

Speakers Recognition Based On Convolution

Khalil I. AL_Saif¹, Nagham Throwat², Esraa Abd Alsalam^{3*}

¹Department of Computer Science, Collage of mathematics and Computer Science, University of Mosul, Mosul, Iraq

^{2,3*} Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Iraq.

E-mail: ¹ khalil_alsarif@uomosul.edu.iq, ² Nagham.th@uomosul.edu.iq, ^{3*} esraa.altoni@uomosul.edu.iq

(Received January 03, 2012; Accepted June 06, 2012; Available online March 01, 2020)

DOI: [10.33899/edusj.2020.172816](https://doi.org/10.33899/edusj.2020.172816), © 2020, College of Education for Pure Science, University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract

In this research, the important features of recognizing speakers' identity by extracting features of speakers voices are used. Voice is considered one of the vital factors adopted algorithm on samples of speakers' voices has been implemented by recording the voice, saving it on a file of the type wav. The voice is treated by using Hamming function to reduce error ratio. Feature of the voice samples have been extracted by taking the spectrum value of the sound signal of all the speakers. Two methods have been used (Eclidian distance and correlation factor) for comparing core of the convolution of the sample of speaker test (a training sample which is being recorded another time) and all the speakers samples who have already been saved on the data base. Applying the algorithm on different voices, ratio of error was very small while matching has increased by the increasing number of speaker. The algorithm has been implemented using matlab.

Keyword: convolution algorithm, Hamming function, Eclidian distance and correlation factor

اعتماد اللافوف الرياضي في تمييز المتكلمين

خليل ابراهيم السيف¹، نغم سعيد ثروت²، اسراء عبد السلام عبدالقادر^{3*}

¹ قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق
^{2,3*} قسم علوم الحاسوب، كلية التربية للعلوم الصرفة، جامعة الموصل، الموصل، العراق

الخلاصة:

تم في هذا البحث استخدام احد السمات الحيوية لتمييز هوية المتكلمين من خلال استخلاص الصفات (feature extraction) لأصوات المتكلمين. حيث يعتبر الصوت من احد العوامل الحيوية المعتمدة في تمييز الهوية الشخصية، وقد تم تنفيذ الخوارزمية المقترحة على عينة من اصوات المتكلمين من خلال تسجيل الصوت وخرنه في ملف صوتي من نوع wav. ومن ثم معالجة الصوت عبر دالة هامينك لتقليل نسبة الخطا، وتم استخلاص خواص العينة الصوتية بأخذ قيمة الطيف (spectrum) للاشارات الصوتية لجميع المتكلمين بعد تحويلها الى المجال الترددي باستخدام تحويل فورييه السريع، وتم اعتماد اللافوف الرياضي

(convolution) كصفة مميزة لتمييز هوية المتكلمين. واستخدامات طريقتين (المسافة الأقليدية و معامل الارتباط) للمقارنة بين قيم قلب اللافوف الرياضي لعينة اختبار المتكلم (وهي عينة تدريب يتم تسجيلها مرة أخرى) وجميع عينات المتكلمين التي تم خزنها مسبقاً في قاعدة البيانات. عند تطبيق الخوارزمية على اصوات مختلفة كانت نسبة الخطأ ضئيلة جداً وازدادت نسبة التوافق كلما ازداد عدد المتكلمين. لقد تم تطبيق الخوارزمية باستخدام matlab7

الكلمات المفتاحية : خوارزمية اللافوف، نافذة هامينك، المسافة الأقليدية، معامل الارتباط.

1. المقدمة

في ظل التطور السريع في شتى المجالات الالكترونية وتقنية المعلومات، أصبح الإنسان كثير التعامل مع الأنظمة المتعددة والمختلفة مثل الهواتف وأجهزة الصرف الآلي والدخول إلى المباني والشركات وغيرها التي تحتاج التأكد من هوية الشخص قبل دخوله أي بمعنى آخر تصريحاً لدخوله إلى أي نظام يحتوي على معلومات سرية وفي غاية الأهمية. فاستخدام الأرقام السرية قد تكون معرضة للسرقة من قبل قراصنة الانترنت وغيرهم لأنها غير آمنة، وإن استخدام السمات الحيوية هي من الطرق الأخرى التي تضمن سرية البيانات وتحافظ على أمانيتها، فهي وسيلة يتم فيها التعرف على هوية المتحدث وذلك عن طريق مقارنة خصائص المتحدث الصوتية مع الخصائص الموجودة (المخزنة) مسبقاً [1].

2. الدراسات السابقة

- 1- قدم كل من (John H. L. Hansen, Gang Liu, Chengzhu Yu) طريق جديدة لتمييز العينات الصوتية باستخدام النموذج UBM حيث يتم تدريب UBM على الميزات الصوتية وتمن ثم تقدير مصفوفة التحويل باستخدام التحليل الخطي التمييزي (LDA) من خلال معالجة كل مزيج من UBM المدربة كفئة مستقلة [2].
- 2- كما طبق الباحث (Shing – Tai Pan) طريقة fuzzy vector quantization على نموذج ماركوف المخفي المتقطع (Discrete Hidden Markov Model) DHMM لإثبات speech recognition rate عن طريق استخدام Mandarin training speech [3].
- 3- قدم الباحثين (K. Sajeer, Paul Rodrigues) نظام التعرف على الكلام باستخدام تقنيتين مختلفتين لاستخراج الميزات ويتم إجراء دراسة مقارنة للتعرف على الكلمات المنطوقة المستقلة عن المتحدثين. الأول هو نهج هجين من الترميز التنبؤي الخطي (LPC) والشبكات العصبية الاصطناعية (ANN) والثاني تستخدم الطريقة مزيجاً من التحويلات الموجة المنفصلة (DWT) والشبكات العصبية الاصطناعية [4].
- 4- كما قدم كل من (Mihalis Siafarikas, Todor Ganchev, Nikos Fakotakis) بحثاً لتمييز المتكلمين في استخلاص صفات الكلام باعتماد التحويل الموجي (Wavelet) [5].
- 5- قدم الباحثون (Haraty A. Ramzi and El Ariss Omeri) شرحاً لتطبيق تمييز الكلام العربي معتمداً على اللهجة اللبنانية. يبدأ النظام بتقطيع الإشارة، وهي عملية تحويل من إشارة مستمرة إلى إشارة مقطعة، ثم يستخلص الصفات باستخدام (MFCC) وهذه الصفات المستخلصة تقارن مع أنموذج النظام المخزون والذي أساسه الفونيم، فتتم مطابقة الصفات المستخلصة من الإشارة الداخلة مع قالب الكلمة [6].

3. الصوت

الصوت هو الأثر الواقع على الأذن البشرية من بعض حركات ذبذبية للهواء يحدثها الجهاز الصوتي للمتكلم. فهو ظاهرة فيزيائية تثير حاسة السمع، ويختلف معدل السمع بين الكائنات الحية المختلفة، فيقع السمع عند البشر عندما تصل الذبذبات الى تردد ما بين (20)-(20000) هيرتز إلى الأذن الداخلية عبر الهواء. أما الأصوات التي يزيد ترددها على (20000) هيرتز فتعرف بالموجات فوق الصوتية ويتعذر على الأذن سماعها. أما الأصوات ذات التردد المنخفض التي تكون أقل من 20 هيرتز لا تستطيع الأذن البشرية إدراكها أو التأثر بها، تعرف بالموجات تحت السمعية [7].

