

Speech Files Compression Based on Signal Feature

Khalil I. Alsaif

Khalil_alsaif@hotmail.com

Computer center

University of Mosul, Iraq

Received on: 03/05/2006

Saja J. Mohammed

Sj_alkado@uomosul.edu.iq

College of Computer Science and Mathematics

University of Mosul, Iraq

Accepted on: 05/03/2007

ABSTRACT

In this research a new algorithm was suggested for compressing speech files added a new style for storing signals, The suggested idea of compression begins with recording the speech via the microphone, then starting the proposed processing steps as follows :

- Removing silent period.
- Select the number of resulted signal samples.
- Segmenting the resulting signal to number of frames.
- Applying one of the curves fitting algorithms and obtaining the coefficients for the mathematical representation.
- Storing the results in a new file format with .ssc (Speech Signal Compression) extension.

While the decompression process consisted of the reversal compression process steps, the signal is reconstructed using curve fitting coefficients which were stored in the new file, followed by returning the selected sample, then returning the silent period to their original location and finally listening to the retrieved speech signal. When the proposed algorithm had been applied on the files with different speech contents, the compression ratio was approximately (16.283%), and the ratio of SEGSNR was approximately (25.195dB).

Keywords: compression, speech signal, feature extraction, Silent Period

كيس ملفات الكلام باعتماد خواص الإشارة

سجى جاسم محمد

جامعة الموصل /كلية علوم الحاسبات والرياضيات

تاريخ قبول البحث: 2007/03/05

خليل إبراهيم السيف

جامعة الموصل /مركز الحاسبة الالكترونية

تاريخ استلام البحث: 2006/05/03

الملخص

أُفترِح في البحث خوارزمية لعملية كبس ملفات الكلام أضفت طابعاً آخر على أساليب خزن إشارة الكلام، تبدأ فكرة الكبس المقترحة باكتساب الكلام من لاقط الصوت، ثم البدء بخطوات المعالجة المقترحة والتي تضمنت:

- العمل على إزالة فترات الصمت (Silent Period).
- اقتباس بعض العينات للإشارة الناتجة.
- تقطيع الإشارة الناتجة ومن ثم تطبيق إحدى طرائق مواعمة منحنى الإشارة (Curve Fitting) والحصول على معاملات التمثيل الرياضي للإشارة.
- خزن النتائج في ملف بصيغة جديدة، أطلق عليه الامتداد (.ssc:Speech Signal Compression).

أما عملية فك الكبس فقد تضمنت خطوات معاكسة لعملية الكبس إذ يتم إعادة بناء كائنات الإشارة من جديد باستخدام معاملات مواعمة المنحنى (Curve Fitting Coefficients) المخزونة في الملف الجديد، تتبعها إعادة العينات المقطعة من الإشارة ثم إرجاع فترات الصمت إلى مواقعها الأصلية والاستماع إلى إشارة الكلام المسترجعة، وبتطبيق الخوارزمية المقترحة على ملفات ذات محتويات كلامية مختلفة تم الحصول على نسبة اختزال في حجم الملف الصوتي بحدود (16.283%) ونسبة SEGSNR بحدود (25.195 dB).

الكلمات المفتاحية: الكبس، إشارة الكلام، استخلاص الخواص، فترات الصمت.

1 - المقدمة

لا يخفى على أحد أهمية الدور الذي تؤديه وسائل الاتصال الخاصة بالكلام في الحياة اليومية، خاصة بعد استحداث الهاتف الخليوي وخدمات الانترنت التي أوجدت إمكانية نقل الصوت بصيغته الرقمية عبر شبكات الاتصال، ومما جعل هذا ممكناً حدوث التطور السريع في الاتصالات بعيدة المدى ومعالجة الإشارة الرقمية، ومع هذا فقد واجهت عملية نقل الصوت عبر شبكات الاتصال مشكلة الحاجة إلى كميات خزن هائلة لبيانات الكلام والصوتيات [6].

إن نقل البيانات الرقمية للكلام عبر هذه الشبكات يتطلب أن يكون عرض الحزمة المستخدمة لنقل البيانات بقدر كافٍ بحيث تنقل البيانات بنوعية جيدة ومقبولة تؤدي الغرض من النقل، وهذا غالباً ما يؤدي إلى زيادة كلفة النقل فضلاً عن زيادة الوقت المستغرق وزيادة الأخطاء

المتوقعة أثناء نقل البيانات. عليه فقد أصبحت الحاجة ملحة إلى وجود طرائق تعمل على خزن إشارة الكلام بأقل حجم ممكن واستحداث خوارزميات خاصة تعمل على كبس هذه البيانات، لأجل التقليل من كمية الخزن المطلوبة للبيانات ومن ثم التقليل من عرض الحزمة المطلوبة لنقلها وزيادة سرعة النقل وتقليل احتمالية حدوث الأخطاء أثناء عملية النقل [8].

2- الصوت والكلام:

يوصف الكلام بأنه يمثل الصوت الناتج من جهاز التكلم الخاص بالإنسان (Speech Production System) لإنشاء تمثيل لغوي ينقل المعلومات من المستهل بالحديث إلى المستقبل للحديث، أما الهدف منه فيتركز في نقل المعنى بين الأشخاص أو لزيادة الرابطة الاجتماعية بين الأشخاص أو الجماعات [3].

كما أن مفهوم الكلام يتضمن إلى جانب المحتويات الواقعية الحقيقية لإشارة الكلام، معلومات عن شخصية المتكلم وهويته (من الذي يتكلم)، الانفعال، التنغيم، فضلاً عن مضمون الرسالة المنقولة بالكلام، كل هذه العوامل مهمة بالنسبة إلى مفهوم الكلام [3].

3 - كبس البيانات بين المحاسن والمساوئ:

إن الهدف الأساسي من استخدام تقنيات الكبس تمثيل المعلومات بأكثر دقة ممكنة باستخدام أقل عدد ممكن من الوحدات التخزينية [11]، ومن الملاحظ أن الملفات التي تستخدم معها تقنيات الكبس تمتاز بما يأتي [1][2]:

- صغر حجم الملف نسبة إلى حجم الملف الأصلي، وهذا ما يجعله يشغل حيزاً أقل في الذاكرة.
- إن صغر حجم الملف يوفر زيادة في سرعة الإرسال عبر وسائل الاتصالات .
- تتميز الملفات المكبوسة بوجود نوع من أنواع ترميز البيانات (Data Coding) التي توفر أمانة للملف بوصفه نتيجة عرضية لعملية الكبس.
- عند إرسال أي ملف مكبوس فإن عرض الحزمة (Bandwidth) المطلوبة لإرسال ذلك الملف تكون أقل من حالة كونه غير مكبوس.

