Journal of Techniques, ISSN: 2708-8383, Vol. 5, No. 2, 2023, Pages 206-213
DOI: https://doi.org/10.51173/jt.v5i2.1058

JOURNAL OF TECHNIQUES

Journal homepage: http://journal.mtu.edu.iq

RESEARCH ARTICLE - MANAGEMENT

Comparison of Some Acoustic Noise Models and Their Effect on the Acoustic Diagnosis of Social Media Fingerprints

Sarah Taha Ali[1*], Wleed Abdalaa Araheemah[1], Mohammed Ahmed Taiye[2]

[1] Technical College of Management - Baghdad, Middle Technical University, Baghdad, Iraq

[2] Linnaeus University, Sweden

[*] Corresponding author E-mail: sarahtaha69@yahoo.com

Article Info.

Article history:

Received
01 November 2022

Accepted
03 February 2023

Publishing
30 June 2023

Abstract

The importance of preserving voiceprints as well as verifying their authenticity has increased, especially since reliance on them in the Corona period made many users rely on them in their work in directing administrative orders. As a result, this research came in an attempt to employ several algorithms for neural networks to verify voiceprints for (50-100 (1 person and for each person) (10-20) samples were taken. The results showed that the wavelet transform was affected by (the number of people, the number of sound signatures, and the noise). It was taken from one of the most famous social networking programs, which is (whatsApp), and the results showed that the design for each the type of noise and the type of filter adopted to reduce the effect of noise, reached the highest rating (99.2%) and it is due to the convergent neural network CNN (Band pass filter), while the worst rating reached 95.5% and it is due to the case of the convergent network CNN (AWGN) as shown Results The ability of some filters to increase the classification accuracy of the convention neural network and reduce the effect of noise.

Publisher: Middle Technical University

Keywords: Voiceprint; Noise Models; Audio Filters; Wavelet Transform; Median Filters; Band Pass Filter.

مقارنة بعض نماذج الضوضاء الصوتية وتأثيرها على التشخيص الصوتي لبصمات برامج التواصل الاجتماعي

سارة طه علي[1] *، وليد عبد الله رحيمة[1]، محمد احمد طايع[2]

[1] الجامعة التقنية الوسطى – كلية التقنيات الادارية – بغداد – العراق

[2] جامعة لينيوس، السويد

* البريد الإلكتروني: sarahtaha69@yahoo.com

معلومات المقالة

تاريخ الاستلام
01 تشرين الثاني 2022

تاريخ القبول
03 شباط 2023

تاريخ النشر
30 حزيران 2023

الخلاصة

تزايدت اهمية المحافظة على البصمات الصوتية فضلاً عن التحقق من عائديتها خاصة وأن الاعتماد عليها في مدة كورونا جعل الكثير من المستخدمين يعتمدونها في أعمالهم في توجيه الاوامر الإدارية نتيجة لذلك جاء هذا البحث في محاولة لتوظيف خوارزميات عدة لشبكات عصبونية للتحقق من البصمات الصوتية لـ(50 – 100) شخص ولكل شخص تم اخذ (10 – 20) عينة اظهرت النتائج تأثر التحويل المويجي بكل من (عدد الاشخاص، عدد البصمات الصوتية، الضوضاء) وقد تم اخذها من احد اشهر برامج التواصل الاجتماعي وهو برنامج (whatsApp)، وقد اظهرت النتائج تأثر التصميم لكل من نوع الضوضاء ونوع المرشح المعتمد لتقليل تأثير الضوضاء وقد بلغ اعلى تصنيف (99,2%) وهو يعود الى الشبكة العصبونية التلافيفية CNN (Band pass filter) في حين بلغ اسوء تصنيف 95.5% وهو يعود الى حالة شبكة التلافيفية CNN (AWGN) كما أظهرت النتائج قدرة بعض المرشحات على زيادة دقة التصنيف العائدة الى الشبكة العصبونية التلافيفية وتقليل أثر الضوضاء.

الكلمات المفتاحية: البصمة الصوتية؛ نماذج الضوضاء؛ الفلاتر الصوتية؛ التحويل المويجي؛ مرشح الوسيط؛ مرشح النطاق الترددي.

1. المنهجية

1.1. المقدمة

ان انتشار البصمات الصوتية خاصة مع وجود التطبيقات المساعدة في مجال التواصل الاجتماعي ادى الى تزايد الاهتمام بهذه البصمات والحفاظ عليها من التلاعب، ولغرض التحقق من البصمة الصوتية تم الاعتماد على انظمة التحقق عالية المستوى، إذ تقوم على خدمة هذا النوع من الملفات، وهناك سلبيات تؤخذ على البصمات الصوتية أنها تحتاج الى مساحات خزنية كبيرة ولذلك فهي تحتاج الى توفير سيرفرات ذات تكلفه عالية، إذ تم في هذا البحث تغطية الجوانب النظرية للبصمات الصوتية في مجال مواقع التواصل الاجتماعي، ففي الجزء (2) سنقوم بشرح المبادى النظرية حول انظمة التحليل الصوتي للملفات الصوتية التي تكون من ضمنها ملفات البصمات الصوتية لشبكات التواصل الاجتماعي، اما في الجزء (3) فسنقوم بشرح النظام

Nomenclature & Symbols	
Per._mo.	Perecentage_Mono
Per._st.	Percentage_Stereo
Acc.	Accuracy
AWGN	Additive White Gaussian Noise
CNN	Convolutional Neural Network

المقترح في هذا البحث، إذ إن النظام يمثل احد انظمة تمييز الانماط Pattern Recognition التي تحتاج الى عمليات عدة متسلسلة تبدأ بالمعالجة الاولية pre-processing وانتهاءً بتقيم النتائج هي مرحلة التقييم evaluation.

ان المراحل جميعاً التي تم تمرير البيانات الصوتية عليها تم شرحه بالتفصيل في الجزء (3) اما بالنسبة للجزء (4) فقمنا بتوضيح الاستنتاجات المستحصلة من تطبيق النظام المقترح على البيانات التي تم جمعها واخيرا قمنا بإعطاء افكارا مستقبلية حول كيفية تطوير النظام المقترح ليتماشى مع مجريات التطور[1].

1.2. مشكلة البحث

ان انتقال البصمة الصوتية عبر الشبكة قد يعرضها الى ضوضاء وتلاعب مما يؤدي الى تغير صفات هذه البصمة ومن ثم حدوث مشاكل في تحديد عائديتها.

1.3. هدف البحث

يهدف البحث الى توظيف التحويل المويجي (Wavelet transform) في:

- التحقق الصوتي للبصمات الصوتية في مواقع التواصل الإجتماعي
- التحقق من تأثيرنماذج الضوضاء الصوتية
- كما يهدف الى توظيف بعض المرشحات الصوتية.

