

تمييز الشخص المدخن اعتماداً على تحليل مثلثات صوتية مُولدة

د. جعفر الخير*

(تاريخ الإيداع 26 / 5 / 2016. قُبِلَ للنشر في 26 / 9 / 2016)

□ ملخص □

تمّ في هذا البحث اقتراح معيار لدراسة خصائص إشارة الصوت لكل من صنفى المدخنين وغير المدخنين، إذ تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 12 شخص سوري (6 أشخاص مدخنين و 6 أشخاص غير مدخنين)، تتراوح أعمار جميع الأشخاص بين 35 و 45 عام، جميعهم ذكور ويعيشون في مناطق ريفية متجاورة. استغرقت عملية تسجيل الملفات الصوتية ثلاث ساعات، حيث لفظ جميع المتحدثين جملة /أنا أحبُّ سُورِيَة /، وهي جملة غنية صوتياً تشمل جميع الأحرف الصوتية الطويلة (الألف والواو والياء) والحروف الصوتية القصيرة (الفتحة والضمة والكسرة).

اعتماداً على التسجيلات الناتجة تمّ توليد المثلثات الصوتية الطويلة والقصيرة، وهي مثلثات هندسية رؤوسها تمثل قيم ترددات توافقيات المجرى الصوتي الممتد من الحبال الصوتية إلى الشفاه وتسمى النغمات ((Formants، حيث تمّ توليد المثلثات الصوتية الطويلة (AAIUUU) في 10 مستويات، وكذلك المثلثات الصوتية القصيرة (AIU) في 10 مستويات؛ وبذلك نتج 20 مثلث صوتي لكل شخص. تمّ في كل مستوي حساب المسافة بين جميع مراكز المثلثات الصوتية واختيار المسافة الأدنى (d)، فيكون المستوي ذو القيمة الأعظمية للمسافة (d) هو المستوي الممثل للمثلث الصوتي الأنسب لتمييز المدخنين من غير المدخنين، وقد تحقق ذلك بالنسبة للمثلثين AIU35 و AAIUUU45.

الكلمات المفتاحية: قاعدة البيانات الصوتية، المثلثات الصوتية، المدخن، خوارزمية MFCC.

* أستاذ مساعد، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية سورية .

Smoker Distinction Based on Analysis of Created Vowel Triangles

Dr. Jaffar Alkhier*

(Received 26 / 5 / 2016. Accepted 26 / 9 / 2016)

□ ABSTRACT □

In this research, a new comparison criterion was proposed to study properties of the audio signal for each of the varieties of smokers and non-smoking persons. For this purpose, a database for smokers has been created. The smoker database contains 12 Syrian native speakers, six of them were smokers and the others were non-smokers. The smokers had been smoking for more than 10 years. All speakers were men and their ages ranging between 35 and 42 years old. They live in rural towns and speak the same dialect.

Syrian vowels can be classified into long vowels and short ones. The long vowels are /AA/, /UU/, /II/ pronounced as ([ي، و، ا]) and the short vowels are /A/, /U/, /I/ pronounced as ([كسرة، ضمة، فتحة]). In this study, the Speakers have to pronounce the following sentence /I love Syria/ pronounced as ([أنا أحب سورية]), and it was spoken during three hours. This sentence is rich with vowels.

For each speaker, a long vowel triangle in ten planes and a short vowel triangle in ten planes as well were generated and analyzed. A new criterion was suggested to determine the most suitable vowel triangle for smoker distinction. This criterion depends on calculating the different distances among all centers of vowel triangles in each plane and determining the minimal distance called d. For each plane, the most suitable vowel triangle had been set as AIU35 short vowel triangle and AAIUU45 long vowel triangle.

Key words: Vowel database, vowel triangles, smoker, MFCC Algorithm.

*Associate Professor, Department of computer and automatic control, Faculty of mechanical and electrical engineering, Tishreen University, Lattakia, Syria.

مقدمة:

الصوت عبارة عن تذبذب في ضغط الهواء أو الوسط الناقل سببه جسم مهتز وإن كان الصوت كلاماً فإن الجسم المهتز الذي سبب الكلام هو الأوتار الصوتية للإنسان، أما أصوات الموسيقى فسببها الاهتزاز الحاصل في الأدوات الموسيقية.

يُشار عادةً لهذه التغيرات المتذبذبة في ضغط الهواء بمصطلح موجة الصوت (Sound Wave)، ولكي تكون هذه التغيرات مسموعة يجب أن تحصل بتردد مناسب لتكون محسوسة وتسبب تأرجح طبلة الأذن، فالترددات التي تتمكن الأذن البشرية من تمييزها يتراوح مداها ما بين (20kHz-20Hz)، ومع تقدم العمر يقلّ هذا المدى فقد يصل إلى (17kHz-20Hz).

يمكن إعطاء تعريف لغوي للكلام على أنه صوت مشتمل على الأحرف الهجائية التي تبدأ بالألف وتنتهي بالياء، وهذا بالنسبة إلى اللغة العربية إذ تختلف الأحرف الهجائية باختلاف اللغات المستخدمة. ويوصف الكلام من الناحية العلمية على أنه يمثل الصوت الناتج من جهاز النطق الخاص بالإنسان (Speech Production System) لإنشاء تمثيل لغوي ينقل المعلومات من المتحدث للمستمع، يتكون جهاز النطق من الرئتين والقصبه الهوائية والحنجرة والوترين الصوتيين والحلق واللهاة والحنك واللثة واللسان والتجويف الأنفي والأسنان والشفتان، إن الحنجرة عبارة عن قناة تصل بين البلعوم والشعب الهوائية وتحتوي الحنجرة على الحبال الصوتية وهم عبارة عن حبلين في وسط الحنجرة، حيث أن التدخين يؤدي إلى دخول هواء ساخن ملوث إلى سطح الحنجرة الداخلي والحبال الصوتية مما يؤدي إلى جفاف وتهيج الغشاء المخاطي الحنجري[1].

تمت في العام 1974 دراسة حول تأثير التدخين على التردد الأساسي للإناث البالغات [2]، وكانت النتائج أن التردد الأساسي (Frequency Fundamental) للمدخنين أقل مما هو لغير المدخنين بشكل واضح، وفي العام 2004 أجريت دراسة حول تأثير التبغ على الصوت في مرحلة مبكرة من عملية التدخين (أقل من عشر سنوات) [3]، حيث تمّ استخلاص 127 سمة لـ 134 متكلم من الأشخاص البالغين والتي تراوحت أعمارهم بين 20 إلى 29 سنة من كلا الجنسين، إن أكثر الاختلافات التي تمّت ملاحظتها في الصفات المتعلقة بالتردد مثل خاصية التردد الأساسي عند الإناث. وفي العام 2011 تمّت دراسة حول كيفية تغيير المعاملات الصوتية بعد مدة من الامتناع عن التدخين [4]، إذ تمّت الدراسة لـ 18 شخص مدخنين بعد أسبوع واحد ثمّ بعد شهر من الامتناع عن التدخين حيث لوحظ أنّ التردد والوميض (Shimmer) قلّ بينما نسبة توافقيات الإشارة إلى الضوضاء (Ratio Harmonic to Noise) ارتفعت بشكل مؤثر.

