بحوث مستقبلية (٣٦) ٢٠١١
ص ص [١٥٥-١٧٦]

تمييز جنس المتكلم باعتماد العزوم السبعة لإشارة الكلام

د.خليل إبراهيم السيف [١]
ميسون خضر النعيمي [٢]

الملخص

تم في هذا البحث اعتماد أسلوب جديد لأجل تمثيل إشارة الكلام من خلال مصفوفة ثنائية وذلك لجعلها ملائمة لأجل حساب العزوم السبعة لتلك الإشارة لكون العزوم السبعة ذات ثباتية عند التدوير والقياس والتزحيف والتي اعتمدت في كثير من مجالات التمييز.

تم اعتماد نماذج كثيرة لأصوات من كلا الجنسين وتم حساب حد عتبة للعزوم السبعة لفئة الذكور وحد عتبة لفئة الإناث بعد اختزال الترددات العالية من إشارة الصوت باستخدام تحويل فورير يتم بعدها إيجاد مدى التقارب بين النماذج الداخلة وحد العتبة لكلا الجنسين. إذ تحتوي الخوارزمية على طورين الأول يعمل على بناء حدود العتبة والطور الآخر يعمل على اختيار جنس متكلم وذلك لحساب مدى التقارب بين عناصر صوت المتكلم وحد العتبة الذي تم الحصول عليه من الطور الأول.

أعطت الخوارزمية المعتمدة نتائج صحيحة ذات مدى ٨٠٪ عند اختبار عينات لأصوات من فئة الذكور و٨٥٪ عند اختبار عينات لأصوات من فئة الإناث.

Abstract

In this research, a proposed approach has been adopted to represent the speech signal in two-dimensional arrays to make it suitable to evaluate the seven moments because the seven moments have stability in rotation, scaling and translation, which was adapted in many recognition fields.

Many samples have been adopted for both male and female and a threshold value for both genders of the seven moments have

(١) استاذ مساعد، كلية علوم الحاسبات والرياضيات، جامعة الموصل.
(٢) مدرس مساعد، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

الاستلام ٢٠١٠/٠٦/١٣
قبول النشر ٢٠١١/٠٥/١٦

been calculated after truncating the high frequencies from the speech signal by applying Fourier transformation, Then finding the correlation between the input samples and the threshold of both genders. The proposed algorithm contains two phases, in the first one the threshold values were built and the other phase works to test the gender of spooking person by evaluating the correlated distance between the speech parameters of the spooking person and the threshold values which were got in the first phase.

The applied algorithm has given correct result up to 80% for male samples and 85% for female ones.

١- المقدمة:

إن تمييز الكلام آلياً تقنية واعدة تؤكد على استعمال السمع والكلام للتحاور من خلال واجهة اللغة الطبيعية مع الأجهزة والمعدات (الشبيهة بالبشر من ناحية الإدراك) أصبح حقيقة وعليه أصبحت أنظمة تمييز الكلام بواسطة الآلة هي واحدة من أكثر الطرق كفاءة لتفاعل الإنسان مع الآلة (الوكيل، ٢٠٠٦).

إن اللغات الطبيعية للإنسان تحتوي على صفات مميزة كثيرة منها إمكانية اعتمادها لتصنيف المتكلمين من حيث الفئة العمرية أو من حيث بيان جنس المتكلم، لذا نشط كثير من الباحثين في هذه الاتجاهات لأجل التمييز الآلي لكثير من الأشياء باعتماد صوت المتكلم.

٢- الصوت:

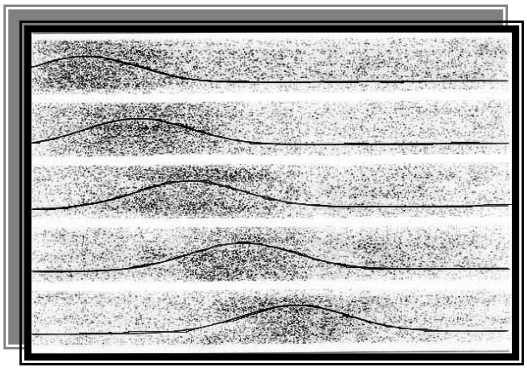
الصوت عبارة عن تذبذب في ضغط الهواء أو الوسط الناقل سبّبه جسم مهتز (Elseom-Cook, 2001)، وإن كان الصوت كلاماً فإن الجسم المهتز الذي سبب

الكلام هو الأوتار الصوتية للإنسان (Saeed, 2002)، أما أصوات الموسيقى فسببها الاهتزاز الحاصل في الأدوات الموسيقية (McEnary, 2001) (Minasi, 1996).