1.4. أهمية البحث

تعّد ملفات الوسائط المتعددة ومنها البصمات الصوتية من الامور المتزايدة الاهمية وأن لاكتشاف العائدية الصوتية فضلاً عن حفظ البصمة الصوتية من التلاعب، بدأت تزداد أهميته مع ازدياد حالات التهديد والمساومات التي تحدث لإنتشار مقاطع صوتية تغيب عائديتها او لا تحدد الوجهة الصحيحة لها.

1.5. فرضيات

- يمكن للضوضاء ان تؤثر على تشخيص البصمة الصوتية
- يمكن لعدد العينات ان يؤثر على تشخيص البصمة الصوتية
- يمكن لعدد الاشخاص ان يؤثر على تشخيص البصمة الصوتية
- قدرة التحويل المويجي على تشخيص البصمة الصوتية

1.6. الدراسات السابقة

- في عام 2019 قام الباحث (Das, Soubhik) باستخدام الطرق الذكية للتعرف على الكلام وذلك من خلال معالجة الصوت المتمثل في نموذج معين للكشف عن مشاكل الجهاز التنفسي. في هذا البحث تُستخدم مجموعة أدوات التعلم العميق من شركة Intel لإجراء عملية التصنيف وتم استخدام مجموعة أدوات Intel Open VINO بشكل اساس. اما بالنسبة للبيانات فإن مجموعة البيانات التي تم استخدامها لمعالجة المشكلة يدويًا قد أنتجت. أثناء التدريب الصوتي، في مرحلة تجميع البيانات تم تدريب النظام على الانماط الصوتية جميعاً، ومن بينها حالة عدم المقدرة على النطق بشكل واضح. وذلك من أجل اخذ التغييرات الرئيسة جميعاً أثناء الكلام في حالة وجود مشاكل الجهاز التنفسي وجعل النظام قادرا على تمييز العوارض التنفسية بدقة. بعد تدريب النموذج تمت عملية الاختبار بأخذ اشخاص عشوائيين. تم استخدام CaffeNet كنموذج تصنيف [1].
- في عام 2020 قدم الباحث (Song, Zhaojuan) باستخدام الشبكة العصبية التلافيفية (ConvNet / CNN) وهي نموذج التعلم العميق الذي يمكن أن يأخذ بيانات الإدخال ويقوم بإجراء عمليات رياضية معقدة عليها وتقوم الشبكة ذاتيا بتعيين الأهمية والأولويات للأوزان القابلة للتعلم لعناصر/ كائنات مختلفة في البيانات وتكون قادرة على التمييز واحد من الآخر. باسو وآخرون. نشر نظرة عامة وشاملة في بحثه ولكنها موجزة إذ تركز على ورقته البحثية على استخدام هذا النوع من الشبكات في مجال تمييز بصمة الصوت، والحد من الضوضاء. في هذا البحث قام الباحثون بتطويرشبكة عصبية عميقة deep neural network وذلك من خلال استخدام المنهجية التوافقية لاسترداد الميزات Features الافضل تمييزا للكلام، إذ تم اختيار نتيجة الطبقة المخفية الثابتة على أنها الافضل في تحديد خاصية الكلام للشبكة التي تم إنشاؤها حديثًا، وتدريب النموذج الصوتي بالميزات الجديدة من خلال إستخلاص المعلومات الصوتية ذات العلاقة [2].
- وفي عام 2021 قدم الباحث (Ashraf Tahseenوأخرون) في بحثهم قاموا بتطويرشبكة عصبية عميقة deep neural network وذلك من خلال استخدام المنهجية التوافقية لاسترداد الميزات Features الافضل تمييزا للكلام، إذ تم اختيار نتيجة الطبقة المخفية الثابتة على أنها الافضل في تحديد خاصية الكلام للشبكة التي تم إنشاؤها حديثًا، وتدريب النموذج الصوتي GMM – HMM بالميزات الجديدة من خلال إستخلاص المعلومات الصوتية ذات العلاقة [3].

اما هذا البحث يتضمن مقارنة البصمات الصوتية لمواقع التواصل الاجتماعي وتحديد فيما اذا حصل تلاعب لهذه البصمات ام لا من خلال تطبيق التحويل المويجي على البصمة الصوتية واستخراج البصمات الصوتية تمهيدا لاستخدامها في الشبكة العصبونية التلافيفية (CNN) وتحديد عائدية البصمة الصوتية للشخص المعني من فرض وجود تلاعب قي البصمة الصوتية من عدمها.

2. الجانب النظري

2.1. التحويل الموجي (Continuous Wavelet Transformation) CWT

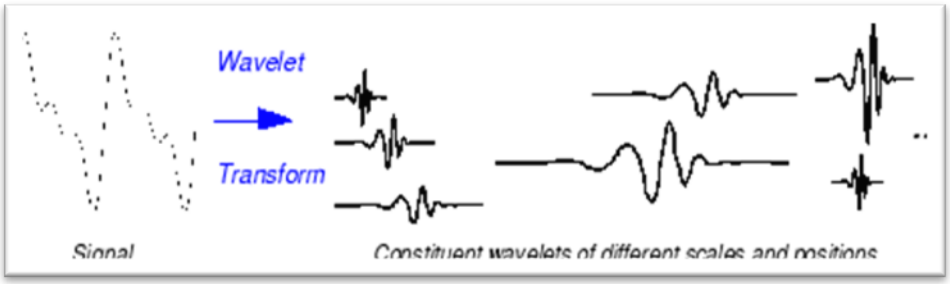
يستخدم التحويل المويجي المستمر (CWT) المنتجات الداخلية لقياس التشابه بين الإشارة ووظيفة التحليل في تحويل فوريير، تكون وظائف التحليل أسية معقدة، على سبيل المثال eJωt التحويل الناتج هو دالة لمتغير واحد، ω في تحويل فوريير قصير الوقت، تكون وظائف التحليل عبارة عن نوافذ أسية معقدة، w (t) ejωt، والنتيجة هي دالة لمتغيرين تمثل معاملات STFT، F (ω, τ) التطابق بين الإشارة والجيوب الأنفية ذات التردد الزاوي ω في فاصل من الطول المحدد المتمركز عند τ في CWT، وظيفة التحليل هي الموجة، ψ يقارن CWT الإشارة بالإصدارات المنقولة والمضغوطة أو الممتدة من الموجة يشار إلى تمديد أو ضغط وظيفة بشكل جماعي على أنه تمدد أو تحجيم ويتوافق مع المفهوم المادي للمقياس بمقارنة الإشارة بالمويجة بمقاييس ومواضع مختلفة، تحصل على دالة من متغيرين التمثيل ثنائي الأبعاد لإشارة 1-D زائد عن الحاجة إذا كانت الموجة ذات قيمة معقدة، فإن CWT هي دالة ذات قيمة معقدة للمقياس والموضع إذا كانت الإشارة ذات قيمة حقيقية، فإن CWT هي دالة ذات قيمة حقيقية للمقياس والموضع بالنسبة إلى معلمة المقياس، a> 0، والموضع، b، يكون CWT سنقدم، لكل عائلة من المويجات، وظيفة التحجيم(t) Φ في وظيفة المويجات Ψ (t) وقيم فلاترها (لاستعادة الفلتر ثنائي القناة) في كل من تمثيل المجال الزمني والمجال الترددي هو [4]:

$$C(a, b; f(t), \Psi(t)) = \int_{-\infty}^{\infty} f(t) \frac{1}{a} \Psi * \left(\frac{t-b}{a}\right) dt \quad (1)$$

