

## دراسة تغيرات ترددات رنين الإشارة الصوتية باختلاف معمارية الجهاز الصوتي للمتحدث

د. جعفر الخير\*

(تاريخ الإيداع 21 / 2 / 2017. قُبل للنشر في 15 / 6 / 2017)

### □ ملخص □

تمّ في هذا البحث دراسة تغيرات بعض خصائص الإشارة الصوتية تبعاً لعامل البنية التشريحية للجهاز الصوتي للمتحدث، إذ تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 57 متحدث بالغ تتراوح أعمارهم بين 35 و 45 عام جميعهم ذكور، جميع المتحدثين من خلفيات اجتماعية وثقافية متقاربة، ولا يعانون من أي عيوب نطقية أو سمعية. سُجّلت الملفات الصوتية المُضمنة في قاعدة البيانات ضمن شروط تسجيل مثالية، حيث تمّ التسجيل في بيئة معزولة صوتياً وبحضور تقني صوت وخبير صوتيات، واستغرق التسجيل لكل متحدث حوالي خمس دقائق لفظ فيها الحروف الصوتية العربية (الألف والواو والياء) من خلال لفظه لكلمة "سألتمونيها" ثلاث مرات متتالية. اعتماداً على تحليل التسجيلات الصوتية الناتجة تمّ رصد التغيرات الحاصلة في قيم ترددات توافقيات المجرى الصوتي الممتد من الحبال الصوتية إلى الشفاه والمسامة النغمات (Formants)، حيث أن الأبعاد والشكل الفسيولوجي للجهاز الصوتي المتعلق بالاصطناع العام للجسم يعتبر من أهم المعلومات التي يمكن التنبؤ بها من تحليل الإشارة الصوتية. لقد أظهرت نتائج هذا البحث وجود تناسباً عكسياً بين قيم ترددات النغمات الثلاثة الأولى  $F_1$ ,  $F_2$ ,  $F_3$  وبين طول المجرى الصوتي، بينما لا توجد علاقة واضحة بين ترددات النغمتين الرابعة والخامسة  $F_4$ ,  $F_5$  وبين طول المجرى الصوتي (Vocal Tract).

**الكلمات المفتاحية:** قاعدة البيانات الصوتية، ترددات النغمات، المتحدث، المجرى الصوتي.

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

## Analysis Study of Formant Frequencies Changes According to the Speaker's Vocal Tract Shape

Dr. Jaffar Alkhier\*

(Received 21 / 2 / 2017. Accepted 15 / 6 / 2017)

### □ ABSTRACT □

In this research, some of audio signal properties have been studied according to the speaker's vocal tract shape. A database of audio files has been recorded. These files belong to 57 men whose age between 35 and 45. All speakers came from the same academic and social culture. Furthermore, they don't suffer from any problems in hearings and utterance.

The vowel database was created in perfect recording conditions. The spent time needed for recording process was about five minutes for each speaker who said the Arabic word "سألتمونيها" three times. That word is very rich of vowel letters. It composes of the whole Arabic long vowel.

Based on the analysis study of the recorded audio signals, the relationship between the formant frequencies and the length of speaker's vocal tract has been studied. The results show an inverse proportion for the first three frequencies F1, f2, F3 and no clear relationship for the two other frequencies F4, F5.

**Key words:** Vowel database, formant frequencies, speaker, vocal tract.

---

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

## مقدمة:

يظهر الصوت نتيجة حدوث تموجات في الهواء ولا ينقل إلا 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 والحنك 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 وتوليد 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 ويطرق هندسية ورياضية السمات الأساسية للأصوات والتي على أساسها يتم تمييز الكلام المنطوق، ومن أمثلة هذه السمات: سعة الموجة ودرجة شدتها [1,2].

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 [3,4].

## أهمية البحث وأهدافه:

يعمل هذا البحث على إنجاز قاعدة بيانات تحوي ملفات صوتية لمتحدثين متنوعين من حيث الاصطناع العام للشكل الفسيولوجي للجهاز الصوتي وخصوصاً المجرى الصوتي، حيث متطوعي قاعدة البيانات الصوتية المنجزة هم متحدثين متباينين طول القامة وطول المجرى الصوتي. تشكل قواعد البيانات الصوتية عنصراً أساسياً لبناء نظم حاسوبية مختلفة كنظم التعرف الآلي على الكلام والنطق الآلي والتعرف على المتحدث والتعرف على اللغات واللهجات والجنس والعمر والحالات النفسية الطويلة الأمد والقصيرة الأمد. يمكن استثمار قاعدة البيانات الصوتية المنجزة في تطوير أدوات مساعدة لدراسة علاقة معمارية الجهاز الصوتي بالبنية الفسيولوجية للمتحدث.

