Detection and Classification of Emotional State Based on Speech Signal

Hiba Adreese Younis **Mrewan Mustafa** **Rahma Raad**
hibaadreese@uomosul.edu.iq
College of Computer Science and Mathematics
University of Mosul, Mosul, Iraq

Received on: 13/03/2018 **Accepted on: 24/01/2019**

ABSTRACT

In this research, an algorithm was proposed to automatically classify the mood of the speaker by referring to his speech. Three moods were adopted in this study, namely joy, sadness and anger in order to distinguish between them.

The principle of the algorithm work includes the initial treatment of the signal of by removing the silence and then cut the signal to a number of sections length of each 512 sample, and then treatment by window (Hamming window) followed by the process of extracting the characteristics such as energy, the basic frequency, resonance frequencies of each section and for all speech signals that Were recorded, which included 30 signals of persons between 15 and 25 years of age in order to prepare the database for the three moods and to draw the characteristic curves and for each mood.

The selection of signals was done from training and testing set for detectingthe mood of these signals by performing the previous steps and then comparing the resulting curve with the previous curves using the correlation coefficient and the Euclidean distance.

The algorithm gave good results when these characteristics were adopted in the classification process and by about 75%.

keywords: emotional states, harmonic product spectrum, linear predictive coding

تحديد وتصنيف الحالة المزاجية بالاعتماد على معاملات إشارة الكلام

هبة أدريس يونس **مريوان مصطفى** **رحمة رعد**
كلية علوم الحاسوب والرياضيات
جامعة الموصل، الموصل، العراق

تاريخ استلام البحث: 13\03\2018 **تاريخ قبول البحث: 24\01\2019**

الملخص

اقترح في هذا البحث خوارزمية تعمل على التصنيف الآلي للحالة المزاجية للمتكلم بدلالة إشارة كلامه. إذ اعتمد ثلاث حالات مزاجية في هذا البحث وهي الفرح والحزن والغضب من أجل التمييز بينها.

مبدأ عمل الخوارزمية يتضمن المعالجة الأولية لإشارة الكلام من إلغاء حالة الصمت ومن ثم تقطيع الإشارة إلى عدد من المقاطع طول كل منها512 عينة، ومن ثم المعالجة بالنافذة (Hamming Window) تليها عملية استخلاص الخصائص المتمثلة بالطاقة، التردد الأساسي، والترددات الرنانة من كل مقطع ولجميع إشارات الكلام

التي سجلت والتي شملت30 إشارة كلام لأشخاص تتراوح أعمارهم بين 15و25 من الإناث، لأجل تهيئة قاعدة البيانات للحالات المزاجية الثلاث و رسم منحنيات الخواص ولكل حالة مزاجية.

تتم عملية اختيار الإشارات من قاعدة البيانات من مجموعة التدريب والاختبار لتحديد حالتها المزاجية بإجراء الخطوات السابقة عليها ومن ثم مقارنة المنحني الناتج مع المنحنيات السابقة باستخدام مقاييس معامل الارتباط والمسافة الإقليدية.

وقد أعطت الخوارزمية نتائج جيدة عند اعتماد تلك الخصائص في عملية التصنيف وبنسبة 75% تقريباً.

الكلمات المفتاحية: الحالات المزاجية، خوارزمية ضرب توافقيات الأطياف، خوارزمية ترميز التنبؤِ الخطي.

1.المقدمة Introduction

تلعب العاطفة دوراً مهماً في حياتنا اليومية الشخصية. وقد أظهرت الأبحاث السابقة أنه يمكن التعرف على العواطف ليس فقط من الوجوه، ولكن أيضا من الأصوات. تماما مثل الميزات المادية في الوجه ، هناك مزيج من الميزات السمعية في الصوت مثل خاصية التردد الأساسي(pitch) وشدة الصوت(intensity) التي يمكن أن تعطي رؤية واضحة عن الحالات العاطفية الكامنة فيالفرد[1].

إن التمييز الآلي للحالة المزاجية من الكلام أصبح ذا أهمية متزايدة في السنوات الأخيرة في حقول ومجالات واسعة من التطبيقات التي استفادت من هذه التقانة[2].

إن عملية تمييز الحالة المزاجية غالباً ما تتضمن تحليل مختلف التعبيرات البشرية في نماذج وصيغ متعددة الوسائط مثل النصوص،أو الصوت، أو الفيديو. يكتشف أنواع مختلفة من الحالات المزاجيةبدمج المعلومات من تعابير الوجه وحركة الجسم والإيماءات والكلام[3].

إن العاطفة هي جزأ لا يتجزأ من قراراتنا اليومية العقلية والذكية، فهي تساعدناعلى التواصلمعبعضنا البعض بالتعبير عن مشاعرنا. هذا الجانب الهام منالتفاعل البشري يحتاج إلى الأخذ بنظر الاعتبار تصميم واجهات بين الإنسانوالحاسبة(Human Computer Interface(HMI)، لبناءالواجهات التيهيأكثر انسجاماً مع احتياجاتالمستخدمين[4].

يمكن تصنيف الطرائق الحالية في التعرف على الحالة المزاجية لتصنيف أنواع معينة منها إلى ثلاث فئات رئيسة: التقانات القائمة على المعرفة والأساليب الإحصائية والطرائق التي تجمع بين الاثنين[3].

2.اشارة الكلام (speech signal)

إشارة الكلام هي عبارة عن إشارة تناظرية(analog)، التي من الممكن تحويلها إلىإشارة رقمية بسهولة . أحد خصائص إشارة الكلام هي السعة التي من الممكن الحصول عليها بسهولة خلال إشارة الكلام الرقمية المسجلة. الخاصية الأخرى هي التردد والتي نحصل عليها عادة باستخدام عدة طرائق منهاترميز التنبؤ الخطي(Linear Predictive Coding).

إنإشارة الكلام التي ينتجها البشر تحوي معلومات أكثر من مجرد الرسالة التي يتم انتقالها فهي تعطي المستمع فكرة عن جنس المتكلم أيضاً وحالته العاطفية، لغة الاتصال ، فضلا عن هوية المتكلم في بعض الحالات، كل هذه المكونات منالمعلومات يعالجها المستمع، وبعد ذلك توضع سوية في الدماغ لتوليد صورة كاملة عن محادثة معينة[5] .

3.أصنافالمعلوماتفيإشارة الكلام

يمكن تصنيف المعلومات في إشارة الكلام إلى ثلاثة مستويات:[5]

1.معلوماتلغوية linguistic information):

وتشملالكلمات(words)،المقاطع(syllables)،العبارات(phrases)التي يتكون منهاالكلام.

2. معلومات لغوية إضافية (paralinguistis information):

هذه المعلومات تمتلك مهمتين:

أولاً: إنهالايمكن استنتاجها من النظير المكتوب وهي تضاف بتعمد لدى المتكلم.

ثانياً: قوة نوعية هذه المعلومات تتباين وتختلف بشكل مستمر ضمن الصنف نفسه.

3. معلومات غير لغوية(non linguistic information):

وهي المعلومات التي لايمكنأن يضيفها المتكلم بتعمد مثل الجنس،العمر،الحالة الصحية.هذه المعلومات يمكن تصنيفها إلى:

أولا: التغييرات الفردية المتعلقة بعلم وظائف الأعضاء وتشمل حجم ووزن وما يتعلق بعلم الأنسجة (عمر) المنطقة الصوتية والتي تؤثر على الرنين في الكلام.

ثانياً: ردود الأفعال التلقائية حسب الحالة العاطفية للمتكلم وتشمل الضحك،البكاء،التثاؤب،التنفس السريع.