ومن ناحية أخرى، فإن لطرائق الكبس مساوئ يمكن إجمالها بما يأتي [1]:

- عند نقل بيانات ملف مكبوس فإن حدوث الأخطاء أثناء عملية النقل تؤثر تأثيراً كبيراً على الملف المكبوس نسبة إلى الملف غير المكبوس، ذلك أنها قد تؤدي إلى تغيير أو فقدان المعلومات الضرورية التي تم الاحتفاظ بها عند عملية الكبس.
- تحتاج الملفات إلى تقنيات خاصة لغرض كبسها مما يؤدي إلى إضافة خوارزميات أخرى إلى العمل المطلوب إنجازه.
- ضياع بعض المعلومات التي قد تكون مهمة للغاية ولاسيما عند استخدام طرائق الكبس بفقدان (Lossy Compression)، وقد تتأثر بعض الملفات المكبوسة بهذه الطرائق تأثيراً كبيراً عند استرجاعها فتصبح الملفات المسترجعة ذات معلومات مشوشة (Noise).

4 - الهدف من البحث:

يهدف البحث إلى اقتراح خوارزمية تعمل على كبس بيانات إشارة الكلام، إذ تعتمد بالدرجة الأساسية على استخلاص خواص بيانات إشارة الكلام المخزونة في الملف الصوتي من نوع (wav)، والعمل على تخزين بيانات إشارة الكلام من خلال استخلاص بعض معاملات إشارة الكلام بحيث تؤدي إلى اختزال جزء كبير من حجم البيانات على ذاكرة الحاسوب، وتوفير إمكانية إعادة تكوين ملف الكلام من جديد بأسلوب واضح وسهل مع التأكيد على حفظ مفهوم الكلام ووضوحه، والعمل قدر الإمكان على تقليل الضوضاء الناتجة عن معالجة إشارة الكلام.

5- نظرة تاريخية إلى معالجة الكلام:

امتلكت عمليات معالجة إشارة الكلام تاريخاً طويلاً وجليلاً بدأ مع بداية القرن العشرين وامتد إلى بداية القرن الحادي والعشرين، فقد بدأ العمل على كبس إشارة الكلام منذ عام 1928 عندما قدمت مهندسة الهواتف هومر دادلي (Homer Dudley) فكرتها التي تحدثت عن كبس إشارة الكلام بعرض حزمة أكثر من 3000Hz إلى 100Hz في سلك تلغراف عبر المحيط الأطلسي، إذ نصت فكرة هومر على ترميز الإشارة ثم نقلها، فبدلاً من نقل إشارة الكلام نفسها يكفي إرسال وصف عن تلك الإشارة بحيث يمكن إعادة تكوينها من جديد عند الجهة المستلمة [9].

من جهة أخرى بدأ في الأربعينات العمل على كبس إشارة الكلام اعتماداً على ترميز صيغة الموجة (Waveform Coding) بطريقة تحميل الدلتا (Delta-Modulation) وذلك في بحوث شركة فليبس الهولندية.

وبعد أن اشتهرت عملية خزن الإشارة الصوتية بطريقة التحميل برمز الإشارة (PCM: Pulse Code Modulation) تم تطوير طريقة أخرى للكبس بهذا النوع وهي طريقة التحميل بالفارق التكيفي لرمز الإشارة (ADPCM: Adaptive Differential Pulse Code Modulation) التي أصبحت الطريقة القياسية والتقليدية لكبس إشارة الكلام في ذلك الوقت [9]. وفي بدايات القرن الحالي برز حقل استخلاص الخواص من بين الحقول المتعددة والمهمة لمعالجة إشارة الكلام، والذي تميز في مجال تمييز الكلام فقد شهدت نهاية القرن العشرين وبدايات القرن الواحد والعشرين تطوراً للعديد من طرائق استخلاص خواص الإشارة من بين هذه البحوث نذكر على سبيل المثال لا الحصر:

قام كل من (جونك هوان لي و هايونك مين بارك و هويونك جنك و سويونك لي) في عام (2002)، بتقديم تقرير تحت عنوان (Feature Extraction Using Independent Component Analysis) والذي تضمن اقتراح طريقة جديدة في مجال استخلاص الخواص استخدمت مبدأ تحليل المكونات غير المعتمدة (Independent Component Analysis)، إذ قامت هذه الطريقة بفصل كفاء لمكونات الضوضاء (التي اكتسبت إلى إشارة الكلام خلال تسجيله) عن إشارة الكلام الأصلية مع العمل على تسهيل العمليات الحسابية، وقد أظهر تطبيق الخوارزمية المقترحة نتائج أداء مقبولة [7].

وقام الباحثان (مارك سكورونسكي و جون هاريس) من جامعة فلوريدا في عام (2003) بنشر بحث تحت عنوان:

Improving the Filter Bank of a Classic Speech Feature Extraction algorithm
الذي قدم طريقة جديدة لتصميم مرشح (Filter Bank) جديد يستخدم في طريقة الـ (MFCC) التي تعمل على استخلاص خواص إشارة الكلام، وقد أظهرت نتائج الاختبار على تلك الطريقة أن استخدام المرشح الجديد أدى إلى زيادة كفاءة أداء أنظمة تمييز الكلام الأوتوماتيكية بشكل واضح [10].

أما من جامعة الموصل فقد قدمت الباحثة إسرائ عبد السلام عبد القادر عام 2001 بحثها الموسوم (كبس الصوت عند الزمن الحقيقي) الذي استخدمت فيه خوارزمية (LZW) في عملية كبس الصوت قبل تخزينه على ذاكرة الحاسوب والعمل على فك الكبس عند محاولة استماع الصوت مجدداً [2].

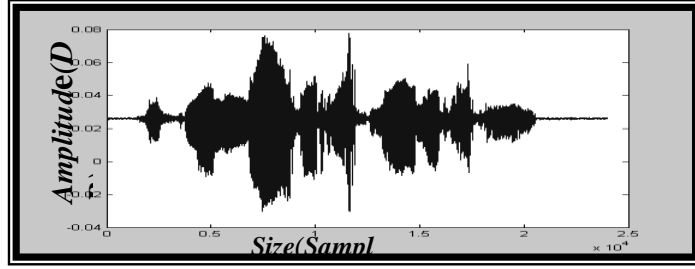
6 - خوارزمية الكبس وفك الكبس المقترحة:

أولاً: مراحل عملية الكبس:

تعتمد الخوارزمية المقترحة على تمثيل الكلام المكتسب إلى الحاسوب بصيغة رقمية في مجال الزمن، ثم اختزال أجزاء منها ومن ثم إيجاد تمثيل رياضي لمعاملات الزمن المتبقية التي يتم تخزينها لغرض اعتمادها عند العمل على استرجاع الكلام لتكوين عينات الإشارة التي تحول إلى الصوت الذي تم استقباله ثم تخزينه.