4. أنواع ملفات الصوت WAV و MP3 و WMA

صيغة WAV

غالبا تستخدم في تسجيل الصوت الخام (الصوت الأصلي الذي لم تدخل عليه أي تعديلات أو أي مؤثرات). وتستخدم الصيغة لإنتاج العمل وتحريره وهندسته والتعديل عليه، ومن ثم يصدر ويحول إلى ملفات مضغوطة مثل : MP3 و WMA و RM و RA كما سيتم شرحها لاحقا [8].

صيغة MP3

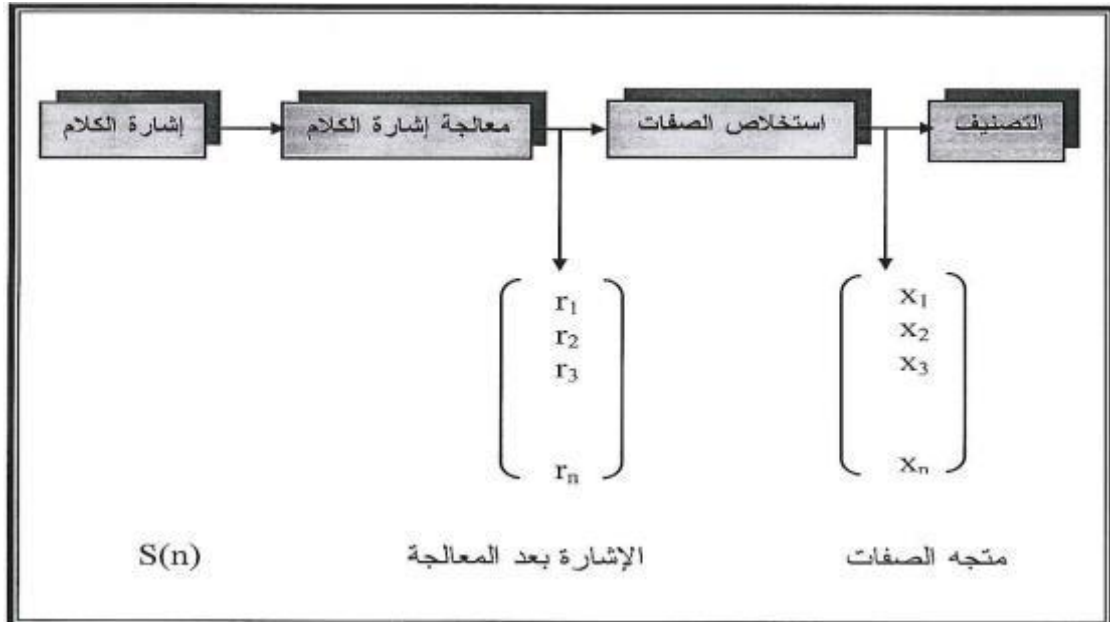
وهي واحدة من مجموعة MPEG مشتقة من عبارة Moving Picture Experts Group ، وهي صيغة مضغوطة الحجم.

صيغة WMA

وهي مختصرة من عبارة Windows Media Audio ، وهي صيغة مضغوطة الحجم. وهي صيغة حديثة مقارنة مع MP3 [9]. وفي هذا البحث تم التعامل مع ملفات من نوع WAV ليتم دراسة العينات الخام قبل اجراء أي تغيير عليها، حيث يتم تحليل اشارة الكلام لاستخلاص صفات المتكلم باستخدام (DCT Discrete Cosine Transform) لحساب اشارة الطيف لعينات الصوت الخام المسجلة، ومن ثم تمييز المتكلم.

5. اشارة الكلام وتحليلها

الكلام هو سلسلة صوتية متصلة ببعضها البعض اتصالا وثيقا ، وان من الممكن وصف الكلام بصيغة إشارة تحمل معلومات معينة أي تمثيلها بشكل موجة سمعية (Wave Format)، في كل أنظمة تمييز الكلام يتم معالجة الإشارة من البداية إلى النهاية مثل المعالجة بالنافذة ، الترشيح ، التعيان والتكميم، أن إشارة الكلام تتمثل بالاعتماد على معالم (Parameters) متنوعة للكلام حسب نوع التطبيق مثل التمثيل باستخدام معالم المجال الزمني (Time domain parameters) للإشارة أو المعالم الطيفية (Spectral parameters) ليتم استخراج خصائص الكلام المطلوبة من قبل النظام المصمم ، إن اختيار نوع التمثيل للكلام يلعب دورا كبيرا ومهما في إظهار خصائص الكلام والشكل (1) يوضح مراحل التمييز [10]



شكل 1. مراحل تمييز الكلام

6. نافذة هامنيك

النافذة عبارة عن دالة رياضية تشبه في وظيفتها عمل النافذة الاعتيادية بإدخال كمية من الضوء . تقوم هذه النافذة بتنعيم الإشارة وإزالة التعرجات الحادة منها ، فالعملية تتضمن التفاف (convolution) إشارة الكلام (Speech signal) S(n) مع نافذة W(n) ، كما موضح في المعادلة (1)، والتي ينتج عنها مجموعة من عينات الكلام ، (Speech samples) معدلة بحسب شكل النافذة حسب المعادلة (2)، والشكل (2) يوضح نافذة هامنيك [11] .

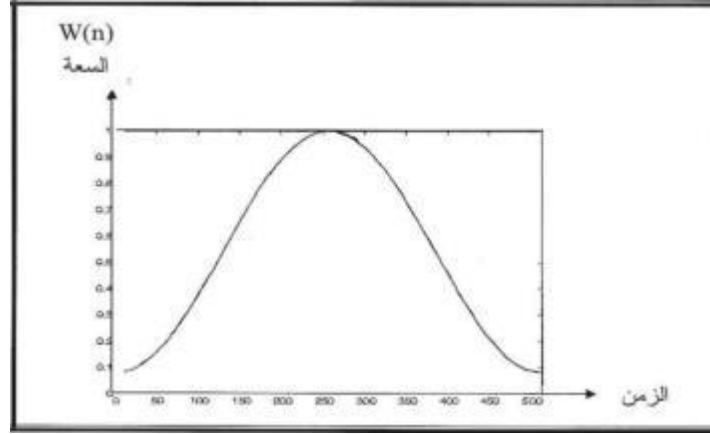
$$ss(n) = s(n) \times w(n) \dots\dots\dots(1)$$

n : عدد العينات الصوتية المسجلة.

* : التفاف الإشارة الصوتية مع نافذة هامنيك.

$$W(n) = \begin{cases} 0.54 - 0.46 \times \text{Cos} (2 \pi n / (N - 1)) & \text{for } 0 \leq n \leq N \\ 0 & \text{Other Wise} \end{cases} \dots\dots\dots(2)$$

N: طول العينة



الشكل 2. نافذة هامينك

7. استخراج صفات الصوت

تحليل الكلام Speech analysis

تحليل الكلام هو الخطوة الأولى في أنظمة تمييز الكلام فهناك مجالان يمكن الاعتماد عليهما لتحليل الكلام هما المجال الزمني (Time domain) والمجال الترددي (Frequency domain).