4.الدراسات السابقة

✓ قام الباحث سلر وآخرون(2002)، بتمييز العاطفة عن طريقة تحليل الكلام وحالة المستخدم لكلتا الحالتينبدراسة معاني الكلمات ومستوى الإشارة.وصنفت إلى سبعحالات(JOY فرح، ANGER غضب,IRRITATION الانزعاج، FEAR خوف، DISGUST الاشمئزاز، SADNESS الحزن،NATURAL, INNER STATE الحالات الاعتيادية) ، واستخلص التردد الأساسيPITCH والطاقة ENERGY، فضلاً عن مدة الخصائص DURATION FEATURE، وكانت نسبة تمييز الحالات هي على التوالي وعلى النحو الآتي : 69%, 73%، 93%، 91%، 88%، 64%[6].

✓ في البحث الي قدمه الباحثان ماريوسو مونيكا ،2012 أوجدت مجموعة من المعاملات المتعلقة لتمييز العاطفة، استخدمت قاعدة البيانات(Sozialberatung Region ObererLeberberg)SROL(لتمييز الحالات وهي الفرح ، الخوف ، الحزن والحالات الاعتيادية، وكانت دقة التميز لتلك الحالات هي63.97% ، وتمت عملية التصنيف للبيانات المدخلة باستخدام خوارزمية (LPC) وكانت الخصائص المستخلصة هي التردد الأساسي F0 والترددات الرنانةF1، F2 وخاصية الزفير (JITTER) والوميض(SHIMMER)[7] .

✓ قام الباحثان (ميانك وتيم ،2012)بتحسين التعرف التلقائي للعاطفة من اشارة الكلام بدمج خصائص الإيقاع والميزات الزمنية. وقد استند البحث على التعرف التلقائي على المشاعر إلى تطبيق ميزات مثل MFCC، التردد الأساسي، الطاقة والكثافة .وتركز الفكرة على اقتراض ميزات الإيقاع من التحليل اللغوي والصوتي وتطبيقها على إشارة الكلام على أساس المعرفة الصوتية فقط .فضلاً عن الاستفادة من مجموعة من الميزات الزمنية والجهارة.

تم استخدام مصنفات مختلفة بعد اختيار أعلى الميزات باستخدام مرشح(Information Gain Ratio (IGR،حيث كانت نسبة التمييز 80.60٪ على قاعدة بيانات برلين المعتمدة على المتكلم[8].

✓ قامالباحث (لي واخرون،2014) باقتراح طريقة حزمة الموجة المعتمدة على مصرف المرشح المحسن(adaptive(filter bank، اذ استخلصت مجموعة من الخصائص السمعية مثل (CWPCC)، وقد أثبتت النتائج أن Coiflet wavelet packet أكثر ملاءمة في تمييز الحالة المزاجية من باقي الحزم الموجية. كما أن الخصائص المقترحة حسنت عملية التمييز مقارنة بالخصائص القياسية[9].

- ✓ قام الباحث(كلوغه وآخرون، 2017) باستخدام ماكنة التعلم الواسع(extreme learning machine) كمحلل آلي لعملية التمييز لإشارة الكلام، إذ طبقت على ثلاث قواعد بيانات مختلفة ، وقورنت النتائج التي تم الحصول عليها مع نتائج كل منsupport vector machine وطريقة deep learning المعتمد على تحليل التمييز العام وحسنت نسبة التمييز بمقدار 3%-14% مقارنة بطريقة support vector machine بينما حسنت نسبة التمييز بمقدار 8%-13% مقارنة بطريقة تحليل التمييز العام[10].
- ✓ قام الباحث (كركيني وآخرون،2018) بمقارنة طرائق مختلفة لعملية تمييز الحالات المزاجية. إذ اقترح حل كفوء بالاعتماد على الدمج بين تلك الطرائق. استخدمت الشبكات العصبية ذات الارتداد العكسي(Recurrent Neural Network) لتصنيف 7 حالات مزاجية وقورن الأداء مع طرائق أخرى مثل Multivariate linear Regression وsupport vector machine واستخلص كل من MFCC وMSFs ، إذكانت نسبة التمييز باستخدام مصنف RNN هي 90.05% بينما كانت نسبة التمييز 82.41% باستخدام مصنف [11] MLR.

ويرى الباحث أن هناك عدة طرائق وخوارزميات لتحديد الحالة المزاجية بالاعتماد على استخلاص خصائص متعددة منها تعمل في مجال الزمن ، ومنها في المجال الترددي. واختلفت عدد الحالات المزاجية التي صنفت وأيضا طرائق التصنيف فضلاً عن قاعدة البيانات المستخدمة، إذ إن لكل دراسة لها معاملات خاصة بها. ودمج في هذا البحث بين الخصائص الزمنية والخصائص الترددية من أجل اعتمادها في المرحلة اللاحقة في عملية التصنيف التي أعطت نتائج جيدة بالاعتماد على قاعدة البيانات المكونة.

5.استخلاص الخصائص

تمثل مرحلة تحليل الاشارة لاستخلاص الصفات الصوتية بدقة التي تمثل المعلومات الصوتية (acoustic information) لتلك الإشارة، إذتعالجالإشارة من البداية إلى النهاية وذلك بتقسيم الإشارةإلىأطر (frames) متساوية في الطول ثم تستخلص الصفات لكل قالب لتكون النتيجة النهائية متجهة من الصفات يمثل ذلك الإطار. وهناك عدة طرائق تستخدم لاستخلاص الصفات لأيةإشارة صوتية فيمكن استخدام الطاقة (energy), معدل السعة (average magnetite)، معدل تقاطع الصفر (zerocrossing) معاملات التنبؤ الخطي ((Lpc) Linear Predictive Coefficients)[12] .

6.خوارزميةproduct spectrumharmonicلايجاد التردد الاساسي

ان التردد الأساسي يحوي معلومات عن المتكلم مثل الجنس، العمر، والحالة العاطفية. وهناك عدة طرائق لحساب قيمة التردد الأساسي، وبالاعتماد على طريقة الحساب يمكن تقسيمها إلى تحديد التردد الأساسي في مجال الزمن، تحديد التردد الأساسي في مجال التردد فضلاً عن تحديد التردد بالاعتماد على معاملات ال(cepstral)[13] وأحد طرائق تحديد وحساب التردد الأساسي في مجال التردد هي خوارزمية ضرب توافقيات الأطياف(**productspectrumharmonic**).

في هذه الخوارزمية تقطع الإشارةإلى مقاطع segments بطول 512عينة، ولكل مقطع يطبق تحويل فورير القصير(Short Fourier Transform) من أجل تحويل الإشارة من مجال الزمن إلى مجالالتردد. وحالما تكون الإشارة في مجال التردد تطبق العمليتان الآتيتان لكل مقطع:

• **عملية الاختزال(downsampling):** ولإجراء هذه العملية نقوم بكبس الطيف لكل مقطعبمقدار مرتين تارة وثلاث مرات تارةأخرى، أي أنه في المرحلة الأولى تختزل قيمتان وتبقى واحدة وفي المرحلة الثانية تختزل ثلاث قيم وتبقى واحدة.

• **عملية الضرب:** وتتضمن عملية ضرب الأطياف الثلاثة سوية ومن ثم إيجاد الترددالذي يقابل أعلى قيمة(peak) والذي يمثل التردد الأساسي[14].

محاسن الخوارزمية

- غير مكلفة حسابياً
- ملائمة لمختلف أنواعالإدخال
- ممكن نغير عدد الاطياف المكبوسة
- ممكن إحلال عملية جمع الأطياف بدل ضرب الأطياف

7.خوارزمية LPCلإيجاد الترددات الرنانة :

تعرف الترددات الرنانة بأنها القمم الطيفية في طيف الصوت. وتشير الترددات الرنانة في علم الكلام والصوتيات الى الرنين الصوتي للقناة الصوتية للشخص، إذ تمثل قمم السعة للطيف الترددي لموجة الصوت.

