مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (1) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (1) 2018

Studying and Comparing the Effectiveness of Audio Features Extraction Algorithms for Recognize Openness Personality

Dr. Jaffar AlKHIER•
Raneem KNAJ••

(Received 29 / 1 / 2018. Accepted 11 / 4 / 2018)

□ ABSTRACT □

Audio signal includes many information about speaker as social background as dialect, gender, age, and mental state of short and long term mental state, which represents the focus of research.

Research aims to design a system to highlight the speaker's personal style, depending on attributes governing long term psychiatric included in audio signal, consisting of a database containing the audio files to a variety of speakers in terms of personal style according to BFI test – 10 and five factor theory, where the audio database Completed recordings included going back to 75 male and 75 female university students, all of the convergent reconstruction and do not suffer from flaws or receive audio.

The research analyzed the voice features speakers recordings using feature extraction algorithms, LPC, MFCC, LSP and found that people loose construction lists be MFCC transaction values in the vehemence of high voice, while the LPC parameters, the LSP did not show distinct signs govern any personality style.

Key words: Personality recognition, Speech database, extracting features, five-Factor theory, BFI– 10.

•Assistant DeProfessor, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria.
••PHd student, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria.

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (1) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (1) 2018

دراسة ومقارنة فعالية خوارزميات استخلاص السمات الصوتية في تمييز نمط الشخصية المنفتحة للمتحدث

د. جعفر الخير•
رنيم كناج••

(تاريخ الإيداع 29 / 1 / 2018. قُبِل للنشر في 11/ 4 / 2018)

□ ملخّص □

تعتبر الإشارة الصوتية حاملة للعديد من المعلومات عن متحدثها كالجنس والعمر والخلفية الاجتماعية واللهجة والحالة النفسية قصيرة الأمد والحالة النفسية طويلة الأمد والتي تمثل محور البحث.

يهدف البحث إلى تصميم نظام لتمييز نمط شخصية المتحدث اعتماداً على السمات الصوتية الناظمة للحالات النفسية طويلة الأمد المُضمنة في إشارته الصوتية، ويعتبر الركن الأول في هذا النظام هو بناء قاعدة بيانات صوتية لمتحدثين سوريين متنوعين من حيث نمط الشخصية وفقاً لنظرية الأبعاد الخمسة وللاختبار BFI– 10، حيث ضمت قاعدة البيانات الصوتية المُنجزة تسجيلات صوتية عائدة إلى 150 متحدث منهم 75 ذكر و75 أنثى، وجميعهم طلاب جامعيون من أعمار متقاربة ولا يعانون من عيوب نطقية أو سمعية.

قام البحث بتحليل السمات الصوتية لتسجيلات المتحدثين باستخدام خوارزميات استخلاص السمات الأشهر وهي LPC, LSP, MFCC ووجد أن الأشخاص المنفتحين تكون قيم معاملات MFCC في إشاراتهم الصوتية مرتفعة، في حين أن معاملات LPC, LSP لم تظهر علامات مميزة ناظمة لأي نمط شخصية.

الكلمات المفتاحية: التعرف على الشخصية، قاعدة بيانات صوتية، استخراج السمات، نظرية الأبعاد الخمسة، استبيان BFI– 10.

• أستاذ مساعد، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية.
•• طالبة دراسات عليا (دكتوراه) ، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية.

مقدمة:

الكلام هو وسيلة الاتصال بين أفراد الجنس البشري، وهو هبة الله للبشر ميزهم بها عن سائر المخلوقات الأخرى، والكلام هو استحداث موجات صوتية بواسطة الحركة الإرادية للتركيب التشريحي في نظام توليد الكلام لدى الإنسان لتنقل المعلومات من المتكلم إلى السامع. هذا التركيب يتألف من قنوات عديدة لتشكل بمجموعها جهاز النطق وأعضاءه وشكلها مختلف من شخص إلى آخر.

يظهر الصوت نتيجة حدوث تموجات في الهواء ولا ينقل إلاَّ خلال وسط مادي سواء كان هواء أو مادة صلبة. والأصوات الكلامية تنتج في جهاز الصوت في الإنسان من خلال الهواء الخارج من الرئتين، فالرئتان تعملان كمنفاخ للهواء الذي عند مروره عبر الحنجرة والحلق والأنف والفم يحدث الصوت، ويتعرض الهواء لتضييق وتعديل في مساره فتنتج أصوات الحروف المختلفة. والصوت هو الوسيلة الطبيعية التي تنقل من خلالها اللغات الطبيعية، فعلى مستوى الصوتيات تبرز عملية تمييز الكلام وتوليده آلياً، وتمثل هذه العملية أحد مواضع الالتقاء بين اللغة وهندسة الإشارات. إذ أن الصوت يأتي على هيئة طيف من الموجات الكهرومغناطيسية، يستطيع الحاسب أن يستخلص منها وبطرق هندسية ورياضية السمات الأساسية للأصوات والتي على أساسها يتم تمييز الكلام المنطوق، ومن أمثلة هذه السمات: سعة الموجة ودرجة شدتها [1].

يعتبر الصوت أحد الظواهر الهامة التي يستعملها الإنسان والحيوان للتخطيط والتفاهم عن طريق حاسة السمع (الأذن) التي يتم بواسطتها تحويل الصوت من موجات صوتية إلى إشارات كهربائية عن طريق الأذن والمخ والتي تتحول إلى معلومات مفهومة وتشمل هذه الظواهر جميع الأصوات على اختلاف مصادرها ووسائلها. فمثلاً سماع الأصوات من الآلات الموسيقية وتعدد وسائل الاتصالات المسموعة التي تعتمد على تحويل الطاقة من شكل إلى أخر وتطور الأجهزة الصوتية التي تأخذ أشكالاً متعددة في تطبيقاتها الحديثة في مجالات الطب والزراعة تجعل العلماء والمهتمين بهذا المجال يكثفون الجهد لفهم الظواهر الموجية من حيث مصادرها وكيفية حدوثها وطرق انتشارها والعوامل التي تتحكم فيها ومدى الاستفادة منها[2].

أهمية البحث وأهدافه:

يعتبر تحديد السمات الصوتية الأنسب لكل نمط شخصية واختيار المصنف الأفضل أمر مهم لناحية تصميم أنظمة صوتية ناظمة لأنواع الشخصية، تستخدم بفعالية وموثوقية في اختيار المتقدمين الأنسب والأكثر ملائمة للتعيين في الوظائف الحساسة التي تتطلب شخصية مثلى للتعامل مع النظام بالطريقة الأفضل، كما هو في أنظمة التفاوض عن بعد، وزيادة قدرة الأطباء النفسيين على التنبؤ بالمعلومات السريرية للمتحدثين ووضع استراتيجيات علاج فعّالة.

طرائق البحث وموارده:

قام البحث ببناء قاعدة بيانات صوتية تضم ملفات صوتية لأشخاص سوريين مختلفين من حيث نمط الشخصية، قاعدة البيانات الصوتية المُنجزَة Syrian Personality Speech Database (SPSD)) تمَّ اعتمادها بموجب ورقة البحث: (Creating a High-Quality Syrian Audio Database for Analysis of Speaker Personality)، المنشورة في مجلة International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 4, Jul – Aug 2017 برقم ISSN: 2347-8578، كما تمَّ

الحصول على قاعدة بيانات صوتية معيارية عالمية من جامعة (Glasgow) وهي القاعدة (SSPNet-Speaker-Personality-Corpus) عن طريق التواصل مع المشرف على المشروع (Vinciarelli Alessandro) من خلال بريده الإلكتروني (vincia@dcs.gla.ac.uk).

تحليل الشخصية من منظور علم النفس:

علم نفس الشخصية هو فرع من فروع علم النفس الذي يدرس الشخصية والفروق الفردية. ونظرية السمات في علم النفس هي نهج لدراسة شخصية الإنسان، وهي من المقاربات الأساسية لدراسة الشخصية حيث يهتم أصحاب هذه النظرية في المقام الأول بقياس السمات والتي يمكن تعريفها بأنها الأنماط المعتادة للسلوك والتفكير، والعاطفة، والسمات مستقرة نسبياً مع مرور الوقت، تختلف بين الأفراد وتؤثر على سلوكهم فمثلاً نجد أن بعض الناس اجتماعيين في حين أن البعض الآخر خجولين، فلكل شخصية سماتها الرئيسية، والتي تحدد خصائص هذه الشخصية ونقاط ضعفها وقوتها وأيضاً مدى مرونتها وقدرتها على التوافق مع الآخرين[3].

ينظر علم النفس إلى الشخصية على أنها تكامل مجموعة من الأبعاد أو العوامل، ولكل بُعد مجموعة من السمات. لكل نظرية عدة نماذج أو مقاييس أو استبيان أو اختبارات، وبالتالي يكون التعريف الإجرائي لمكونات الشخصية هو مجموع الدرجات التي يحصل عليها الفرد على كل عامل على حدة في استبيانات الشخصية، وفي الجدول (1) أبرز وأحدث النظريات المعتمدة في علم النفس:

الجدول (1) أحدث النظريات المعتمدة في علم النفس.