المجال الزمني (Time domain)

استخلاص الصفات يتم من خلال التعامل بشكل مباشر مع الإشارة (الشكل الموجي للإشارة) ، من الملاحظ أن إشارة الكلام ليست ثابتة مع الزمن حيث أن الصفات للإشارة تتغير نسبيا ببطء مع الزمن ، وان هذا الافتراض يقود إلى شيء أساسي ومهم وهو تحليل الوقت القصير (Short time analysis).

المجال الترددي (Frequency domain)

يستخدم بشكل واسع في تمييز الكلام وان الشكل الموجي للكلام يحلل باستخدام التحليل الطيفي Spectral analysis فمن خلاله يمكن وصف ميكانيكية إنتاج الصوت لعدد من الأشخاص من خلال الترددات الرنانة كذلك فان الخصائص التي تميزها الإذن البشرية تكون متضمنة بصورة رئيسة في المعلومات الطيفية ، التحليل الطيفي بصورة عامة يعد قلب المعالجة [12] .

تم في هذا البحث استخدام تقنية تحليل الطيف (Spectral Analysis) في استخراج خواص الكلام لتمييز هوية المتكلم. وتم حساب الطيف من خلال تطبيق تحويل فورير المتقطع (DFT) (Discrete Fourier Transform) على الإشارة لحساب قيمة طيف الإشارة (Magnitude Spectrum) ، و حساب لوغارتيم (log) لتحويل معكوس الجيب تمام المتقطع (Inverse Discrete Cosine Transform) (IDCT) كتهيئة لاجراء استخراج الخواص من الإشارة عرفت على نحو واسع في صفات الكلام، كما في المعادلات (3) (4) (5) [13] .

$$F(k) = \sum_{n=0}^{N-1} s(n) \exp(-j2\pi/N)nk \dots \dots \dots (3)$$

$$FL(k) n = 0 = \log (F(k)) \dots \dots \dots (4)$$

$$IDF(n) = (1 / N) \sum_{k=0}^{N-1} FL(k) \exp (j2 \pi / N) nk \dots \dots \dots (5)$$

S(n): إشارة الصوت الداخلة

N : طول الإشارة الداخلة (s)

n = 0 N-1

F(k) : دالة الطيف .

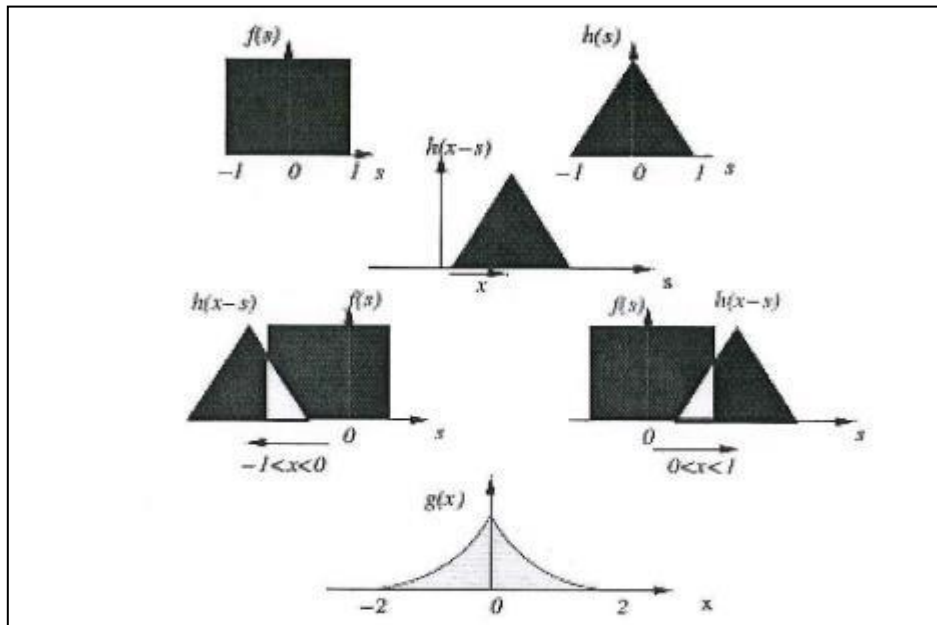
8. اللافوف الرياضي (CONVOLUTION)

عبارة عن مؤثر رياضي يستعمل في تحليل الدوال ويقوم بتحويل الدالة المخرجة الالتفاف من دالتين مدخلتين بحيث تكون قيمة الخرج في أي لحظة متأثرة بكل قيم الدخل السابقة وتستخدم هذه الأداة الرياضية في عدة تطبيقات مثل:

- معالجة الإشارة
- معالجة الصور
- الإحصاء

و تعريف دالة اللافوف الرياضي في معالجة إشارة الصوت (إشارة رقمية) كحاصل مجموع المضروب دالتين الدالة الأولى f(i) والدالة الثانية h(j) تتزاح احدهما زمنيا نحو الأخرى ،كما موضح في المعادلة (6) والشكل (3) [14].

$$g(k) = \sum f(j) \times h(k + 1 - j) \dots \dots \dots (6)$$



الشكل 3. دالة اللافوف الرياضي

9. المسافة الأقليدية:

يتم حساب المسافة بين نقطتين عن طريق ايجاد الجذر التربيعي لمربع الفرق بين النقطتين (p,q).

$$P = (p_1, p_2 \dots p_n) \text{ and } Q = (q_1, q_2 \dots q_n)$$

المعادلة (7) توضح المسافة الأقليدية لـ (p,q) مجموعة من النقاط .

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad \dots(7)$$

حيث يتم التعرف على هوية المتكلم من خلال الحصول على اقل مسافة أقليدية بين قلب اللافوف الرياضي لعينة الاختبار وبين العينات المخزونة في قاعدة البيانات .

10. معامل الارتباط Correlation

تم قياس مقدار التقارب بين عينات الصوت في قاعدة البيانات وعينة الاختبار باستخدام معامل الارتباط (r) ، كما في المعادلة (8). اذا كان معامل الارتباط (قريب من الواحد) يعني أن هناك تقارب بين العينات [15].

$$r = \frac{\sum(X-X')(Y-Y')}{\sqrt{\sum(X-X')^2} \sqrt{\sum(Y-Y')^2}} \quad \dots(8)$$

X: الوسط الحسابي لعينة الاختبار الصوتية X_1, X_2, \dots, X_n

Y: الوسط الحسابي لعينة صوتية في قاعدة البيانات Y_1, Y_2, \dots, Y_n

11. الخوارزمية المقترحة:

تم اقتراح خوارزمية لتمييز هوية المتكلم باعتماد طيف الاشارة في تحليل اشارة الكلام الداخلة، واجراء الالتفاف للإشارة الداخلة مع معدل الإشارات في قاعدة البيانات باستخدام اللافوف الرياضي (convolution) لاجل تحديد الصفات التي يتم اعتمادها كأساس للتمييز وفق الخطوات التالية....