تعد تقانة التنبؤ الخطي (LPC) احدى التقانات المستخدمة لتحليل الكلام للحصول على الصفات المميزة لإشارة الكلام وتعد أنموذجا تحليلياً سهلاً مقارنة بطرق تحليل أخرى, وهو انموذجا دقيق رياضياً وسهل التنفيذ، كما أنها تعطي أنموذجاً جيداً لإشارة الكلام, فالطريقة التي بها تحلل إشارة الكلام تؤدي الى فصل معقول للمناطق الصوتية للكلام على الرغم من أن هناك طرقاً عديدة لمعالجة الإشارة من البداية إلى النهاية إلا أإن ترميز التنبؤ الخطي أثبت نفسه وذلك بتحقيقه لنتائج جيدة بهذا المجال وهي طريقة رياضية دقيقة و بسيطة ويستطاع استخدامها برمجياً في الكيانات المادية وللحسابات التابعة لها تكون على العموم أقل منالتي تحتاجها المرشحاتويعد من أهمالأساليب المستخدمة لمعالجة الإشارة. يستخدم مرشح التوقع الخطي لوصف علاقة إخراج متغير في النظام إلى واحد أوأكثر من الإدخالات التي تعد محددات لوصف*النظام والتي تحكم بمعادلات تفاضلية اعتيادية.*

إن طريقة ترميز التنبؤ الخطي المتغير مع الزمن والذي يمثل المنطقة الصوتية تستخدم خوارزمية LPCلإيجاد الترددات الرنانة لإشارة الكلام عن طريق دالة الاختلاف التي تمثل كل عينة من الإشارة على أنها مجموعة خطية من العينات السابقة، ويتم إيجاد معاملاتدالة الاختلاف (معاملات التوقع)عن طريق إيجاد جذور متعددة الحدود والتي تمثل الترددات الرنانة،إذأن كل زوج من الجذور المعقدة تستخدم لحساب الترددات الرنانة وعرض الحزمة المقابل لها. وتخمن هذه المعاملات عن طريق جعل معدل مربع الخطأ ما بين الإشارة المتوقعة والحقيقية أقل ما يمكن[15].

8. الخوارزمية المقترحة لتصنيف الحالة المزاجية بالاعتماد على معاملات إشارة الكلام

تتكون الخوارزمية المقترحة من مرحلتين رئيستين:

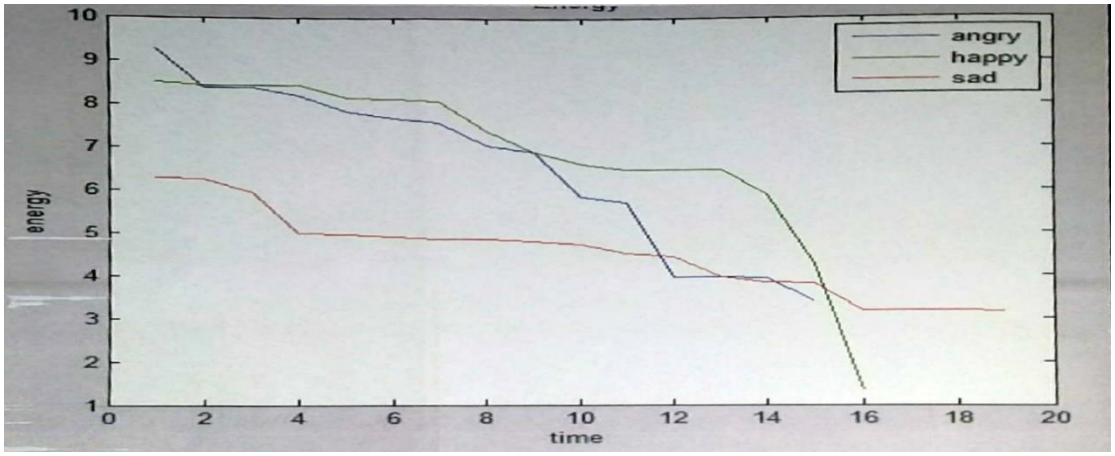
المرحلة الأولى: مرحلة إعداد قاعدة البيانات

تتكون هذه المرحلة من عملية تسجيل مجموعة من إشارات الكلام العشوائية لأصوات المتكلمين(الإناث) تتراوح أعمارهم بين(15−25) سنة، إذ تمت عملية التسجيل في غرفة ذات ظروف ملائمة (نسبة الضوضاء والصدى

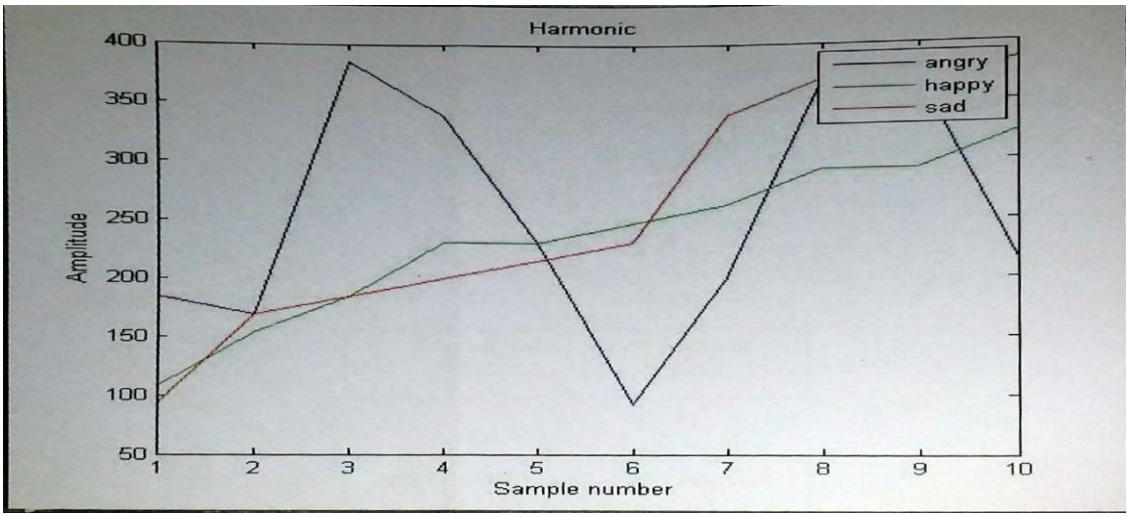
قليلة) باستخدام جهاز الموبايل النقال ذات المواصفات والدقة العالية استخدمت الإعدادات التالية في عملية التسجيل الصوتي:

- تردد التعيان: 8000 عينة/ثانية
- نوع البيانات المستخدمة للتسجيل:double
- مدة التسجيل: 3 ثانية
- مادة التسجيل:
  - ✓ نجحت في الامتحان(في حالة الفرح)
  - ✓ رسبت في الامتحان(في حالة الحزن)
  - ✓ من فعل هذا؟ (في حالة الغضب)
- نوع التسجيل: أحادي (mono)