يهدف هذا البحث إلى إستخلاص العلاقة بين معمارية الجهاز الصوتي والبنية الفسيولوجية للمتحدثين عن طريق ملاحظة وتتبع التغيرات في قيم ترددات رنين الإشارة الصوتية (Formants Frequencies) والتي تعتبر من أهم سمات الإشارة الصوتية للمتحدث والمستخلصة وفق خوارزمية (Mel Frequency Cepstral Coefficients) (MFCC). باعتماد تحليل (MFCC) سنحصل على قيم ترددات النغمات الخمس الأولى  $F_1, F_2, F_3, F_4, F_5$  وبملاحظة اختلاف هذه القيم باختلاف طول المجرى الصوتي للأشخاص المشكلين لقاعدة البيانات الصوتية المنجزة ستظهر العلاقة بين قيم هذه السمات وبين طول المجرى الصوتي للمتحدث.

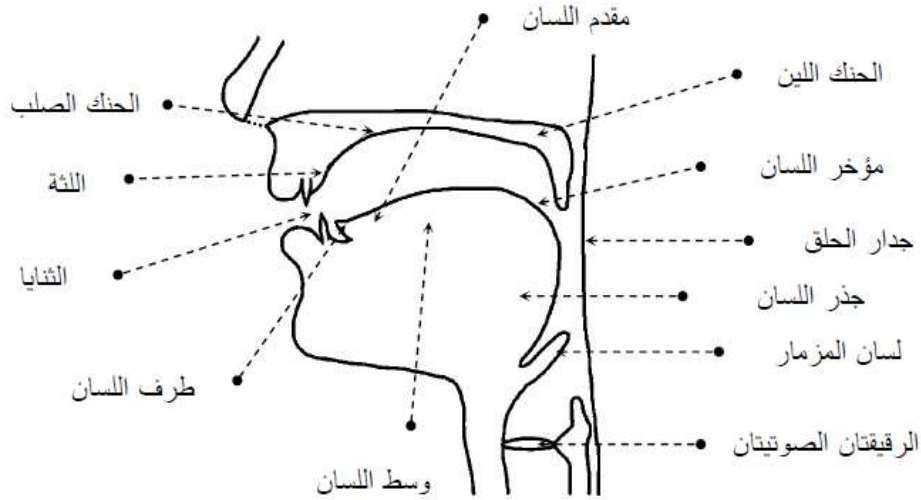
## طرائق البحث وموارده [5,6,7]:

تضمن البحث استخلاص سمات النغمات من الإشارة الصوتية للمتحدثين وحساب مُعامل التردد الموافق لهذه النغمات المُستخلصة وفق خوارزمية (Mel Frequency Cepstral Coefficients) (MFCC). وبمعنى آخر فقد تمّ تسجيل كلمة غنية صوتياً من قبل متحدثين متنوعي طول المجرى الصوتي وهي كلمة "سألتمونيها" ثلاث مرات متتالية، وحساب مُعاملات ترددات النغمات الخمسة الأولى للحروف الصوتية العربية الثلاث الموجودة ضمن كلمة

"سألتُمُونِيهَا" وهي الألف والواو والياء من أجل مقارنة تغير هذه القيم بتغير طول المجرى الصوتي للمتحدث (Vocal Tract). هذه الترددات موافقة لقمم الاستجابة الترددية لطيف المجرى الصوتي، وذلك لتقدير ترددات توافقيات المجرى الصوتي المعروفة بالنغمات (Formants).

### الجهاز الصوتي [8,9]:

يدعى الجهاز المسؤول عن توليد الأصوات اللغوية عند الإنسان بالمجرى الصوتي (Vocal Tract) والذي تخرج منه جميع الأصوات اللغوية، ويتكون كما في الشكل (1) من الأعضاء الواقعة بين الرقيقتين الصوتيتين من جهة وبين فتحتي الأنف والشفيتين من الجهة الأخرى، بما في ذلك الرقيقتين الصوتيتين والشفيتين. . يحتاج إخراج الأصوات من الجهاز الصوتي إلى مصدر للطاقة وبدونه لا يمكن أن يُصدر أصواتاً، فهو يشبه المذياع دون طاقة كهربائية لا يمكن أن يعمل. وبالتالي يعتبر الجهاز الصوتي محول للطاقة من نوع إلى آخر. بمعنى آخر يحول الطاقة الهوائية إلى طاقة فيزيائية مصدراً أصواتاً متباينة الخصائص من حيث التردد والشدة. إن المصدر الذي يزود الجهاز الصوتي بالطاقة هو الجهاز التنفسي الذي يتكون من القفص الصدري بما فيه من رئتين وقصبة هوائية. والجهاز التنفسي هنا يمتد من الرقيقتين الصوتيتين إلى الحويصلات الهوائية داخل الرئتين، فالهواء داخل الجهاز التنفسي هو مخزون الطاقة للجهاز الصوتي.