اسم النظرية	عدد العوامل	الاستبيان المعتمد	تعريب الاستبيان
نظرية كاتل Cattel	16	Clinical Analysis Questionnaire(CAQ)	محمد السيد وصالح أبو عبادة
نظرية آيزنك Eysanck	4	Eysanck Personality Questionnaire(PEQ)	أحمد عبد الخالق
نظرية جولدبيرج Goldberg	5	NEO-PI-R, 240 items NEO-FFI, 60 items BFI, 44 items BFI, 10 items	محمد أبو هاشم، بدر الأنصاري

1- نظرية العوامل الخمسة الكبرى (Big Five Factors Model):

وهي النظرية المعتمدة لتصنيف أنماط شخصية متحدثي البحث، يعد نموذج العوامل الخمسة الكبرى للشخصية من أهم النماذج وأحدثها التي فسرت سمات الشخصية، وهذا النموذج الهرمي يتكون من خمسة عوامل رئيسية هي [4]:

(1) المقبولية: ((A) Agreeableness) يعكس هذا العامل كيفية التفاعل مع الآخرين، فالدرجة المرتفعة تدل على أن الأفراد يكونون أهل ثقة ويتميزون بالود والتعاون والإيثار والتعاطف والتواضع، ويحترمون مشاعر وعادات الآخرون.

(2) الضمير الحي ((C) Conscientiousness): يعكس هذا العامل المثابرة والتنظيم لتحقيق الأهداف المرجوة، فالدرجة المرتفعة تدل على أن الفرد منظم ويؤدى واجباته باستمرار وبإخلاص.

(3) الانبساطية ((Extraversion(E)): يعكس هذا العامل التفضيل للمواقف الاجتماعية والتعامل معها، فالدرجة المرتفعة تدل على أن الأفراد مرتفعي الانبساطية يكونون نشطين ويبحثون عن الجماعة.

(4) العصابية ((Neuroticism(N)): يعكس هذا العامل الميل إلى الأفكار والمشاعر السلبية أو الحزينة، فالدرجة المرتفعة تدل على أن الأفراد يتميزون بالعصابية فهم أكثر عرضة لعدم الأمان، والأحزان.

(5) الانفتاح على الخبرة ((Openness to Experience (O)): يعكس هذا العامل النضج العقلي والاهتمام بالثقافة، والدرجة المرتفعة تدل على أن الأفراد خياليون، ابتكاريون، يبحثون عن المعلومات بأنفسهم.

2- استبيان العوامل الخمسة الكبرى ((Big Five Inventory (BFI-10):

والجداول التالية توضح الاستبيان بنسختيه العربية والأجنبية، حيث يتكون من عبارتين لكل من السمات الخمسة ليصل عدد العبارات إلى عشر عبارات[5,6]:

English version.
Instruction: How well do the following statements describe your personality?

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
... is reserved	(1)	(2)	(3)	(4)	(5)
... is generally trusting	(1)	(2)	(3)	(4)	(5)
... tends to be lazy	(1)	(2)	(3)	(4)	(5)
... is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
... has few artistic interests	(1)	(2)	(3)	(4)	(5)
... is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
... tends to find fault with others	(1)	(2)	(3)	(4)	(5)
... does a thorough job	(1)	(2)	(3)	(4)	(5)
... gets nervous easily	(1)	(2)	(3)	(4)	(5)
... has an active imagination	(1)	(2)	(3)	(4)	(5)

Scoring the BFI-10 scales:
Extraversion: 1R, 6; Agreeableness: 2, 7R; Conscientiousness: 3R, 8; Neuroticism: 4R, 9; Openness: 5R; 10 (R = item is reversed-scored).

تعريب المقياس على يد بدر الأنصاري أنتج النسخة التالية:

أرى نفسي كشخص...	أخالف بشدة (1)	أخالف قليلاً (2)	لا أخالف ولا أوافق (3)	أوافق قليلاً (4)	أوافق بشدة (5)
كتوم ومتحفظ					
يتمتع بثقة عامة					
يميل إلى أن يكون كسول					
مسترخي، ويعالج الضغوطات التي تتعرض له بشكل فعال					
لديه اهتمامات فنية					
اجتماعي وودّي					

يميل إلى إيجاد أخطاء الآخرين					
يقوم بأعمال فعالة وشاملة					
سريع العصبية					
لديه خيال نشط					

2- قواعد البيانات الصوتية:

يعتبر وضوح الكلام المنطوق بواسطة أي جهاز آلي للنطق مقياساً لكفاءته، وعلى الرغم من أن معظم الأجهزة المستخدمة حالياً لقراءة النصوص قد تنتج كلاماً مفهوماً، لكنه آلي وبعيد عن خصائص الكلام الطبيعي. لذا برزت الحاجة الملحة إلى توفر مولد آلي للنبرة عالي الجودة في نظم توليد الكلام، وهناك العديد من الأبحاث والمشاريع الحديثة في هذا المجال، والتي في مجملها تعتمد على التحليل اللفظي والنبري للمتن الصوتي المسجل والمصنف، ونظراً لتواضع الأعمال المُنجزة في هذا المجال بالنسبة للغة العربية، قدم هذا البحث المراحل المتبعة لإنجاز قاعدة معلومات صوتية نبرية لمتحدثين سوريين كتلك المتوفرة للغات أخرى.

تشكل قواعد بيانات الكلام (Speech Database) الركن الأساسي لبناء نظم حاسوبية مختلفة كنظم التعرف الآلي على الكلام والنطق الآلي والتعرف على المتحدث والتعرف على اللغات واللهجات، وتتكون قواعد بيانات الكلام عادة من ملفات صوتية (Wave Files) سبق أن سجلت لكلام متكلمين باللغة المطلوبة، وكلما كانت قاعدة بيانات الكلام غنية وثرية في محتواها كلما ساعد ذلك على إخراج نظم حاسوبية ذات أداء متميز[7].

تتنوع قواعد البيانات الصوتية بتنوع الهدف المرجو من بنائها، حيث تحوي عادةً ملفات صوتية تمثل عينات من الحالات المراد دراستها، وبما أن هدف البحث هو إيجاد آلية لتمييز أنماط الشخصية فقد تمَّ تضمين قاعدة البيانات الصوتية تسجيلات صوتية متنوعة لمتحدثين متنوعي نمط الشخصية، ومن ثمَّ تحليل التسجيلات المختلفة.

2-1- قواعد البيانات الصوتية المتوفرة لسمات الشخصية:

الجدول(2) يبين قواعد البيانات الصوتية المتوفرة لسمات الشخصية [8].

اسم قاعدة البيانات	اللغة / اللهجة	عدد المتحدثين	ملفات الإشارة	المصدر
aetobi	American English	17	Audio	Beckman et al (2005); Pitrelli et al (1994); Silverman et al (1992)
ae	Australian English	7	Audio, spectra, formants	Millar et al (1997); Millar et al (1994)
andosl	Australian English	200	Audio, formants	Millar et al (1997); Millar et al (1994)
ema5 (ema)	Standard German	20	Audio, EMA	Bombien et al (2007)
epgassim	Australian English	60	Audio, EPG	Stephenson & Harrington (2002); Stephenson (2003)

epgcoutts	Australian English	2	Audio, EPG	Passage from Hewlett & Shockey (1992)
epgdorsal	German	45	Audio, EPG, formants	Ambrazaitis & John (2004)
epgpolish	Polish	40	Audio, EPG	Guzik & Harrington (2007)
gt	German	9	Audio, f0	Utterances from various sources

2-2- قاعدة البيانات الصوتية العالمية المُعتمدة:

تحتاج أي منهجية بحث مُقترحة إلى مقارنة نتائجها مع نتائج منهجيات بحث عالمية معيارية من أجل درجة مصداقية عالية، كذلك الأمر بالنسبة لهذا البحث فالخوارزميات المقترحة في متنه سيتم تطبيقها على قاعدة بيانات عالمية معيارية معتمدة ومقارنة النتائج مع نتائج خوارزميات أخرى مطبقة على نفس قاعدة البيانات العالمية وهي القاعدة (SSPNet Speaker Personality Corpus) ، التي تحوي 640 ملف صوتي تعود لـ 322 متحدث، مدة كل ملف حوالي 10 ثانية، صيغة كل ملف من نوع (wav) ذو مواصفات (bit, mono, 8kHz 16) وقد تمَّ الحصول عليها من خلال التواصل مع المشرف على إنجازها في جامعة (Glasgow) وهو الدكتور (Alessandro Vinciarelli) من خلال بريده الإلكتروني (vincia@dcs.gla.ac.uk).