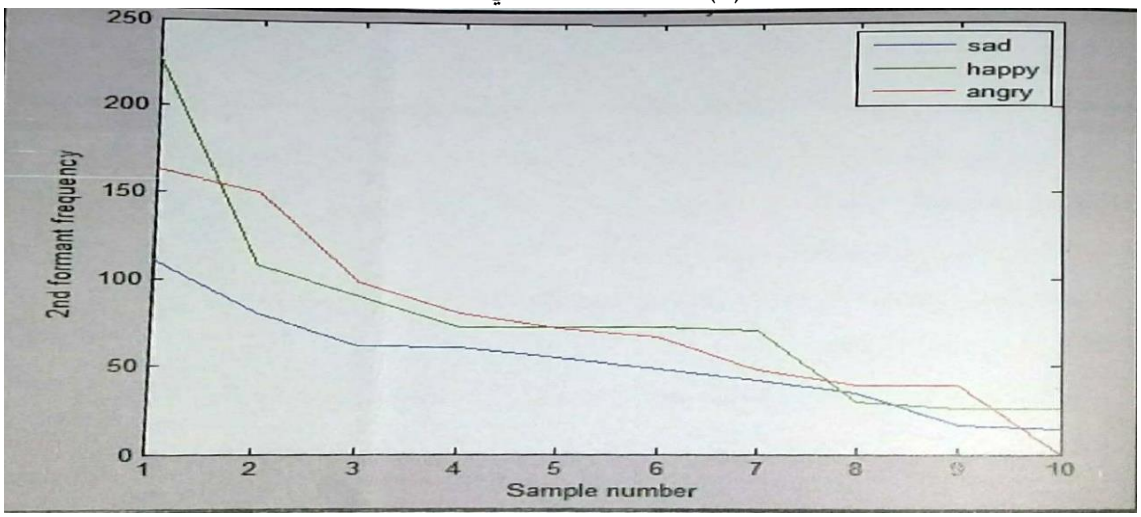
وقسمت في هذه المرحلة من الخوارزمية الإشارة بعد إلغاء حالة الصمت باستخدام برنامج(cool edit pro2) إلى عدد من الإطارات(frames) طول كل إطار 512عينة، وذلك لأن إشارة الكلام تتغير مع الزمن لكنها ثابتة نسبياً لمقطع قصير منها، ومن ثم المعالجة بالنافذة (نافذة هامينك)، من أجل تنعيم الإشارة وتقليل عدم استمرارية الإشارة بعد تقسيمها إلى مقاطع، تليها عملية استخلاص الخواص(الطاقة، التردد الأساسي، الترددات الرنانة) لكل إطار، إذ أستخدمتخوارزمية ترميز التنبؤ الخطي لإيجاد الترددات الرنانة وخوارزمية ضرب توافقيات الأطياف لإيجاد التردد الأساسي. وتتم تكرار العمليات أعلاه على كافة إشارات الكلام التي تسجل من أجل رسم المنحنيات لتلك الخواص ولكل حالة مزاجية. الأشكالمن(1) إلى(4) توضح منحنيات الخصائص بعد ترتيبها تنازليا للحالات الثلاثة.



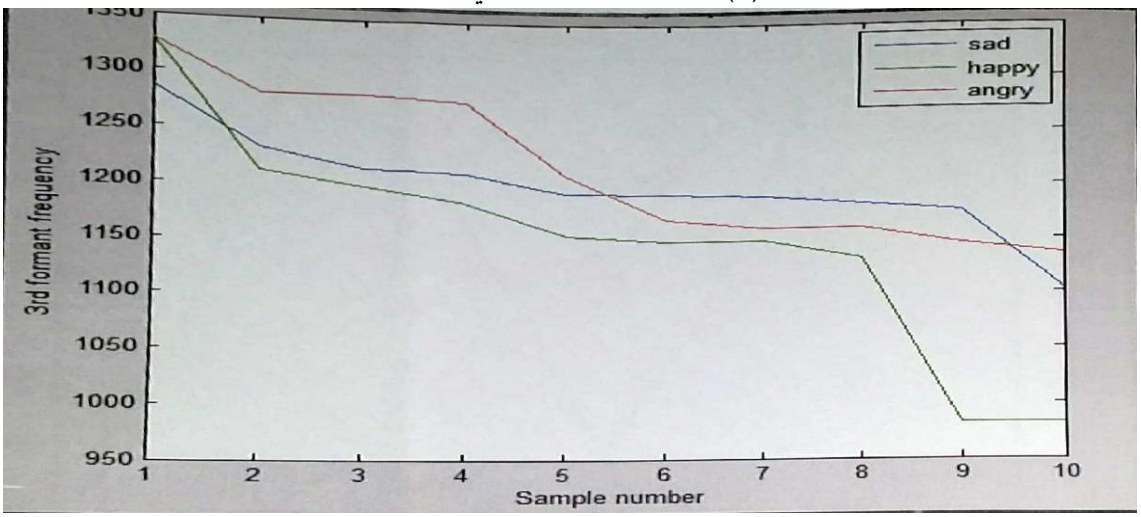
الشكل (1): قيمة الطاقة للحالات الثلاث



الشكل (2): قيمة التردد الأساسي للحالات الثلاث



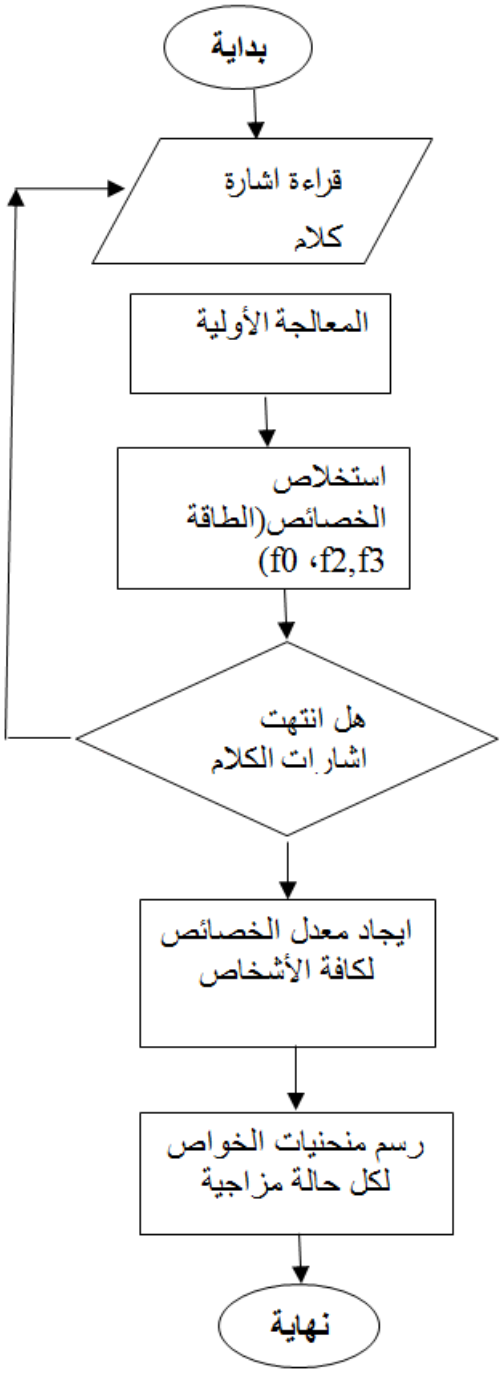
الشكل (3): قيمة التردد الرنان الثاني للحالات الثلاث



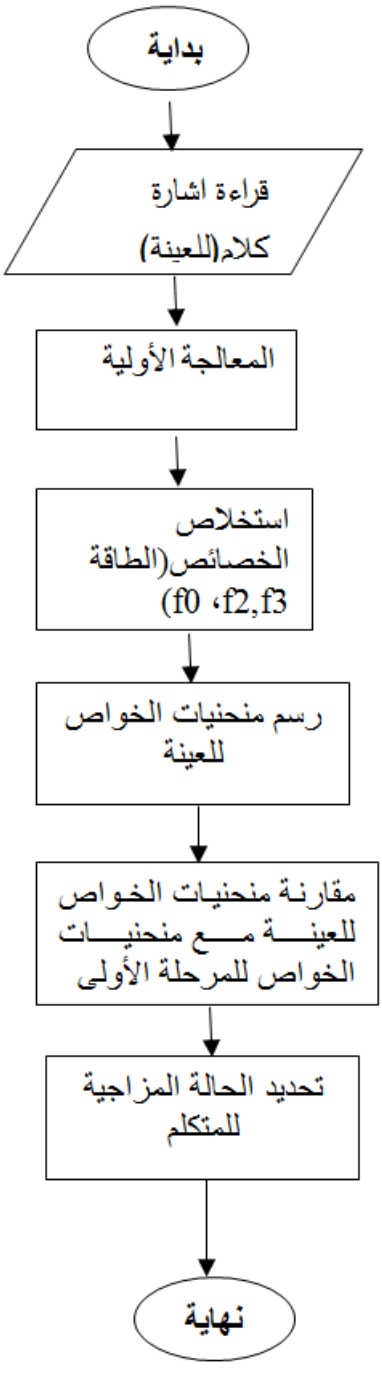
الشكل (4): قيمة التردد الرنان الثالث للحالات الثلاث