الشكل (1) أعضاء النطق في الجهاز الصوتي

عند انتفاخ الرئتين أثناء عملية الشهيق يزيد حجم الهواء داخلهما فيستخدمه الجهاز الصوتي أثناء الزفير في إصدار الأصوات اللغوية. رغم أن الجهاز الصوتي يمكن أن يصدر أصواتاً أثناء الشهيق إلا أنه لا تُعرف لغة بشرية تستخدم هذه الطريقة في إخراج أصواتها.

يعتبر الجهاز التنفسي المصدر الأساس لمعظم الأصوات اللغوية لكافة اللغات البشرية، كما أنه المصدر الوحيد لجميع أصوات اللغة العربية. تسمى الأصوات التي تعتمد على الجهاز التنفسي بالأصوات الرئوية (Pulmonic) وهناك لغات تستخدم مصادر أخرى للطاقة غير الرئتين في إخراج بعض أصواتها.

تعتبر لغة الهوسية، على سبيل المثال، اللغة الرئيسية في شمال نيجيريا، كما ويوجد لغات أفريقية وهندية أمريكية أخرى تستخدم الحنجرة بمعزل عن الرنين في التحكم في الهواء الداخل والخارج عبر الجهاز الصوتي، وذلك بغلق الرقيقتين الصوتيتين، ومن ثم عزل الهواء داخل الرنين عن الهواء فوق الرقيقتين الصوتيتين، ثم رفع أو خفض الحنجرة الحاوية على الرقيقتين الصوتيتين، فتعمل الحنجرة عمل المكبس، وتسمى هذه الوسيلة بالحنجرية (Glottalic). فعند رفع الحنجرة يندفع الهواء إلى خارج الجهاز الصوتي، وبالتالي يمكن التحكم فيه لإصدار أصوات قذفية (Ejectives)، وعند خفض الحنجرة يمكن إصدار أصوات هي أصوات انفجارية داخلية (Implosives).

تستخدم لغات أخرى مثل لغة الزولو في أفريقيا مصدراً آخر للطاقة يسمى طبقي (Velaric) نسبة إلى انطباق مؤخر اللسان على الحنك اللين، إذ يُدفع اللسان بأكمله إلى أعلى لينطبق على الحنك ثم يُخفض وسط اللسان أو مقدمه أو كلاهما معاً، ويبقى على مؤخرة اللسان ملامسة الحنك اللين مما يؤدي إلى خفض ضغط الهواء داخل الفم، فيصدر عن ذلك أصوات تسمى طقطقة (Clicks).

### قاعدة البيانات الصوتية:

تشكل قواعد بيانات الكلام (Speech Database) الركن الرئيسي في بناء النظم الحاسوبية، حيث تشكل البنية التحتية لبناء نظم التخاطب مع الحاسب. تتكون قواعد بيانات الكلام عادة من ملفات صوتية (Wave Files) 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 [10].

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 العلاقة بين قيم ترددات النغمات للإشارة الصوتية وبين طول المجرى الصوتي خصوصاً وطول المتحدث عموماً فقد تم تضمين قاعدة البيانات الصوتية تسجيلات صوتية متنوعة لأشخاص مختلفي طول القامة ومن ثم تحليل التسجيلات المختلفة للمتحدثين.

### مراحل إعداد قاعدة البيانات الصوتية:

يُسجل مهندس الصوت كلمة "سألتمونيها" ثلاث مرات متتالية، والتي تشمل مختلف الحروف الصوتية العربية (الألف والواو والياء).

- يُجزئ النظام الكلمة إلى مقاطع صوتية بزمان 20 ميلي ثانية.
- يُحلل النظام المقاطع الصوتية ويحدد طبقتها وتوقيتها وعلوها والتي تقاس بالديسيبل  $loudness = 3 \text{ dB}$ .
- يُفَرَن برنامج تعرف الكلام كل مقطع صوتي مسجلاً بالتمثيل النصي الموافق.
- يُخزن الحاسوب المقاطع الصوتية رقمياً مبيناً حدودها، أي النقطة التي يبدأ عندها المقطع وتلك التي ينتهي عندها، ويصنفها جاهزة للاستعمال في قاعدة بيانات.