أهمية البحث وأهدافه:

يسمح هذا البحث باعتماد معيار مناسب لتمييز الأشخاص المدخنين من غير المدخنين باستخدام مثلثات صوتية يتم توليدها وتحليلها، وذلك من خلال معرفة المثلث الصوتي الأنسب للتمييز، وبالتالي معرفة أفضل النغمات التي يحقق استخلاصها تمايز بين المثلثات الصوتية للمدخنين والمثلثات الصوتية لغير المدخنين.

يهدف البحث إلى إيجاد معيار للتمييز بين المدخنين وغير المدخنين، وذلك عن طريق توليد المثلثات الصوتية الطويلة والقصيرة الخاصة بسنة متحدثين مدخنين والمثلثات الصوتية الطويلة والقصيرة الخاصة بسنة متحدثين غير مدخنين، وإيجاد مراكز ثقل (COG (Center Of Gravity لجميع المثلثات الصوتية لكافة المتحدثين في كل مستويات

النعمة، وحساب المسافات بين هذه المراكز في كل مستوي؛ فمستوي النعمة الذي يحقق تباعد أكبر بين مراكز المثلثات الصوتية للمدخنين عن مراكز مثلثات غير المدخنين هو مستوي النعمة الأفضل لإجراء دراسات مقارنة بين خصائص الصوت للمدخن وخصائص الصوت لغير المدخن.

طرائق البحث وموارده:

تضمن البحث توليد وتحليل مثلثات الحروف الصوتية السورية الطويلة (الألف والواو والياء) والقصيرة (الفتحة والضمة والكسرة)، أي أنه يتم تسجيل جملة غنية صوتياً من قبل متحدثين مدخنين وغير مدخنين، وحساب معاملات السرعة والتسارع والانحراف المعياري للحروف الصوتية من أجل تشكيل مصفوفات الترابط، كما يتم حساب ترددات رنين المجرى الصوتي (Tract Vocal) والتي تحسب بالاعتماد على طيف ترميز التنبؤ الخطي (Predictive Linear LPC (Coding))، هذه الترددات موافقة لقمم الاستجابة الترددية لطيف المجرى الصوتي، وذلك لتقدير ترددات توافقيات المجرى الصوتي المعروفة بالنغمات ((Formants، وتستخدم الشبكات العصبونية ذات التغذية الأمامية للتدريب والاختبار للملفات الصوتية المشكلة لقاعدة البيانات الصوتية، حيث تشكل الحروف الصوتية المسجلة في قاعدة البيانات الصوتية دخل للشبكة العصبونية، مع الأخذ بعين الاعتبار أن استخلاص السمات كان وفق خوارزمية معاملات تردد الميل (MFCC (Mel Frequency Cepstral Coefficients).

وبذلك سيقدم البحث أدوات مساعدة لتحليل الصوتيات لغرض تمييز المدخنين عن غير المدخنين، عن طريق توليد المثلثات الصوتية التي تم إنشائها في بيئة تطويرية مناسبة وهي برنامج (Matlab) وتحليل هذه المثلثات عن طريق حساب المسافات بين مراكزها في كل مستوي نعمة، والذي تم بكتابة معادلات رياضية في بيئة (Matlab).

تأثير التدخين على الحنجرة وبالتالي صوت المتكلم:

يؤدي التدخين إلى جفاف الغشاء المخاطي لمجري التنفس، بداية من الشفاه والأنف، ومروراً باللسان والتجويف الفموي والحنجرة ثم الجهاز التنفسي كله، ويؤدي هذا الجفاف إلى ضمور الغدد المخاطية المسؤولة عن ترطيب هذا الغشاء، ولهذا الترطيب أهمية كبيرة لميكانيكية عمل هذه الأعضاء، وقد ثبت علمياً بأن جفاف أي غشاء مخاطي وتعرضه لمواد كيميائية مثيرة، يتسبب في تحوله إلى نسيج آخر أقرب إلى نسيج الجلد فيكون جافاً، وقد يصبح العضو عرضة للإصابة بالسرطان، ويأتي التدخين على رأس قائمة مسببات سرطان الحنجرة، كما أن الدخان يسبب تمدد الشعيرات الدموية بالحنجرة وبالتالي سهولة انفجارها وحدوث النزيف تحت الغشاء المخاطي المغلف للثنايا الصوتية، مؤدياً إلى تكوين حبيبات أو لحميات أو أكياس أو تجمع مائي أو دموي بها مما يؤدي إلى اختلاف في بعض خصائص الصوت للشخص المدخن [9,8].

إعداد قاعدة البيانات:

تشكل قواعد بيانات الكلام (Database Speech) الركن الرئيسي في بناء نظم حاسوبية، وتشكل البنية التحتية لبناء نظم تخاطب مع الحاسب، وتتكون قواعد بيانات الكلام عادة من ملفات صوتية (Files Wave) سبق أن سجلت لأشخاص باللغة المراد التعامل معها. وكلما كانت قاعدة بيانات الكلام شاملة في محتواها وغنية بتنوع عناصرها، كلما ساعد ذلك على إخراج نظم حاسوبية ذات جودة عالية في أدائها.

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 للتمييز بين المدخنين وغير المدخنين فقد تم تضمين قاعدة

البيانات الصوتية تسجيلات صوتية لأشخاص مدخنين وتسجيلات صوتية لأشخاص غير مدخنين.

المتحدثون:

تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 12 شخص سوري، نصفهم مدخنون والنصف الآخر غير مدخن، جميع الأشخاص المدخنين يمارسون عادة التدخين منذ أكثر من 10 أعوام، وجميع الأشخاص من شرائح عمرية متقاربة حيث تتراوح أعمارهم بين 35 و45 عام، جميعهم ذكور ويعيشون في مناطق ريفية متجاورة، أي يتماثل الجميع بعامل العمر واللهجة والجنس وبالتالي لن يكون لهذه العوامل تأثير على القيم الحاصلة بنتيجة البحث، ويتميزون فقط في صنف المدخنين وغير المدخنين.

المادة اللغوية:

تختلف المواد اللغوية التي يقوم المتحدثون بتسجيلها باختلاف الغاية من بناء قاعدة البيانات الصوتية، ونظراً لأنّ 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 الهستوغرامي (Histogram) وبالتالي تظهر بوضوح في الرسم الطيفي، فقد لفظ جميع المتحدثين جملة /أنا أحبُّ سُورِيَّةَ /، وهي جملة غنية صوتياً تشمل جميع الأحرف الصوتية الطويلة أي الألف والواو والياء والحروف الصوتية القصيرة أي الفتحة والضمة والكسرة. هذه الجملة كُتبت على أوراق A4، كل شخص من الأشخاص الاثني عشر يلفظ الجملة ثلاث مرات، يتم اللفظ بشروط عادية من السرعة ومستوى رفع الصوت.