المرحلة الثانية: تحديد الحالة المزاجية للمتكلم

يختار في هذه المرحلة عدد من إشارات الكلام العشوائية، قسم منها من داخل قاعدة البيانات والأخرى خارجية، إذ تجرى المعالجة التي تمت في المرحلة الأولى بالخطوات نفسها وتستخلص الخصائص نفسها ومن ثم إيجاد كلا من معامل الارتباط والمسافة الأقليدية من أجل التوصل إلى الحالة المزاجية للمتكلم. والشكلان(5)،(6) توضحان المخططان الصندوقين لمراحل الخوارزمية.



الشكل(5): المخطط الصندوقي للمرحلة الأولى (إعداد قاعدة البيانات)

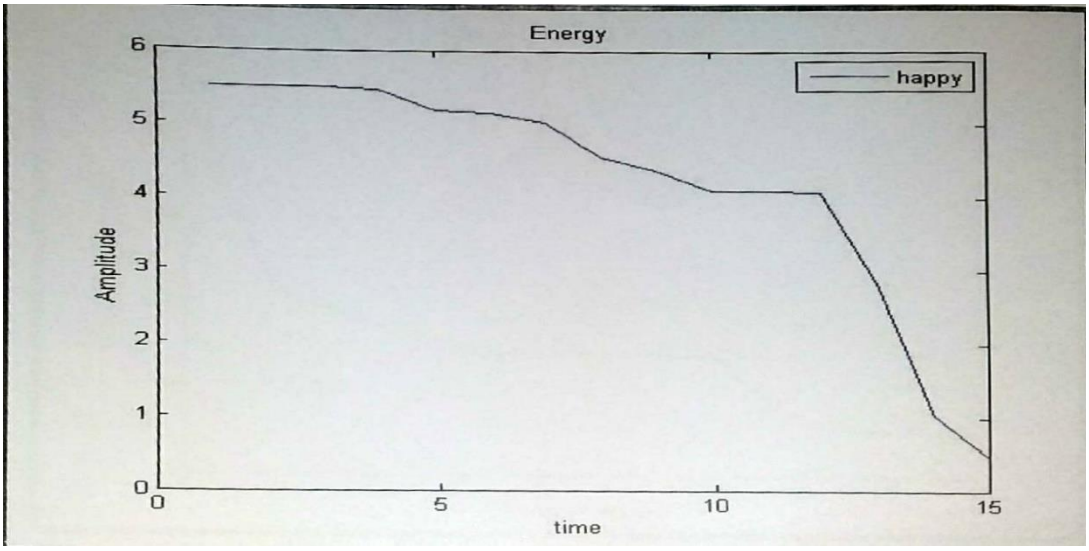


الشكل(6): المخطط الصندوقي للمرحلة الثانية (التصنيف)

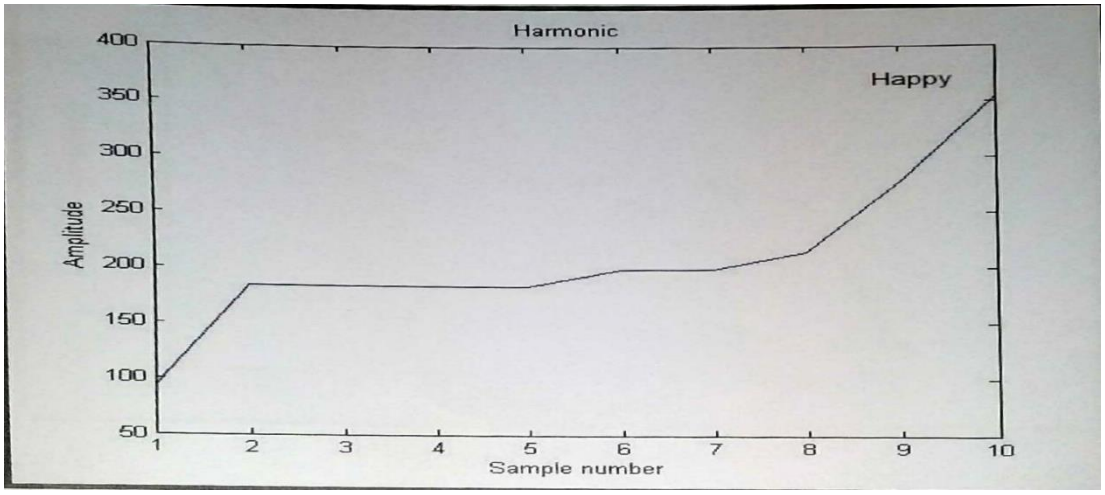
9. النتائج

❖ حالة الفرح

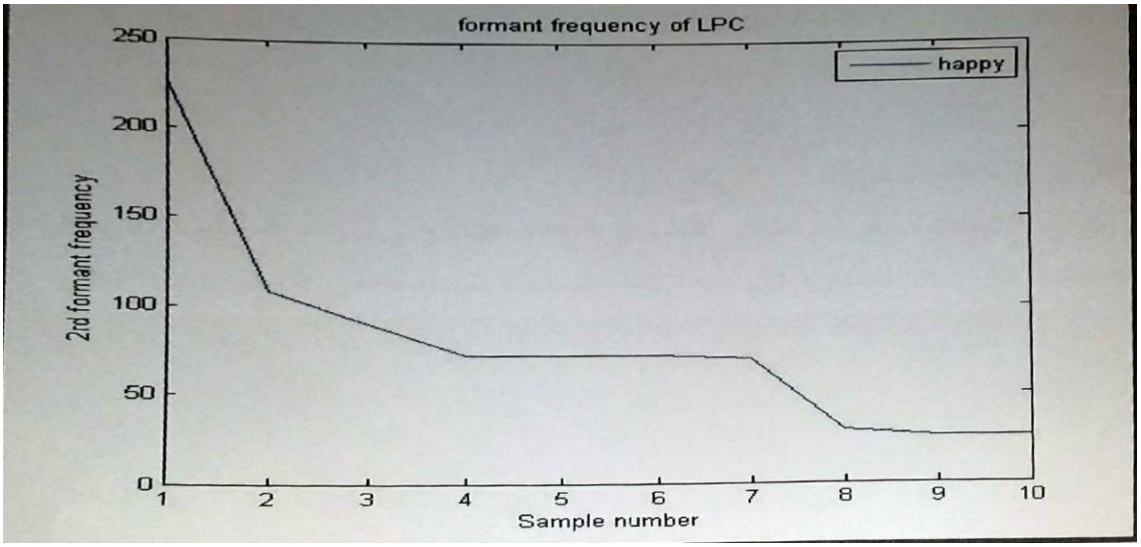
قرئ في هذا المثال ملف صوتي في حالة الفرح(happy.wav) وأجريت المعالجة الأولية عليه ومن ثم يتم تقطيع الإشارة إلى عدد من المقاطع طول كل مقطع 512عينة ليتم فيما بعد استخلاص الخواص منها والمتمثلة بالطاقة والتردد الأساسي والترددات الرنانة والأشكال من(7) إلى(10) تمثل منحنيات تلك الخواص.