يبنى الباحثون اعتماداً على قاعدة بيانات المتحدثين نموذجاً إحصائياً يفيد في استنتاج الخواص العامة التي تحكم ارتفاع طبقات الصوت أو انخفاضها إضافة إلى توقيت وعلو الكلام. يستفيد النظام لاحقاً من هذه الخواص ليُجعل مسمع كلام النظام يبدو قريباً من كلام الإنسان. وفي النهاية يتم تخزين بيانات كل مقطع صوتي في قاعدة البيانات وتمييزه بشعاع مميز (رقم فريد) عن الآخر حيث يسمى بالوحدة (Unit) وتتضمن معلومات حول (prosodic/phonetic) الخاصة بالوحدة.

**المتحدثون:**

تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 57 رجل. جميع المتحدثين من شرائح عمرية متقاربة حيث تتراوح أعمارهم بين 35-45 عام، وهم أشخاص أكاديميين ومن خلفيات ثقافية وحاصلين على شهادات جامعية ومن السكان الأصليين لمركز مدينة اللاذقية السورية. جميعهم ذكور ويعيشون في مناطق متجاورة، أي يتماثل الجميع بعامل العمر واللهجة والجنس وبالتالي لن يكون لهذه العوامل تأثير على القيم الحاصلة بنتيجة البحث، ويتميزون فقط في المواصفات الجسمية، فهم أشخاص متنوعي طول القامة مما يعني تنوع في طول المجرى الصوتي.

**المادة اللغوية:**

تختلف المواد اللغوية التي يقوم المتحدثون بتسجيلها باختلاف الغاية من بناء قاعدة البيانات الصوتية، ونظراً لأنّ حوالي 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 [11]، فقد تمّ انقضاء المقاطع الصوتية من المقابلة التي أجريت لكل متحدث، وهي مقاطع غنية صوتياً تشمل جميع الأحرف الصوتية الطويلة أي الألف والواو والياء. المقابلات أجريت مع كل متحدث على حدا، بحيث تمّ 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 كان الاستوديو خالياً من الشبائيك التي قد تسمح بدخول الضوضاء، وباب الاستوديو عازل للصوت أيضاً.

يتمّ التسجيل لكل متحدث على حدا، حيث تُجرى مقابلة مع كل متحدث من المتحدثين السبعة والخمسين، يلفظ كل متحدث كلمة "سألتُمُونيها" ثلاث مرات متتالية، بشروط عادية من السرعة وعلو الصوت، وبالاستعانة بمجموعة من الأدوات كالميكروفون الديناميكي نوع (Yamaha)، والقناة الصوتية (Stereo Channel)، وبرنامج التسجيل (Adobe Audition 3). تمّ الانتهاء من جلسة التسجيل لجميع المتحدثين بتردد أخذ عينات 44.1 kHz خلال خمس ساعات.

**المراحل العملية للبحث:**

يتضمن البحث عدة مراحل عملية يمكن تلخيصها على الشكل التالي:  
في المرحلة الأولى، تم دراسة تصنيف الحروف اللغوية السورية إلى صوتية وصامتة، وتحديد خصائص الحروف الصوتية وجمع الخوارزميات المناسبة لمعالجة الخصائص المقترحة، ثم بناء قاعدة بيانات تتضمن معلومات شخصية عن أشخاص النظام.  
في المرحلة الثانية، تم اختيار المشتركين بناءً على معايير جغرافية واجتماعية، فأخذت العينات من مناطق مجاورة لتقييد تأثير عامل اللهجة على دقة النتائج. ركز البحث على تنوع الأشخاص من ناحية طول القامة مما يعني تنوع في الاصطناع العام للجسم وبالتالي تنوع في طول المجرى الصوتي للمتحدثين. كما تم تقسيم الأشخاص إلى مجموعات وفق الجدول 1 بحيث كل مجموعة لها مجال طول معين:

جدول (1): تصنيف المتحدثين حسب الطول

المجموعة A	المجموعة B	المجموعة C	المجموعة D	
160 – 168 سم	169 – 176 سم	177 – 184 سم	185 – 190 سم	طول القامة
13 شخص	16 شخص	18 شخص	10 أشخاص	عدد العينات

تعتمد تطبيقات البيانات الصوتية في عملها على مبدئين مهمين:

المبدأ الأول: أن لكل إنسان جهازاً صوتياً مميزاً لا يشابهه فيه أحد، ويقصد بالجهاز الصوتي هنا جميع الأعضاء التي تساهم بإخراج الأصوات عند الإنسان كالقفص الصدري والحنجرة واللسان والأنف والشفيتين والأسنان من حيث شكل وحجم الأعضاء وارتباطها ببعضها البعض.

المبدأ الثاني: أن لكل إنسان نظاماً عصبياً فريداً يتحكم في جهازه الصوتي؛ والنظام العصبي هنا يختص بمكونات الجهاز العصبي التي تتعلق بالتحكم في عضلات أعضاء الجهاز الصوتي، والسبب في ذلك أن لكل إنسان نشأة خاصة به ينفرد بها عن الآخرين في اكتساب اللغة وتشكيل شخصيته مما يكون له طريقته المميزة في إنتاج الكلام.