المرحلة الأولى: تبدأ مراحل الخوارزمية المقترحة بخطوة التهيئة، إذ يتم تهيئة متطلبات العمل بتحديد القيم المطلوبة لتنفيذ البرنامج وقد تمت تهيئة المثال المعتمد في وصف الخوارزمية حسب القيم الآتية:

- الوقت المستغرق للتسجيل: 3 ثوانٍ.
 - معدل التعيان (Sampling Rate): 8000KHz.
 - الحد الأدنى والأعلى لاتساع موجة صمت: 0.02dB، 0.03dB على التوالي.
 - عدد العينات المتكررة (العتبة) لاحتساب فترة الصمت: 10.
 - عدد النبضات (Sample) المعتمدة في انتخاب العينات : 2.
 - درجة المعادلة المعتمدة للمواءمة : 10 (والتي اختيرت بعد إجراء عدد من التجارب العملية وإيجاد الأفضل)
 - حجم المقطع الواحد للإشارة: 20 عينة لكل مقطع.
- المرحلة الثانية:** تبدأ عملية التسجيل باستخدام لاقط الصوت (الميكروفون) المربوط على بطاقة الصوت داخل الحاسوب، ومن البديهي معرفة إن حجم الإشارة بهذه الشروط يكون 24000 عينة (Sample)، الشكل (1) ناتج من رسم إشارة كلام مسجلة اعتماداً على ما تم ذكره.



الشكل(1): إشارة مسجلة بصوت رجل عند التلفظ بالبسملة

المرحلة الثالثة: تبدأ عملية خزن بيانات الإشارة الداخلة إلى ملف من نوع (wav)، ثم قراءة البيانات من الملف إلى مصفوفة أحادية البعد، وتجدر الإشارة إلى أن حجم هذه المصفوفة عند خزن الإشارة المذكور أنفأً فيها يساوي 187KB ولفترة زمنية مقدارها 3 ثوانٍ، وهي تمثل البيانات الحقيقية للكلام، ومن البديهي إن يزداد حجم الملف الصوتي عند زيادة الفترة الزمنية للتسجيل والعكس صحيح.

المرحلة الرابعة: تبدأ في هذه المرحلة أولى خطوات المعالجة المقترحة لغرض كبس بيانات الكلام الداخلة، إذ تتم إزالة فترات الصمت (Silent Period) الموجودة داخل إشارة الكلام التي تحقق الشروط الآتية:

- أن تكون ضمن مدى اتساع موجة الصمت المعتمدة (الحد الأدنى والأعلى المعتمدين).
- أن تساوي قيمة العتبة (عدد التكرارات) المثبتة أو تتجاوزها.

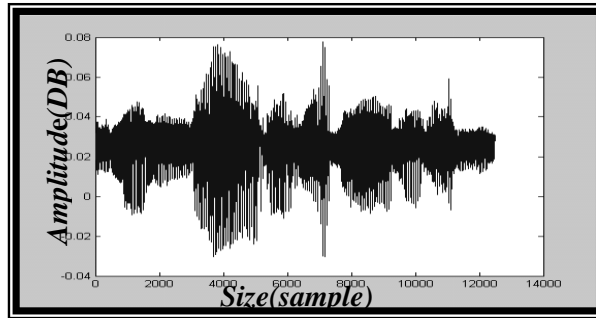
مع الاحتفاظ بعناوين بداية كل فترات الصمت المقطعة وحجمها في مصفوفة ذات بعدين (Indx)، علماً أن هذه المصفوفة تكون قد جهزت بقيمة ابتدائية (صفر) تحسباً لحالة عدم توافر فترات الصمت في الملف.

إن هذه المرحلة من المراحل المهمة جداً في خوارزميات كبس ملفات الكلام نسبة إلى طبيعة كلام الإنسان الذي يحوي العديد من الوقفات بين الكلمات أو العبارات أو فقرات الحديث، خاصة إذا كان وقت الكلام طويلاً نوعاً ما مثال ذلك أثناء إلقاء الخطابات أو أثناء عملية تغيير المتحدث [5].

عليه بالاعتماد على الإشارة السابقة وعندما تكون قيمة العتبة (10 عينة) سيكون ناتج إخراج هذه المرحلة المصفوفة الموضحة في الجدول (1)، والإشارة الموضحة في الشكل (2):

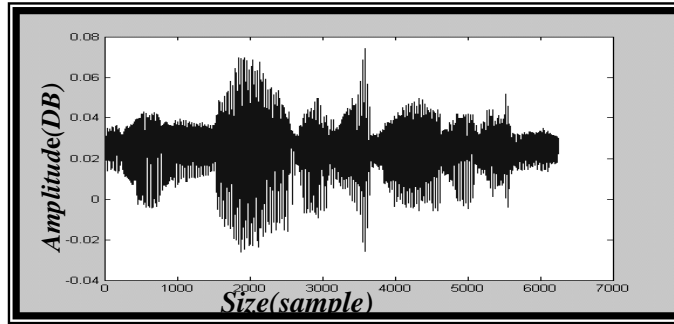
الجدول (1) هيئة المصفوفة الناتجة (indx)

عنوان بداية فترة الصمت	طول فترة الصمت
1	1914
1918	33
1952	10
⋮	⋮
20449	3558



الشكل (2): الإشارة بعد إزالة فترة الصمت منها وبقيمة عتبة مقدارها 10 عينة/Sample

المرحلة الخامسة: يتم في هذه المرحلة اقتطاع عينات محددة من الإشارة الناتجة لغرض معالجتها في المراحل القادمة عوضاً عن معالجة الإشارة كاملة، إذ تمثل الإشارة الجديدة الناتجة سلسلة من قيم جديدة كل قيمة فيها مساوية لمعدل عدد محدد من العينات الأصلية للإشارة والذي قد تم تثبيته في مرحلة التهيئة وخرنه في المتغير الافتراضي (Smpl)، الشكل (3) يوضح ناتج هذه المرحلة عندما تكون قيمة المتغير (Smpl) = 2:



الشكل(3): الإشارة الناتجة بعد اقتطاع عينات منها

في الشكل أعلاه يمكن وصل حجم الإشارة إلى ما يقارب 6000 عينة أي قد تم فقدان ما يعادل ثلثي حجم الإشارة الأصلية، فضلاً عن كون هذه المرحلة ذات تأثير واضح في مرحلة الكبس القادمة.