يشار عادة 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 (20Hz-20KHz)، ومع تقدم العمر يقل هذا المدى (McEnary, 2001) فقد يصل إلى (20Hz-17KHz) (Elseom-Cook, 2001) (Kivimaki, 2000).

وتجدر الإشارة إلى أن الترددات التي تكون أقل من هذا المدى تدعى بـ (Infrasound)، والتي تكون أعلى منه تدعى بـ (Ultrasound) (Kivimaki, 2000).

ينتقل الصوت في الوسط على شكل موجات (Waves)، إذ تحدث إزاحات في جزيئات الوسط بعيدة نسبيا عن مواقع التعادل، هذه الإزاحات تحدث تخلخل وتضاغط في جزيئات الوسط والذي بدوره يؤدي إلى انتقال الطاقة من مصدر الصوت الأصلي إلى جزيئات الوسط المجاورة لها، وهكذا بشكل دوري حتى تصل إلى جهة المستلم (Kientzle, 1997)، وكما هو موضح في الشكل (١):



شكل (١) حركة موجات الصوت

ففي الشكل السابق تشير المناطق المظللة إلى أن الجزيئات المتضاغطة مع بعضها (تشكل مساحة ذات ضغط عالٍ) والمناطق ذات اللون الفاتح تشير إلى الجزيئات المتناثرة نسبياً، فعندما يحصل تضاغط بين الجزيئات تتكون قوة جزئية تؤدي إلى إنضغاط جزيئات الوسط المجاورة، وهذه بدورها تنقل قوتها الجزئية إلى مجاوراتها (عبد القادر، ٢٠٠١).

٣- الصوت والكلام:

يمكن إعطاء تعريف لغوي للكلام على أنه صوت مشتمل على الأحرف الهجائية التي تبدأ بالألف وتنتهي بالياء (عبد الحميد، ١٩٩٦)، وهذا بالنسبة إلى اللغة العربية إذ تختلف الأحرف الهجائية باختلاف اللغات المستخدمة.

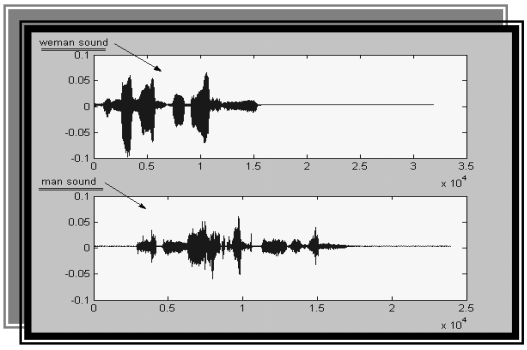
ويوصف الكلام (من الناحية العلمية) على أنه يمثل الصوت الناتج من استخدام الأوتار الصوتية لإنشاء تمثيل لغوي ينقل المعلومات من المستهل بالحديث إلى المستقبل للحديث (Brooks, 2004)، أما الهدف منه فتتركز في نقل المعنى بين الأشخاص أو لزيادة الرابطة الاجتماعية بين الأشخاص أو الجماعات، ولهذا فإن المتحدثين عادة يستخدمون لغة واحدة بمجموعة محددة وعامة من الكلمات لكي يتم نقل المعلومات بينهم بشكل صحيح (Abdul-Majeed, 2000).

أما من ناحية الحاسوب فإن الصوت والكلام كلاهما لديه خواص ، فالمعالجة الأولية التي تجرى على الصوت ذاتها تجرى على الكلام، من تقطيع (تعيين Sampling) وتكميم وترميز وخزن، إلا أن بعض برامج معالجة الصوت تختلف عنها بالكلام مثال ذلك الكبس، تعود هذه الاختلافات إلى أهمية الاحتفاظ ببعض خواص الكلام من وضوح ونبرة ومفهوم إذ يجب أن يبقى الكلام مفهوماً ومعبرا عن شخصية قائله حتى بعد أن تم إجراء الكبس وفتح الكبس عليه (Brooks, 2004).

