات الصور المتحركة مثلاً ، أو إدخال الشبكات العصبية في عملية تطوير البرنامج للحصول على ملفات صوت تركب تركيبا للاستعانة بها في بعض الأعمال الخاصة بإنتاج الأفلام أو المونتاج الصوتي في الإذاعات.

وكل ما سبق ينضوي تحت لواء التحديث والتطوير للاستفادة مما قدم في هذه الدراسة في نواح شتى تساهم في تقدم المعلومات ولو بشيء يسير . والله ولي التوفيق .

قائمة المراجع والمصادر

- Donald G. Kimber, Lynn D. WIICOX, Francine R. Chen, Thomas P. Moran , **Speaker Segmentation for Browsing Recorded Audio**, 1995 , CHI '95 Mosaic of creativity. Publisher ACM Press .Identifier :oai:ACMDL: articles 223528 .

- George F. Meyer, Jeffrey B. Mulligan, Sophie M. Wuerger ,**a Continuous audio–visual digit recognition** .Published by Elsevier B.V. Nov 15,2003 INFFUS 121, doi 10.1016.

- George Tzanetakis, Perry Cook **Multifeature Audio Segmentation for Browsing and Annotation**.Application of signal processing to Audio and Acoustics New Palts ,New York , Oct 17-20,1999 .Ieeexplore.ieee.org(vol 10, No. 5).

- Jim Baumann , **Voice Recognition** , 2000. Angel Investor News .

- Keith Bain , Sara H. Basson ,Mike Wald , **Speech Recognition in University Classrooms : Liberated Learning Project** .ASSETS 2002. Pages (192-196). Publisher ACM Press New York.NY.USA.

- Lie Lu, Hao Jiang ,HongJiang Zhang , **A Robust Audio Classification and Segmentation Method**,2001. ACM Multimedia . Citeseerist ,pages (203-211).

- Michael Hoffman ,Laraine Coleman **,Voice Recognition Technology** ,2000 .TMA Associates.

- Olivier Pietquin, Laurent Couvreur, Pierre Couvreur , **Applied Clustering for Automatic Speaker-Based Segmentation of Audio Material**, 2002. Belgain Jornal of Operations Research Statistics and Computer Science (JORBEL) special issue orland statistics in the universities of Mons, Volume 41 n^{o},1-2,01. TCTS lab Publications.

- Renato De Mori , Fabio Brugnara , **HMM Methods in Speech Recognition,**1999. Publisher Cambridge University Press .Pages (21-30) Publication year 1997 .

- Sharath narayana , **Voice recognition** . 2002. Mistral Software Pvt.ltd .

- Silvia Pfeiffer , **Pause concepts for audio segmentation at different semantic levels** ,2001 .Publisher ACM Press . ACM Multimedia ,Pages (187-193).

- Stephen Cook , **Speech Recognition HOWTO ,** Revision V2.0 April 19,2002 by SCC . Changed license information (Now GFDL).

- Til T. Phan and Thomas Soong, **Text-Independent speaker Identification , Mapping from Speech to Images using Continuous state space.** Dec 8,1999 .

- Timothy J. Hazen, Kate Saenko, ChiaHao ,La, James R. Glass , **A Segment Based Audio Visual Speech Recognizer : Data Collection, Development, and Initial Experiments**,2004. ICMI ' 04, State college , Pens .USA. ACM 1-58113-954-3/04/0010.

- Tong Zhang and C.-C. Jay Kuo, **Heuristic Approach for Generic Audio Data Segmentation and Annotation**, 1999.ACM Multimedia '99;Orlando,FL; USA .pp 67-76. portal.acm.org.

قائمة مراجع الإنترنت

- **http://www.lightlink.com/tjweber/StripWav/WAVE.html , Last visited date Feb 3, 2006.**
- **http://www.sonicspot.com/guide/wavefiles.html#data , Last visited date Feb 20, 2006.**
- **http://www.acm.org , Last visited date Dec 22, 2005.**

ABSTRACT

Multimedia had played a main roll in the development of the computer usage. It had given many forms of data other than the written or read text form of data. It provides the possibility of motion, sounds and images. All of that can be active with simple and available instruments and drivers attached with the computers.

Many researches had look forward to improve both sides of multimedia, software and hardware, to face the desires and demands of the different users.

Audio and audio recordings were one of multimedia applications. It compensated the usage of tape recorders and gave the chances for using the audio as another important form of data.

Audio files recordings and retrievals took a big concern from the workers in this field. They make their efforts to have a full benefit from the files they have saved, the way they can get the files back, and what to do with these files.

Some researchers had produced different strategies about file segmentation and speech recognition.

This research had given a new algorithm for wave files segmentation to recognize the voices contained in these files. It had converted a part of dummy huge number of bits of the voice file into a smaller wave file containing a part of the voice that was in the original file. The segmented files can be run on different audio players and can be heard obviously.

This research had contained some algorithms to help building the segmentation algorithm, such as Normalization and Minimization algorithms.

Hopping this new algorithm adds some additional information to the huge pool of knowledge.

تمت بحمد الله