اي f(t) هي إشارة ليتم تحويلها، φ(t) هو الاقتران المعقد للدالة الأم، a هو مقياس التحليل، b وهو الوقت الذي أخذ التحويل F(t) [5]. ما يعادل التفاف الإشارة مع مكافئ لارتداد بدافع الاستجابة h(t) عندما:

$$h_{(t)} = \frac{1}{\sqrt{a}}\ \varphi\left(\frac{t-b}{a}\right) \quad (2)$$

يستخدم العامل $\frac{1}{\sqrt{a}}$ للحفاظ على القاعدة. الآن، تستخدم الدالات في توسعة البلاط بتكرار الوقت بالنسبة إلى الصغيرة a(a< 1)، h(t) سوف تكون قصيرة وذات تردد عالٍ، بينما a(a> 1) تكون كبيرة، h(t) وتكون طويلة وذات تردد منخفض. كما رأينا في المعادلة (28.2)، تعّد الإشارة المحولة $\varphi(t)$ دالة لمتغيرين، ومعلمات المقياس والترجمة على التوالي لاحظ الشكل(1).



شكل (1) الموجات المكونة من مقاييس ومواقع مختلفة [5]

2.2. البصمة الصوتية

تستخدم بعض أقسام الشرطة، في الولايات المتحدة الأمريكية، بصمات الصوت، كدليل في القضايا الجنائية؛ وتستخدم الاجهزة الامنية بصمة الصوت في تتبع الاشخاص المطلوبين بمجرد مطابقة بصمات اصواتهم عبر الهواتف مع تسجيلات سابقة لهم ومن ثم يتم تحديد أماكنهم. لكن بعض الخبراء، يعتقدون أنه من الصعب تفسير بصمات الصوت، وأنها ليست دقيقة، بما يكفي لاستخدامها في تتبع الاشخاص وخاصة بعد ظهور التقنيات الحديثة التي من شأنها تزيف نبرة الصوت.

تعتمد البصمة الصوتية على مبدأين هما أن لكل إنسان جهازا صوتيا فريدا لا يشابهه أحد فيه، الجهاز الصوتي هي أعضاء الجسم التي تساعد في إخراج الصوت مثل: الفم اللسان، القفص الصدري... الخ من حيث شكل وحجم الأعضاء وارتباط بعضها ببعض، وأن لكل إنسان نظاما عصبيا فريداً يتحكم في الجهاز الصوتي. [7,6].

2.3. مميزات البصمة الصوتية

- لا تحتاج البصمة الصوتية إلى أجهزة متخصصة لالتقاط البيانات الحيوية من الشخص مثل بصمة الأصبع وبصمة العين، فسماعة الهاتف أو لاقط الصوت المرفق مع أجهزة الحاسوب يقوم بالمهمة.
- عالجت البصمة الصوتية مشكلة السرقة والتزوير للأرقام السرية والبطاقات.
- باستخدام البصمة الصوتية يمكن التعرف والتحقق من الشخص عن بعد.
- ساعدت البصمة الصوتية الأشخاص على عدم حفظ الأرقام السرية [8].

2.4. الضوضاء الصوتية Noise

هناك العديد من أنواع ومصادر الضوضاء أو التشوهات وتشمل:

- ضوضاء إلكترونية مثل الضوضاء الحرارية وطلقات الرصاص الضوضاء.
- الضجيج الصوتي الناجم عن الحركة والاهتزاز أو مصادر الاصطدام مثل الآلات الدوارة، المركبات المتحركة، نقرات لوحة المفاتيح، الرياح والأمطار.
- الضوضاء الكهرومغناطيسية التي يمكن أن تتداخل مع إرسال واستقبال الصوت.

تشويه الإشارة هو المصطلح الذي يستخدم غالبًا لوصف تغيير منهجي غير مرغوب فيه في إشارة ويشير إلى التغييرات في إشارة من الخصائص غير المثالية من قناة الاتصال، تتلاشى الإشارة أصداء [9].

وانعكاسات متعددة المسارات وعينات مفقودة [10]. بحسب ترددها، خصائص الطيف أو الوقت، عملية الضوضاء مصنفة كذلك في فئات عدة:

2.4.1. الضوضاء البيضاء

الضوضاء العشوائية البحتة لها دافع وظيفة الارتباط التلقائي وطيف القدرة المسطح تحتوي الضوضاء البيضاء نظريًا على الترددات جميعاً بتنسيق قوة متساوية.

2.4.2. ضجيج النطاق الضيق

إنها عملية ضوضاء مع عرض النطاق الترددي الضيق مثل 60/50 هرتز.

2.4.3. الضوضاء الملونة

هي ضوضاء غير بيضاء أو أي ضوضاء ضوضاء النطاق العريض طيفها غير مسطح. ومن الأمثلة على ذلك الضوضاء الوردية والضوضاء البنية وضوضاء الانحدار التلقائي.

2.4.4. الضجيج المندفع

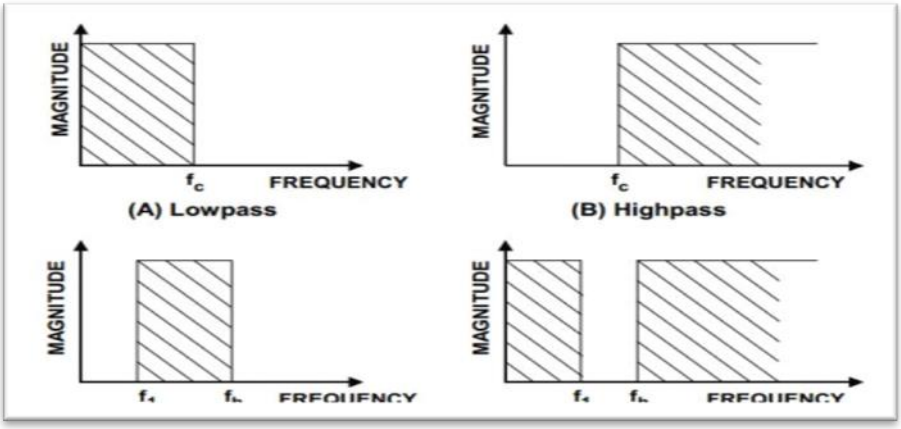
يتكون من نبضات قصيرة المدة السعة العشوائية ووقت الحدوث والمدة الزمنية.

2.4.5. نبضات ضوضاء عابرة:

تتكون من طويلة نسبياً مدة نبضات الضوضاء مثل النقرات والضوضاء المتفجرة وما إلى ذلك.