ظروف التجربة:

تمّ التسجيل داخل استديو معزول صوتياً، حيث تمّ إعداده وتجهيزه على نحو خاص يحقق إنتاجاً صوتياً واضحاً ودقيقاً وأميناً، حيث أن جدران الاستديو وأرضيته وسقفه معالجة بمواد ماصة للصوت كما تبطن الفراغات بمثل هذه المواد، كذلك الاستوديو خالياً من الشبائيك التي قد تسمح بدخول الضوضاء، وباب الاستوديو عازل للصوت أيضاً. يتم التسجيل لكل شخص على حدا، حيث يقرأ الجملة المكتوبة على ورقة A4، بشروط عادية من السرعة وعلو الصوت، وبالإستعانة بمجموعة من الأدوات كالميكروفون الديناميكي نوع (Yamaha)، والقناة الصوتية (Stereo Channel)، وبرنامج التسجيل (Adobe Audition 3)، وخلال ثلاث ساعات تمّ الانتهاء من جلسة التسجيل لجميع المتحدثين بتردد أخذ عينات 44.1 kHz

توليد المثلثات الصوتية:

المثلث الصوتي هو مثلث هندسي، يلزم لتوليده معرفة إحداثيات رؤوسه، وبما أن رؤوس المثلث الصوتي هي قيم ترددات نغمات الحروف الصوتية، فالعمل البرمجي ينصب على معرفة قيم ترددات هذه النغمات، وبما أنّ النغمات هي إحدى سمات الكلام (Speech Features) فيتم استخلاصها وفق الخوارزمية MFCC وهي خوارزمية جاهزة برمجياً في (MathWorks)) يتم تغيير بارامتراتهما بما يتناسب مع عدد العينات وتردد النقطيع المستخدم في هذا البحث. بنتيجة تطبيق خوارزمية MFCC نحصل على النغمات متضمنة العديد من البارامترات (Formant Parameters) مثل عرض الحزمة (Bandwidth) والميل (Slope) والتردد (Frequency)، ولحساب التردد تمّ اعتماد خوارزمية LPC المتوفرة أيضاً في (MathWorks)) بعد ضبط رتبة التنبؤ فيها بما يتناسب مع خصوصية العينات المأخوذة، وبنتيجة هذه الخوارزمية نحصل على قيم الترددات F1, F2, F3, F4, F5 لكافة الحروف الصوتية المُضمنة في الجملة الملفوظة لجميع المتحدثين.

وبالتالي لتوليد المثلث الصوتي الطويل (أي المثلث الخاص بالحروف الصوتية الطويلة وهي الألف والواو والياء والمرمزة بالحروف /AA/، /UU/، /II/؛ يتم استخدام التعلية البرمجية (Plot) بلغة الماتلاب لرسم هذه النقاط الثلاثة في المستويات العشر التالية F3, F2F4, F2F5, F3F4, F3F5, F4F5, F1F2, F1F3, F1F4, F1F5, F حيث أنه كل من النقاط /AA/، /II/، /UU/ تُحدد بإحداثيين هما قيمتي النغمتين الممثلتين كمحورين للمستوي المطلوب، وبعد رسم رؤوس المثلث يتم الاستعانة بالتعلية (Line) للوصل بخط بين هذه النقاط وتشكيل أضلاع المثلث، وكذلك الأمر بالنسبة للمثلثات الصوتية القصيرة (أي المثلثات الخاصة بالحروف الصوتية القصيرة وهي الفتحة والضمة والكسرة والمرمزة بالحروف /A/، /I/، /U/؛ وبالتالي ينتج المثلث الصوتي، ولحساب مركز ثقل المثلث الذي هو نقطة تقاطع متوسطات المثلث يتم حساب المتوسط الحسابي لرؤوس هذا المثلث بكتابة المعادلة الرياضية المعيرة عن المتوسط الحسابي بلغة الماتلاب.

المعيار المقترح:

لكل شخص من الأشخاص الاتني عشر المعتمدين في نظام قاعدة البيانات الصوتية تمّ توليد عشرة مثلثات صوتية طويلة وعشرة مثلثات صوتية قصيرة وبالتالي نتج لكل شخص عشرين مثلث صوتي، وكان عدد المثلثات الصوتية الكلية الناتجة هو $240=20*12$ مثلث صوتي، يسمى كل مثلث صوتي باسم من خمس خانوات، ترمز الخانات الثلاثة الأولى إلى الحروف الصوتية الطويلة أو القصيرة التي تتم معالجتها، وترمز الخانتين الباقيتين إلى ترددي النغمتين المُستخلصتين، وعليه تكون المثلثات العشرون الناتجة لكل متحدث هي: AIU12, AIU13, AIU14, AIU15, AIU23, AIU24, AIU25, AIU34, AIU35, AIU45, AAIUU12, AAIUU13, AAIUU14, AAIUU15, AAIUU23, AAIUU24, AAIUU25, AAIUU34, AAIUU35, AAIUU45 المثلثات الصوتية الناتجة تمّ حساب إحداثيات مركز الثقل COG والذي هو عبارة عن نقطة تقاطع متوسطات المثلث وتقع داخله دائماً، ولحساب مركز ثقل المثلث الصوتي يتم حساب المتوسط الحسابي لرؤوس المثلث الصوتي وكتابة المعادلة الرياضية المعيرة عن هذا المتوسط.

في كل مستوي من المستويات العشر ومن أجل كل نوع من المثلثات الصوتية سواء الطويلة أو القصيرة تمّ تمثيل مراكز ثقل المثلثات الصوتية لجميع المتحدثين الاتني عشر. ومن تمّ حساب المتجه d الذي هو عبارة عن متجه يصل بين مركزي مثلثين في نفس المستوي، وهذا الشعاع يتجه من مركز ثقل المثلث الصوتي للمتحدث الحالي إلى كافة مراكز ثقل المثلثات الصوتية لبقية المتحدثين، ويتم التعبير عن قيم المتجهات d في كل مستوي من المستويات العشرين (عشر مستويات تخص المثلثات الصوتية الطويلة وعشر مستويات تخص المثلثات الصوتية القصيرة) عن طريق مصفوفة M كما توضح المعادلة (1):

$$M = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1N} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2N} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \dots & d_{NN} \end{bmatrix}$$

المصفوفة السابقة M تمثل قيم المتجه d لجميع المتحدثين في كل مستوي، وتمثل N عدد المتحدثين الكلي وهو 12، وبما أنّ عدد المتحدثين 12 فقد تمّ تمييز كل متحدث بأحد الأرقام التسلسلية من 1 حتى 12، وتحديدًا تمّ إعطاء