ومن الجدير بالذكر أن زيادة عدد العينات المقتطعة لحساب المعدل تؤدي إلى زيادة نسبة التشوه في الإشارة الناتجة بعد عملية فك الكبس، لهذا كان اختيار قيمة المتغير (Smpl) كما هي أعلاه ناتجا عن عدد من المحاولات مع دراسة التأثير في عامل التشوه في الإشارة (SEGSNR).
المرحلة السادسة: يتم في هذه المرحلة تقطيع الإشارة إلى عدد من المقاطع (Segment) وبحجم 20 عينة لكل مقطع، وخرن الناتج في مصفوفة ذات بعدين.

المرحلة السابعة: يتم في هذه المرحلة تطبيق طريقة مواعمة المنحنيات (Curves Fitting) لإيجاد أقرب منحنى يتواءم مع منحنى الإشارة الناتجة من الإشارة الأصلية وباستخدام إحدى الدوال الجاهزة في لغة (Matlab) التي استخدمت في برمجة الخوارزمية المقترحة، وتتضمن المرحلة في تطبيقها إجراء عملية كبس أخرى، فبعد عملية تقطيع الإشارة يتم اخذ كل مقطع وتطبيق أسلوب المواعمة عليه بمعادلة من الدرجة العاشرة، أي ستكون المعادلة الناتجة كالآتي:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \dots + a_9x^9 + a_{10}x^{10} \dots (1)$$

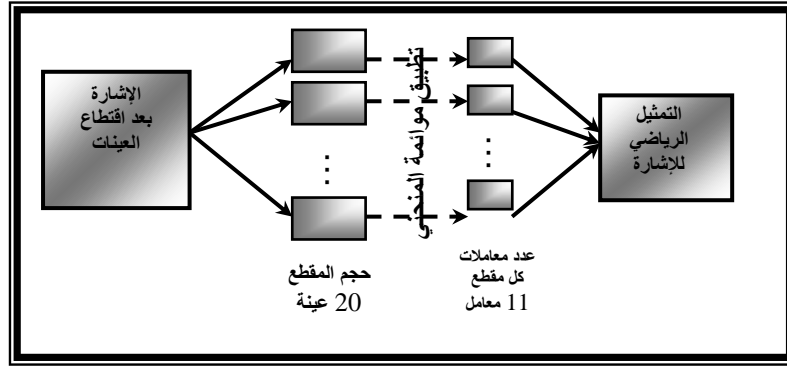
إذ أن a : تمثل قيم المعاملات المطلوب إيجادها.

$f(x)$: تمثل قيمة العينة عند النقطة x .

عليه سيكون ناتج هذه المرحلة لكل مقطع من الإشارة سلسلة من معاملات مواعمة المنحني التي يبلغ عددها 11 معاملاً، وبهذا يكون قد تم التعويض عن كل 20 عينة من الإشارة

بـ11 معاملاً، وفي هذه المرحلة يكون الملف قد فقد ما يقارب الـ45% من حجمه الناتج من المراحل السابقة عليه يكون قد تم اختزال في حجم الملف أي عملية كبس أخرى على البيانات، فضلاً عن ما تحققه هذه المرحلة من اختزال آخر في حجم الملف إلا إنها تضيف أسلوباً جديداً ذا سرية عالية غير مطروقة في أساليب ترميز الكلام وتشفيره ، كذلك أدت إلى دعم طرائق تخزين الملفات الصوتية من خلال تخزين معاملات مواءمة الإشارة عوضاً عن تخزين قيم الإشارة الصوتية نفسها أو باستخدام طرائق ترميز الكلام المعروفة.

تُخزن نتائج هذه العملية في مصفوفة تحت اسم (poly_coeff)، وبحجم (11*n)، إذ يمثل n عدد مقاطع الإشارة الكلي وكما موضح في الشكل(4).



الشكل(4): تطبيق مواءمة الإشارة

المرحلة الثامنة: وتتضمن عملية تخزين نتائج العمل وفق الخطوات الآتية:

- ترتيب نواتج العمل المهمة في مصفوفة موحدة تضم أولاً بيانات المصفوفة (indx) التي تم فيها تخزين عناوين وأطوال بدايات فترات الصمت المقطعة في المرحلة الرابعة من مراحل العمل، تليها نواتج عملية مواءمة منحنى الإشارة المخزونة في المصفوفة (poly_coeff).
- إعداد القيم التي سوف تمثل بادئة الملف الجديد، كما في الجدول(2):

الجدول(2): بادئة الملف (.SSC).

حجم الحقل	الاستخدام
8Byte	طول بادئة الملف
8Byte	طول بيانات الصوت الحقيقية
8Byte	نسبة التعيان
8Byte	الحد الأدنى لمدى اتساع موجة الصمت
8Byte	الحد الأعلى لمدى اتساع موجة الصمت
8Byte	عدد العينات المتكررة (العتبة) لاحتساب فترة الصمت
8Byte	عدد فترات الصمت المقطعة
8Byte	عدد النبضات (Sample) المعتمدة في انتخاب العينات
8Byte	عدد مقاطع الإشارة
8Byte	حجم كل مقطع
8Byte	عدد العينات المتبقية من تقطيع الإشارة
8Byte	درجة المعادلة المعتمدة للمواءمة
32Byte	محجوزة للاستخدام المستقبلي

- خزن نواتج الخطوتين السابقتين في ملف جديد وبصيغة خزن جديدة خاصة بالأسلوب المقترح، وقد تم اختيار الرمز (.ssc: Speech Signal Compression) ليكون امتدادا لهيئة ملف الخزن المقترح، يوضح الشكل(5) هيكلية الملف الناتج:

بادئ الملف الناتج (File Header)	16	24000	8000	0.02
	0.03	10	192	2
	312	20	9	10
	0	0	0	0
عناوين فترات الصمت المقطعة من الإشارة	1	1914	1918	33
	1952	10	1979	20
	⋮			
	20376	66	20449	3558
معاملات مواعمة منحنى الإشارة التي تمثل بيانات الكلام	666.32	3934.3-	5183.4	4594.7-
	1869.2	40.077	366.37	151.46
	⋮			
	91.097	4.8921-	0.11807	0

الشكل(5): هيكلية الملف الناتج

ثانياً: عملية فك الكبس:

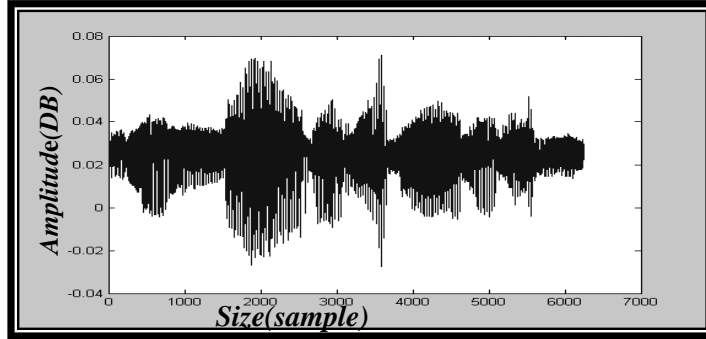
إن خطوات خوارزمية فك الكبس تأخذ مجرى التنفيذ المعاكس لخطوات خوارزمية الكبس، التي يمكن تلخيصها بما يأتي:

المرحلة الأولى: تبدأ عملية فك الكبس بقراءة الملف المكبوس من نوع (SSC). والحصول على المعلومات المهمة لإجراء عملية فك الكبس التي قد تم خزنها في بادئة الملف، واسترجاع المصفوفتين (indx) و (poly_coeff) من الملف بأبعادهما الأصلية.

المرحلة الثانية: تتم إعادة بناء الإشارة واسترجاع مكوناتها باستخدام معاملات مواعمة المنحني المسترجعة من الملف الناتج من عملية الكبس، إذ يتم تزويد المعادلة (1) بقيم معاملات المواعمة المسترجعة لكل مقطع مع قيم (x) التي تساوي:

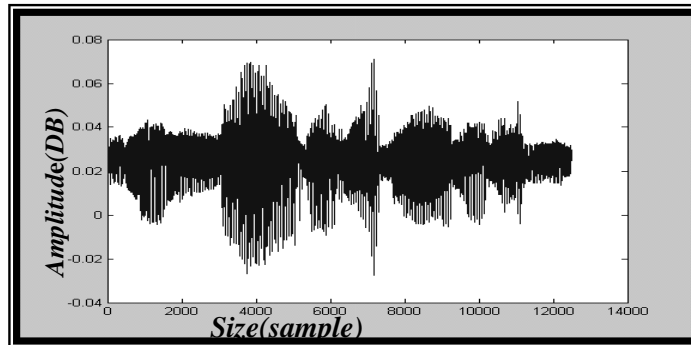
$$X = 0.05, 0.1, 0.15, 0.2 \dots 1.0$$

المرحلة الثالثة: تجميع المقاطع المسترجعة من المرحلة السابقة و تخزينها في مصفوفة ذات بعد أحادي لأغراض المعالجة اللاحقة لفك الكبس، يوضح الشكل (6) الإشارة الناتجة من عملية إعادة بناء الإشارة وتجميعها :



الشكل(6): الإشارة بعد إعادة بنائها وتجميعها

المرحلة الرابعة: استرجاع العينات التي تم اختزالها من الإشارة من خلال تكرار كل قيمة عينة مسترجعة من المرحلة السابقة لعدد من المرات مساوٍ لـ (Smpl-1)، وعليه يتم الحصول على الإشارة الموضحة في الشكل (7).

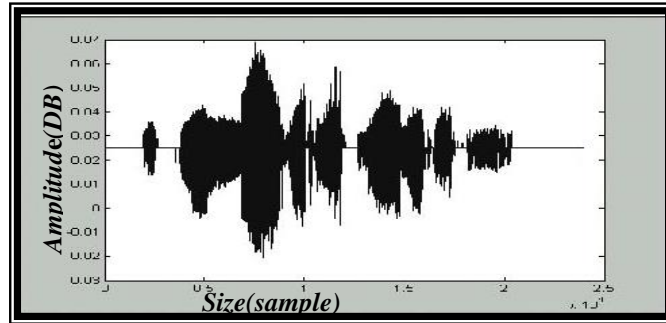


الشكل(7): استرجاع العينات المقطعة من الإشارة

المرحلة الخامسة: تبدأ عملية إرجاع فترات الصمت المقطعة إلى الإشارة، ويتم الاستعانة بشكل كلي على المصفوفة (indx) المسترجعة من الملف (.ssc)، وتمثل قيم عينات الصمت المسترجعة معدل

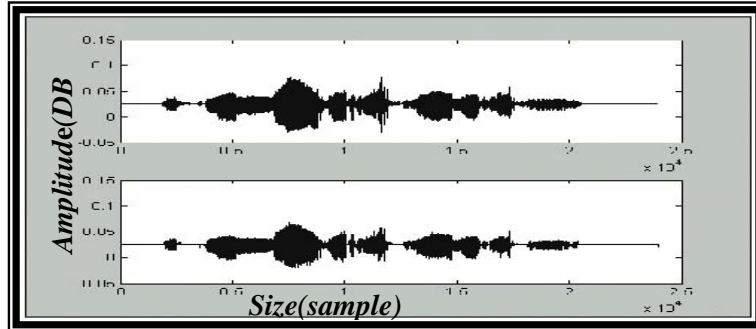
القيمتين الدنيا والعليا لاتساع موجة الصمت في الإشارة الأصلية اللتين قد تم تثبيتهما في بادئة الملف المستحدث.

المرحلة السادسة: بعد تنفيذ المراحل أعلاه يكون قد تم الحصول على الإشارة النهائية المسترجعة التي يوضحها الشكل(8)، وفي هذه المرحلة يتم خزن ناتج عملية فك الكبس المقترحة في ملف من نوع (wav) ثم إعادة تشغيل الملف واستماع إشارة الكلام الناتجة.