2.5. المرشحات الصوتية Filters

يمكن تطوير الوظيفة الرئيسة للمرشح بواسطة اختبار الطبيعة المعتمدة على التردد مقاومة المحرّضات والمكثفات تردد تغيير قيم كلا من رد الفعل تتغير المعاوقة، كما تتغير نسبة مقسم الجهد على التوالي. ينتج عن هذه العملية التغيير في الإدخال/ الإخراج تعتمد وظيفة النقل على التردد، فهي كذلك المعروف باسم استجابة التردد المكمل الوظيفي لمرشح الترددات المنخفضة هو مرشح الترددات العالية. هنا، الترددات المنخفضة نطاق التوقف، والترددات العالية فرقة التمرير. يوضح الشكل رقم (2) التمريرة العالية المثالية مرشح، مرشح تمرير منخفض، مرشح تمرير النطاق، رفض النطاق مرشح [12, 11].



الشكل (2) انواع المرشحات filters [13]

2.6. مقاييس تقييم الاداء

في هذا الجزء سنتكلم عن اهم المقاييس المتبعة لتقييم اداء انظمة تمييز الانماط ومن ضمنها انظمة التمييز الصوتي والمرئي وهي كالآتي:

Accuracy - الدقة هي مقياس الأداء الأكثر بديهية وهي ببساطة نسبة الملاحظة المتوقعة بشكل صحيح إلى إجمالي الملحوظات. قد يعتقد المرء إنه إذا كانت لدينا دقة عالية فإن نموذجنا هو الأفضل. نعم، تعَد الدقة مقياسًا رائعًا ولكن فقط عندما يكون لديك مجموعات بيانات متماثلة حيث تكون قيم السلبيات الزائفة والموجبة الزائفة متماثلة تقريبًا. لذلك، عليك أن تنظر إلى معلمات أخرى لتقييم أداء النموذج الخاص بك.

precision – هذا المقياس ايضا يسمى الدقة ويمثل نسبة القيم المتوقعة بشكل صحيح إلى إجمالي القيم الإيجابية المتوقعة.

Recall الاستدعاء (الحساسية) - الاستدعاء هو نسبة القيم المتوقعة بشكل صحيح إلى القيم جميعاً اي القيم الايجابية والسلبية.

F1-score - درجة F1 هي المتوسط المرجح للدقة والاستدعاء. لذلك، تأخذ هذه النتيجة في الحسبان الإيجابيات الزائفة True Negative والسلبيات الخاطئة False Negative. حدسيًا، ليس من السهل فهم الدقة، لكن F1 عادةً ما تكون أكثر فائدة من الدقة، خاصةً إذا كان لديك توزيع فئة غير متساوٍ. تعمل الدقة بشكل أفضل إذا كانت الإيجابيات الكاذبة False Positive والسلبيات الكاذبة False Negative لها التأثير نفسه. ما إذا كانت تكلفة الإيجابيات الكاذبة والسلبيات الخاطئة مختلفة تمامًا، فمن الأفضل النظر إلى كل من الدقة والاستدعاء [14].

2.7. انواع الامتدادات الصوتية

تُستخدم ملفات الصوت بشكل شائع لتخزين الصوت الرقمي مثل الموسيقى أو المؤثرات الصوتية التي يتم تشغيلها بعد ذلك من خلال البرامج المرتبطة بها. نظرًا لأن الصوت معقد نوعًا ما، فإن تخزين البيانات بتنسيق رقمي يمكن أن ينتج عنه أحجام ملفات كبيرة. من الشائع ضغط بيانات الصوت لتقليل الحجم الكلي للملف.هناك نوعان رئيسان من الضغط ؛ بدون فقدان الذي يضغط الصوت دون فقدان الجودة وينتج عنه نسبة 2: 1 في حجم الملف مقارنة بتخزين البيانات بتنسيق خام وفقدان مما يقلل من حجم الملف بشكل أكبر ولكنه يقلل من جودة الصوت. فضلا عن تنسيقات تخزين الصوت، تحتوي القائمة التالية أيضًا على التنسيقات المرتبطة بأجهزة وبرامج الصوت. لكل نوع ملف صوتي مزايا وعيوب فريدة. فيما يلي سبعة أنواع شائعة من الملفات الصوتية وبعض الاختلافات الفريدة بينها:

2.7.1. نوع ملف الصوت M4A

M4A هو ملف صوتي MPEG-4. إنه ملف مضغوط صوتي يستخدم في الإعداد الحديث بسبب زيادة الطلب على الجودة نتيجة للتخزين السحابي ومساحة أكبر لمحرك الأقراص الثابتة في أجهزة الكمبيوتر المعاصرة. جودته العالية تجعله ملائمًا، إذ سيحتاج المستخدمون الذين يحتاجون إلى سماع أصوات مميزة في ملفات الصوت إلى ذلك على أنواع الملفات الأكثر شيوعًا. تستخدم برامج تنزيل الموسيقى مثل Apple iTunes M4A بدلاً من MP3 لأنها أصغر حجمًا وجودة أعلى. تأتي قيودها في شكل توافق، حيث أن الكثير من البرامج غير قادرة على التعرف على M4A، مما يجعلها مثالية لنوع محدد فقط من المستخدمين [15] .

2.7.2. FLAC

ملف الصوت FLAC هو Free Lossless Audio Codec. إنه ملف صوتي مضغوط إلى حجم أصغر من الملف الأصلي. إنه نوع ملف معقد أقل استخدامًا بين تنسيقات الصوت. هذا لأنه على الرغم من أن له مزايا، إلا أنه غالبًا ما يحتاج إلى تنزيلات خاصة ليعمل. عندما تفكر في أن الملفات الصوتية تتم مشاركتها كثيرًا، فقد يتسبب ذلك في إزعاج كل مستخدم جديد يتلقى ملفًا. ما يجعل FLAC مهمًا للغاية هو أن الضغط بدون فقدان يمكن أن يوفر الحجم ويعزز مشاركة ملف صوتي مع القدرة على العودة إلى معيار الجودة الأصلي. تبلغ مساحة التخزين المطلوبة تقريبًا لملف الصوت الأصلي ستين من المائة - وهذا يوفر الكثير من مساحة محرك الأقراص الثابتة والوقت المستغرق في التحميل أو التنزيل.[16] .