المدخنين الأرقام من 1 إلى 6 والأشخاص غير المدخنين الأرقام من 7 إلى 12. حيث أنه من أجل التمثيل AIU12 أي تمثيل المثلثات الصوتية القصيرة في المستوى F1F2 تكون المصفوفة M معبرة عن القيم 18d وهي قيمة المتجه الواصل بين مركز ثقل المثلث الصوتي للشخص الأول وهو مدخن مع مركز ثقل المثلث الصوتي للشخص الثامن وهو غير مدخن، ويعبر المتجه 93d مثلاً عن المتجه الواصل بين مركز ثقل المثلث الصوتي للشخص التاسع وهو غير مدخن مع مركز ثقل المثلث الصوتي للشخص الثالث وهو مدخن وهكذا بالنسبة لبقية القيم، وتكون عناصر القطر الرئيسي في المصفوفة M هي قيم صفرية نظراً لانعدام المسافة بين النقطة وذاتها رياضياً بمعنى أن قيمة 11d مثلاً هي الصفر لأنه لا توجد مسافة بين مركز ثقل المثلث الصوتي للشخص وبين نفسه، كذلك تهمل قيم المتجهات الواصلة بين مراكز المثلثات الصوتية للأشخاص المدخنين فيما بينهم، ويتم تجاهل قيم المتجهات الواصلة بين مراكز المثلثات الصوتية للأشخاص غير المدخنين فيما بينهم؛ لأن الهدف الرئيسي للبحث هو إيجاد المثلث الصوتي الأنسب للتمييز بين المدخنين وغير المدخنين وليس إيجاد المثلث الصوتي الأنسب لتمييز الأشخاص أو التعرف عليهم. وتكون المصفوفة المعبرة عن القيم المحسوبة هي المصفوفة (1) حيث يعبر عن القيمة التي يتم تجاهلها بالرمز - كما تظهر المعادلة (2):

$$M = \begin{bmatrix} 0 & - & - & - & - & - & d_{17} & d_{18} & d_{19} & d_{1,10} & d_{1,11} & d_{1,12} \\ - & 0 & - & - & - & - & d_{27} & d_{28} & d_{29} & d_{2,10} & d_{2,11} & d_{2,12} \\ - & - & 0 & - & - & - & d_{37} & d_{38} & d_{39} & d_{3,10} & d_{3,11} & d_{3,13} \\ - & - & - & 0 & - & - & d_{47} & d_{48} & d_{49} & d_{4,10} & d_{4,11} & d_{4,12} \\ - & - & - & - & 0 & - & d_{57} & d_{58} & d_{59} & d_{5,10} & d_{5,11} & d_{5,12} \\ - & - & - & - & - & 0 & d_{67} & d_{68} & d_{69} & d_{6,10} & d_{6,11} & d_{6,12} \\ d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & 0 & - & - & - & - & - \\ d_{81} & d_{82} & d_{83} & d_{84} & d_{85} & d_{86} & - & 0 & - & - & - & - \\ d_{91} & d_{92} & d_{93} & d_{94} & d_{95} & d_{96} & - & - & 0 & - & - & - \\ d_{10,1} & d_{10,2} & d_{10,3} & d_{10,4} & d_{10,5} & d_{10,6} & - & - & - & 0 & - & - \\ d_{11,1} & d_{11,2} & d_{11,3} & d_{11,4} & d_{11,5} & d_{11,6} & - & - & - & - & 0 & - \\ d_{12,1} & d_{12,2} & d_{12,3} & d_{12,4} & d_{12,5} & d_{12,6} & - & - & - & - & - & 0 \end{bmatrix} \quad (2)$$

من أجل كل مصفوفة M تم اختيار القيمة الأدنى للمتجه d بمعنى أنه من أجل كل مستوي من المستويات العشرين تمت مقارنة أطوال جميع المتجهات وانتقاء القيمة الأدنى للمتجه d وتوثيقها في جدول تحت مسمى d_{min} حيث تمثل هذه القيمة أقرب مركزي مثلثين صوتيين في كل مستوي، وبعد حساب قيم d_{min} لجميع المستويات تم اختيار أعلى قيمة من بين جميع قيم الـ d_{min} الناتجة ووثقت باسم d_{max} ، ويكون المستوي الممثل لقيمة الـ d_{max} هو المستوي الأنسب لتمييز الأشخاص المدخنين من غير المدخنين، وتفسير ذلك أن المستوي ذو قيمة الـ d_{max} هو مستوي يحقق أكبر تباعد بين مراكز ثقل المثلثات الصوتية للمدخنين ومراكز ثقل المثلثات الصوتية لغير المدخنين، وبالتالي هو المستوي الذي يحقق أكبر تمايز بين المثلثات الصوتية للمدخنين عن غير المدخنين ويكون هو الأنسب من أجل كافة الدراسات المهمة بتأثير التدخين على إشارة الصوت.

التعرف على الحرف الصوتي المحكي باستخدام مصنف الشبكات العصبونية ذات التغذية الأمامية والانتشار الخلفي للخطأ (FFBPNN):

تعتبر الشبكات العصبونية ذات التغذية الأمامية واحدة من أهم الطرق الحديثة التي لها كفاءة عالية في إعطاء نتائج مرضية وجيدة في التعرف، وجزء هام من بناء الشبكة العصبونية هو استخدام خوارزمية دقيقة وقوية في التعلم، والأكثر انتشاراً هي الشبكة العصبونية ذات التغذية الأمامية والانتشار الخلفي للخطأ، وفيها يكون اتجاه الإشارات الداخلة في الشبكة دوماً إلى الأمام، وبذلك تكون الإشارة الخارجة من أي عصبون تعتمد على الإشارات الداخلة فقط، وتحتاج الشبكات العصبونية ذات التغذية الأمامية إلى وجود زوجين من المتجهات هما متجه الإدخال ومتجه الإخراج المطلوب، تبدأ عملية التدريب بمتجه الإدخال حيث يطبق على الشبكة فينتج الإخراج الحقيقي، ويقارن مع ما يقابله من متجه الإخراج المتوقع والفرق بينهما يمثل الخطأ الذي يستخدم لتعديل الأوزان طبقاً لخوارزمية التعلم، ويستمر التدريب إلى أن يصل الخطأ إلى أقل ما يمكن.

يكون دخل هذه المرحلة هو شعاع السمات الخاص بالحرف الصوتي، وخرجها هو الحرف الصوتي الذي تمّ التعرف عليه.

البنية المقترحة للشبكة العصبونية (FFBP):

اعتماداً على النتائج التي خلصت إليها الدراسات المرجعية التي اعتمدت الشبكات العصبونية كمصنفات للتعرف على الحروف الصوتية؛ تم اقتراح البنية التالية للشبكة العصبونية مع البارامترات اللازمة لعملية التدريب:

- طبقة خفية واحدة.
- عدد عصبونات الطبقة الخفية 1000.
- عدد عصبونات طبقة الخرج 10.
- تابع الـ 'tansig' للطبقة الخفية.
- تابع الـ 'logsig' للطبقة الثانية (طبقة الخرج).
- تابع الـ 'trainscg' كتابع تدريب (تفعيل).
- معدل تعلم ضمن المجال (0.01-0.05) وتم اختيار القيمة 0.035 كقيمة نهائية.
- قيمة الخطأ 0.00001.