2-3- قاعدة البيانات الصوتية السورية المُنجَزة:

قام البحث ببناء قاعدة بيانات صوتية تضم ملفات صوتية لأشخاص مختلفين من حيث نمط الشخصية، قاعدة البيانات الصوتية المُنجزَة ((Syrian Personality Speech Database (SPSD تمَّ اعتمادها بموجب ورقة البحث: (Creating a High-Quality Syrian Audio Database for Analysis of Speaker Personality)، المنشورة في مجلة International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 4, Jul – Aug 2017 برقم ISSN: 2347-8578، وقد أُعتمدت المعايير العالمية في بناء قاعدة البيانات (SPSD) كمايلي:

- **<u>المتحدثون:</u>**

عند أخذ عينة لتمثيل مجموعة معينة فإنه من المهم أن يكون التمثيل دقيقاً إلى أقرب درجة ممكنة، وفي حالة هذا البحث فإنَّ أفضل تمثيل هو اختيار عينات تمثل مختلف أنماط الشخصية. إنَّ متحدثي هذا النظام هم طلاب جامعة طرطوس، حيث قام البحث باستهداف هذه العينات لتسجيل أصواتهم لعدة أسباب منها: كون طلاب الجامعة يمثلون خلفيات اجتماعية متباينة، فهم ينحدرون من محافظات سورية متنوعة تمثل لهجات متنوعة مما يغني قاعدة البيانات الصوتية، ولا يعانون من عيوب نطقية أو سمعية، فضلاً عن كونهم ينتمون إلى شريحة عمرية واحدة فالفروق العمرية بين أصغرهم وأكبرهم لا تتجاوز الخمس سنوات مما ينفي تأثير عامل العمر على دقة النتائج، فضلاً عن سهولة التواصل معهم وجمعهم ونقلهن من وإلى مكان التسجيل كون الباحثة مدرستهم ويوجد تواصل يومي لها معهم من خلال المحاضرات التدريسية. تمَّ انتقاء المتحدثين من الجنسين حيث ضمت قاعدة البيانات الصوتية المُنجزة (SPSD) تسجيلات صوتية عائدة إلى 150 متحدث منهم 75 ذكر و75 أنثى. وقد نُظمت قاعدة بيانات معلوماتية التالية تخص متحدثي النظام، وهي عبارة عن جدول مكون من 150 سطر كل سطر يخص متحدث تمَّ تنظيمهن في الأعمدة التالية:

التسلسل	الاسم	الجنس	العمر	المنشأ

حيث العمود الأول يوثق رقم المتحدث التسلسلي، والعمود الثاني يدون الاسم الثلاثي للمتحدث، والعمودين الثالث والرابع لإسناد جنس المتحدث وعمره، أما العمود الخامس فيحوي المحافظة السورية التي ينتمي لها المتحدث.

- **المادة اللغوية:**

تتنوع المادة اللغوية الملفوظة بين دراسة وأخرى، فأحياناً يتم لفظ الحرف الصوتي بشكل منفصل ولكن بذلك يُهمل تأثير الحرف المجاور للحرف الصوتي على هذا الصوتي، وأحياناً يتم لفظ الحرف الصوتي ضمن مقطع صوتي، وأحياناً كلمة وأحياناً جملة. ونظراً لأنَّ حوالي من 60% إلى 70% من الكلام السوري هو صوتي حيث تكثر الحروف الصوتية وحركات التشكيل النحوية في هذا الكلام، ونظراً لأن الحروف الصوتية تكون أكثر طاقة طيفية في التمثيل الهيستوغرامي (Histogram) وبالتالي تظهر بوضوح في الرسم الطيفي، وبالتالي تعتبر الحروف الصوتية خير مُقيّم لخصائص اللهجات الملفوظة[9].

وقد أُعتمِدَت الجملة الثرية صوتياً التالية كمادة لغوية لهذا البحث: "أهلاً بكم في سورية، بلد الخير والجمال والمحبة، نتمنى لكم قضاء وقت ممتع في ربوعنا." حيث لفظها كل متحدث ثلاث مرات بثلاثة تكرارات عشوائية.

- **بيئة العمل:**

تمَّ في البداية تنزيل إعلانات على صفحات طلاب الجامعة لمن يرغب بالتطوع لتسجيل صوته، وبدأ المتطوعين بالاستجابة، وتمَّ التنسيق والاتفاق معهم على كيفية ومواعيد التسجيل، حيث أنه كان المركز الإذاعي والتلفزيوني بمدينة طرطوس قد خصص استديو لخدمة هذا البحث أيام السبت 2017/9/16 والاثنين 2017/9/18 والسبت 2017/9/23 والاثنين 2017/9/25 لمدة ساعة ونصف من كل يوم من الساعة التاسعة والنصف صباحاً إلى الساعة الحادية عشرة صباحاً، وبالتالي يكون الوقت الإجمالي الذي استغرقه البحث في تسجيل قاعدة بياناته هو ست ساعات.

تمّ تقسيم المتحدثين المئة والخمسين إلى أربع مجموعات على أيام التسجيل الأربعة، وقبل أن يحين الموعد المخصص للتسجيل في كل يوم يمضي المتحدثون حوالي ساعة من الوقت في الإجابة على أسئلة استبيان (BFI-10) بنسخته العربية، حيث قامت الباحثة بتوزيع نسخ ورقية من الاستبيان المعرّب على المتحدثين وقاموا بالإجابة على عباراته العشرة، ثمّ حُسبت درجة كل عامل من العوامل الخمس الكبرى لدى كل متحدث بإشراف أخصائيتين في علم النفس مجازتين بدرجة الماجستير في علم النفس، ثمَّ تمت عملية تسجيل الملفات الصوتية بإشراف مهندستي الصوت صبا سليمان وولاء محمد من المركز الإذاعي والتلفزيوني في محافظة طرطوس.

تم الاستعانة بالأدوات التالية أثناء التسجيل في الاستديو بميكروفون ديناميكي نوع Yamaha وميكسر Soundcraft Digital وبرنامج التسجيل Adobe Audition 3 والقناة الصوتية Stereo Channel وكان تردد أخذ العينات 44100 Hz. حيث تمَّ التسجيل داخل استديوهات معزولة صوتياً، جدران الاستديو وأرضيته وسقفه معالجة بمواد ماصة للصوت كما تبطن الفراغات بمثل هذه المواد، كذلك الاستوديو خالياً من الشبابيك التي قد تسمح بدخول الضوضاء، وباب الاستوديو عازل للصوت أيضاً. ويتصل الاستوديو بغرفة المراقبة من خلال العديد من الأجهزة، توجد في الحائط المتصل بغرفة المراقبة نافذة كبيرة من الزجاج، يراقب من خلالها الشخص الذي يشرف على التسجيل ما يقوم به الشخص الذي يقوم بالتسجيل، تمَّ التسجيل لكل شخص على حدا، حيث قرأ الجملة المكتوبة على ورقة A4، بشروط عادية من السرعة وعلو الصوت.

3- معالجة الإشارة:

تتضمن معالجة الإشارة كافة العمليات الرياضية والحسابية التي يتم تطبيقها على إشارة ما بهدف تخفيض الضجيج فيها وزيادة فائدتها في تمييز البيانات التي تحملها. من بين كل العمليات الممكنة سنطبق ما يلي :[10,11,12]

3-1- التصحيح حسب المتوسط Mean correction:

الهدف من تعديل قيم الإشارة تبعاً للمتوسط هو تخفيض أثر أي مركبة مستمرة تنتجها أجهزة التسجيل، حيث نختار عتبة معينة من المتوسط ونطرحها من قيم الإشارة حيث أن قيمة هذه العتبة صغيرة جداً ولا تتجاوز قيمتها 0.05، تعدل هذه العملية المطالات بشكل طفيف ولكنها لا تغير في شكل الإشارة.

3-2- تحليل طيف الإشارة Spectrum analysis:

تكمن نقاط أهمية تحليل طيف الإشارة في ما يلي، أولاً: أن الإشارة في مجال الزمن لا تعطي صورة واضحة عن المركبات المهمة في الإشارة، وثانياً: يتيح لنا التعرف على ماهية الترددات المكونة لهذه الإشارة وحذف كل الترددات التي قد تكوف ذات أثر سلبي على هذه الإشارة، حيث أن تردد إشارة الكلام عند الإنسان يقع في المجال بين 300 Hz و 3700 Hz فأي تردد خارج حدود هاتين العتبتين يمكن اعتباره ضجيج على هذه الإشارة مما يدفعنا لترشيح الإشارة للتخلص منه.

إن تحليل طيف الإشارة هو عملية تطبيق لتحويل فورييه (Fourier Transform) على قطاعات جزئية من الإشارة تسمى إطارات، حيث يعتبر كل إطار إشارة مستقلة ويطبق عليه تحويل فورييه بشكل مستقل عن بقية الإطارات، وغاية هذه العملية هو معرفة الترددات الموجودة في كل إطار زمني، وتسمى هذه العملية بتحويل فورييه قصير الحد Short-Time Fourier Transform (STFT) ويكون ناتج STFT هو رسم ثنائي البعد يمثل الإطارات الزمنية والترددات المقابلة لكل إطار.

3-3- تحسين الإشارة Pre-emphasis:

تنتج البيئة المحيطة أثناء عملية التسجيل ترددات غير مرغوبة، هذه الترددات هي ترددات منخفضة لا تعتبر من ترددات الكلام المنطوق من قبل المتحدث لذلك يجب حذفها من الإشارة.