2.7.3. MP3

ملف الصوت MP3 هو تنسيق ملف MPEG Audio layer 3. الميزة الرئيسة لملفات MP3 هي الضغط الذي يوفر مساحة قيمة مع الحفاظ على جودة شبه خالية من العيوب لمصدر الصوت الأصلي. هذا الضغط يجعل MP3 شائعًا جدًا لجميع أجهزة تشغيل الصوت المحمولة، وخاصة Apple iPod. ما يزال MP3 مناسبًا للمشهد الرقمي اليوم لأنه متوافق مع كل جهاز قادر على قراءة الملفات الصوتية تقريبًا. ربما يكون أفضل استخدام لملف MP3 هو مشاركة الملفات الصوتية على نطاق واسع نظرًا لحجمها القابل للإدارة. كما أنه يعمل بشكل جيد مع مواقع الويب التي تستضيف ملفات صوتية. أخيرًا، ما يزال MP3 شائعًا بسبب جودة الصوت الإجمالية. على الرغم من أنها ليست أعلى جودة، إلا أن لها فوائد أخرى كافية للتعويض .[17]

2.7.4. MP4

غالبًا ما يُخطئ ملف صوتي MP4 على أنه نسخة محسنة من ملف MP3. ومع ذلك، هذا لا يمكن أن يكون أبعد عن الحقيقة. كلاهما مختلفان تمامًا وأوجه التشابه تأتي من الاسم نفسه بدلاً من وظيفتهما. لاحظ أيضًا أنه يُشار أحيانًا إلى MP4 على أنه ملف فيديو بدلاً من ملف صوتي. هذا ليس خطأ، لأنه في الواقع ملف صوتي وفيديو. نوع ملف الصوت MP4 هو امتداد وسائط شامل، قادر على الاحتفاظ بالصوت والفيديو والوسائط الأخرى. يحتوي MP4 على بيانات في الملف، بدلاً من رمز. من المهم ملاحظة أن ملفات MP4 تتطلب برامج ترميز مختلفة لتنفيذ الكود بشكل مصطنع والسماح بقراءته [17] .

2.7.5. WAV

ملف الصوت WAV هو ملف صوتي موجي يخزن بيانات شكل الموجة. تقدم بيانات شكل الموجة المخزنة صورة توضح قوة الحجم والصوت في أجزاء معينة من ملف WAV. من الممكن تمامًا تحويل ملف WAV باستخدام الضغط، على الرغم من أنه ليس قياسيًا. أيضًا، يتم استخدام WAV عادةً على أنظمة Windows. أسهل طريقة لتصور هذا المفهوم هي التفكير

في أمواج المحيط. يكون الماء أعلى صوت، ويمتلئ وأقوى عندما تكون الموجة عالية. وينطبق الشيء نفسه على شكل الموجة في WAV. تكون المرئيات عالية وكبيرة عندما يزيد الصوت في الملف. عادةً ما تكون ملفات WAV ملفات صوتية غير مضغوطة، على الرغم من أنها ليست من متطلبات التنسيق [18].

WMA .2.7.6

WMA (Windows Media Audio) هو بديل يستند إلى Windows لنوع ملف MP3 الأكثر شيوعًا وشعبية. ما يجعله مفيدًا للغاية هو ضغطه الخالي من الضياع، مع الاحتفاظ بجودة الصوت العالية في أنواع عمليات إعادة الهيكلة جميعاً. على الرغم من أنه تنسيق صوتي عالي الجودة، إلا أنه ليس الأكثر شيوعًا نظرًا لحقيقة أنه لا يمكن الوصول إليه من العديد من المستخدمين، وخاصة أولئك الذين لا يستخدمون نظام التشغيل Windows. إذا كنت من مستخدمي Windows، فما عليك سوى النقر نقرًا مزدوجًا فوق أي ملف WMA لفتحه. سيفتح الملف باستخدام Windows Media Player (إلا إذا قمت بتغيير البرنامج الافتراضي). إذا كنت لا تستخدم Windows، فهناك بعض البدائل. الخيار الأول هو تنزيل نظام جهة خارجية يقوم بتشغيل WMA [18].

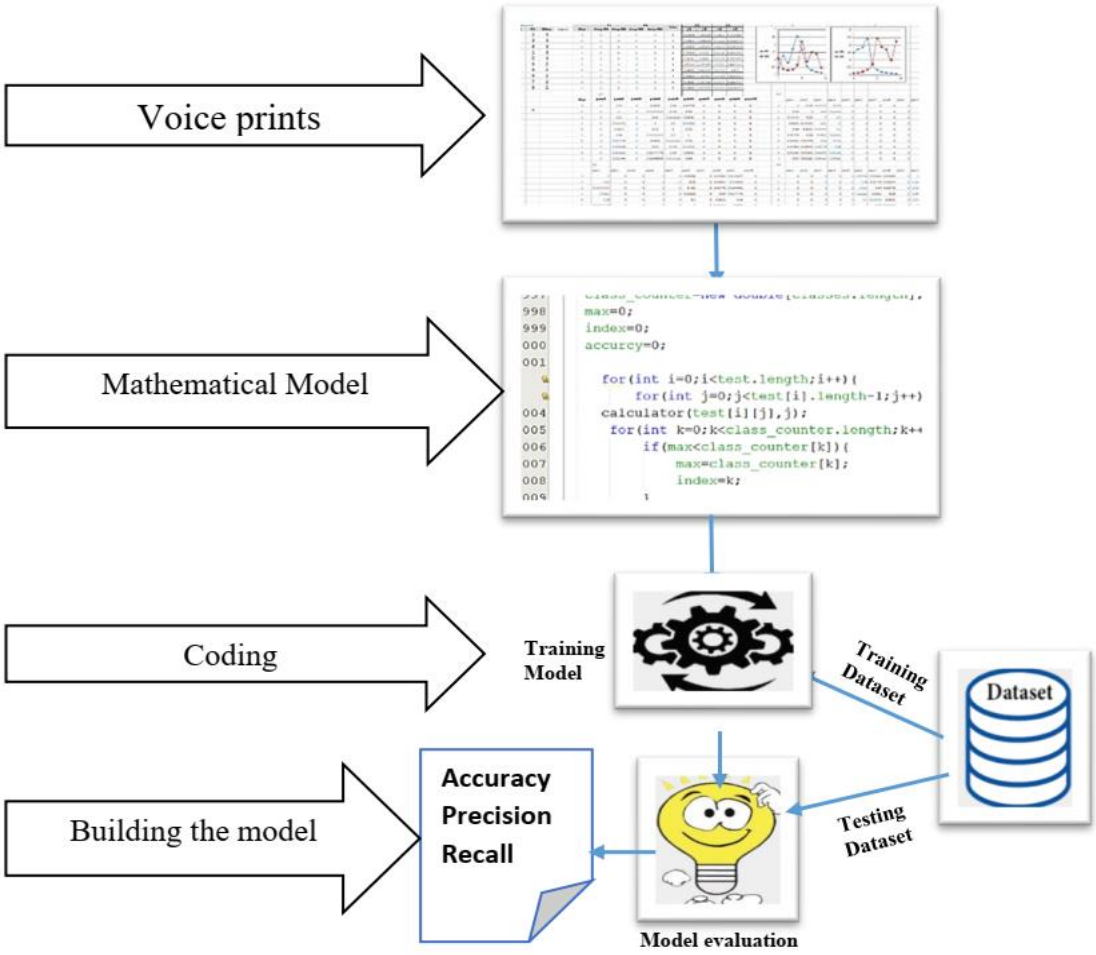
AAC .2.7.7

AAC (ترميز الصوت المتقدم) هو ملف صوتي يوفر صوتًا عالي الجودة بشكل لائق ويتم تحسينه باستخدام الترميز المتقدم. لم يكن أبدًا أحد أكثر تنسيقات الصوت شيوعًا، خاصةً عندما يتعلق الأمر بملفات الموسيقى، لكن AAC ما يزال يخدم بعض الأغراض للأنظمة الرئيسة. يتضمن ذلك الأجهزة المحمولة الشائعة ووحدات العاب الفيديو، حيث يعّد AAC مكونًا صوتيًا قياسيًا. لفتح ملف AAC، يكون التنسيق الأكثر شيوعًا والمباشر لمعظم المستخدمين من خلال iTunes. كل هذا يستلزم تشغيل نظام iTunes وفتح ملف AAC من جهاز الكمبيوتر في قائمة "ملف" [18].