علماً أن دخل الشبكة العصبونية هو شعاع السمات الناتج من مرحلة استخلاص السمات ويتألف من 65 عنصر تمثل سمات حرف صوتي واحد، أما الخرج فهو عبارة عن شعاع مؤلف من 10 عناصر.

يمكن استخدام التعليمات التالية لبناء هذه الشبكة باستخدام الماتلاب:

```
('net=newff(minmax(P),[1000,200],{'tansig','logsig'},'trainscg
net.trainParam.show=50;
net.trainParam.lr=0.035;
net.trainParam.epochs=1000;
net.trainParam.goal=0.00001;
```

ويمكن تدريبها وفق التعليمات التالية:

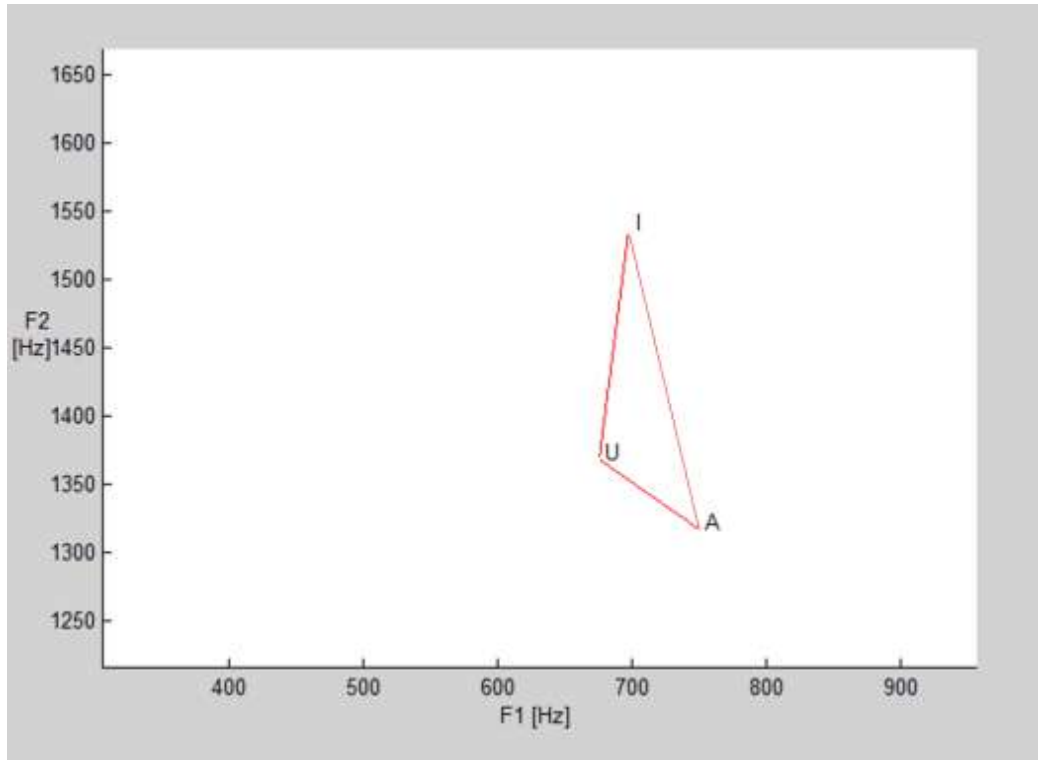
```
(net tr]=train(net,P,T]
```


النتائج والمناقشة:

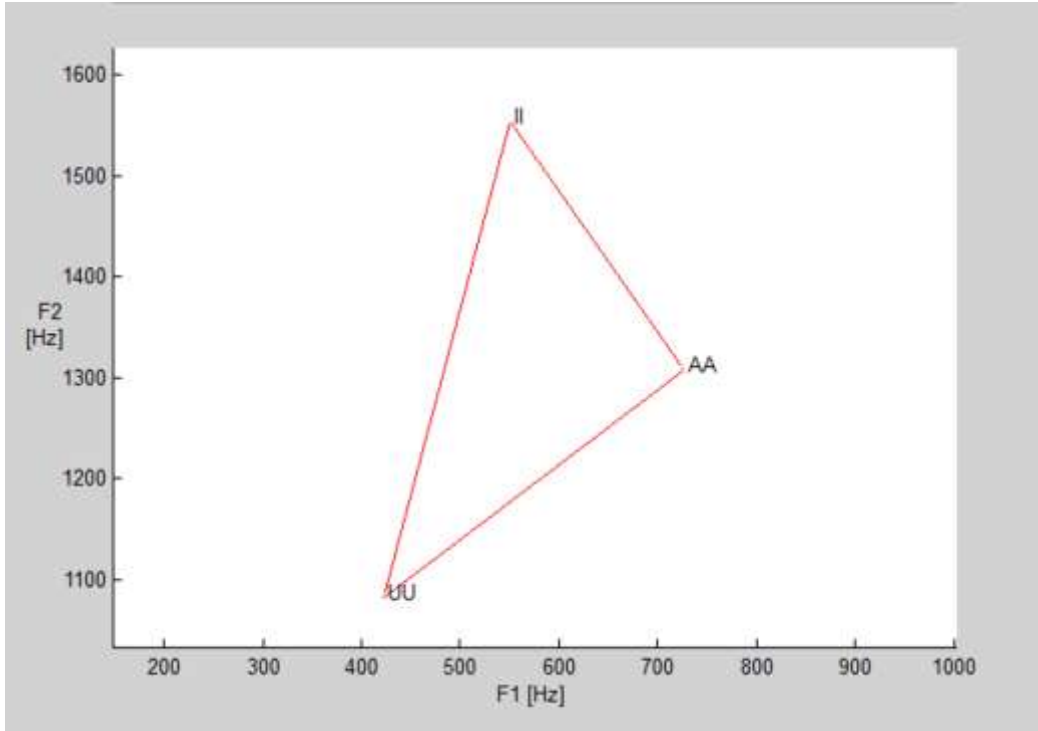
تضمن البحث عدة مراحل، فتمّ في المرحلة الأولى بناء قاعدة بيانات لأصوات أشخاص سوريين تكون مناسبة لتصميم المشروع، حيث تضمنت قاعدة البيانات تسجيلات صوتية لأشخاص سوريين من نفس العمر والجنس واللهجة نصفهم يمارسون عادة التدخين ونصفهم لا يدخنون، وتمّ في المرحلة التالية إجراء تحليل MFCC على التسجيلات من أجل استخراج السمات، وبناء مصفوفات ارتباط من معاملات السرعة والتسارع والانحراف المعياري، ومن ثمّ بناء الشبكة العصبونية وتدريبها وحفظها، حيث اعتمدت الشبكة العصبونية كمصنّف يقوم بإيجاد الرسم التخطيطي (Mapping) بين تسلسل أشعة السمة وبين عنصر الكلام المتعرف عليها، ثمّ إجراء الدراسة على نتائج تنفيذ الكود البرمجي للتسجيلات، حيث يتمّ رسم مثلثات الحروف الصوتية الطويلة والقصيرة، وهي عبارة عن مثلثات هندسية رؤوسها تمثل قيم الترددات للحروف الصوتية، وحساب مراكز المثلثات الصوتية لكافة المتحدثين في كل مستوى، وقياس المسافة بين هذه المراكز.