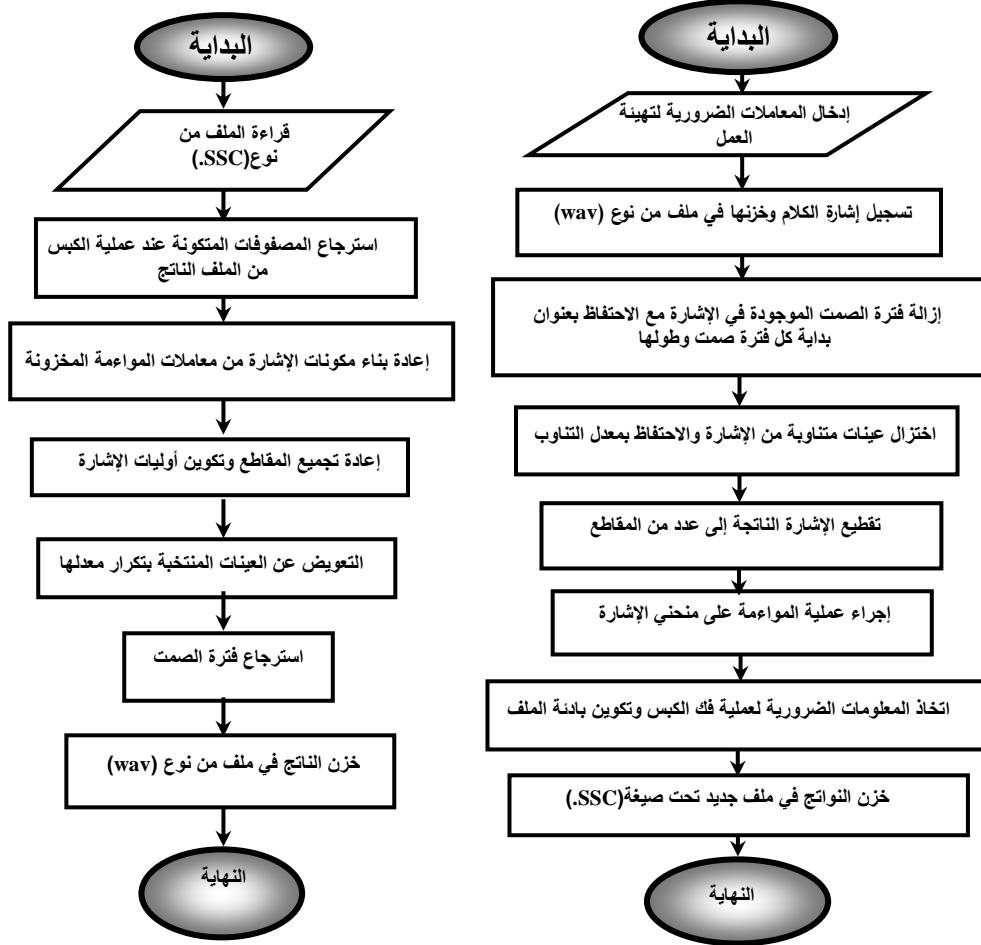
الشكل(8): الإشارة المسترجعة بعد إعادة فترات الصمت

وعند مقارنة النتائج من خلال رسم الإشارتين الأصلية والمسترجعة كليهما يتبين مقدار التقارب العالي بين شكلي الإشارتين، وكما هو موضح في الشكل (9) إذ يمثل الرسم الأول الإشارة الأصلية المسجلة للكلام، أما الرسم الثاني فيمثل الإشارة المسترجعة من عملية فك الكبس وفك الكبس.



الشكل(9): الإشارة الأصلية والإشارة المسترجعة من عملية فك الكبس وفك الكبس

المخطط الصندوقي للخوارزمية المقترحة:



(ب) عملية فك الكبس

(أ) عملية الكبس

الشكل (10) المخطط الصندوقي للخوارزمية المقترحة

7- مناقشة النتائج : تمت دراسة مدى كفاءة الخوارزمية المقترحة على ملفات مختلفة وذلك باعتماد الاختبارات الآتية التي تعمل على دراسة نسبة الكبس وجودة الكلام المسترجع:

أولاً: اختبار مدى كفاءة الكبس:

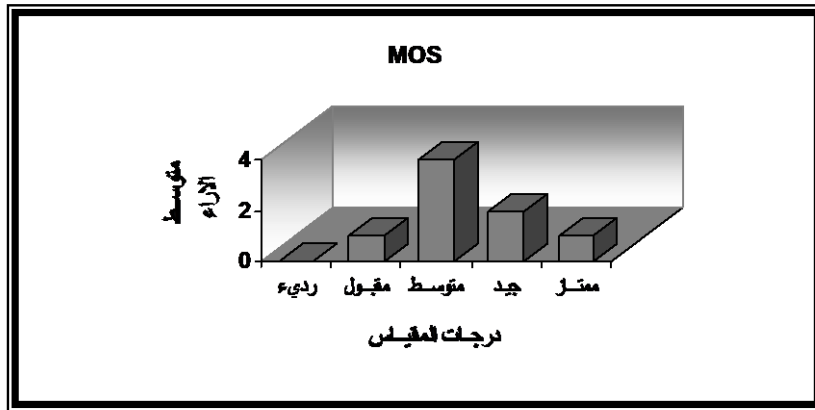
من المعروف إن قياس مدى كفاءة أي خوارزمية للكبس يعرف من خلال حساب نسبة الكبس من المعادلة (2).

$$\text{نسبة الكبس} = (\text{حجم الملف بعد الكبس} / \text{حجم الملف قبل الكبس}) * 100\% \dots (2)$$

عليه تكون نسبة الكبس الناتجة من تطبيق الخوارزمية على المثال السابق 22.45%، أما نسبة الكبس التي يوفرها ملف الخزن (wav) على المثال السابق فهي 24.9%

ثانياً: اختبار جودة الكلام المسترجع:

يمكن معرفة جودة الكلام الناتج من تطبيق الخوارزمية المقترحة بالاعتماد على المقاييس الشخصية والموضوعية لقياس جودة الكلام المسترجع، وقد تم تطبيق مقياس (MOS: Mean Opinion Score) على الإشارة، (وهو مقياس يحوي خمس درجات يعتمد فيه على آراء المستمعين للإشارة الأصلية والمسترجعة واعطاء نسبة لمدى جودة الكلام المسترجع) [4] [8] [11] والحصول على النتائج الموضحة في الشكل (11):



الشكل(11): قياس درجة جودة الإشارة المسترجعة باستخدام مقياس MOS

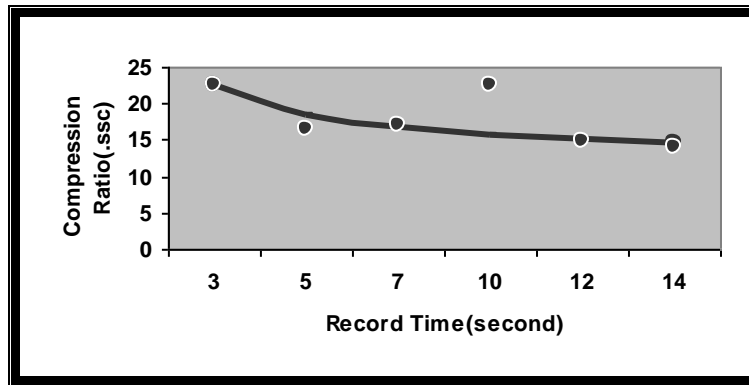
أما عند استخدام طريقة الاختبار الموضوعي لقياس جودة الكلام المسترجع من اعتماد المثال السابق فإن نتيجة حساب (SEGSNR :Signal to Noise Ratio) [4] تكون مساوية لـ(30.3011dB).