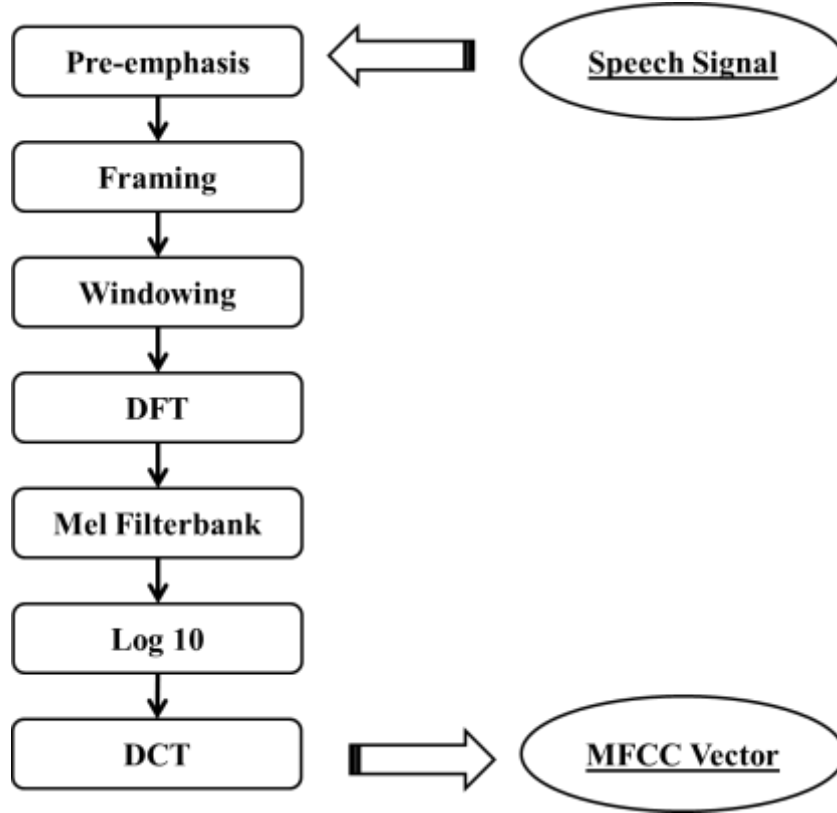
وعليه يمكن القول إن العينات المأخوذة لبناء قاعدة البيانات الصوتية ملائمة لهدف المشروع.

في المرحلة التالية، تم تطبيق خوارزمية الـ (MFCC (Mel Frequency Cepstral Coefficients) على التسجيلات من أجل استخراج السمات، وهي خوارزمية جاهزة برمجياً في بيئة المحاكى (MathWorks)، حيث يتم تغيير بارامتراتهما بما يتناسب مع عدد العينات وتردد التقطيع المستخدم في هذا البحث.

تعتبر خوارزمية (MFCC) من الطرق السائدة المستخدمة في استخراج السمات وذلك بسبب حساسية مرشحاتها لخواص إشارة الصوت البشرية، وتستخدم معاملات (MFCC) بشكل كبير في مجال التعرف على الكلام، حيث قُدمت معاملات تردد الميل في عام 1980 ومازالت رائدة في هذا المجال منذ ذلك الوقت.

تُشرح الأصوات المولدة من قبل الإنسان حسب شكل المسلك الصوتي (Vocal Tract)، وبالتالي إذا أمكن تحديد شكل المسلك الصوتي بدقة فإنه يمكن تحديد الصوت (Phoneme) الذي يتم إنتاجه. يتجلى شكل المسلك الصوتي في غلاف طيف طاقة الزمن القصير (Short Time Power Spectrum) وبالتالي تكون الغاية من استخدام الخوارزمية (MFCC) هو تمثيل هذا الغلاف بشكل دقيق.

تعتمد الخوارزمية (MFCC) على التغيرات في عرض الحزمة الترددية للأذن البشرية، والتي تستخدم من أجل النقاط الصفات الرئيسية للكلام، حيث تمتلك (MFCC) تباعداً خطياً على الترددات الأقل من 1000 هرتز وتباعداً لوغاريتمياً على الترددات الأعلى من 1000 هرتز، ويبين الشكل (5) خطوات عمل الخوارزمية (MFCC).