النتائج العملية:

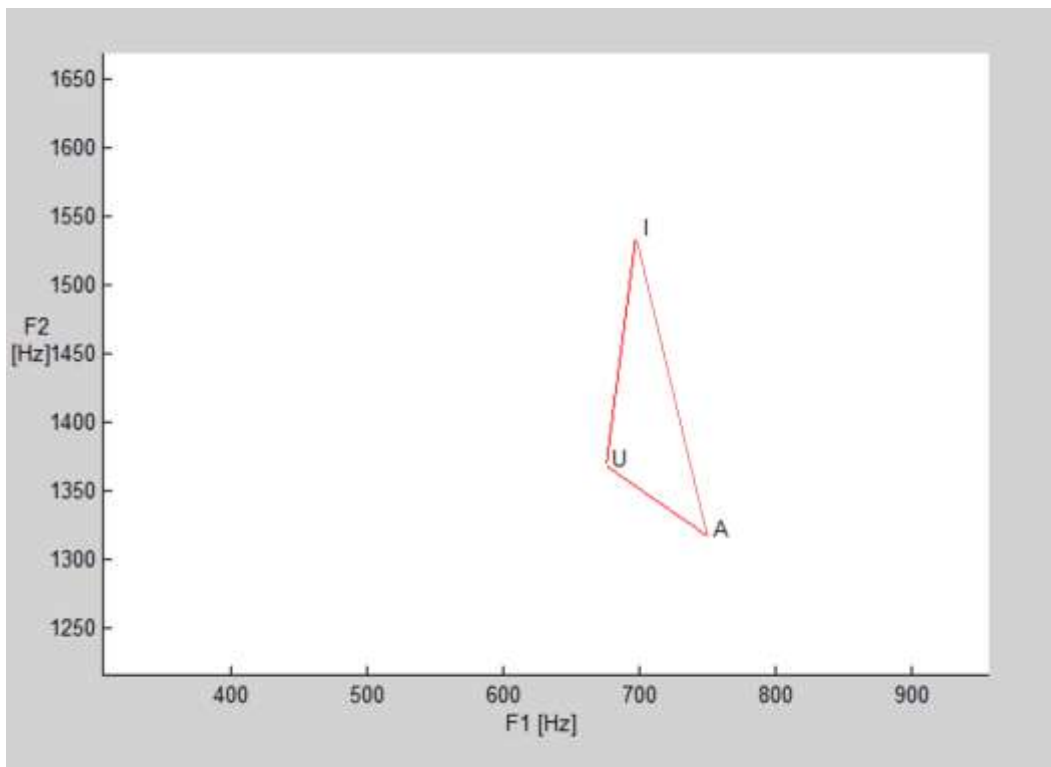
وقد أعطى تطبيق الكود البرمجي المثلثات الصوتية الطويلة والقصيرة للأشخاص المدخنين والمثلثات الصوتية الطويلة والقصيرة للأشخاص غير المدخنين، والأشكال التالية توضح المثلثات الصوتية الطويلة والقصيرة لأحد المدخنين وهو الشخص رقم 1 وللشخص رقم 7 غير المدخن:



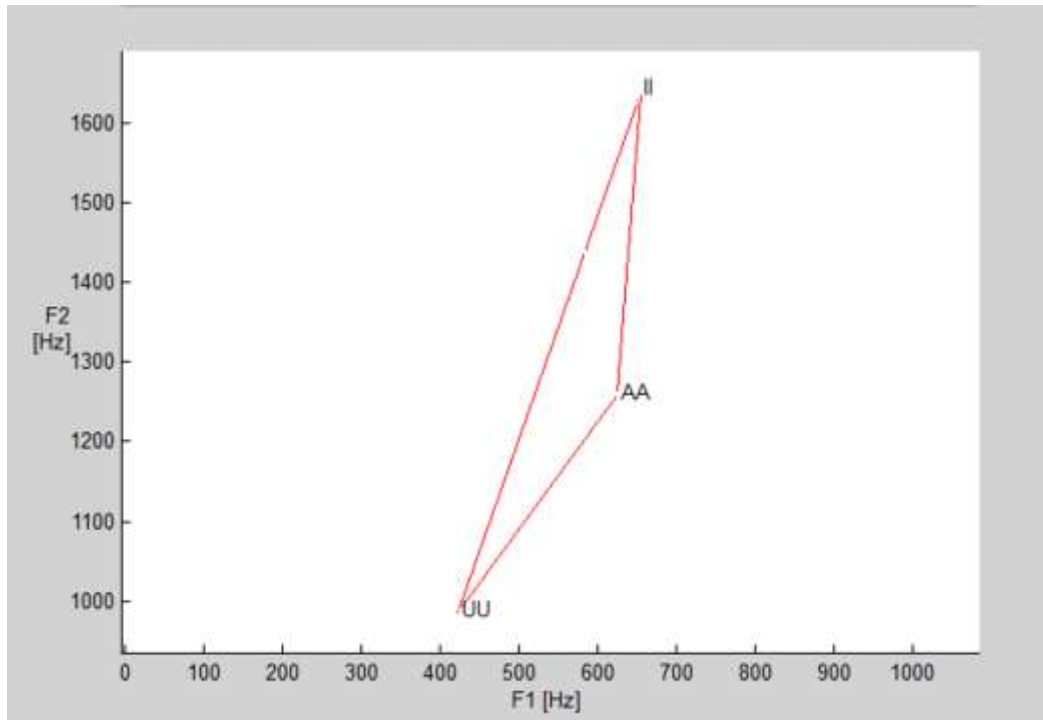
الشكل(1): مثلث صوتي قصير خاص بالمتحدث المدخن رقم 1