3. الجانب العملي

يتألف النظام المقترح وكما موضوع في الشكل من الخطوات التالية:

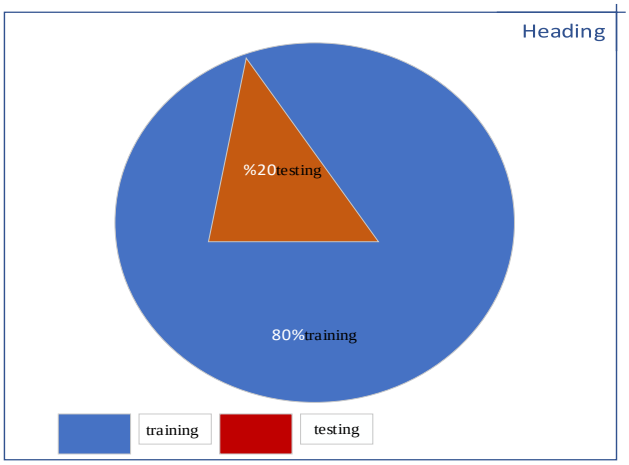
- مرحلة ادخال البيانات الصوتية او البصمة الصوتية الى النظام وتحويلها بواسطة فلاتر معينة الى نظام ثنائي الابعاد.
- مرحلة التمثيل الرياضي للخوارزميات التي سيتم استخدامها على هذه البيانات وذلك من خلال دراسة المعادلات التي سيتم برمجتها لاحقا.
- مرحلة البرمجة التي تبدا بعمل الشبكة العصبية التي سيتم تطبيقها على البيانات المدخلة.
- مرحلة بناء الموديل وهي بداية عملية التدريب وايضا تتضمن عملية الاختبار والتقييم كما موضح في الشكل (3) فإن المراحل الـ4 تم تلخصيها في هذا الشكل.



الشكل (3) خطوات الطريقة المقترحة

3.1. تطبيق التصنيف بواسطة الشبكة العصبية التلافيفية (CNN)

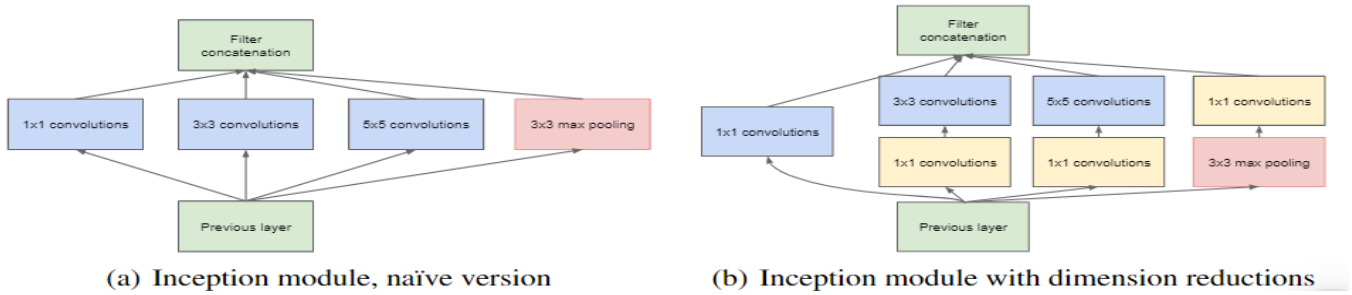
بعد عملية التحويل الى النمط ثنائي الابعاد يتم تقسيم البيانات كما موضح في الشكل التالي الـ training و testing ويتم ذلك بتقسيم البينات كما موضح في الشكل (4) مرحلة الاختبار تقييم النموذج إذا كان صحيحًا أم لا من خلال حساب الدقة والدقة والاسترجاع وقياس F ومعدل الخطأ (Loss) اعتمادًا على مصفوفة الارتباك سيتم توقع كل مثيل لمجموعة بيانات الاختبار لمعرفة ما إذا كان التنبؤ صحيحًا أم خطأ، على سبيل المثال، لنفترض أن لدينا مصفوفة باسم "اختبار" تحتوي على بيانات لنموذج الاختبار المكون من 4 مثيلات مع ميزتين وفئتين.



الشكل (4) تقسيم البيانات الى 20% بيانات خاصة بالاختبار، و80% بيانات خاصة بالتدريب

تم توصيل طبقة الإدخال بمزيج من طبقات الالتفاف والتجميع طبقات لتعلم ميزات الصورة في هذه الطريقة المقترحة، تم اقتراح ثلاث طبقات تلافيفية، حيث تبعت كل طبقة التفاف بطبقة تجمع تحدد امتداد أهم الميزات من منطقته الاستقبالية ويقلل من عدد المعلمات المطلوبة لتدريب النموذج في نموذجنا، تم استخدام وظيفة max-pooling، الذي ينتج الحد الأقصى للقيمة في حيز المستطيل الناتج من الطبقة السابقة ربطت الطبقات المتصلة بالكامل في نهاية CNN الذي يكون ناتجا من التواء طبقة الاخراج في متجه كثيف Dense vector تم تمرير هذا الناتج إلى مرحلة التنبؤ النموذجي optimal predicationتم بعد ذلك تحويل المتجه الكثيف إلى مخرجات نموذج من خلال الارتباط بالكامل في الطبقة التي تليها التي تحتوي على 92 ناتجًا تشير إلى عدد الفئات المستخدمة كانت هذه الطبقة الناتجة عبارة عن طبقة Sigmoid، التي نتجت توزيعاً ثنائي الابعاد 2-dimensional في المدى من 0-1، حيث K هوعدد epochs (K = 2). تم اقتراح Google Net (أو Inception V1) من خلال بحث في Google (بالتعاون مع جامعات مختلفة) في عام 2014 في ورقة بحثية بعنوان " Going Deep with Convolutions" كانت هذه البنية هي الفائز في تحدي تصنيف الصور ILSVRC 2014 لقد قدم انخفاضًا كبيرًا في معدل الخطأ مقارنة بالفائزين السابقين AlexNet (الفائز في ILSVRC 2012) و ZF-Net (الفائز في ILSVRC 2013) ومعدل خطأ أقل بكثير من VGG (المركز الثاني في 2014).