ويجب التأكيد على أن مفهوم الكلام يتضمن إلى جانب المحتويات الواقعية الحقيقية لإشارة الكلام فهناك أيضاً معلومات عن شخصية المتكلم وهويته (من الذي يتكلم)، الانفعال، التنغيم، فضلاً عن مضمون الرسالة المنقولة بالكلام، كل هذه

العوامل مهمة بالنسبة إلى مفهوم الكلام (Abdul-Majeed, 2000) (Brooks, 2004).

في الشكل (٢) يمكن ملاحظة صورة الموجة الصوتية عند النطق بعبارة (بسم الله الرحمن الرحيم) بصوت أنثى مرة وبصوت رجل مرة أخرى، يمكن ملاحظة الفرق بين الموجتين على االرغم من تشابه العبارة.



شكل (٢) يبين موجة الصوت عند لفظ

غالباً تعتمد النسب (8KHz, 11.25KHz, 22.05KHz, 44.1KHz) في تعيين الكلام أثناء التسجيل والتي تمثل عدد العينات المأخوذة في الثانية الواحدة، ويعتمد اختيار أحد هذه القيم على المعالجة التي تتم على إشارة الكلام (حليم، ٢٠٠٣).

٤- استخلاص خواص الصوت:

يتطلب فهم الصوت فهم شيء عن الإشارات (Signals) ومنها المصطلحات الآتية: (Minasi, 1996)

- **التردد Frequency**: هو عدد المرات التي تعيد فيها موجة الصوت نفسها في فترة محددة أو هو مقياس عدد التأرجحات الكاملة لموجة الصوت والتي تحدث لكل ثانية، ويقاس التردد عادة بالهيرتز (Hz: Hertz).

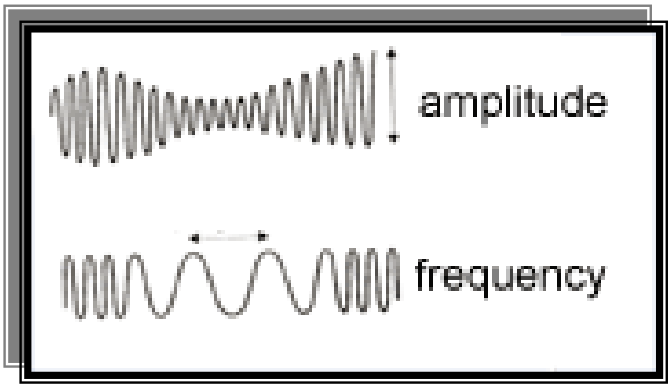
ويتعلق بهذا المصطلح الفيزيائي مصطلح آخر خاص بالإدراك السمعي

للإنسان وهو المجال (Pitch) (McEnary, 2001) إلا أن هذا المصطلح تسيطر عليه محددات سمع الإنسان وهما: (Elseom-Cook, 2001)

- محدودية المدى الطبيعي للأصوات التي يمكن للإنسان أن يدركها وكما مر ذكره سابقاً.
- مع الموجات المعقدة ولأكثر من موجة صوت فإن هذه الترددات لا يمكن إدراكها بصورة مستقلة لأنها تصبح ذات (pitch) واحدة عامة.

- **سعة الموجة Amplitude**: تسمى كمية التغيرات الحاصلة في ضغط الهواء نسبة إلى الضغط الجوي الطبيعي بـ(Amplitude) أو سعة الموجة وغالبا ما يستخدم هذا المصطلح للإشارة إلى قمة أو ذروة الاتساع أي اكبر تغيير حاصل في ضغط الهواء والذي سببته موجة الصوت (Dobrian, 1997) (Minasi, 1996).

وترتبط أيضاً مع هذه الخاصية خاصية محددة للصوت وهي الصخب (Loudness) ولكنها لا تطابقها تماماً لأن حساسية الأذن البشرية تختلف باختلاف التردد وإن هذه الخاصية تعتمد على سعة الموجة وترددها (Elseom-Cook, 2001) (قدو، ٢٠٠٤). يبين الشكل (٣) الفرق بين تردد الموجة وسعتها.