الشكل (2) خطوات عمل الخوارزمية (MFCC)

إن تطبيق خوارزمية الـ MFCC يمكننا من الحصول على نغمات متضمنة العديد من البارامترات (Formant Parameters) مثل عرض الحزمة (Bandwidth) والميل (Slope) والتردد (Frequency)، وبالنتيجة نحصل على قيم الترددات  $F1, F2, F3, F4, F5$  لكافة الحروف الصوتية المضمنة في الكلمة الملفوظة لجميع المتحدثين. تستخدم خوارزمية الـ MFCC مرشح تمرير عالي وذلك من أجل تعويض جزء التردد العالي الذي تم فقدته أثناء آلية إنتاج الكلام (زيادة الطاقة النسبية للطيف عالي التردد) ومن ثم يتم تقطيع الإشارة الصوتية إلى عدد من الإطارات (frames)، زمن كل إطار من 20 إلى 40 ميلي ثانية (هنا تم تحديد زمن الإطار بـ 20 ميلي ثانية) مع وجود تداخل اختياري يساوي إلى نصف أو ثلث حجم الإطار وذلك من أجل تسهيل الانتقال من إطار إلى آخر، ليتم بعدها إخضاع كل إطار لعملية النوفذة (windowing) باستخدام نافذة هامينغ (Hamming window) وذلك من أجل القضاء على الانقطاعات عند الحواف. 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. تعمل خوارزمية الـ MFCC على ترشيح طيف الإشارة الصوتية (short time power spectrum) عن طريق مجموعة من المرشحات المثلثية (Mel filter bank) (التي صممت كمحاكاة لمرشحات تمرير الحزمة التي تظهر في النظام السمع) المتباعدة بانتظام وفقاً لمقياس ميل الترددي (Mel frequency scale). يتم بعد ذلك حساب اللوغاريتم لطيف مجال ميل (Mel scale spectrum)، ومن ثم يستخدم تحويل جيب تمام المتقطع DCT لإعادة تحويل طيف مجال ميل اللوغاريتمي إلى مجال الزمن ونتيجة هذا التحويل هو الحصول على شعاع MFCC.

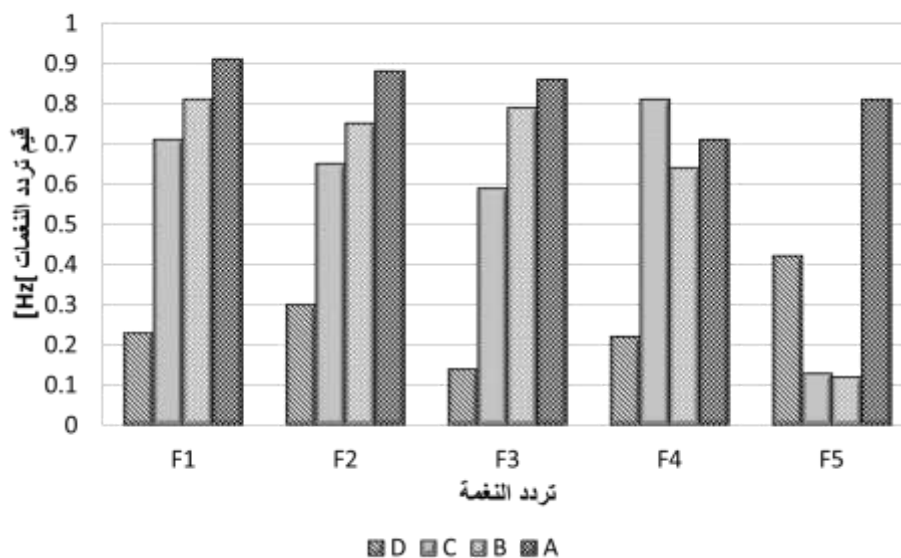


المرحلة الأخيرة هي مرحلة تقييم ومناقشة النتائج التي تم الحصول عليها واستخلاص العلاقة التي تربط بين الترددات الخمسة الأولى لنغمات الإشارات الصوتية المسجلة وطول المجرى الصوتي للمتحدثين.