إن الإشارة الممثلة للكلام تكون فيها الترددات المرتفعة ذات مطالات منخفضة (مثل الأحرف اللا صوتية كالأحرف الانفجارية) والعكس بالعكس (أي الصوتيات (phonemes)) ذات التردد المنخفض تكون بمطالات مرتفعة، وأيضاً الأحرف الصوتية تكوف بمطالات مرتفعة، والهدف من عملية تحسين الإشارة هو رفع مطالات الترددات المرتفعة مقابل مطالات الترددات المنخفضة بحيث نزيد من القدرة على التنبؤ بالصوتيات التي تمثلها هذه الترددات.

4- استخراج السمات الصوتية:

الكلام هو أكثر أشكال التواصل شيوعاً بين البشر وقد حاول الباحثون تطوير نظام يمكنه أن يحلل ويصنف الإشارات الصوتية حيث أن الأنظمة الحاسوبية التي تفهم اللغات المحكية يمكن أن يكون لها استخدامات في الكثير من المجالات.

إن الإشارات الصوتية هي إشارات متغيرة (quasi-stationary) فعندما يتم اختبار الإشارات الصوتية خلال فترة قصيرة (5~100 msec) فإن خصائصها تكون ثابتة، ولكن من أجل زمن طويل فإن السمات الصوتية تصبح

مختلفة وبالتالي فإن عملية استخلاص الخصائص تعد أهم مرحلة من مراحل أنظمة معالجة الكلام. إن استخلاص السمات الكلامية هو متطلب أساسي لأي نظام تعرف على الكلام، وهو تمثيل رياضي لملف الكلام[13].

4-1- خوارزمية ((Mel Frequency Cepstral Coeffiecents (MFCC)):

MFCC من الطرق السائدة والمهيمنة المستخدمة في استخراج السمات وذلك بسبب حساسية مرشحاتها لخواص إشارة الصوت البشرية. تستخدم معاملات MFCC بشكل كبير في التعرف على الكلام، حيث تم تقديم هذه المعاملات من قبل العالمين Davis Mermelste، في عام 1980 ومازالت متقدمة في هذا المجال منذ ذلك الوقت[14].

إن الأصوات التي تولد من قبل الإنسان يتم ترشيحها حسب شكل المسلك الصوتي (vocal tract)، فإذا تمكنا من تحديد شكل المسلك الصوتي بدقة فإنه يمكن تحديد الصوت (phoneme) الذي يتم إنتاجه. شكل المسلك الصوتي يتجلى في غلاف طيف طاقة الزمن القصير (short time power spectrum) وهدف MFCC هو تمثيل هذا الغلاف بدقة والشكل التالي هو المخطط الصندوقي لعمل الخوارزمية.

الشكل (1) المخطط الصندوقي لعمل الخوارزمية MFCC

تعتمد MFCC على التغيرات المعروفة في عرض حزمة الترددات للأذن البشرية، حيث أن لمرشحاتها تباعداً خطياً على الترددات المنخفضة ولوغاريتمياً على الترددات المرتفعة وهي تستخدم من أجل التقاط الصفات الرئيسية للكلام، حيث تمتلك MFCC تباعداً خطياً على الترددات الأقل من 1000 هرتز وتباعداً لوغاريتمياً على تردد أكبر من 1000 هرتز.

- pre-emphasis

يتم تطبيق عملية pre-emphasis (وهي عملياً مرشح تردد عالي high pass filter) على الإشارة وذلك من أجل تعويض جزء التردد العالي الذي تم فقده أثناء آلية إنتاج الكلام (زيادة الطاقة النسبية للطيف عالي التردد)، حيث يتم إعادة تقييم كل قيمة في إشارة الكلام باستخدام الصيغة (1).

$$s2(n) = s(n) - a*s(n-1) \qquad (1)$$

حيث: s(n): إشارة الكلام

s2(n): إشارة الخرج بعد عملية الـ pre-emphasis

a: ثابت تتراوح قيمته بين 0.9 و o.1

- التأطير (framing)

إشارة الكلام هي إشارة متغيرة باستمرار لذلك من أجل تبسيط الدراسة نعتبر أنه من أجل نطاق زمني قصير (short time scale) فإن إشارة الصوت لا تتغير كثيراً لهذا السبب يتم تقطيع الإشارة إلى عدد من الإطارات

(frames)، زمن كل إطار من 20 إلى 40 ميلي ثانية مع وجود تداخل اختياري يساوي إلى نصف أو ثلث حجم الإطار وذلك من أجل تسهيل الانتقال من إطار إلى آخر.

- النوفذة (windowing)

كل إطار(frame) سوف يخضع لعملية النوفذة (windowing) باستخدام نافذة هامينغ (Hamming window)، وذلك من أجل القضاء على الانقطاعات عند الحواف. تعطى نافذة هامينغ بالعلاقة (2):

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right), \quad 0 \le n \le N \qquad (2)$$

حيث w(n) هو مطال العينة الجديد.

n هو ترتيبها في النافذة.

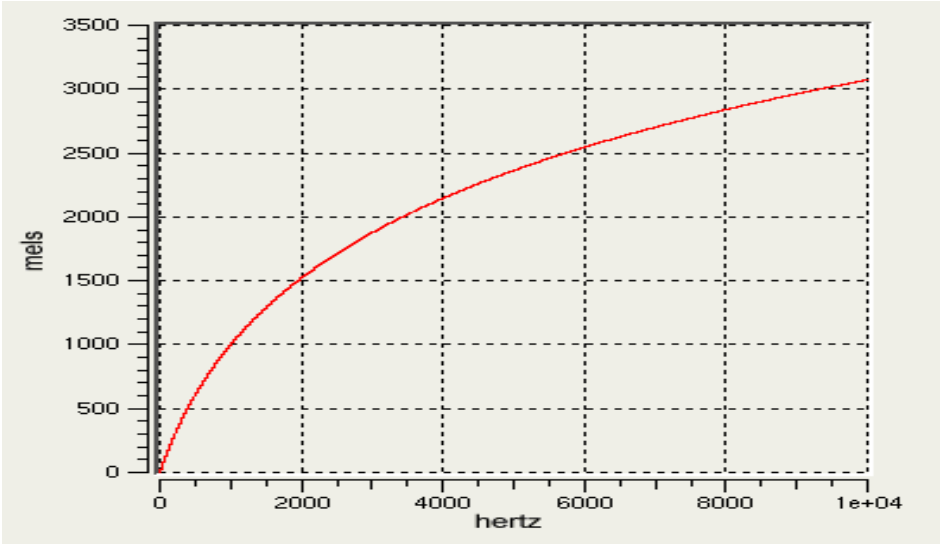
N هو الطول الكلي للنافذة.

بعد عملية النوفذة windowing سوف يتم تطبيق تحويل فورييه السريع FFT من أجل كل إطار وذلك من أجل استخراج مركبات التردد للإشارة في مجال الزمن.

- ترشيح الإشارة وفقاً لتردد ميل (Mel frequency filtering)

تعمل MFCC على ترشيح طيف الإشارة الصوتية (short time power spectrum) عن طريق مجموعة من المرشحات المثلثية (Mel filter bank) (التي صممت كمحاكاة لـ مرشحات تمرير الحزمة band pass filtering التي تظهر في النظام السمعي)، المتباعدة بانتظام وفقا لمقياس ميل الترددي (Mel frequency scale) الذي يعبر عن علاقة تربط التردد الملاحظ لنغمة صافية (m) (perceived frequency of pure tone) بترددها المقاس الأصلي (f) ويعطى بالعلاقة (3):

الإنسان يستطيع أن يميز التغيرات الصغيرة في الـ pitch (وهي الارتفاع أو الانخفاض النسبي لنغمة (tone) كما تدركها الأذن، والتي تعتمد على عدد الاهتزازات التي تنتجها الحبال الصوتية في الثانية) وبشكل أفضل عند الترددات الصغيرة من الترددات الكبيرة، بالتالي فإن تضمين هذا المقياس يجعل سماتنا أقرب إلى سمع الإنسان.

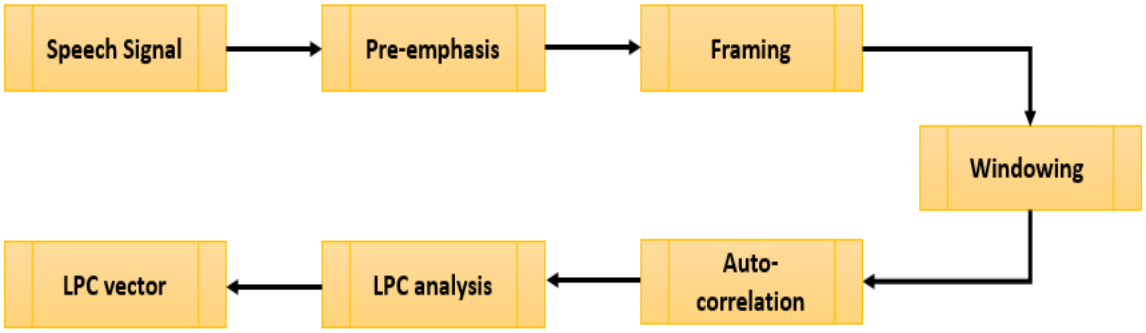


الشكل(2)، نطاق ميل mel scale

يتم بعد ذلك حساب اللوغاريتم لطيف مجال ميل (Mel scale spectrum)، ومن ثم يستخدم تحويل جيب التمام المتقطع DCT لإعادة تحويل طيف مجال ميل اللوغاريتمي إلى مجال الزمن ونتيجة هذا التحويل هو الحصول على شعاع MFCC.