الخطوة الاولى :اكتساب الصوت

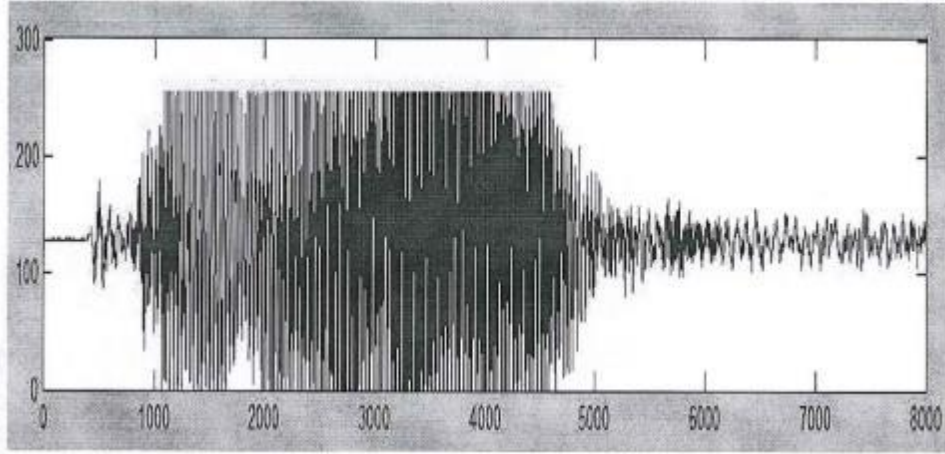
بعد اكتساب الصوت الى الحاسوب عن طريق لاقط الصوت، وتم اختيار بعد المسافة بين اللاقط والمتكلم حوالي 30سم، وفي غرفة هادئة إلى حد ما. وتمت مرحلة التهيئة لمتطلبات العمل بتحديد القيم المطلوبة لتنفيذ البرنامج، مثل الوقت المطلوب، عدد العينات ونوعية البيانات المحددة والمطلوبة لتسجيل الصوت، ثم خزن الصوت في ملف من نوع (Wav) يتم التسجيل الصوتي باستخدام الدالة التالية:

```
X=wavrecord(time*fs,fs,'uint8')
```

Time: هي الفترة المحددة لتسجيل الكلام او أي نص مضروباً في قيمة تردد العينة مثلاً ثانيتين.

FS: هي (frequency sample) وتم اخذها (8000) عينة بالثانية كلما زادت كبر حجم الملف واستغرق زمناً اطول في المعالجة| .

X: هو متجه يمثل القيم الحقيقية التي نتجت عن تسجيل الصوت . قراءة مجموعة أصوات المتكلمين لنص ثابت مثلاً كلمة (عراق) والشكل (4) يمثل عينة صوتية للشخص مثلاً تسلسله (24).



شكل 4. تسجيل صوت المتكلم لكلمة العراق

الخطوة الثانية: بناء قاعدة البيانات

يتم تسجيل اربعة عينات لكل متكلم، ومن ثم اخذ معدل العينات باستخدام المعادلة (9) الآتية:

$$X1 = (a1 + a2 + a3 + a4)/4$$

$$X2 = (b1 + b2 + b3 + b4)/4$$

·
·
·

$$Xn = (z1 + z2 + z3 + z4)/4 \quad \dots\dots (9)$$

حيث أن $X1, X2, \dots, Xn$ تمثل معدل اصوات لمجموعة متكلمين سيتم اجراء خطوات الخوارزمية المقترحة على تلك المعدلات ومن ثم خزنها في قاعدة البيانات. بينما تمثل (z,b,a) عينات مقروءة لكلمة (العراق) للشخص الأول ثم الثاني وهكذا إلى الأخير.

الخطوة الثالثة:

تطبق دالة هامينك (hamming) لتقليل تأثير الحواف، على عينات الاختبار وعلى العينات التي تم خزنها في قاعدة البيانات وهذا ما نلاحظه في معادلات (10) و (11) والشكل (5).

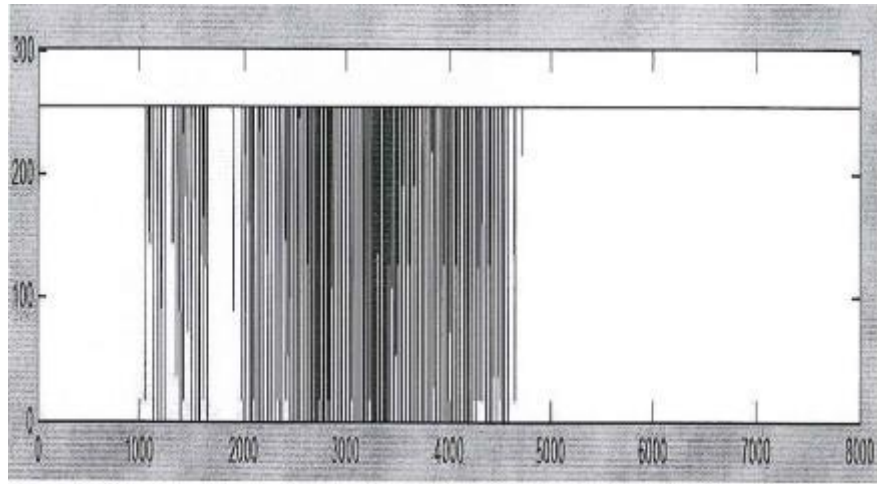
$$Ss(n) = s(n) \times w(n) \quad \dots\dots (10)$$

S : يمثل عينة صوتية اختبارية للشخص ما.

W : دالة هامينك

$$datass(n) = X(n) \times w(n) \quad \dots\dots (11)$$

X : العينات التي تم خزنها في قاعدة البيانات



الشكل 5. تطبيق دالة هامينك على صوت المتكلم 24

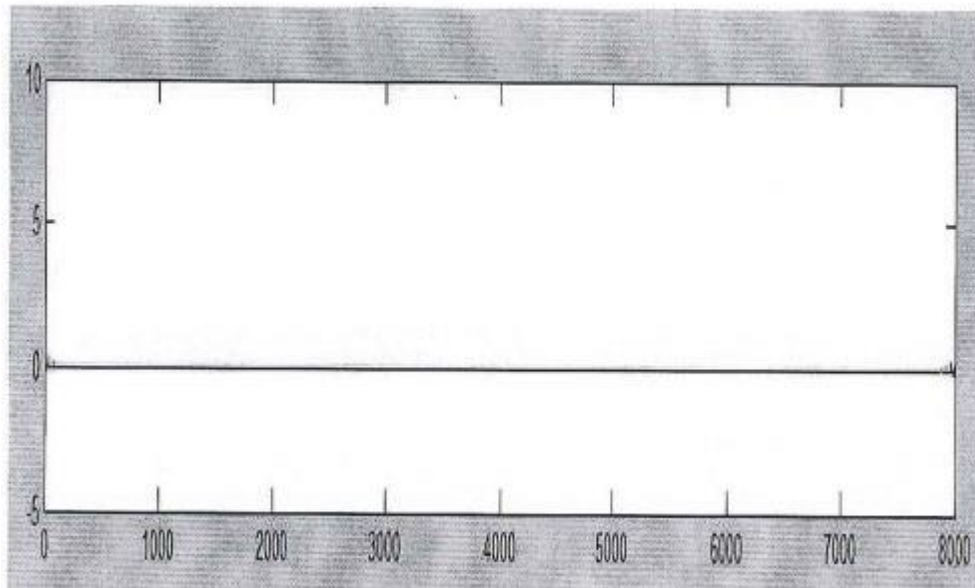
الخطوة الرابعة:

تم حساب الطيف من خلال تطبيق المعادلات (3) (4) (5) المذكورة أعلاه على عينات الاختبار التي سيتم تخزينها في قاعدة البيانات وهذا ما نلاحظه في الشكل (6). طبقت على عينات الاختبار وعلى العينات نفس الشخص تسلسله (24) كما في الدوال التالية.

$$Scp(n)=\text{spectrum}(ss(n))$$

$$\text{datass}(n)=\text{spectrum datass}(n)$$

$\text{spectrum}(ss(n))$: تطبيق معادلات الطيف على العينات التي تم تخزينها في قاعدة البيانات بعد معالجتها بنافذة هامينك.



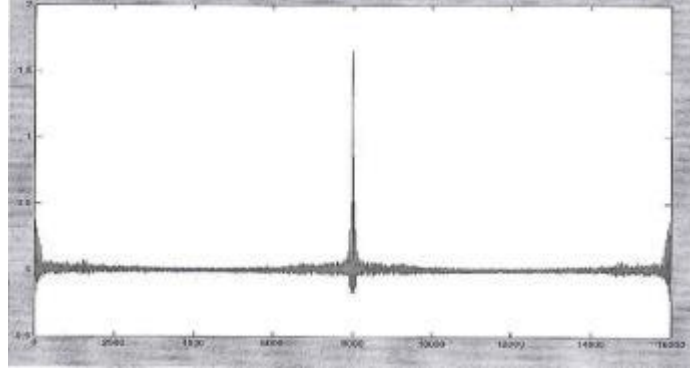
الشكل 6. تطبيق معادلات الطيف على إشارة الاختبار الصوتية

الخطوة الخامسة :

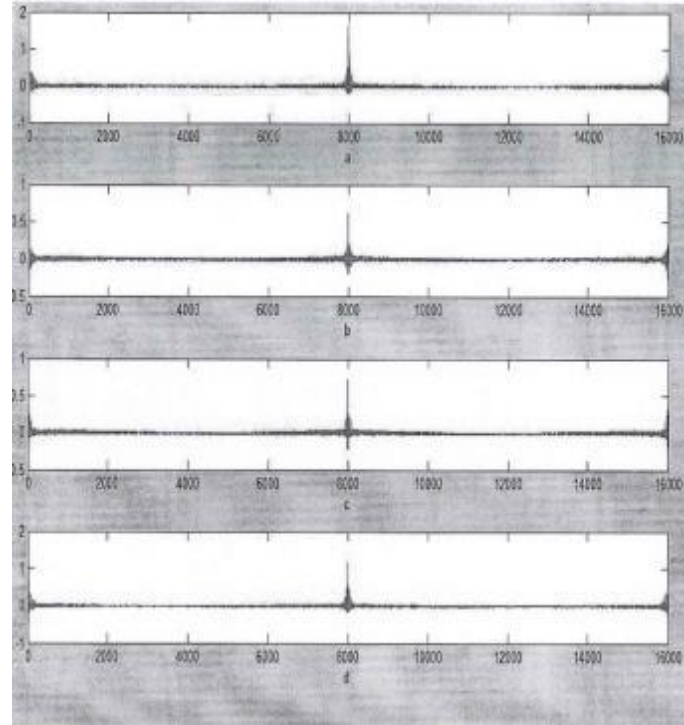
سيتم على قلب اللافوف الرياضي (convolution kernal) الناتج من التفاف عينة الاختبار الناتجة من الخطوة الثالثة مع نفسها، ومن ثم مع جميع العينات في قاعدة البيانات ليوضح أن قلب اللافوف يتطابق في قيمته كما في الدوال الاتية، والشكل (7) والشكل (8).

$$\text{Convs}(n)=\text{Scp}(n)* \text{Scp}(n)$$

$$\text{Conv}(n)=\text{Scp}(n)* \text{datass}(n)$$



شكل 7. اللافوف الرياضي (convolution kernal) الناتج من التفاف من الخطوة الثالثة مع نفسه



الشكل 8. الشكل (a) يمثل التفاف عينة الاختبار الناتجة من الخطوة الثالثة مع نفسها، أي تمثل (الشخص 24)، الشكل (b) التفاف عينة الاختبار مع عينة الشخص الأول في قاعدة البيانات، الشكل (c) التفاف عينة الاختبار مع عينة الشخص الثاني في قاعدة البيانات، الشكل (d) التفاف عينة الاختبار مع عينة الشخص (24) في قاعدة البيانات.

بعد الانتهاء من الالتفاف سيتم اخذ قلب الالتفاف لجميع المتكلمين في قاعدة البيانات ومن ثم مقارنتها مع قلب التفاف عينة الاختبار مع نفسها وهذا ما ستوضحه الخطوة السادسة.

الخطوة السادسة:

يتم حساب اقرب قيمة للمسافة الأقليدية للقلب اللافوف الرياضي بين عينة الاختبار و عينات المتكلمين المخزونة في قاعدة البيانات، قيمة قلب اللافوف الرياضي لعينة الاختبار مع نفسها .باستخدام المسافة الأقليدية. كما موضح في الجدول (1) لبعض العينات المسجلة والتي مرت بخطوات الخوارزمية.

الجدول 1: قيمة معامل الارتباط لقلب اللافوف الرياضي بين عينة الاختبار و عينات المتكلمين