### النتائج والمناقشة

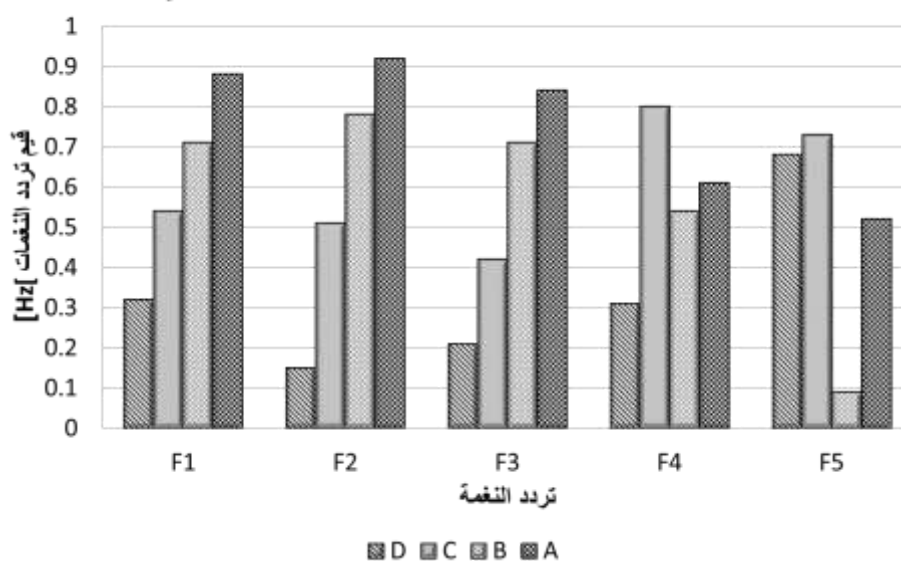
تم تطبيق خوارزمية الـ MFCC على الإشارات الصوتية المسجلة من أجل الحصول على قيم الترددات الخمسة لكل حرف صوتي والتي تتراوح قيمها ضمن المجال  $[0.2, 1]$  هرتز، حيث توضح المخططات التالية العلاقة بين قيم الترددات الخمسة الأولى للإشارات الصوتية للمتحدثين وطول المجرى الصوتي عند نطق كل حرف صوتي:

العلاقة بين تردد النغمات وعامل الطول عند نطق الحرف الصوتي الألف



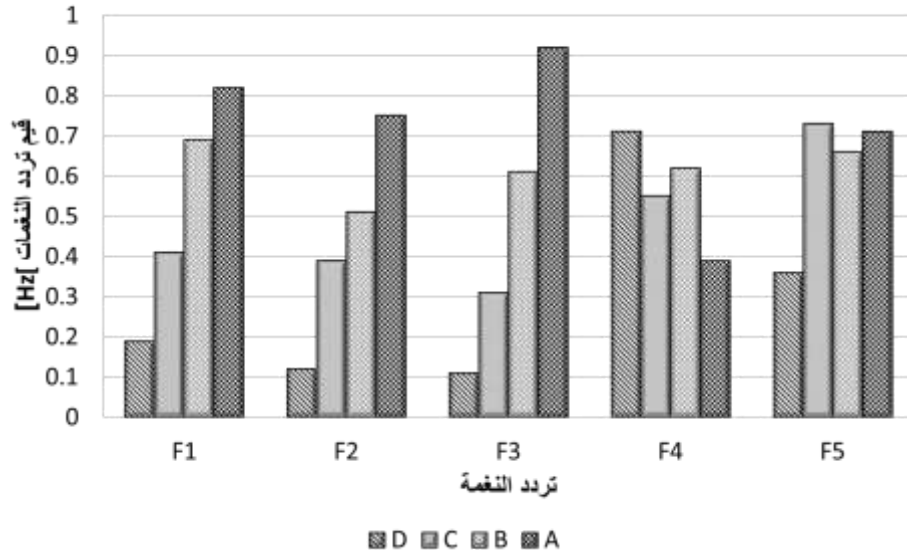
الشكل (3) العلاقة بين تردد النغمة وعامل الطول عند نطق الحرف الصوتي الألف

العلاقة بين تردد النغمات وعامل الطول عند نطق الحرف الصوتي الواو



الشكل (4) العلاقة بين تردد النغمة وعامل الطول عند نطق الحرف الصوتي الواو

العلاقة بين تردد النغمات وعامل الطول عند نطق الحرف الصوتي الياء



الشكل (5) العلاقة بين تردد النغمة وعامل الطول عند نطق الحرف الصوتي الياء

نلاحظ من المخططات البيانية الثلاثة الموضحة لقيم نغمات التردد الخمس الأولى عند نطق الحروف الصوتية العربية الثلاثة، أن هذه القيم منتظمة في تباينها بالنسبة للنغمات الثلاث الأولى F1, F2, F3 وغير منتظمة بالنسبة للنغمتين الرابعة والخامسة F4, F5. بمعنى آخر لدينا أربعة مجموعات من المتحدثين A, B, C, D كل مجموعة تمثل الأشخاص الذين يتراوح طولهم ضمن مجال معين، بحيث تكون المجموعة A هي الأقصر طولاً والمجموعة D تمثل المتحدثين الأكثر طولاً. نلاحظ أنه بالنسبة لجميع المتحدثين وعند نطق جميع الحروف الصوتية تكون قيم ترددات النغمات الثلاث الأولى F1, F2, F3 متناسبة عكساً مع طول الشخص وبالتالي مع طول المجرى الصوتي له. الأشخاص من المجموعة A هم الأقصر طولاً لكن قيم ترددات نغماتهم الثلاث الأولى تكون الأعلى وهكذا بالنسبة للمجموعة B حيث متحدثيها أطول من متحدثي المجموعة A لكن قيم ترددات نغماتهم أخفض. أما بالنسبة للنغمتين الرابعة والخامسة فلا نجد أن هناك علاقة منتظمة تربط طول المجرى الصوتي بقيمة تردد النغمة، فتارةً يكون التناسب طردي وطوراً يكون التناسب عكسي؛ لهذا يمكن القول أن العلاقة ضعيفة بين قيم ترددات النغمات F4, F5 وبين طول المجرى الصوتي والعلاقة عكسية بالنسبة للنغمات الثلاث الأولى F1, F2, F3.