4-2- خوارزمية (Linear Predictive Coding (LPC))[15] :

المخطط الصندوقي لعمل الخوارزمية مبين في الشكل (4)، الفكرة الرئيسية من ترميز التنبؤ الخطي (Linear Predictive Coding) والذي يعتمد على آلية إنتاج الكلام (أي أنه يستخدم نموذج مرشح-مصدر source-filter التقليدي)، هي أن عينة محددة من الكلام في الوقت الحالي يتم تقريبها كمزيج خطي من عينات الكلام السابقة.



الشكل (3) المخطط الصندوقي لعمل الخوارزمية LPC

من خلال تقليص مجموع مربعات الاختلافات (على فترة زمنية محدودة) بين عينات الكلام الفعلية وقيم التوقع (التنبؤ) الخطية، سوف يتم تحديد مجموعة فريدة من البارامترات (معاملات التنبؤ الخطية)، هذه المعاملات تشكل أساساً لتحليلات التنبؤ الخطي للكلام، في الواقع إن عوامل التنبؤ الفعلية لا تستخدم في التعرف على الكلام لأنها نموذجية تظهر التباين العالي، لذلك يتم تحويل معاملات التنبؤ هذه إلى مجموعة أقوى من البارامترات هي (Cepstral Coefficients) بواسطة المعادلات الرياضية التالية (4).

$$c_{\mathrm{LPC}}(m)=\begin{cases}-a(m)-\sum_{i=1}^{m-1}(1-\frac{i}{m})a(i)c_{\mathrm{LPC}}(m-i), & 1\le m\le P\\ -\sum_{i=1}^{m-1}(1-\frac{i}{m})a(i)c_{\mathrm{LPC}}(m-i), & m>P,\end{cases} \qquad (4)$$

حيث: m=1،...،P،a(m) : تعبر عن معاملات التنبؤ الخطي LPC coefficients.

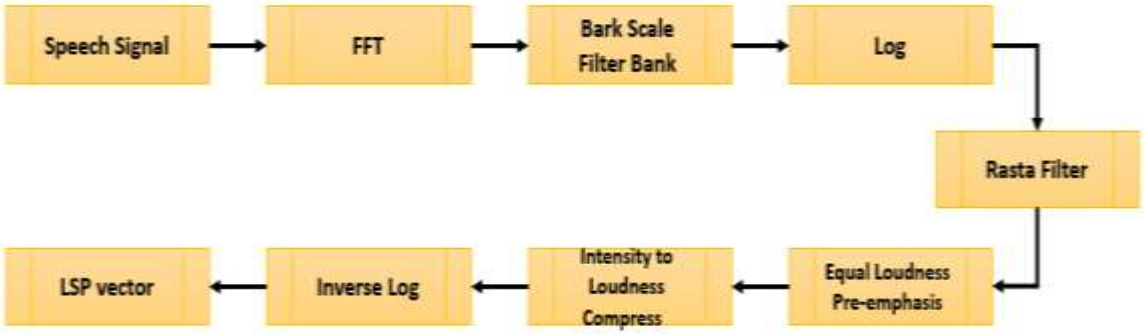
P : يعبر عن ترتيب النموذج model order.

تجري عمليات pre-emphasis، التأطير framing، النوفذة windowing نفسها المطبقة في MFCC.

4-3- خوارزمية (Line spectrum pairs (LSP)):

تعتمد هذه التقنية على psychophysics of hearing (العلاقة بين مؤثر فيزيائي والإداراكات المؤثرة)، تقوم باستبعاد المعلومات التي ليس لها صلة بالكلام بالتالي تحسين عملية التعرف[16]. LSP مماثلة لـ LPC إلا أن خصائصها الطيفية تم تحويلها لتتناسب مع خصائص النظام السمعي عند الإنسان، حيث أن LSP تقارب ثلاث جوانب رئيسية:

- The critical-band resolution curve
- The equal loudness curve
- The intensity-loudness power-law relation



الشكل (4) المخطط الصندوقي لعمل الخوارزمية LSP

بعد معالجة إشارة الكلام يتم حساب تحويل فورييه السريع FFT للحصول على طيف الإشارة (power spectrum)، ومن ثم يتم استخدام مرشحات بشكل شبه منحرف تعتمد على نطاق bark الذي يعطى بالعلاقة:

$$\Omega(w) = 6\ln\left\{w/1200\pi + \left[(w/1200\pi)^2 + 1\right]^{0.5}\right\} \quad (5)$$

من أجل دمج طيف الطاقة (power spectrum) داخل حزم حرجة متداخلة overlapping critical (bands) يتم بعد ذلك الترشيح بواسطة (RASTA filter)، وهي تقنية تعتمد على ترشيح تمرير الحزمة (band pass time filtering) المطبقة على لوغاريتم الطيف الممثل للكلام.

الشكل (5) RASTA filter

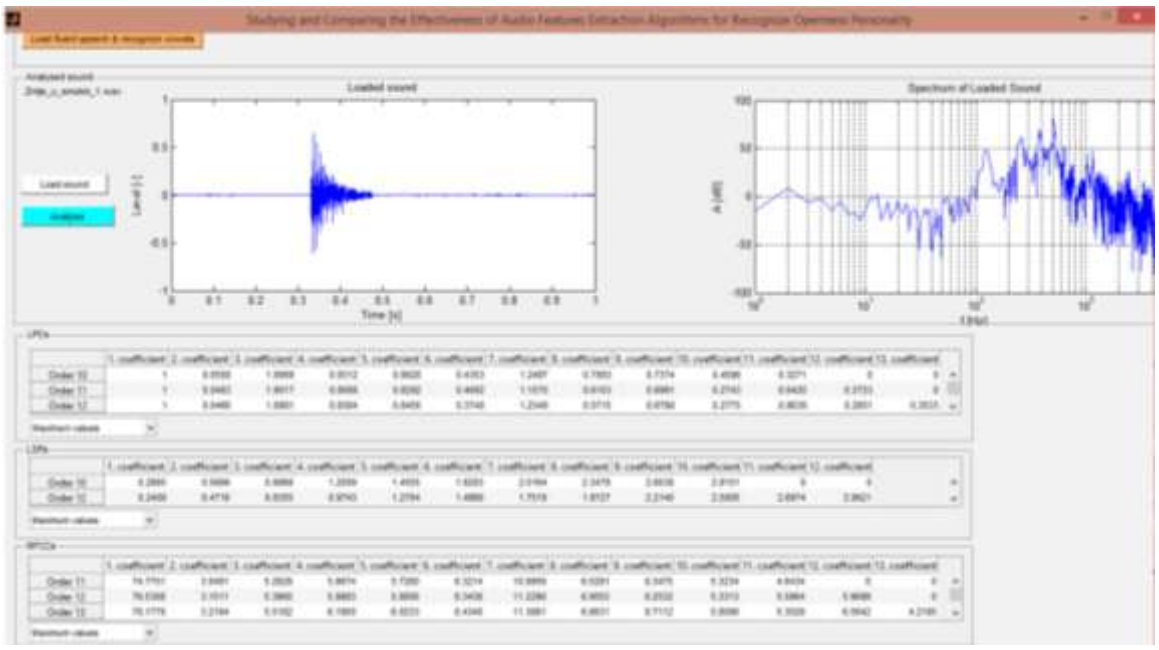
تتم بعدها عملية pre-emphasis لطاقة الطيف (power spectrum) بواسطة منحني قياس ارتفاع الصوت (equal-loudness curve)، والتي تقارب الحساسية غير المتساوية لسمع الإنسان عند ترددات مختلفة، عند حوالي 40 dB، حيث أن كل معامل لطاقة الطيف (power spectrum coefficient) سوف يتم ضربه بالوزن E الذي يعطي بالعلاقة:

$$E(w) = \frac{(w^2 + 56.8\times10^6)w^4}{(w^2 + 6.3\times10^6)^2(w^2 + 0.38\times10^9)} \quad (6)$$

والذي يسبب انخفاضاً في الحساسية في نطاق التردد العالي وهذا ما يجعل ملائمة الطيف لـ التنبؤ الخطي LP أكثر تجانساً على نطاق الترددات العالية. تتم بعد ذلك عملية ضغط لطيف الكلام ويتم تنفيذ هذه العملية وفق (power–law of hearing) والذي ينمذج العلاقة غير الخطية بين الكثافة (intensity) والشدة (loudness) الملاحظة لها، حيث يتم في هذه المرحلة تقليل التغيرات الديناميكية وتسطيح قمم الطيف حيث أن الطيف الناتج مسطح أكثر (smoother) مع قمم (peaks) أقل وضوحاً، يتم بعد ذلك حساب المعاملات التنبؤية، وتحويلها إلى معاملات (cepstral) ومن ثم الحصول على شعاع LSP.