متكلم 1	متكلم 2	متكلم 3	متكلم 4	متكلم 5	متكلم 6	متكلم 7	متكلم 8	متكلم 9	متكلم 10	متكلم 11	متكلم 12	متكلم 13	متكلم 14
0.5	0.5	0.4	0.4	0.4	0.5	0.3	0.5	0.5	0.5	0.4	0.4	0.6	0.7
9	5	1	8	6	7	6	7	5	3	0	700	3	4
0.6	0.7	0.4	0.7	0.5	0.7	0.5	0.7	0.6	0.7	0.6	0.7	0.5	0.7
7	3	9	1	2	3	8	2	3	2	5	1	3	2
0.5	0.4	0.7	0.5	0.5	0.5	0.5	0.5	0.4	0.5	0.4	0.5	0.5	0.4
0	9	9	6	2	58	2	2	9	5	9	2	58	9
0.6	0.6	0.4	0.6	0.7	0.7	0.7	0.5	0.5	0.6	0.6	0.6	0.5	0.7
0	5	5	8	20	30	20	30	90	90	7	8	10	30
0.4	0.4	0.5	0.6	0.5	0.5	0.5	0.6	0.5	0.5	0.5	0.4	0.5	0.5
8	900	600	0	400	100	0	400	900	700	100	700	00	6
0.5	0.7	0.4	0.5	0.4	0.6	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0.7
9	6	80	90	20	90	20	90	20	400	10	8	30	30
0.5	0.6	0.3	0.6	0.5	0.7	0.6	0.5	0.7	0.6	0.6	0.6	0.5	0.7
30	70	890	30	90	10	90	10	80	300	80	8	10	30
0.6	0.5	0.4	0.7	0.6	0.7	0.5	0.6	0.5	0.7	0.7	0.5	0.5	0.4
0	0	7	10	30	50	9	30	9	000	9	2	10	9
0.5	0.6	0.5	0.7	0.6	0.6	0.7	0.6	0.6	0.6	0.6	0.5	0.5	0.7
7	20	1	00	30	30	30	10	900	60	10	10	58	30
0.7	0.6	0.4	0.7	0.6	0.6	0.6	0.5	0.7	0.8	0.6	0.6	0.6	0.6
0	70	09	100	800	700	800	100	400	200	400	700	30	9
0.5	0.6	0.4	0.5	0.4	0.5	0.5	0.4	0.5	0.6	0.5	0.5	0.6	0.5
89	50	50	700	400	600	10	400	700	200	700	20	800	30
0.5	0.5	0.4	0.5	0.4	0.5	0.5	0.4	0.5	0.6	0.4	0.5	0.4	0.5
3	90	3	90	70	60	60	6	90	80	60	600	80	4

0.6	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.5	0.4	0.6	0.5	0.6	0.6	متكلم
30	300	300	100	100	900	000	310	70	2	30	00	000	67	13
	0				0		0							
073	0.8	0.5	0.5	0.6	0.4	0.5	0.4	0.5	0.5	0.6	0.4	0.5	0.6	متكلم
00	100	800	60	00	90	90	20	50	40	1	20	800	9	14

مثال: سيتم طرح مثال توضيحي للخوارزمية المقترحة

1 -اكتساب الصوت:-تمر عملية اكتساب الصوت بمرحلتين

المرحلة الاولى بناء قاعدة البيانات : تسجيل عينات صوتية لمجموعة من الاصوات لعدد من المتكلمين باستخدام الميكرفون ولكلمة "العراق" سنأخذ على سبيل المثال المتكلم الثاني وكما مذكور في الجدول (1)، ستكون العينات المقروءة كما في الدوال التالية.

$$Y1= 126 133 139 141 142 138 134 130 \dots$$

$$Y2= 165 164 146 124 115 129 156 186 187 \dots$$

$$Y3=187 196 206 207 202 195 191 187 186 \dots$$

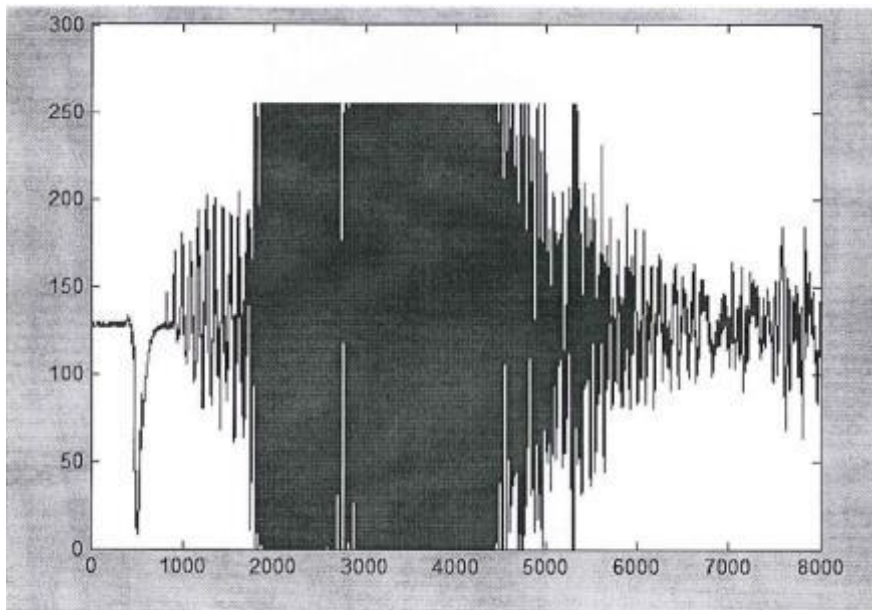
$$Y4=139 133 127 122 120 119 117 112 \dots$$

Y1، Y2، Y3، Y4 : العينات المسجلة لكلمة "العراق" للمتكلم الثاني

ومن ثم تكرار عملية التسجيل أربعة مرات واخذ الوسط الحسابي للتسجيلات الأربعة وتخزين في قاعدة البيانات.

المرحلة الثانية: تسجيل كلمة "عراق" للمرة الخامسة (لنفس المتكلم) وهنا أسمينها عينة الاختبار، كما في الدالة التالية والشكل (9).