### الاستنتاجات والتوصيات:

تضمنت هذه الدراسة إيجاد العلاقة التي تربط بين بعض خصائص الإشارة الصوتية وعامل البنية الفسيولوجية للجهاز الصوتي للمتحدث، إذ تمّ بناء قاعدة بيانات تضم ملفات صوتية تعود إلى 57 شخص سوري، تتراوح أعمارهم بين 35 و 45 عام، جميعهم ذكور ويعيشون في مناطق ريفية متجاورة، أي يتماثل الجميع بعامل العمر واللهجة والجنس، ويتميزون فقط في طول القامة وبالتالي طول المجرى الصوتي. استغرقت عملية تسجيل الملفات الصوتية خمس ساعات داخل أستوديو معزول صوتياً، حيث لفظ جميع المتحدثين كلمة "سألتمونيها" ثلاث مرات متتالية، وهي كلمة غنية صوتياً تشمل جميع الأحرف الصوتية الطويلة أي الألف والواو والياء الموجودة في اللغة العربية.

اعتماداً على التسجيلات الناتجة تم إجراء تحليل (MFCC (Mel Frequency Cepstral Coefficients) على التسجيلات من أجل استخراج السمات، للحصول على نغمات متضمنة العديد من البارامترات (Formant Parameters) بهم منها التردد (Frequency)، وبالنتيجة نحصل على قيم الترددات F1, F2, F3, F4, F5 لكافة الحروف الصوتية المنطوقة من قبل جميع المتحدثين. وقد أظهرت نتائج هذا البحث أنّ هناك تناسباً عكسياً بين قيم ترددات النغمات الثلاثة الأولى F1, F2, F3 وبين طول المجرى الصوتي، بينما العلاقة ضعيفة بين ترددات النغمتين الرابعة والخامسة F4, F5 وبين طول المجرى الصوتي (Vocal Tract) وبالتالي يمكن القول أن تردد الرنين الثالث F3 هو التردد الحرج الذي يمكن اعتماده لتوصيف العلاقة السابقة.

تجدر الإشارة أنه سيكون من المفيد جداً دراسة تأثير الاضطراب الفسيولوجي لباقي أجزاء جهاز النطق على ترددات رنين الإشارة الصوتية للمتحدث ودراسة تأثيرها كذلك على باقي سمات الإشارة الصوتية بالإضافة إلى استخلاص العلاقة الرياضية التي تصف هذا التأثير.

### المراجع:

- [1] M. E. Ayadi , M. S. Kamel , F. Karray, “Survey on Speech Emotion Recognition: Features, Classification Schemes, And Databases”, Pattern Recognition 44, PP.572-587, 2011.
- [2] I. Chiriacescu , “Automatic Emotion Analysis Based On Speech” , M.Sc. THESIS Delft University of Technology, 2009.
- [3] Nitin Thapliyal , Gargi Amoli “Speech based Emotion Recognition with Gaussian Mixture Model” international Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 5, July 2012.
- [4] T.Vogt, E. Andre and J. Wagner, “Automatic Recognition of Emotions from Speech: A review of the literature and Recommendations for practical realization”, LNCS 4868, PP.75-91, 2008.
- [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] Muller, E. M.; and Brown, W. , Variation in the Supraglottal Air Pressure Waveform and Their Articulatory Interpretation, *in Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass, (Academic, New York), Vol. 4., 2012.
- [9] Al-Mozainy, Hamza Q. Vowel Alternations in a Beduin Hijazi Arabic Dialect: Abstractness and Stress. *The University of Texas at Austin*, 2013.
- [10] P.Shen, Z. Changjun, X. Chen, “Automatic Speech Emotion Recognition Using Support Vector Machine”, International Conference On Electronic And Mechanical Engineering And Information Technology, 2011.
- [11] E. Bozkurt, E, Erzin, C. E. Erdem, A. Tanju Erdem, “Formant Position Based Weighted Spectral Features for Emotion Recognition”, Science Direct Speech Communication, 2011.