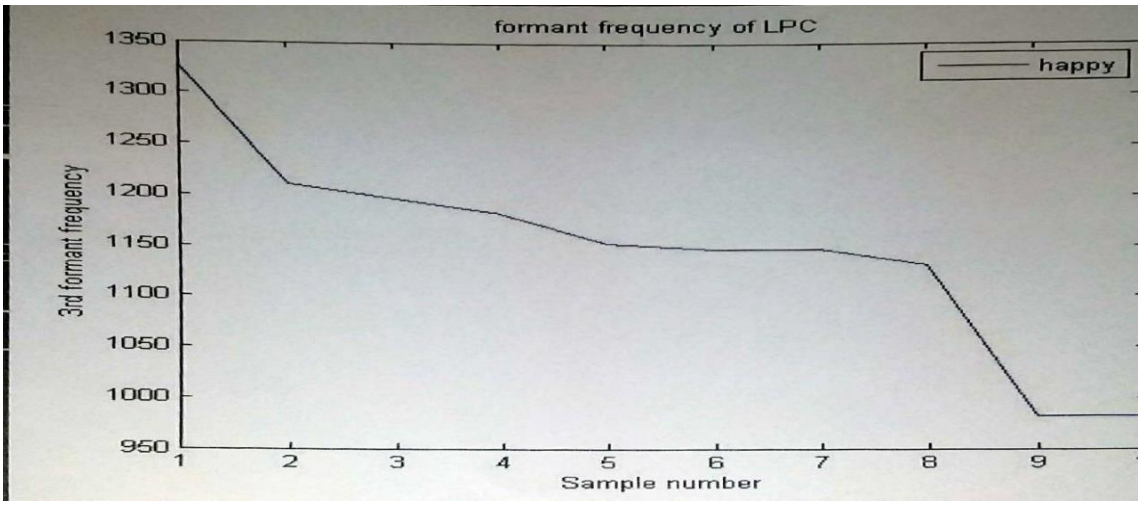
الشكل(7): قيمة الطاقة للعينة



الشكل(8): قيمة التردد الأساسي للعينة



الشكل(9): قيمة التردد الرنان الثاني للعينة

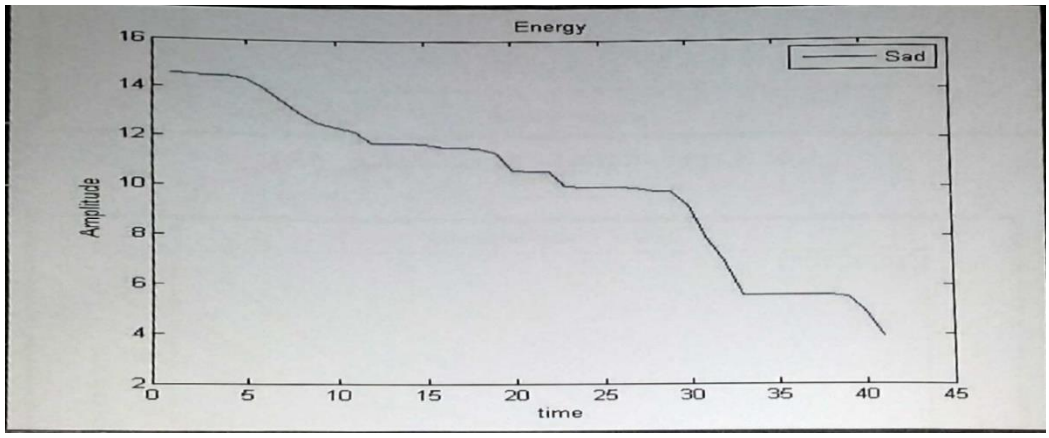


الشكل(10): قيمة التردد الرنان الثالث للعينة

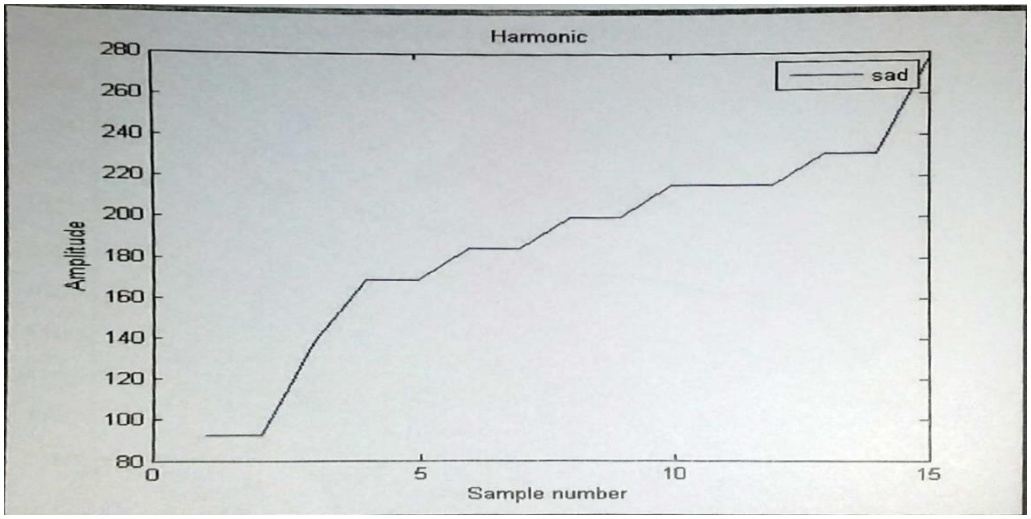
وبعد إيجاد قيم تلك الخصائص للشخص المراد تحديد حالته المزاجية يتم إيجاد كل من معامل الارتباط والمسافة الإقليدية بين قيم تلك الخصائص والقيم المقابلة لها في المنحنيات السابقة التي تم الحصول عليها في مرحلة إعداد قاعدة البيانات التي كانت تمثل أعلى ترابط وأقل مسافة إقليدية مع المنحني الذي يمثل حالة الفرح.

❖ حالة الحزن

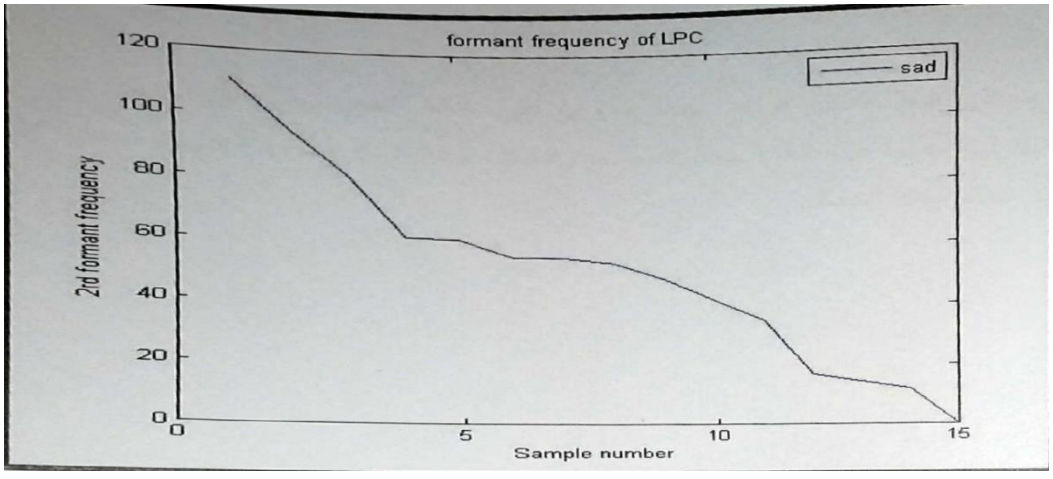
قرئ في هذا المثال ملف صوتي في حالة الفرح(sad.wav) وأجريت المعالجة الأولية عليه ومن ثم تقطع الإشارة إلى عدد من المقاطع طول كل مقطع 512عينة ليتم فيما بعد استخلاص الخواص منها والمتمثلة بالطاقة والتردد الأساسي والترددات الرنانة والأشكال من(11) إلى(14) تمثل منحنيات تلك الخواص.