$$SAMPLE=133 139 141 142 138 134 130 \dots$$

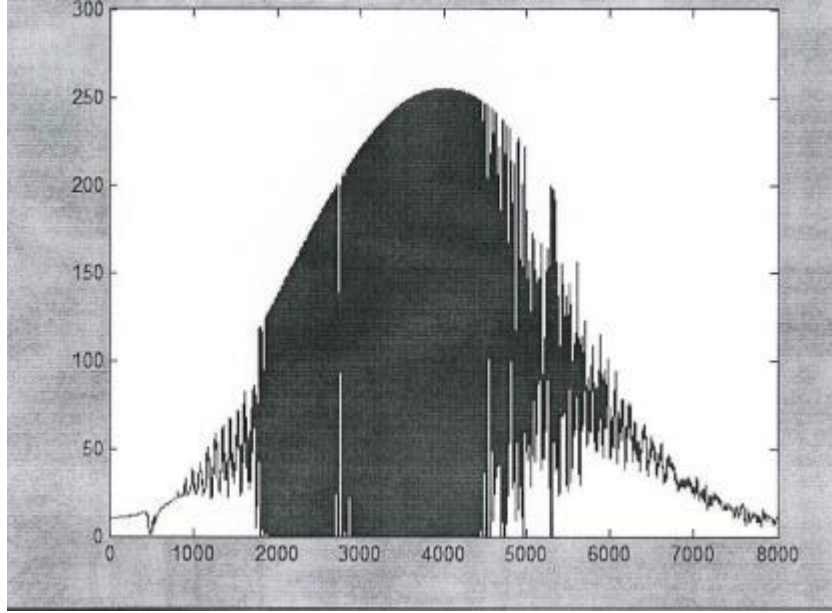


الشكل 9. تسجيل عينة صوتية لكلمة العراق للمتكلم الثاني

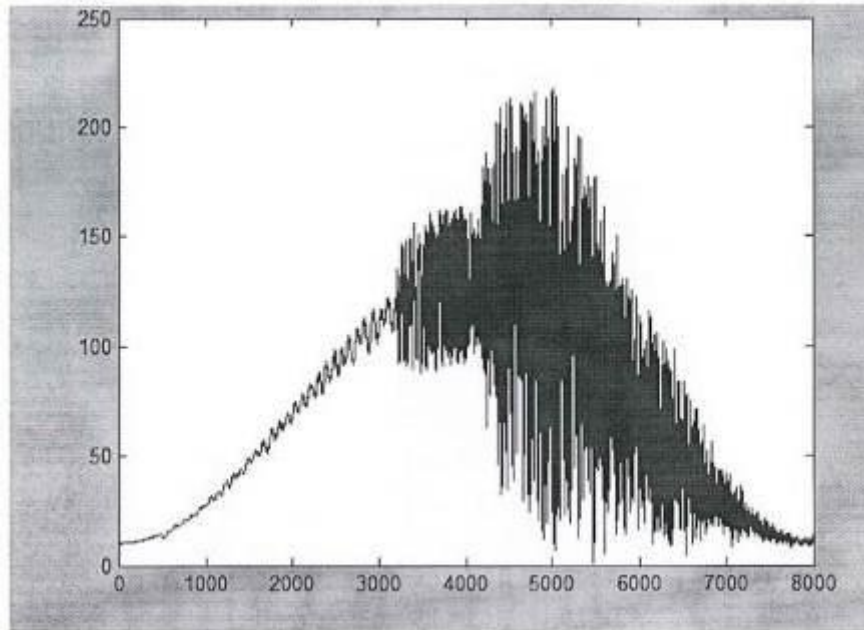
2- اجراء التقاف عينات الصوتية المسجل للمرة الخامسة (عينة الاختبار) مع نافذة هامينك والشكل (10) يوضح ذلك كذلك اجراء التقاف الوسط الحسابي الذي ذكر في الخطوة الأولى مع نافذة هامينك وهذا ما يوضحه الشكل (11)، وكما في الدوال التالية.

$$S=h*\text{sample}$$

$$S_m=h*\text{mean}$$



الشكل 10. التقاف عينة الاختبار مع نافذة هامينك

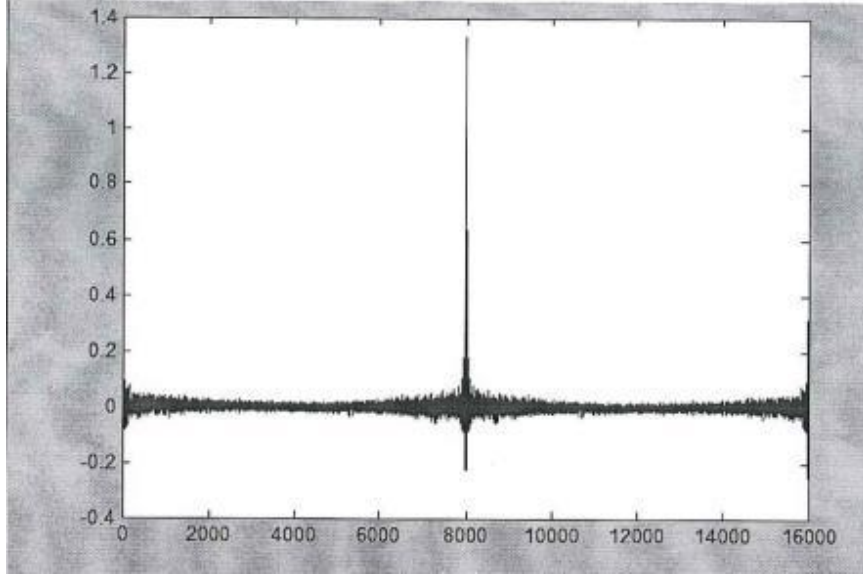


الشكل 11. التقاف الوسط الحسابي مع نافذة هامينك

3-تحليل الطيفي:-

يتم اجراء التحليل الطيفي كما مذكور في المعادلات (3و4و5) للقيم الناتجة من الخطوة الثانية.

4- اللافوف الرياضي:-تطبيق اللافوف الرياضي بين عينة الاختبار مع نفسها كما مذكور في المعادلة (6)، كذلك الشكل (12) يبين ذلك ومن ثم تطبيق اللافوف الرياضي بين عينة الاختبار والوسط الحسابي الشكل (13) يبين ذلك.



الشكل 12. اللافوف الرياضي بين عينة الاختبار مع نفسها

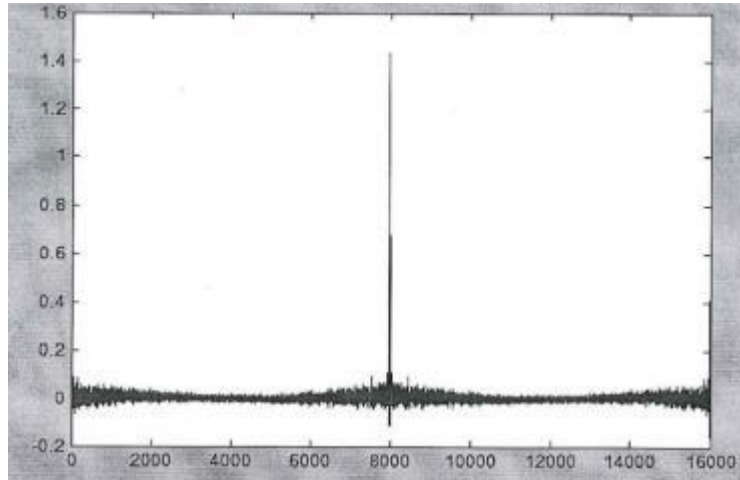
4- استخدام عاملين لتمييز قلب اللدوف هما:

المسافة الأقليدية:

حساب المسافة الأقليدية لقلب اللافوف الرياضي بين عينة الاختبار ونفسها والوسط الحسابي وعينة الاختبار كما ذكر في المعادلة (7) وهذا موضح في الجدول (1) .

معامل الارتباط:

حساب معامل الارتباط اللافوف الرياضي بين عينة الاختبار ونفسها والوسط الحسابي و عينة الاختبار كما ذكر في المعادلة (8) وهذا موضح في الجدول (1) .



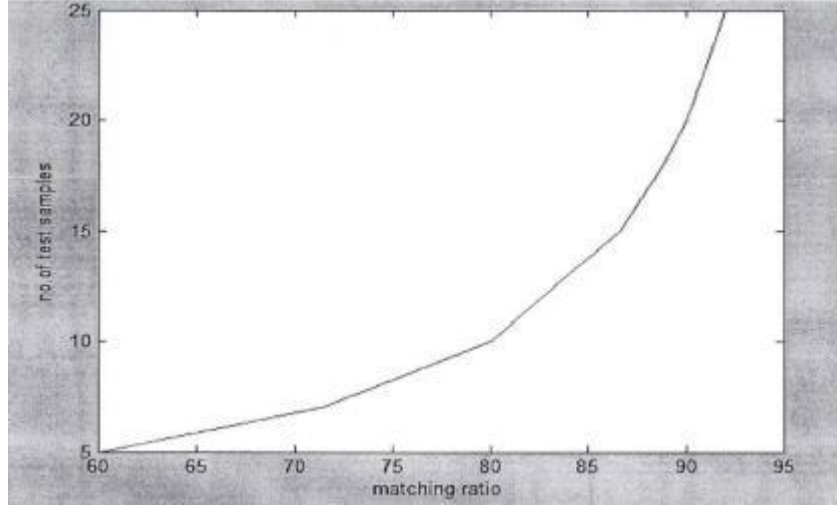
الشكل 13. اللافوف الرياضي بين عينة الاختبار مع الوسط الحسابي