تختلف بنية GoogLeNet اختلافًا كبيرًا عن الشبكات السابقة مثل AlexNet و ZF-Net يستخدم العديد من أنواع الطرق المختلفة مثل الالتواء 1 × 1 وتجميع المتوسط العالمي الذي يمكّنه من إنشاء بنية أعمق، في هذا النوع يتم استخدام التلافيفات لتقليل عدد المعلمات (الأوزان والتحيزات) للشبكة من خلال تقليل المعلمات، نزيد أيضًا من عمق البنية لنلق نظرة على مثال التفاف 1 × 1 كما موضح في الشكل (5) على سبيل المثال، إذا أردنا إجراء التفاف 5 × 5 باستخدام 48 مرشحًا يمكن تمثيل هذا النوع من الشبكات كما موضح في الجدول (1).



الشكل (5) شبكة عصبية من نوع google net

الجدول (1) المتغيرات الخاصة بخوارزمية ال Google Net

type	Patch siz/stride	Output size	depth	#1*1	#3*3 reduce	#3*3	#5*5 reduce	#5*5	Pool proj	params	ops
convolution	7*7/2	112*112*64	1							2.7k	34m
Max pool	3*3/2	56*56*64	0								
convolution	3*3/1	56*56*192	2		64	192				112k	360m
Max pool	3*3/2	28*28*192	0								
Inception(3a)		28*28*256	2	64	96	128	16	32	32	159k	128m
Inception(3b)		28*28*480	2	128	128	192	32	96	64	380k	304m
Max pool	3*3/2	14*14*480	0								
Inception(4a)		14*14*512	2	192	96	208	16	48	64	364k	73m
Inception(4b)		14*14*512	2	160	112	224	24	64	64	437k	88m
Inception(4c)		14*14*512	2	128	128	256	24	64	64	463k	100m
Inception(4d)		14*14*528	2	112	144	288	32	64	64	580k	119m
Inception(4e)		14*14*832	2	256	160	320	32	128	128	840k	170m
Max pool	3*3/2	7*7*832	0								
Inception(5a)		7*7*832	2	256	160	320	32	128	128	1072k	54m
Inception(5b)		7*7*1024	2	384	192	384	48	128	128	1388k	71m
Avg pool	7*7/1	1*1*1024	0								
Dropout(40%)		1*1*1024	0								
linear		1*1*10001	1							1000k	1m
softmax		1*1*1000	0								

3.2. تجربة رقم (1)

تم اخذ أعداد مختلفة للأشخاص الذين تعود لهم البصمة الصوتية لتكون 50 شخصاً مرة و100 شخص مرة ثانية ولكل عدد مأخوذ من الاشخاص تم اخذ عينات بصمات صوتية مختلفة لتكون 10 بصمات مرة و20 بصمة مرة ثانية ولكل الحالات السابقة تم اخذ البصمة الصوتية مرة والبصمة الصوتية بعد اضافة ضوضاء additive white Gaussian noise مرة واستخدام Band pass filter على البصمة الصوتية مرة ومرشح الوسيط median filter على البصمة الصوتية مرة اخيرة وبذلك تم تطبيق الشبكة العصبونية التلافيفية (CNN) لاستخراج مستوى الدقة والوقت المستغرق وكما في الجدول (2).

جدول رقم (2) لعينة من 50 شخصاً

simple	Dataset	Per._mo.	Per._st.	Time	Acc.	Recall	Pre.	TNR	Spe.	F.
10	normal	99.823	0.177	21.951	97.5	1	1	1	1	1
	noise	99.818	0.181	24.943	95.5	1	1	1	1	1
	Filter1b	99.818	0.181	17.078	98.5	1	0.600	0.993	0.993	0.750
	Filter2	99.819	0.180	24.761	96.8	1	1	1	1	1
20	normal	99.910	0.089	25.518	97.9	1	1	1	1	1
	noise	99.909	0.090	21.793	95.7	0.900	0.692	0.995	0.995	0.782
	Filter1b	99.818	0.181	20.504	98.6	1	1	1	1	1
	Filter2	99.910	0.089	29.145	96.7	1	1	1	1	1

يهدف الاختبار تجربة الموديل باستخدام عينات ل 50 شخصاً ورؤية تأثير الطرق المقترحة على هذا العدد من العينات. كان من الملاحظ ان الدقة باستخدام المقاييس المذكرة جميعاً بقيت محافظة على الوتيرة نفسها. ان اختبار الموديل بقيم مختلفة لأشخاص مختلفين ولعدد عينات مختلف في كل مرة يؤكد مدى استقرارية الطريقة المقترحة. وذلك من خلال انتظام النتائج في اغلب الاحيان.

بملاحظة الجدول رقم (1) الذي يظهر مستوى الدقة والوقت المستغرق للمجموعة الاولى تبين لنا بأن طريقة Filter1b اي طريقة الـ Band pass filter هي اكبر دقة تنصنف وهي 98.6 وحصلت الطريقة التي تضمنت اضافة (additive white Gaussian noise) فيها على اسوأ دقة إذ بلغت نسبة الدقة في هذه الطريقة 95.5.

اما بالنسبة للوقت فإن افضل طريقة حققت الوقت الامثل خلال تنفيذ هذه التجربة هي ايضا طريقة Filter1b اي طريقة الـ Band pass filterإذ سجلت هذه الطريقة وقت تنفيذ اقل من الطرق الاخرى حيث بلغ وقت التنفيذ 17.078 دقيقة وذلك مقارب لنتائج الوقت الذي استغرته هذه الخورازمية في التجربتين السابقتين. اما بالنسبة لبقية الطرق من ناحية الوقت كان ادائها متقارب نسبيا. ونلحظ ان بقية المقايس وهي(precision, recall, F_score, TNR) قد بلغت اعلى قيمة لها هو(1)وفي بعض الطرق يقترب من (1) وهذا يدل على جودة ودقة الموديل .

3.3. تجربة رقم (2)

تم اخذ 100 شخص في هذا الاختبار وتم استخدام الـ normal, Additive white Gaussian noise , Median Filter, Band Pass Filter وتمثل هذه الاربع طرق الحالات التي تم معالجة البيانات المدخلة من خلالها. ان الجدول (3) يوضح فكرة استخدام مقاييس الدقة عليها. وهذه المقاييس تتضمن sensitivity , accuracy, precision, recall, time, TNR and other ان الجدول وضح التغاير بالقيم بعد استخدام هذه المقاييس. من الجدير بالذكر ان هناك تشابها نسبيا في الاختبارات جميعاً. يلاحظ ان هناك تاثيرا لعدد العينات فكلما كان عدد العينات أكبر كانت النتيجة افضل. وان الدقة بشكل عام باستخدام الخلايا العصبية للتميز هي مقياس جيد لعكس الصورة الحقيقة لتنفيذ الموديل المقترح من خلال التمحص بالجدول يمكن القول ان الدقة بشكل عام مع العينات الثلاثة بلغ اكثر من 98 %. وبملاحظة الجدول الذي يظهر مستوى الدقة والوقت المستغرق للمجموعة الاولى من التجارب تبين لنا بأن طريقة Filter1b اي طريقة ال Band pass filter حققت اعظم دقة تنصنف وهي 98.1 وحصلت الطريقة التي تضمنت اضافة noise فيها على اسوأ دقة التي بلغت 95.5. وذلك بديهي من الناحية العملية حيث ان ال noiseتعمل على تشويش البيانات مما ينعكس سلبا على عمل المصنف اما بالنسبة للطريقتين الاخريتين فان عملهما يكاد يكون متماثل من ناحية الدقة.