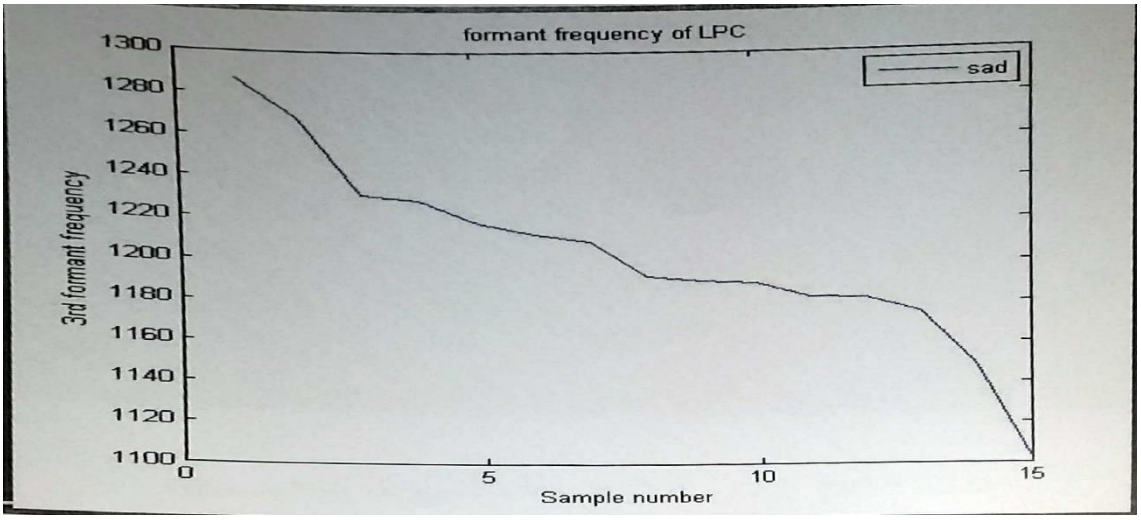
الشكل(11): قيمة الطاقة للعينة



الشكل(12): قيمة التردد الأساسي(f0) للعينة



الشكل(13): قيمة التردد الرنلن(f2) للعينة

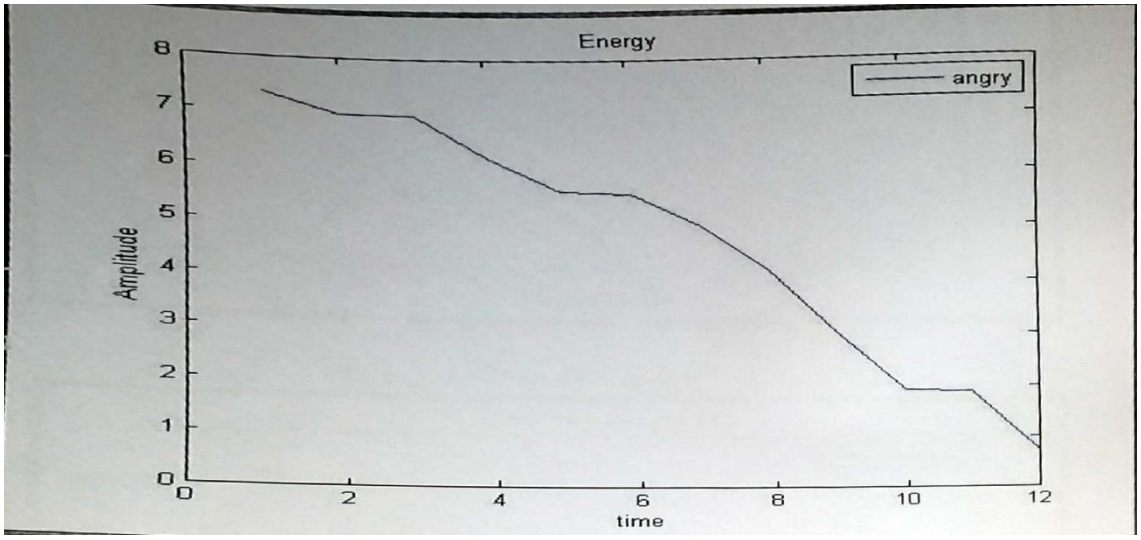


الشكل(14): قيمة التردد الرنين(f3) للعينة

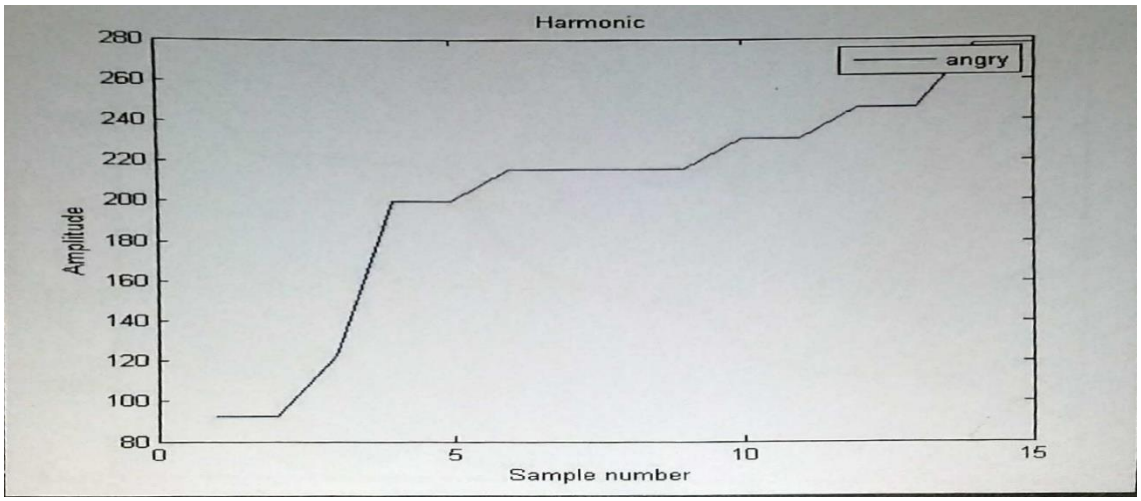
وبعد إيجاد قيم تلك الخصائص للشخص المراد تحديد حالته المزاجية يتم إيجاد كل من معامل الارتباط والمسافة الإقليدية بين قيم تلك الخصائص والقيم المقابلة لها في المنحنيات السابقة التي تم الحصول عليها في مرحلة إعداد قاعدة البيانات، ,والتي كانت تمثل أعلى ترابط وأقل مسافة إقليدية مع المنحني الذي يمثل حالة الحزن.