الشكل(2): مثلث صوتي طويل خاص بالمتحدث المدخن رقم 1

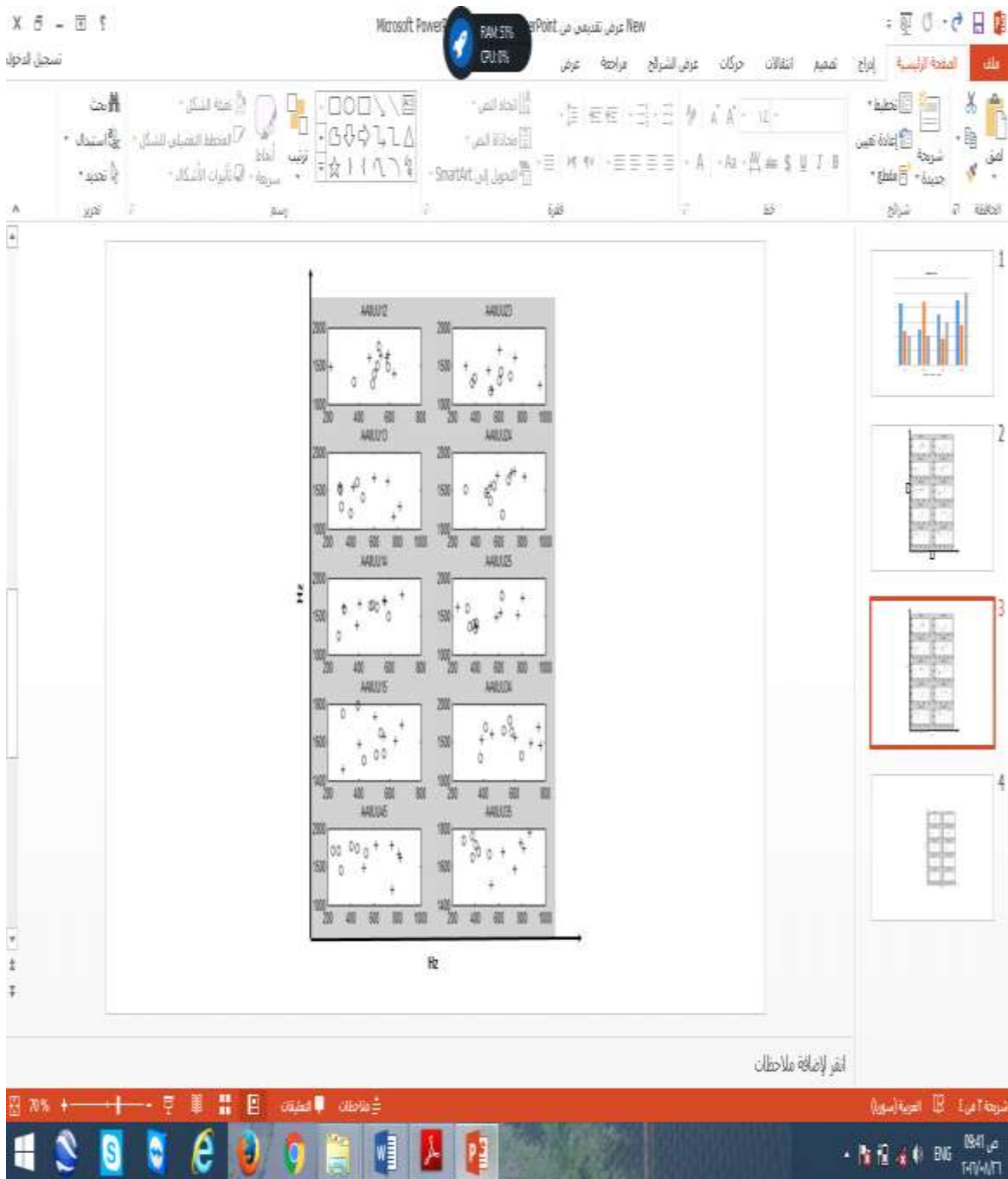


الشكل(3): مثلث صوتي قصير خاص بالمتحدث غير المدخن رقم 7

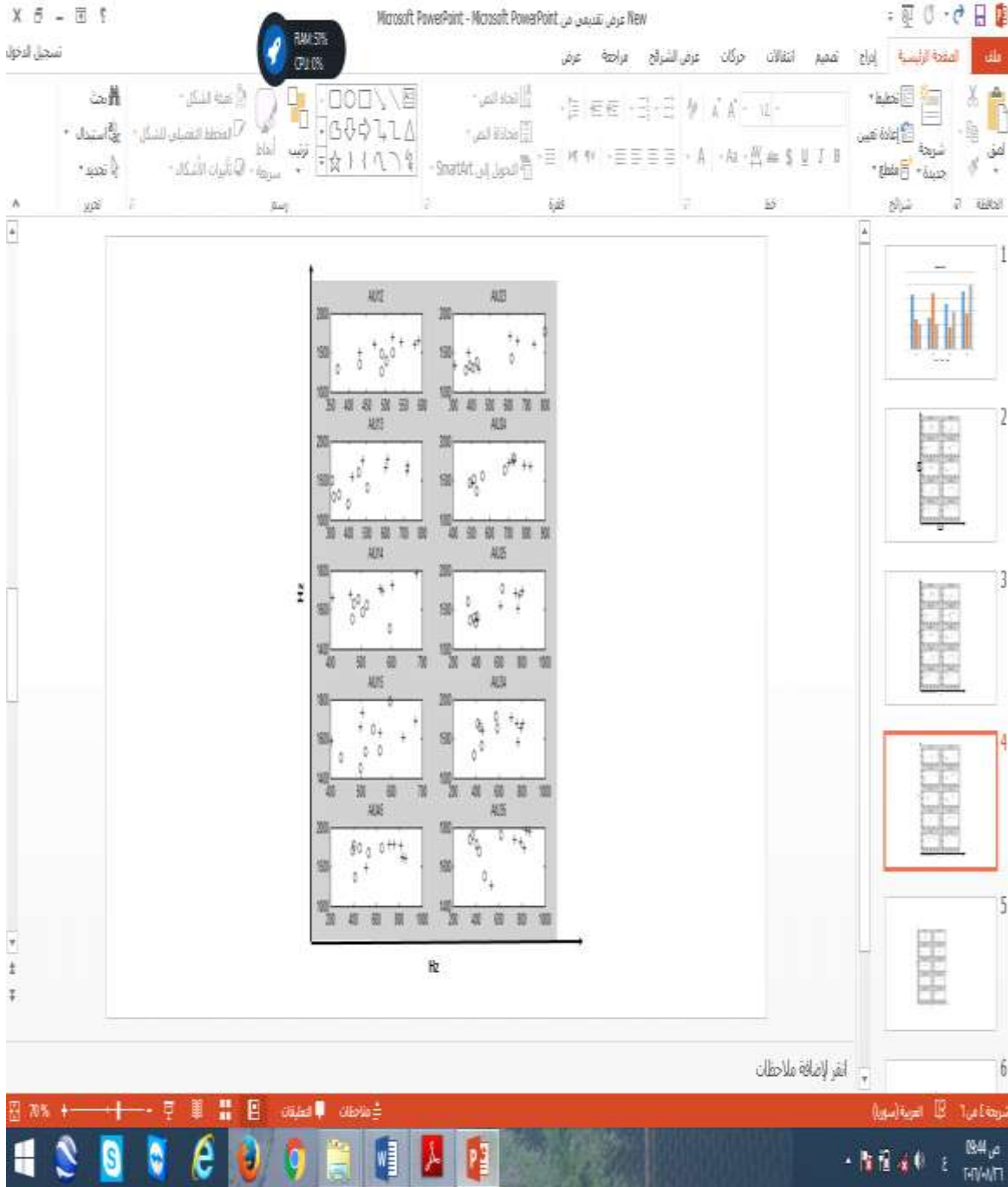


الشكل(4): مثلث صوتي طويل خاص بالمتحدث غير المدخن رقم 7

وبعد الحصول على مراكز جميع المثلثات الصوتية لجميع المتحدثين؛ تمّ رسم هذه المراكز لجميع المتحدثين في كل مستوى، وملاحظة توزيع مراكز ثقل المثلثات الصوتية للأشخاص المدخنين المُعبر عنها بالرمز (0) وبالرمز (+) للأشخاص غير المدخنين، كما يظهر في الشكلين (5,6)، حيث تمثل قيم المحاور قيم ترددية:



الشكل (5): توزيع مراكز المثلثات الصوتية الطويلة لجميع المتحدثين في المستويات العشر



الشكل(6): توزيع مراكز المثلثات الصوتية القصيرة لجميع المتحدثين في المستويات العشر

وبمقارنة هذه النتائج وحساب قيم المتجه d_{min} لكل مستوي والتي تمثل أصغر مسافة بين مراكز المثلثات الصوتية المُولدة في كل مستوي؛ تمّ الحصول على القيم الموضحة في الجدول (1):

الجدول (1) قيم المتجه d_{min} لكل مستوى.