النتائج والمناقشة:

تضمن البحث عدة مراحل، فتمَ في المرحلة الأولى تحديد نظرية علم نفس الشخصية التي سيتمّ الاستناد إليها لتمييز المتحدثين واختيار الاستبيان المناسب والحصول على نسخته المعرّبة، ثم بناء قاعدة بيانات لأصوات أشخاص سوريين تكون مناسبة لتصميم المشروع، ومن ثم بناء قاعدة بيانات تتضمن معلومات شخصية عن أشخاص النظام، حيث تمَ اختيار متحدثين مختلفي أنماط الشخصية، فأُخذت العينات من كافة أنماط الشخصية لإغناء قاعدة البيانات الصوتية، وتمَّ في المرحلة التالية إجراء تحاليل MFCC، LSP،LPC على التسجيلات من أجل استخراج السمات، ,ولهذه الغاية تمّ تصميم واجهة مستخدم رسومية GUI ببرنامج الماتلاب لإظهار قيم معاملات الخوارزميات الثلاث MFCC، LSP،LPC من أجل رتب (معاملات) متنوعة، والشكل التالي يبين واجهة البرمجية التي تمَّ تصميمها:

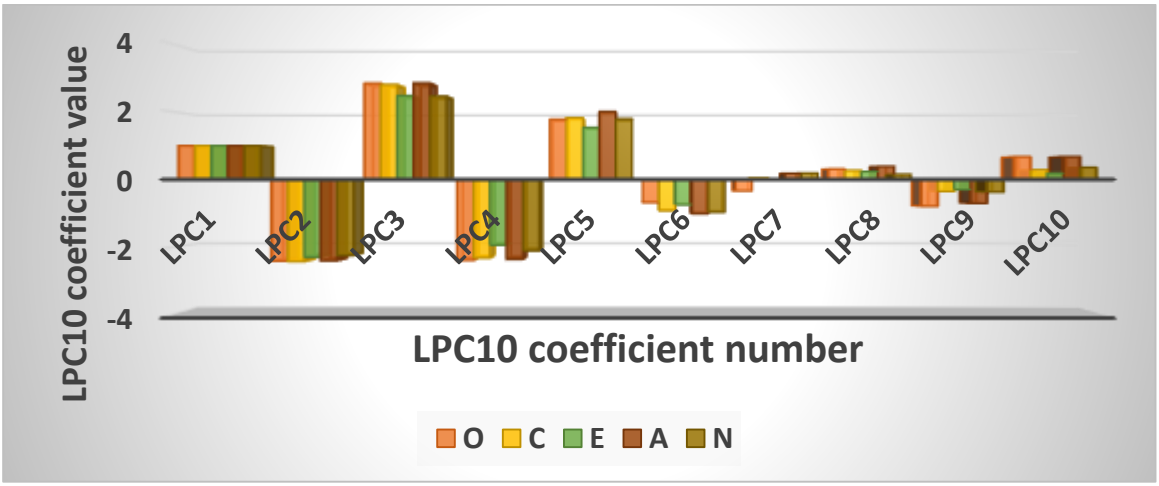


الشكل (6) واجهة الـ GUI الخاصة بالبحث والحاملة لاسمه

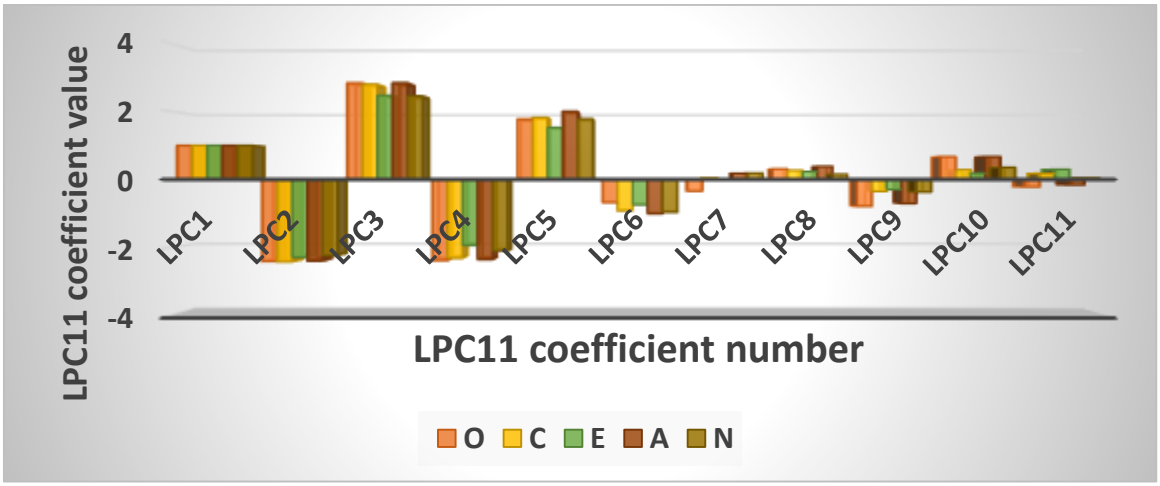
نلاحظ أنَّ الواجهة في الشكل السابق تحمل في شريط عنوانها عنوان المقال، وتحوي على اليسار زر (load sound) بمجرد النقر عليه يمكننا استعراض الملف الصوتي من مساره وتحميله إلى البرمجية، ويظهر مباشرةً

رسمه في المجال الزمني على اليسار تحت العنوان (loaded sound) ورسمه الطيفي على اليمين تحت العنوان (Spectrum of loaded sound)، ونلاحظ تحت الزر (load sound) يوجد زر (Analysis) عند النقر عليه تظهر قيم معاملات الخوارزميات الثلاث MFCC، LSP،LPC، حيث تمت البرمجة لإظهار 13 معامل للخوارزميتين MFCC، LPC و12 معامل للخوارزمية LSP برتب 12،10 بينما تمّ استخدام الرتب 10، 11، 12 لأجل الخوارزمية LPC والرتب 11، 12، 13 للخوارزمية MFCC.

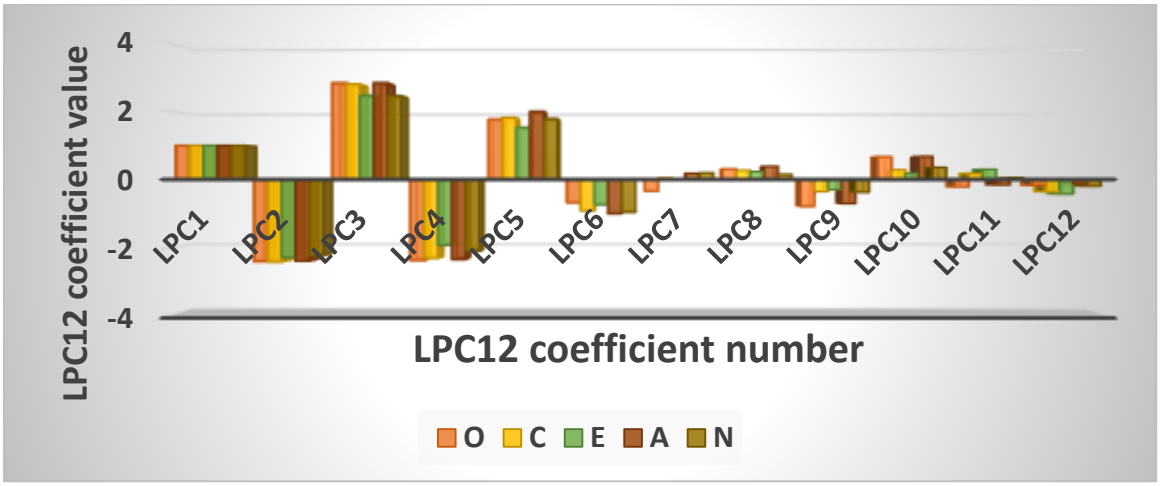
وبنتيجة تنفيذ الكود البرمجي المكتوب ضمن بيئة الماتلاب وإجراء التحليل حصلنا على قيم السمات (المعاملات) للتسجيلات الصوتية جميعها، وبالمقارنة مع قاعدة البيانات المعلوماتية التي تتضمن نمط الشخصية الخاص بكل متحدث له ملف صوتي ضمن قاعدة البيانات الصوتية حيث أنه من إجمالي عدد الأشخاص البالغ 150 المكونين لقاعدة البيانات الصوتية هناك 92 شخص لديه عامل الانفتاح Openness مرتفع مما يعني أن الشخص يندمج في خبرات جديدة، وينفتح على التجارب الغريبة، ولديه سعة أفق. وأن هناك 73 شخص لديهم عامل الوعي Conscientiousness مرتفع أي هم أشخاص دقيقون، موثوقون، فعالون في التخطيط. كذلك يوجد 88 شخص لديه عامل الاجتماعية Extroversion عالي ما يدل على أن هؤلاء الأشخاص اجتماعيون، نشيطون، مستقلون. كما بلغ عدد الأشخاص الذين لديهم عامل التوافق Agreeableness مرتفع 87 شخص أي هم أشخاص متفقون مع الآخرين يثقون بهم، ويتعاطفون معهم. كما أنه 64 شخص يتمتعون بنمط شخصية عصبية Neuroticism غير مستقرة عاطفياً خجولة، تُصدم بسهولة، عديمة الثقة بنفسها. بالمقارنة نصل إلى النتائج الموثقة بالمخططات التالية:



الشكل (7) متوسط قيم معاملات LPC10 من أجل الأبعاد الخمسة

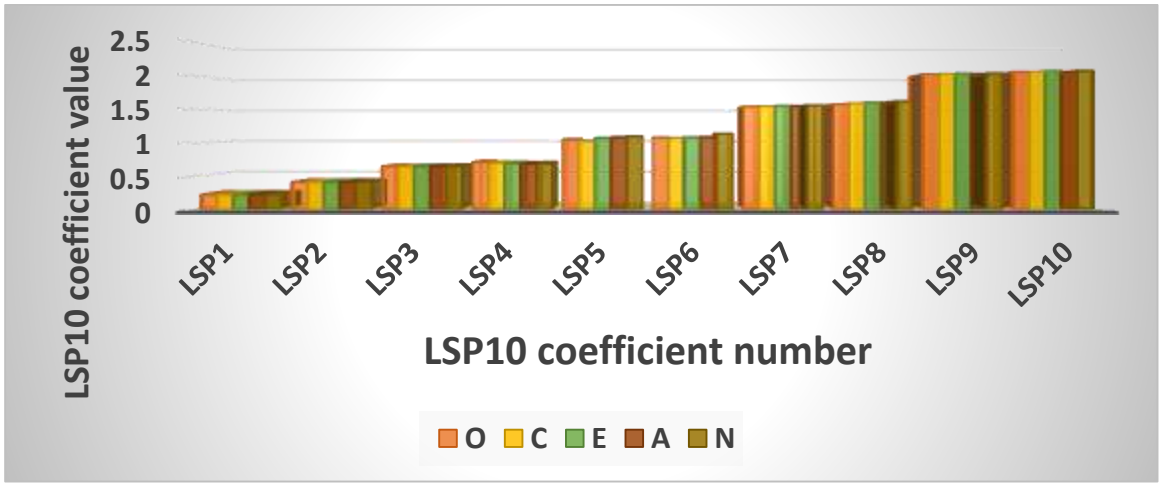


الشكل (8) متوسط قيم معاملات LPC11 من أجل الأبعاد الخمسة

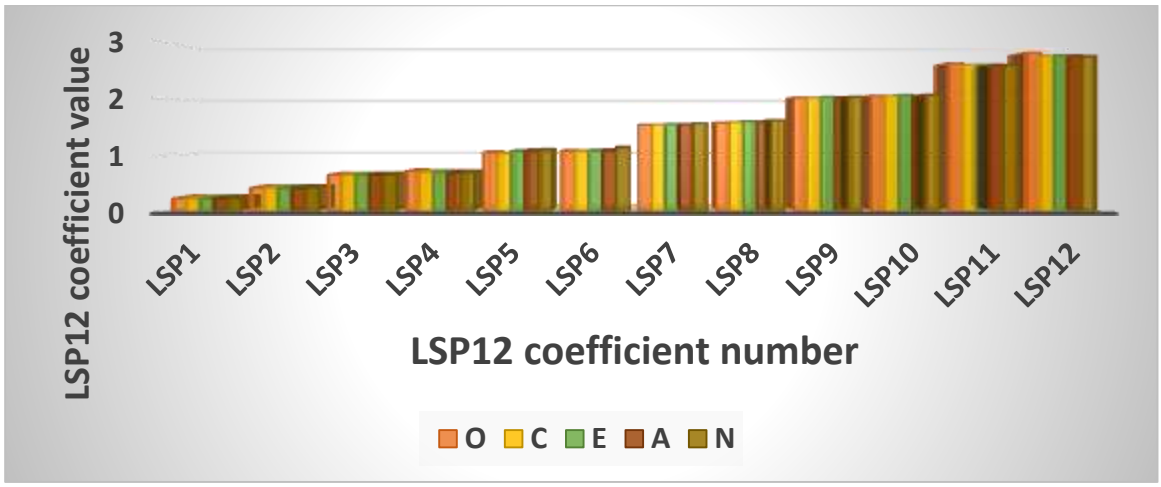


الشكل (9) متوسط قيم معاملات LPC12 من أجل الأبعاد الخمسة

تبين المخططات السابقة في الأشكال 7، 8، 9 متوسطات قيم معاملات LPC المختلفة الرتبة (10، 11، 12) من أجل المتحدثين الذين يحملون الأبعاد الخمسة (O, C, E, A, N) في أنماط شخصياتهم، ويتبين من المخططات الواردة أعلاه أن القيم من أجل كل معامل LPC متقاربة بين الأبعاد الخمسة ويصعب تمييز بُعد أو نمط شخصية من خلال تميزه بقيم معاملات (سمات) LPC منفردة. وكذلك هو الحال بالنسبة لخوارزمية LSP في استخلاص السمات حيث أن قيم معاملاتها متقاربة بين الأبعاد الخمسة مما يجعلها غير مجدية في مجال تمييز أي بُعد (نمط) شخصية، كما هو موضح بالمخططات التالية:



الشكل (10) متوسط قيم معاملات LSP10 من أجل الأبعاد الخمسة



الشكل (11) متوسط قيم معاملات LSP12 من أجل الأبعاد الخمسة

أما خوارزمية MFCC المستخدمة لاستخلاص السمات، فإنها تمتلك عند تطبيقها شعاع سمات بـ 13 معامل هي:

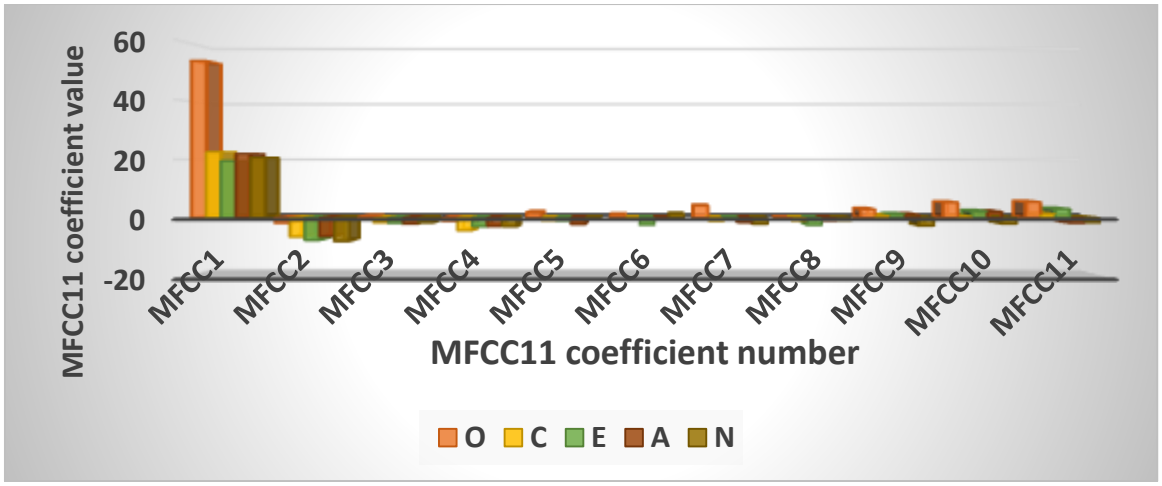
0	Z	P	I	S	PL	CSMG	H	F	Freq	Sl	B	E

تشير الرموز إلى المعاملات التالية:

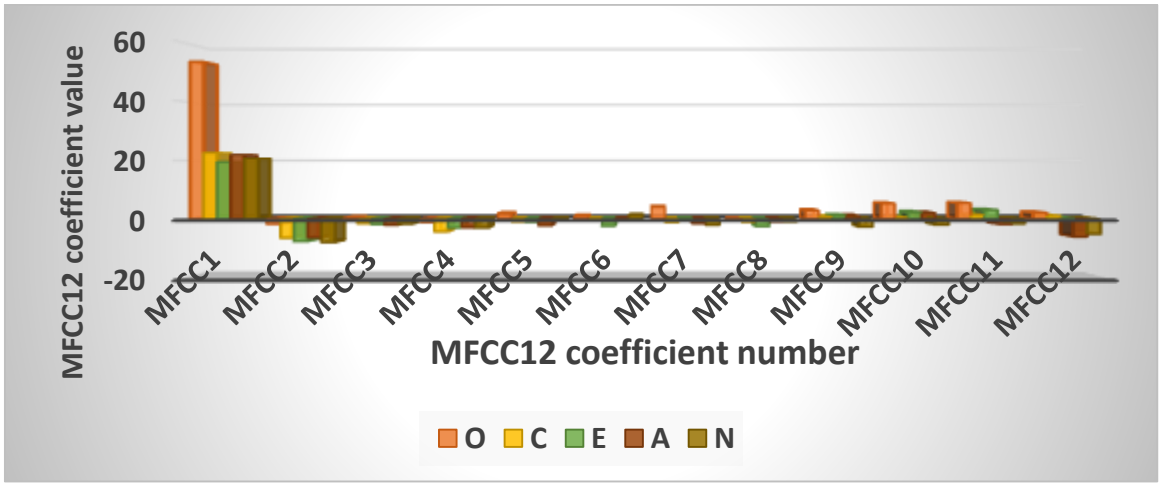
F0: Fundamental Frequency, Z: Zero Crossing Rate, P: Pitch, I: Intensity, S: Spectral Flux, PL: Perceptual Loudness, CSMG: Centre of Spectral Mass Gravity, H: Harmonics to Noise Ratio, F: Formant, Freq: Frequency, Sl: Slope, B: Bandwidth, E: Energy.