ومن خلال تطبيق الخوارزمية المقترحة على عدد من الأمثلة يمكن ملاحظة النتائج التي حققتها الخوارزمية المقترحة مقارنة مع ملف الصوت (wav) التي يوضحها الجدول (3):

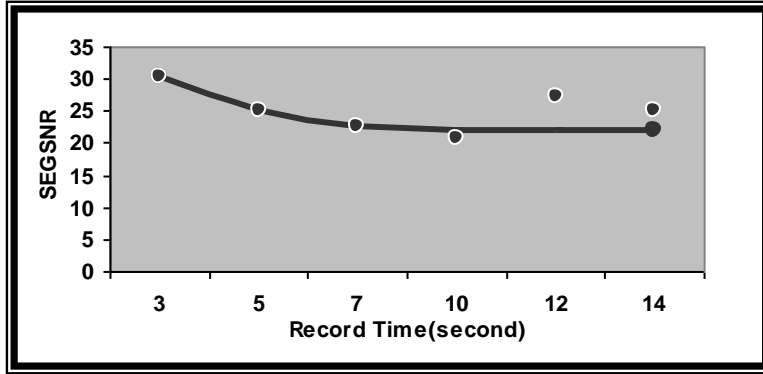
الجدول (3): النتائج التي حققتها خوارزمية الكبس المقترحة

مقياس التشوه (dB)	نسبة الكبس		حجم الملف			مدة التسجيل (sec)	ت
	(.wav)	(.ssc)	.wav	بعد الكبس بالطريقة المقترحة	قبل الكبس		
30.3	24.9%	22.6%	46.9	42.4	188	3	1
25.07	24.9%	16.6%	78.1	51.7	312	5	2
22.6	24.9%	17.0%	109	74.3	497	7	3
20.8	24.9%	22.6%	156	141	625	10	4
27.3	24.9%	14.8%	187	111	750	12	5
25.1	24.9%	14.1%	218	123	875	14	6

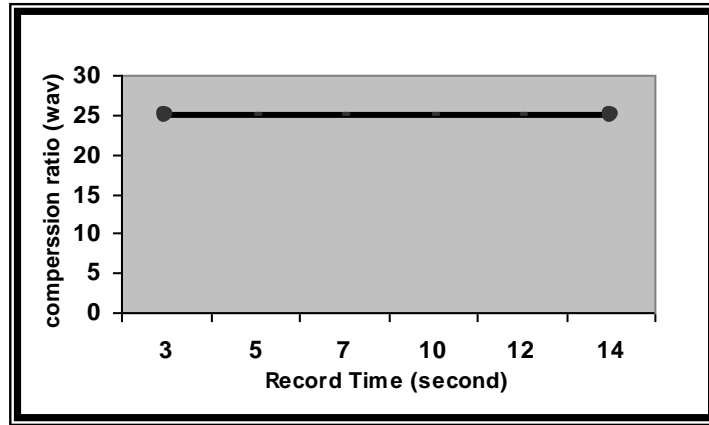
ويمكن إعطاء توضيح أكثر للنتائج التي تم التوصل إليها عن طريق المخططات الآتية:



الشكل (12): نسبة الكبس المحققة وعلاقتها بالحجم



الشكل(13): نسبة التشوه وعلاقتها بحجم الملف

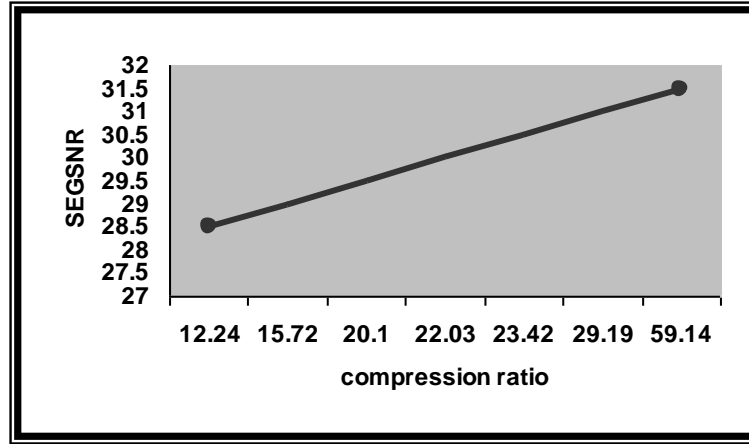


الشكل(14): نسبة الكبس في الملف (wav) وعلاقتها بالحجم

ثم طبقت الخوارزمية على مثال واحد بقيم مختلفة في مرحلة التهيئة وتم الحصول على النتائج التي يوضحها الجدول(4) والشكل (15):

الجدول(4): العلاقة بين نسبة الكبس ونسبة التشوه المحققتين

ت	نسبة الكبس (%)	SEGSNR(dB)
أولاً	12.24	28.03
ثانياً	15.72	29.72
ثالثاً	20.1	29.43
رابعاً	22.03	30.42
خامساً	23.42	30.42
سادساً	29.19	30.63
سابعاً	59.14	31.79



الشكل (15): العلاقة بين نسبة الكبس ونسبة التشوه المحققتين

وعند دراسة المنحنيات المذكورة في الأشكال (12)، (13)، (14) و (15) فضلاً عن الجدولين (3) و(4) يتضح أن الخوارزمية المقترحة قد أعطت نتائج جيدة جداً، إذ تبين أن نسبة الكبس تؤول إلى الاستقرار عندما يبدأ حجم الملف بالاستقرار مما يجعل منها ذات فائدة كبيرة مع الملفات ذات الأحجام الكبيرة التي تكون الحاجة ملحة إلى كبسها، فقد طبقت الخوارزمية المقترحة على أحجام مختلفة تراوحت بين KB188 إلى KB875 وكانت نسبة الكبس مستقرة تقريباً.

كذلك لوحظ من النتائج المبيّنة أن نسبة الكبس ذات علاقة خطية مع نسبة التشوه، وكون ميل تلك العلاقة قليل جداً إذ أن نسبة التشوه تزداد بمقدار قليل قد لا يصل إلى ثلث قيمة الزيادة بالحجم وهذا ناتج إيجابي فمن المتوقع أن تكون هناك زيادة في نسبة التشوه مصاحبة للزيادة في حجم الملف لكن بسبب اختزال فترات الصمت وإعادتها بصورة قريبة جداً من قيمها الأصلية أدى إلى ضغط نسبة التشوه وجعلها لا تتسارع تسارعاً عالياً مع زيادة حجم الملف مما يعطي دفعاً إيجابياً على النتائج العملية التي تم الحصول عليها عند تطبيق الخوارزمية.