المثلث الصوتي	d_{min} [Hz]
AIU12	1
AIU13	5
AIU14	6
AIU15	5
AIU23	7
AIU24	11
AIU25	18
AIU34	19
AIU35	23
AIU45	21
AAIIUU12	2
AAIIUU13	2
AAIIUU14	6
AAIIUU15	7
AAIIUU23	9
AAIIUU24	14
AAIIUU25	17
AAIIUU34	20
AAIIUU35	21
AAIIUU45	23

وبملاحظة قيم المتجه الدنيا d_{min} لكل مستوى ومقارنتها؛ نجد أنّ القيمة العظمى d_{max} تحققت في المستوى F3-F5 بالنسبة للمثلثات الصوتية القصيرة وفي المستوى F4-F5 بالنسبة للصوتيات الطويلة، وعليه تكون المثلثات الصوتية الأنسب لتمييز المدخنين عن غير المدخنين من خلال خصائص إشارة الصوت هي المثلثين AIU35 و AAIIUU45، حيث أنه عند معالجة الحروف الصوتية الطويلة يُفضل استخلاص النغمتين الرابعة والخامسة، وعند معالجة الحروف الصوتية القصيرة يُستحسن استخلاص النغمتين الثالثة والخامسة، فالنغمات الموصى باستخلاصها هي النغمات التي تبرز اختلافات واضحة وتمايز جلي بين المدخنين وغير المدخنين.

الاستنتاجات والتوصيات:

تضمنت هذه الدراسة اقتراح معيار مناسب يمكن من دراسة خصائص إشارة الصوت لكل من صنف المدخنين وغير المدخنين، إذ تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 12 شخص سوري، نصفهم مدخنون والنصف الآخر غير مدخن، جميع الأشخاص المدخنين يمارسون عادة التدخين منذ أكثر من 10 أعوام، تتراوح أعمار جميع

الأشخاص بين 35 و 45 عام، جميعهم ذكور ويعيشون في مناطق ريفية متجاورة، أي يتماثل الجميع بعامل العمر واللهجة والجنس، ويتميزون فقط في صنف المدخنين وغير المدخنين.

استغرقت عملية تسجيل الملفات الصوتية ثلاث ساعات داخل استوديو معزول صوتياً، حيث لفظ جميع المتحدثين جملة /أنا أحب سُورِيَةَ /، وهي جملة غنية صوتياً تشمل جميع الأحرف الصوتية الطويلة أي الألف والواو والياء والحروف الصوتية القصيرة أي الفتحة والضمة والكسرة.

اعتماداً على التسجيلات الناتجة تم توليد المثلثات الصوتية الطويلة والقصيرة، وهي مثلثات هندسية رؤوسها تمثل قيم ترددات توافقيات المجرى الصوتي الممتد من الحبال الصوتية إلى الشفاه وتسمى النغمات ((Formants، حيث تم توليد المثلثات الصوتية الطويلة (AAIIUU) في 10 مستويات، وكذلك المثلثات الصوتية القصيرة (AIU) في 10 مستويات؛ وبذلك نتج 20 مثلث صوتي لكل شخص. تم في كل مستوي حساب المسافة بين جميع مراكز المثلثات الصوتية واختيار المسافة الأدنى (d)، فيكون المستوي ذو القيمة الأعظمية للمسافة (d) هو المستوي الممثل للمثلث الصوتي الأنسب لتمييز المدخنين من غير المدخنين، وقد تحقق ذلك بالنسبة للمثلثين AIU35 و AAIIUU45. حيث أنه عند معالجة الحروف الصوتية الطويلة يُفضل استخلاص النغمتين الرابعة والخامسة، وعند معالجة الحروف الصوتية القصيرة يُستحسن استخلاص النغمتين الثالثة والخامسة، فالنغمات الموصى باستخلاصها هي النغمات التي تبرز اختلافات واضحة وتمايز جلي بين المدخنين وغير المدخنين.

وتجدر الإشارة أنه سيكون من المفيد جداً تطبيق المعيار المقترح لتمييز المدخنين من غير المدخنين اعتماداً على مثلثات صوتية للغات أخرى، ودراسة معايير إضافية لتكون عامل مميز في التصنيف للذين يمارسون عادة التدخين.

المراجع

- [1] LOUISE D, ANGELIKA B, *Voice Parameter Changes in Smokers During Abstinence from Cigarette Smoking*, 2011, Proceedings of the Icphs xvii, p588-590.
- [2] GRAY G, HARVEY R, *The Effects of Smoking on the Speaking Fundamental Frequency of Adult Women*, Journal of Psycholinguistic Research, Issn 1573-6555, vol. 3 (3. 1974), p. 225-231.
- [3] AMPARO G, *Early Effects of Smoking on the Voice: A multidimensional Study*, 2004, *International Medical Journal of Experimental and Clinical Research*, 2004, p649-56.
- [4] LOUISE D, ANGELIKA B, *Voice Parameter Changes in Smokers During Abstinence from Cigarette Smoking*, 2011, Proceedings of the Icphs xvii, p588- 590.
- [5] STANEK, M., POLAK, L. Algorithms for Vowel Recognition in Fluent Speech Based on Formant Positions. *In Proc. 36th International Conference on Telecommunication and Signal Processing*. Rome (Italy), 2013, pp. 521-525.
- [6] HU, R. and DAMPER, R. *Fusion of two classifiers for speaker identification: removing and not removing silence*, the 8th International conference on Information fusion, vol 1, 2006, 429-436.
- [7] STANEK, M., SIGMUND, M. Speaker Dependent Changes in Formants Based on Normalization of Vowel Triangle. *In Proc. 23rd International Conference RADIOELEKTRONIKA. Pardubice*. Czech Republic, 2013, pp. 337-341.
- [8] SIGMUND, M., ZELINKA, P., *Analysis of Voiced Speech Excitation Due to Alcohol Intoxication*. *Information Technology and Control*, 2011, vol. 40, pp. 145-150.
- [9] ALLAM M, *Speech Segmentation in Synthesized Speech Morphing Using Pitch Shifting*, The International Arab Journal of Information Technology, 2011, vol. 8, no. 2, p221-226.