عندما نلحظ وقت التنفيذ نلحظ ايضا ان افضل طريقة حققت الوقت الاقل هي ايضا طريقة Filter1b اي Band pass filter حيث سجلت هذه الطريقة وقت تنفيذ اقل من الطرق الاخرى وبلغ وقت التنفيذ 17.082 دقيقة. اما بالنسبة لبقية الطرق من ناحية الوقت كان ادائها متقاربا نسبيا.

ونلحظ ان بقية المقاييس وهي (precision, recall, F_score, TNR) قد بلغت اعلى قيمة لها هو (1) وفي بعض الطرق يقترب من (1) وهذا يدل على جودة ودقة الموديل.

جدول (3) لعينة 100 شخص

simple	Dataset	Per._mo.	Per._st.	Time	Acc.	Recall	Pre.	TNR	Spe.	F.
10	normal	99.911	0.088	27.997	97.2	1	0.357	0.982	0.982	0.526
	Filter1b	99.909	0.090	17.082	98	1	1	1	1	1
	Filter2	99.911	0.088	24.954	96.5	1	0.722	0.995	0.995	0.838
20	normal	99.954	0.045	29.997	97.4	1	1	1	1	1
	noise	99.954	0.045	30.297	95.7	0.769	1	1	1	0.896
	Filter1b	99.954	0.045	21.601	98.1	0.800	0.761	0.997	0.997	0.780
	Filter2	99.955	0.045	27.836	96.4	1	0.880	0.998	0.998	0.936

4. الاستنتاجات

بعد ظهور نتائج عملية لعينة البحث ظهرت لنا عدة من الاستنتاجات أهمها:

- قدرة التحويل المويجي على تقديم مميزات صوتية (sound features) تمتلك قدرة عالية على التعرف وتميز الاصوات العائدة لكل شخص ضمن العينة.
- تأثر البصمات الصوتية بضوضاء (white noise)
- قدرة المرشحات المقدمة على تقليل تأثير الضوضاء المصاحبة للبصمات الصوتية.
- أن زيادة عدد الاشخاص المطلوب التميز بينهم في النظام يؤدي الى تقليل دقة التمييز بين البصمات الصوتية وهذا طبيعي بسبب زيادة التشابه بين هذه البصمات.
- يمكن اقتراح نماذج ضوضاء (الملح والفلفل، ضوضاء كاوس) لملاحظة تأثر دقة التمييز عند اختلاف نموذج الضوضاء المقدم.
- اقتراح مرشحات أخرى (مرشح الوسط الحسابي، مرشح لابلاس) كمرشحات مكانية (special filter) لملاحظة الاختلاف في فقدرة هذه المرشحات على أزالة تأثير نماذج الضوضاء المختلفة.
- يمكن اعتماد تحويلات أخرى (تحويل Fourier , تحويل لابلاس) لملاحظة قدرة هذه التحويلات على التمييز بين البصمات الصوتية.

References

[1] Das, Soubhik "A Machine Learning Model for Detecting Respiratory Problems using Voice Recognition " 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) IEEE, 2019.

[2] Song, Zhaojuan "English speech recognition based on deep learning with multiple features " Computing 102 3 (2020): 663-682.

[3] Ali, Ashraf Tahseen, Hasanen S. Abdullah, and Mohammad N. Fadhil. "Voice recognition system using machine learning techniques." Materials Today: Proceedings (2021).

[4] Guido, Rodrigo Capobianco, et al. "CWT× DWT× DTWT× SDTWT: Clarifying terminologies and roles of different types of wavelet transforms." International Journal of Wavelets, Multiresolution and Information Processing 18.06 (2020): 2030001.

[5] Barburiceanu, Stefania, Romulus Terebes, and Serban Meza "3D texture feature extraction and classification using GLCM and LBP-based descriptors " Applied Sciences 11 5 (2021): 2332.

[6] Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society 2017: 1–14.

[7] Turow J (2021) The Voice Catchers: How Marketers Listen In to Exploit Your Feelings, Your Privacy, and Your Wallet. New Haven: Yale University Press.

[8] D. a. a. tawisi (2017),"voiceprint its features and use" , Arab journal for security studies and training.

[9] Alegre F, Soldi G and Evans N (2014) Evasion and obfuscation in automatic speaker verification. In: ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings, Florence, Italy, 4–9 May 2014, pp.749–753.

[10] Amoore L (2020) Cloud Ethics: Algorithms and the Attributes of Ourselves and Others. Durham: Duke University Press. Andrejevic M (2012) Exploitation in the data mine. In: Fuchs C, Boersma K, Albrechtslund A and Sandoval M (eds) Internet Jansen et al. 11 and Surveillance. New York and London: Routledge, pp.91– 108.

[11] Valentino-DeVries J (2020) How the police use facial recognition, and where it falls short. New York Times, 12 January

[12] Ponraj, Abraham Sudharson. "Speech Recognition with Gender Identification and Speaker Diarization." 2020 IEEE International Conference for Innovation in Technology (INOCON). IEEE, 2020.

[13] Botchkarev, Alexei. "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology." arXiv preprint arXiv:1809.03006 (2018).

[14] Mustafa, Wan Azani, et al. "Image enhancement based on discrete cosine transforms (DCT) and discrete wavelet transform (DWT): A review." IOP Conference Series: Materials Science and Engineering. Vol. 557. No. 1. IOP Publishing, 2019.

[15] Kofman A (2018) Interpol rolls out international voice identification database using samples from 192 law enforcement agencies. The Intercept, 25 June. Leese M (2020) Fixing state vision: Interoperability, biometrics, and identity management in the EU. Geopolitics: 1–21.

[16] Alegre F, Soldi G and Evans N (2014) Evasion and obfuscation in automatic speaker verification. In: ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings, Florence, Italy, 4–9 May 2014, pp.749–753

[17] Karen wan,(2009)," the use of multimedia in education : design – production – evaluation" Dar al-Shuaa for publishing and science halb, Syrian.

[18] Z. R. Tawfiq, "Voice Based Authentication Using Artificial Neural Network ", Software Engineering science, Iraqi Commission for Computers and Informatics, IRAQ, 2012.