12. الاستنتاجات:

1- نلاحظ كلما زادت عينات الاختبار كلما زادت نسبة التطابق وقلت نسبة الخطاء وهذا يبينه الشكل (14) .

$$\text{نسبة التطابق} = (\text{عدد العينات المتطابقة} | \text{عدد العينات الكلية}) * 100$$

$$\text{نسبة الخطأ} = (\text{عدد العينات الغير المتطابقة} | \text{عدد العينات الكلية}) * 100$$



الشكل 14. تزداد نسبة التطابق كلما زادت عدد عينات الاختبار

- 2- اخذ قيم العينات المسجلة موجبة وصحيحة كان له تأثير واضح على نسبة التطابق .
- 3- اخذ دالة هامينك للإشارة الصوتية لإزالة الضوضاء والترددات الغير مرغوب فيها كان له تأثير على دقة النتائج.
- 4- الحالة النفسية للمتكلم والوضع الهادئ لمكان التسجيل أيضا له تأثير واضح على وضوح الصوت المسجل و على نسبة الضوضاء .
- 5- كان هناك فرق بين التمييز هوية المتكلم عند اعتماد معامل الارتباط والمسافة الأقليدية اظهرت المسافة الأقليدية ضعفا في التمييز بين العينات المتقاربة في القيم، بينما معامل الارتباط كانت النتائج افضل عند اعتماده في تمييز هوية المتكلم، نلاحظ هذا واضحا في الجدول (1) الذي اعتمد المسافة الأقليدية اظهر تمييزا خاطئ لعشرة متكلمين من مجموع (60) متكلم
- 6- كلما كان التباين في قيمة قلب اللافوف الرياضي لأصوات المتكلمين بسيطا كان استخدام المسافة الأقليدية في تمييز هوية المتكلم ضعيفا، بينما معامل الارتباط لا يتأثر بالعينات المتقاربة ولذلك يكون التمييز بنسب عالية.
- 7- تم بناء قاعدة البيانات بأخذ معدل (mean) لأربع حالات تسجيل لكل متكلم تتضمن اوضاع مختلفة للمتكلم مما ادى لتقليل الخطأ في تمييز هوية المتكلم اثناء تطبيق الخوارزمية المقترحة.

شكر وتقدير

نود ان نعبر عن شكرنا وتقديرنا لكلية التربية للعلوم الصرفة /قسم الحاسوب في جامعة الموصل على التسهيلات التي كانت مفيدة للغاية في جودة عملنا .

المصادر

- [1] S. Ghai and R. Sinha ,“Exploring the effect of differences in the acoustic correlates of adults’ and children’s speech in the context of automatic speech recognition,” *Eurasip J. Audio & Speech & Music Process.* ,vol. 2010 ,2010 ,doi: 10.1155/2010/318785.
- [2] C. Yu ,G. Liu ,and J. H. L. Hansen ,“Acoustic Feature Transformation using UBM-based LDA for Speaker Recognition” , *Conference: Interspeech 2014* ,pp.1851-1854,2014 ,doi: [10.21437/Interspeech.2014-420](https://doi.org/10.21437/Interspeech.2014-420).
- [3] S. T. Pan ,“Fuzzy vector quantization on the modeling of discrete hidden markov model for speech recognition,” *Int. J. Fuzzy Syst.* ,vol. 13 ,no. 2 ,pp. 130–139 ,2011 ,doi: 10.30000/IJFS.201106.0008.
- [4] K. P. J. SONIA SUNNY ,DAVID PETER S ,“RECOGNITION OF SPEECH SIGNALS: AN EXPERIMENTAL COMPARISON OF LINEAR PREDICTIVE CODING AND DISCRETE WAVELET TRANSFORMS,” *Int. J. Eng. Sci. Technol.* ,vol. 4 ,no. 4 ,2012.
- [5] M. Siafarikas ,T. Ganchev ,and N. Fakotakis ,“Wavelet packet based speaker verification,” *ODYSSEY04-The Speak. Lang. Recognit. Work.* ,2004.
- [6] O. Haraty ,Ramzi and El Ariss ,“Colloquial Arabic speech recognition application ” ,*American Journal of Applied Sciences* ,vol.4 ,no. 1 ,pp. 23-32 ,2007.
- [7] R. H. and H. A. N. Achmad Rizal ,“Signal Domain in Respiratory Sound Analysis: Methods , Application and Future Development,” *J. Comput. Sci.* ,vol. 11 ,pp. 1005–1007 ,doi: : 10.3844/jcssp.2015.1005.1016.
- [8] F. Alías ,J. C. Socoró ,and X. Sevillano ,“A review of physical and perceptual feature extraction techniques for speech ,music and environmental sounds,” *Applied Sciences* ,vol. 6 ,no. 5. 2016 ,doi: 10.3390/app6050143.
- [9] N. C. Abbas Khosravani ,Cornelius Glackin ,Nazim Dugan ,Gerard Chollet ,“THE INTELLIGENT VOICE 2016 SPEAKER RECOGNITION SYSTEM.” 2016 ,doi: : <https://www.researchgate.net/publication/309631582>.
- [10] C. Ittichaichareon ,“Speech recognition using MFCC,” ... *Conference on Computer* pp. 135–138 ,2012 ,[Online]. Available: <file:///home/benny/Downloads/9712576.pdf>0A<http://psrcentre.org/images/extraimages/712576.pdf>.
- [11] G. Gautam ,S. Shrestha ,and S. Cho ,“Spectral Analysis of Rectangular ,Hanning ,Hamming and Kaiser Window for Digital Fir Filter,” *International journal of advanced smart convergence* ,vol. 4 ,no. 2. pp. 138–144 ,2015 ,doi: 10.7236/ijasc.2015.4.2.138.
- [12] C. Lee ,D. Hyun ,E. Choi ,J. Go ,and C. Lee ,“Optimizing feature extraction for speech recognition,” *IEEE Trans. Speech Audio Process.* ,vol. 11 ,no. 1 ,pp. 80–87 ,2003 ,doi: 10.1109/TSA.2002.805644.
- [13] T. Pihlajamäki ,“Multi-resolution Short-time Fourier Transform Implementation of Directional Audio Coding,” p. 79 ,2009.

- [14] A. S. Thakur and N. Sahayam ‘“Speech Recognition Using Euclidean Distance”’ *Int. J. Emerg. Technol. Adv. Eng.* ‘vol. 3 ‘no. 3 ‘pp. 587–590 ‘2013.
- [15] P. S. Sahoo ‘Soumya ‘Nilamadhab Das ‘“Word Extraction from Speech Recognition using Correlation Coefficients”’ *Int. J. Comput. Appl.* ‘vol. 51 ‘no. 13 ‘2012.