كما أتضح من نتائج تطبيق الخوارزمية على عدد من الأمثلة المختلفة أن الخوارزمية لا تتأثر بصفة المتكلم أو نوع الكلام وصيغته المعتمدة في التطبيقات العملية للخوارزمية، إذ انتخبت نماذج مختلفة من الأصوات لرجل وامرأة وطفل فضلاً عن اختلاف أساليب إلقاء الكلام التي تباينت بين الأسلوب الخطابي والتلاوة القرآنية والقراءة الاعتيادية واختلاف نبرات الصوت ومخارج نطق الحروف، ولم يسجل على الخوارزمية أي اختلاف في نسبة الكبس أو نسبة التشوه مع اختلاف نوع الكلام أو صيغته وإنما أعطى نتائج متقاربة من بعضها البعض.

8- الاستنتاجات:

- من خلال التطبيق العملي للخوارزمية المقترحة في هذا البحث تم التوصل إلى الاستنتاجات الآتية :
- اعتماد فكرة استخلاص خواص الكلام لغرض كبسه أعطت نتائج كبس عالية على الرغم من وجود تشوه مقبول نوعاً ما عند كبس كلام لأغراض التراسل.
- إن الخوارزمية قد طرقت الباب على أسلوب غير تقليدي لأجل كبس إشارة الكلام المسموع.
- إن الخوارزمية المقترحة ذات إستقرارية مع الملفات الكبيرة الحجم.
- أعطت الخوارزمية المقترحة نتائج إيجابية من حيث نسبة التشوه عند تغيير حجم الملف المكبوس.
- لقد حافظت الخوارزمية المقترحة على نسبة عالية جداً من الكلام، إذ أن النماذج والفترات التي يتم اختزالها دون معالجة يتم استرجاعها بشكل قريب جداً من قيمها الأصلية عند عملية إعادة بناء الكلام.

- من النقاط المهمة التي أعطتها الخوارزمية المقترحة سرية الكلام المرسل أو المخزون إذ يكون من الصعب اكتشاف ان محتويات الملف تعود إلى كلام مسموع، عليه من الممكن تطويرها لكي تعتمد كأحدى وسائل سرية المعلومات.
- أضافت الخوارزمية المقترحة أسلوباً جديداً إلى أساليب خزن إشارة الكلام بخزن معاملات موءمة منحني الإشارة فضلا عن الطرائق التقليدية المعروفة.
- يلاحظ في ملف الخزن (wav) انه لا يستغل كونه حاوياً على بيانات كلام وهذا من نقاط الضعف التي تسجل على الملف.

9- العمل المستقبلي:

- فيما يأتي بعض المقترحات للعمل المستقبلي الذي يمكن أن يبني على ما تم اقتراحه وبناءؤه ومن ثم تطبيقه :
- إن الخوارزمية المقترحة تفتح باباً يمكن استغلاله لغرض تشفير الكلام المسموع والحفاظ على سرية المعلومات التي يحتويها ذلك الملف المكبوس.
 - يمكن العمل على تحسين أداء الخوارزمية من خلال اعتماد طرائق أخرى في موءمة الملفات لأجل الحصول على معادلات رياضية قريبة جداً من العلاقة الرياضية التي تمثل إشارة الكلام الأصلية.
 - العمل على تطوير الوسائل الديناميكية الخاصة بتحديد واكتشاف فترة الصمت للحصول على إمكانيات عالية لأجل اختزال فترات الصمت.
 - العمل على تحسين نسبة الكيس من خلال إيجاد النسبة المثالية للتقطيع وذلك بدراسة ملفات ذات مواصفات مختلفة.

المصادر

- [1] الحديدي، إبراهيم احمد صالح احمد (2003) "كبس الصور الرقمية باستخدام تحويلات الموجة والمكتم الإتجاهي"، بحث ماجستير، العراق ، جامعة الموصل، كلية علوم الحاسبات والرياضيات.
- [2] عبد القادر، إسراء عبد السلام (2001) "كبس الصوت عند الزمن الحقيقي"، بحث ماجستير، العراق ، جامعة الموصل، كلية علوم الحاسبات والرياضيات.
- [3] Brooks, David W.; Carr, Adam and Edkins, Keith (2004) “data Compression”, Boston, U.S.A., **Wikipedia the Free Encyclopedia, GNU Free Documentation License**, <http://en.wikipedia.org/wiki/>
- [4] Hanzo, Lajos; Somerville, F. and Woodard, Jason P. (2001) “Voice Compression and Communications”, **IEEE press, Wiley Inter science**, NY 10016-5997.
- [5] Kientzle, Tim (1997) **programmer’s Guide to Sound**, Addison_ Wesley Developers Press, ISBN 0-201-41972-6.
- [6] Kivimaki, Jukka (2000) “Very low bit rate speech coding using speech recognition, analysis and synthesis” , M. Sc. Thesis, **Tampere university of technology, Department of information Technology**.
- [7] Lee, Jong-hwan ; Park, Hyung Min and Jung, Ho-Young (2002) “Feature Extraction Using Independent Component Analysis”, Korea, **Brain Science Research Center (BSRC)**. [http://bsrc.kaist.ac.kr/braintech/image/reports/1-year/eng0101a01/eng0101a01-14 .htm](http://bsrc.kaist.ac.kr/braintech/image/reports/1-year/eng0101a01/eng0101a01-14.htm).
- [8] Liesenborgs, Jori (2000) “Voice over IP in networked virtual environments”, Ph.D. Thesis, Belgium, **Limburgs University Centrum, Limburgs**, May.
- [9] Schroeder, Manfred R. (1999) “Speech Processing”, Germany, University of Gottingen , **IOS press**. [http://www.cs.umb.edu/~asi/multimedia/papers/Schroeder .pdf](http://www.cs.umb.edu/~asi/multimedia/papers/Schroeder.pdf).

- [10] Skowronski, Mark D. and G. Harris, John (2003) “Improving The Filter Bank Of Classic Speech Feature Extraction Algorithm”, Bangkok, Thailand, **IEEE Intl Symposium on Circuits and Systems**, Vol. IV, PP. 281-284, May 25 - 28, 2003, ISBN: 0-7803-7761-31.
- [11] Stein, Jonathan Y. (2000) **Digital Signal Processing**, John Wiley & Sons, inc., USA.