- **حالة الغضب**

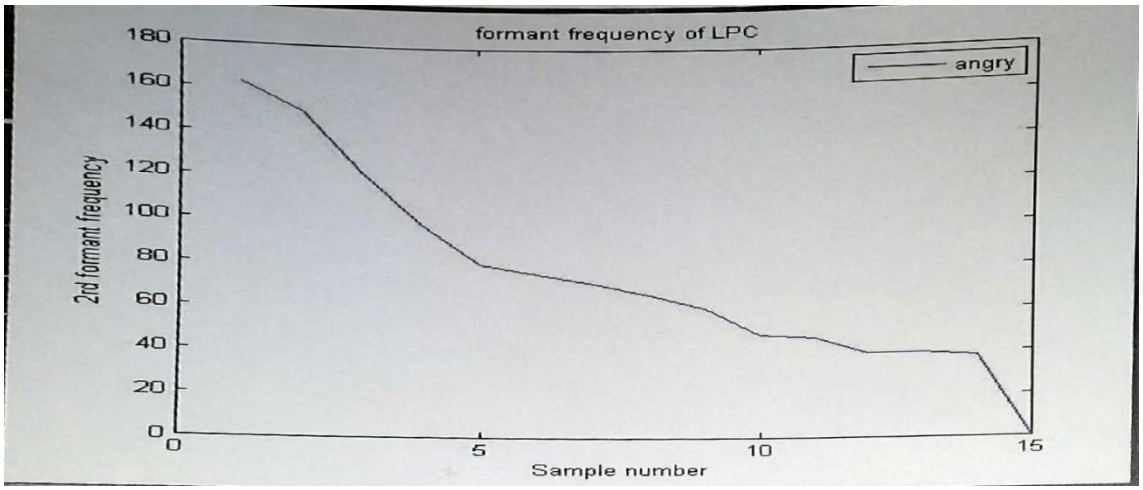
قرئ في هذا المثال ملف صوتي في حالة الغضب(angry.wav) وأجريت المعالجة الأولية عليه من إلغاء حالة الصمت والمعالجة بالنافذة ومن ثم يتم تقطع الإشارة إلى عدد من المقاطع طول كل مقطع 512عينة ليتم فيما بعد استخلاص الخواص منها والمتمثلة بالطاقة والتردد الأساسي والترددات الرنانة والأشكال من(15) إلى (18) تمثل منحنيات تلك الخواص.



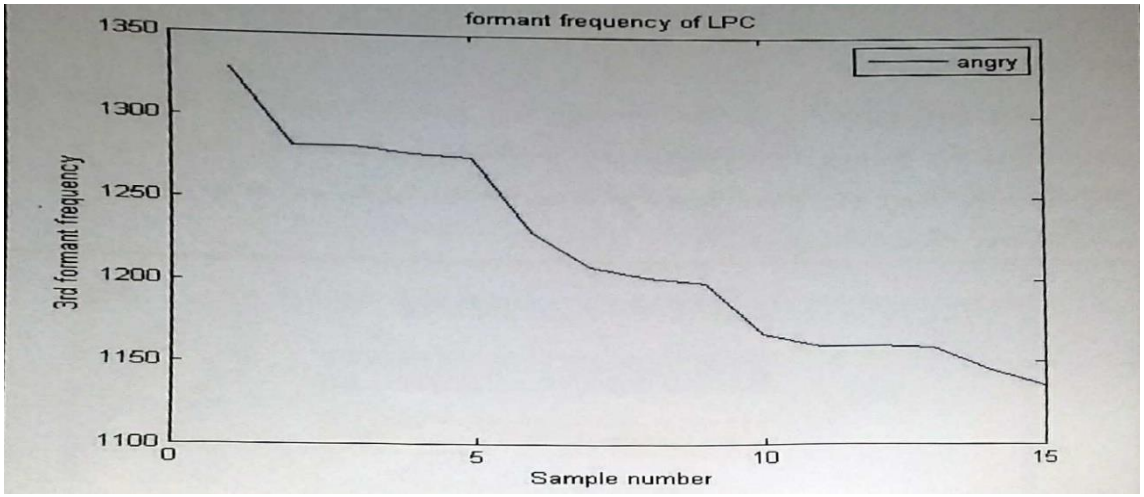
الشكل(15) قيمة الطاقة للعينة



الشكل(16): قيمة التردد الأساسي(f0) للعينة



الشكل(17): قيمة التردد الرنان(f2) للعينة



الشكل(18): قيمة التردد الرنان(f3) للعينة

وبعد إيجاد قيم تلك الخصائص للشخص المراد تحديد حالته المزاجية يتم إيجاد كل من معامل الارتباط والمسافة الإقليدية بين قيم تلك الخصائص والقيم المقابلة لها في المنحنيات السابقة التي تم الحصول عليها في مرحلة إعداد قاعدة البيانات، إذ أعطت أعلى ارتباط وأقل مسافة إقليدية مع المنحني الذي يمثل حالة الغضب.

10. الأستنتاجات

1. إن خواص اشارة الكلام المتمثلة بالطاقة(مجال الزمن)، التردد الأساسي، الترددات الرنانة(مجال التردد) أعطت نتائج جيدة بدقة تحديد الحالة المزاجية للمتكلم.
2. إن استخدام معامل الارتباط فضلاً عن المسافة الإقليدية يعتبر عامل مهم في عملية التصنيف عن طرائق عملية إجراء المقارنة.
3. إن تقسيم إشارة الكلام إلى عدد من المقاطع بطول512عينة للمقطع الواحد واستخلاص الخصائص لكل مقطع أعطت نتائج جيدة عند اعتماد هذا الطول.

11. الأعمال المستقبلية

1. إمكانية استخدام طرائق أخرى في التصنيف كالطرائق الذكائية.
2. إمكانية استخلاص خصائص أخرى لإشارة الكلام مثل خاصية النرفز(jitter) وخاصية الوميض(shimmer) ودراسة تأثيرها على دقة التصنيف.
3. إمكانية اعتماد عدد أكبر من العينات في قاعدة البيانات ودراسة تأثيرها على عملية التصنيف.

المصادر

[1] Kuhn Lisa Katharina, 2014,"emotion recognition in the human face and voice", Doctor thesis of Philosophy.

[2] S. D´Mello et. al, 2011,"Automatic Emotion Recognition from Speech", PhD Research Proposal (Eds.), ACII 2011, Part II, LNCS 6975, pp. 191–199, © Springer-Verlag Berlin Heidelberg 2011.

[3] From Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Emotion_recognition

[4] Anagnostopoulos, Christos Nikolaos, et.al, 2015,"Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011",Springer Science+Business Media Dordrecht ,Volume 43, Issue 2, pp 155–177.

[5] عبد النبي، هبه أدريس يونس، 2011، "تحديد الفئة العمرية للمتكلم باستخدام القيم المميزة"، بحث ماجستير، علوم حاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل.

[6] SchullerBjörn, et. al , 2002, "automatic emotion recognition by the speech signal", from SCI 2002, IIIS.

[7] Zbancioc Marius, Feraru Monica, 2012, "a study about the statistical parameters used in the emotion recognition",11th international conference on development and application systems.

[8] BhargavaaMayank., Polzehlb Tim, 2012, "Improving Automatic Emotion Recognition from speech using Rhythm and Temporal feature",In Proceedings of ICECIT-2012 Published by Elsevier.

[9] S.Li,et. al, 2014,"speech emotion recognition based on coiflet wavelet packet cepsetralcoefficients",part II, © Springer-Verlag Berlin Heidelberg 2014, CCIS 484, pp 436-443.

[10] Gluge,S., et. al, 2017,"emotion recognition from speech using representation learning in extreme learning machines", in proceeding of the 9th international joint conference on computational intelligence, pp179-185, ISBN:978-989-758-274-5.

[11] Kerkeni,L.,et. al, 2018,"speech emotion recognition:methods and cases study",in proceedings of the 10th international conference on agents and artificial intelligence(ICAART 2018), vol 2, p 175-182.

[12]قدو، سجى جاسم محمد، 2004،" كبس اشارة الكلام بواسطة استخلاص الخواص"، بحث ماجستير، علوم حاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل.

[13] Pavol Partial, et.al,"impact of emotions on fundamental speech signal frequency", latest trends in information technology, ISBN: 978-1-61804-134-0, p.409-414.

[14] Middleton, Gareth,2003,"pitch detection algorithms", openstax-CNXmodule:m11714. http://cnx.org/content/m11714/102/.

[15] MousaAllam, 2011,"speech segmentation in synthesized speech morphing using pitch shifting", The international Arab journal of information technology, vol.8, no.2, p221-226.