من أجل رتبة MFCC11 تكون K=11 أي عدد المعاملات هو 11 وبالتالي شعاع الخرج الناتج يحوي 11 قيمة تدل على أول 11 معامل، وهكذا من أجل MFCC12 فإنَّ شعاع الخرج الناتج تُضاف له سمة B: Bandwidthليصبح شعاع بـ 12 قيمة، وكذلك الأمر بإضافة سمة الـ E: Energy نحصل على MFCC13.

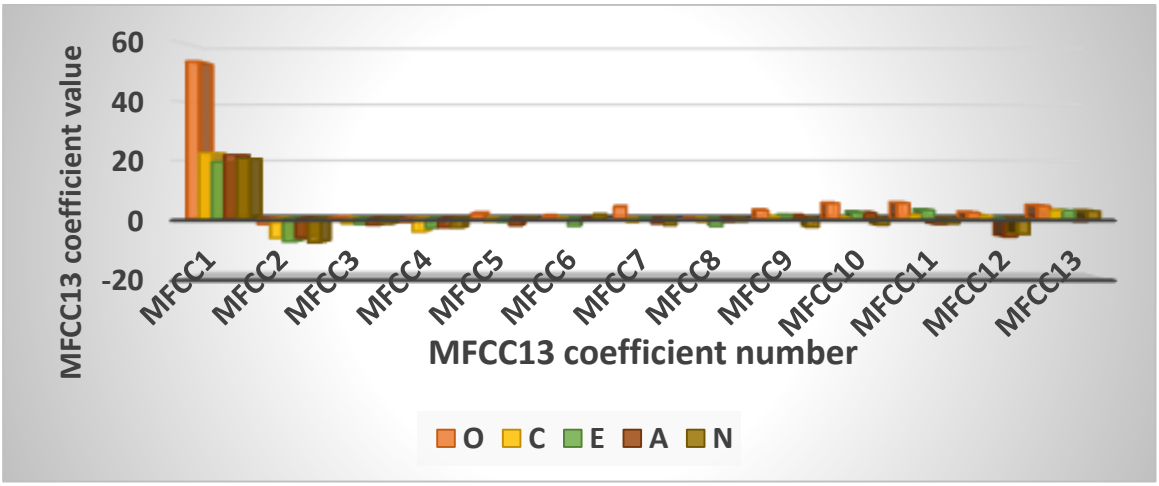
تبين المخططات الناتجة من إجراء تحليل MFCC وجود تباين واضح بين قيم المعاملات بالنسبة لنمط الشخصية المنفتحة (البعد O) وبين قيم المعاملات بالنسبة لباقي أنماط الشخصية (C, E, A, N)، حيث تثبت النتائج ارتفاع قيم المعاملات بالنسبة للشخصية المنفتحة عن باقي أنماط الشخصية وخاصة المعاملات MFCC1, MFCC2, MFCC7 مما يعتبر علامة مميزة وناظمة لتمييز هذا النمط من الشخصيات، حيث بالإمكان استنتاج أن الأشخاص الذين يحملون بُعد الانفتاح (O) يتميزون بقيم معاملات MFCC مرتفعة عن أنماط الشخصية، والعكس صحيح أي يمكن الاستدلال على نمط شخصية المتحدث بأنه منفتح من خلال القيم المرتفعة لمعاملاته في تحليل MFCC.



الشكل (12) متوسط قيم معاملات MFCC11 من أجل الأبعاد الخمسة



الشكل (13) متوسط قيم معاملات MFCC12 من أجل الأبعاد الخمسة



الشكل (14) متوسط قيم معاملات MFCC13 من أجل الأبعاد الخمسة

الاستنتاجات والتوصيات:

تضمنت هذه الدراسة إنشاء قاعدة بيانات لأنماط الشخصية لمتحدثين سوريين، تضمنت 150 متحدث ومتحدثة من مناطق مختلفة من الجمهورية العربية السورية، كانت تلك العناصر متنوعة من حيث الجنس، فاحتوت ذكور وإناث من شرائح عمرية متقاربة، أي أنه تمَّت مراعاة أثر الجنس، وإغفال أثر العمر؛ لذلك يوصى بتوسيع قاعدة البيانات الصوتية المُنجزة بإضافة تسجيلات صوتية لأشخاص من فئات عمرية مختلفة، ودراسة تأثير عامل العمر على السمات الناظمة لأنماط الشخصية.

وتجدر الإشارة أنه سيكون من المفيد جداً تطبيق الكود البرمجي المُنجز في بيئة الماتلاب على دخل من قاعدة البيانات المعيارية المعتمدة العالمية التي تمَّ الحصول عليها وذلك لاختبار دقة النتائج.

المراجع

[1] Abdo, M. S., Kandil, A. H., El-Bialy, A. M. Automatic detection for some common pronunciation mistakes applied to chosen Quran sounds. Biomedical Engineering Conference (CIBEC), 5th Cairo International, 2010, pp. 219-222. DOI: 10.1109/CIBEC. 5716073. (2010).

[2] KOOLAGUDI, S. G. Emotion recognition from speech: a review. International journal of speech technology, 99-117. (2012).

[3] Seddiq, Y. M., Alotaibi, Y., Formant-based analysis of vowels in Modern Standard Arabic—Preliminary results, 11th international conference on information science, signal processing and their applications (ISSPA), pp. 689-694. DOI: 10.1109/ISSPA.2012.6310641. (2012).

[4] MOHAMMADI, G., VINCIARELLI, A.: Automatic personality perception: Prediction of trait attribution based on prosodic features. IEEE Transactions on Affective Computing 3(3), 273–284 (2012).

[5] UAN Z., and X, Z.: Automatic Chinese Personality Recognition Based on Prosodic Features. X. He et al. (Eds.): MMM 2015, Part I, LNCS 8935, pp. 180–190, (2015).

[6] DEIVIDAS E, GINTAUTAS T, Improving Speech Recognition Rate through Analysis Parameters, Electrical, Control and Communication Engineering. Volume 5, Issue 1, Pages 61–66, ISSN (Online) 2255-9159, May (2014).

[7] HOMAYOON B, Fundamentals of speaker Recognition-Springer Sience ISBN: 978-0-387-77591-3. (2011).

[8] Alghamdi, M., EL Hadj, Y. O. M., Alkanhal, M. A Manual System to Segment and Transcribe Arabic Speech, Proceedings of IEEE International Conference on Signal Processing and Communications (ICSPC), pp: 233–236. DOI: 10.1109/ICSPC.2007.4728298. (2007).

[9] POLZEHL, T., MOLLER, S., METZE, F.: Automatically assessing acoustic manifestations of personality in speech. In: 2010 IEEE Spoken Language Technology Workshop (SLT), pp. 7–12. IEEE (2010).

[10] POLZEHL, T., MOLLER, S., METZE, F.: Automatically assessing personality from speech. In: 2010 IEEE Fourth International Conference on Semantic Computing (ICSC), pp. 134–140. IEEE (2010).

[11] MOHAMMADI, G., VINCIARELLI, A., MORTILLARO, M.: The voice of personality: mapping nonverbal vocal behavior into trait attributions. In: Proceedings of the 2nd International Workshop on Social Signal Processing, pp. 17–20. ACM (2010).

[12] BJÖRN S, Voice and Speech Analysis in Search of States and Traits. In Computer Analysis of Human Behavior, Albert Ali Salah, Theo Gevers (eds.), Advances in Pattern Recognition, Springer, pp. 227-253, (2011).

[13] KOMARRAJU M, KARAU J, SCHMECK R, AVDIC A. The big five personality traits, learning styles, and academic achievement. Personality and Individual Differences 51(4), 472–477 (2011).

[14] Muda, L., Begam, M., & Elamvazuthi, I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083. (2010).

[15] Wijoyo, S. Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot. In Proceedings of 2011 International Conference on Information and Electronics Engineering (ICIEE 2011) (pp. 28-29). (2011).

[16] Morris, R. W., & Clements, M. A. Modification of formants in the line spectrum domain. IEEE Signal Processing Letters, 9(1), 19-21. (2002).