شكل (٣) الفرق بين التردد والسعة

هناك خواص متعلقة بالصوت هي:

- **النبرة Timbre: تعد النبرة ثالث المكونات الأساسية للطبيعة المدركة للصوت ومن خلال هذه الخاصية يمكن التمييز بين صوت شخص عن شخص آخر ومعرفة صوت آلة عن آلة أخرى (McEnary, 2001)، ولا يوجد صفة فيزيائية متعلقة بها" لأنها عبارة عن ناتج لأشياء متعددة (Elseom-Cook, 2001).**
- **الغلاف Envelope: يعبر الغلاف غالباً عن شكل التغييرات الحاصلة في جهارة الصوت (Volume) (Dobrian, 1997)، ويمكن تقسيم الغلاف إلى أربعة أقسام (Elseom-Cook, 2001):**

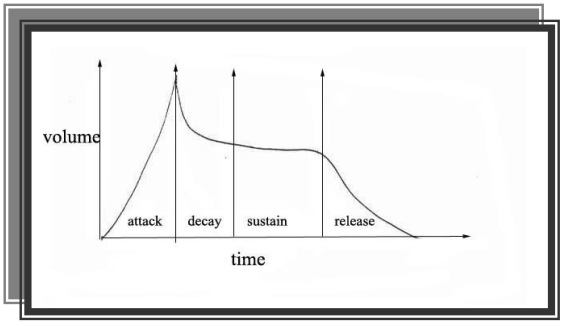
– الاستهلال Attack: هي الفترة الزمنية التي تبدأ مع بداية الصوت إلى الوصول إلى القمة أو ذروة سعة الموجة.

– الاضمحلال Decay: يمثل الاضمحلال الفترة الزمنية التي يتناقص فيها حجم الصوت أو جهارته من الذروة إلى مستوى آخر أدنى منه الذي قد يحتفظ بمستوى جهارة الصوت نفسه لفترة من الزمن.

– الثبات Sustain: هي فترة استمرار جهارة الصوت على مستوى واحد.

– الإطلاق Release: فترة تقليل جهارة الصوت من الثبات إلى حالة الصفر أي انتهاء موجة الصوت بصورة كاملة.

الشكل (٤) يوضح هذه الحالات الأربع في موجة صوتية.



شكل (٤) غلاف لموجة صوتية

٥- العزوم السبعة:

تعدّ طريقة العزوم من أوائل الطرق التي اعتمدت في التمييز البصري للحروف ولازالت تعتمد من قبل برامجيات التمييز المجهزة منها مع أجهزة الماسح الضوئي والتي تعتمد مبدأ الـ (off- Line).

تم في هذا البحث اعتماد العزوم التي أصبحت وسيلة لتصنيف الأشياء خلال الثلاثين سنة المنصرمة (Bozinovic, 1989). وقد تمت الاستفادة من العزوم ووظائف العزوم كنمط في عدد من التطبيقات لتحقيق تمييز ثابت.

منذ العام ١٩٦٢ تم الاعتماد على إدخال ثوابت عزمية بالاستناد على طرق الثوابت الجبرية وباستخدام ربط غير خطي من العزوم المنتظمة يشار لها على أنها عزوم هندسية، وقد تم اشتقاق مجموعة من العزوم الثابتة التي لها خاصية مرغوبة في كونها ثابتة تحت التزحيف والتقييس والدوران، ولهذا السبب فإن هذا النوع من العزوم يلائم بشكل أفضل مختلف الأساليب والطرائق.

يمكن تعريف العزم لأية مرتبة $(p+q)$ لدالة مستمرة $f(x,y)$ بالعلاقة الآتية:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$p, q = 0,1,2,...$

إذ

X,Y يمثلان الإحداثيين السيني والصادي للوحدة الصورية على التوالي.

p,q يمثلان مرتبة العزم المنتظم .

$f(x,y)$ تمثل قيمة دالة شدة إضاءة الصورة (وتكون قيمتها صفر أو واحد في الصور الثنائية).

x وy يمثلان المراكز القياسية للصورة باتجاه محور x وy.

اعتمد (Papoulis, 1965) على نص نظرية الأحادية Uniqueness, والمتضمنة على أنه إذا كانت f(x,y) مستمرة كقطع Piecewise continuos ولها قيم غير الصفر فقط في جزء محدد في المستوى x-y، تكون العزوم من كل الدرجات موجودة، عندئذ يتحدد تتابع العزوم (m_{Pq}) بـ f(x,y) بشكل فريد. ويمكن أن يعبر عن العزوم المركزية كما يأتي:

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q f(x,y)dxdy \ldots\ldots\ldots\ldots(2)$$

إذ

$$\bar{x} = \frac{m_{10}}{m_{00}} \ , \ \bar{y} = \frac{m_{01}}{m_{00}}$$

مع الصورة الرقمية تصبح المعادلة (2) كما يأتي:

$$\mu_{pq} = \sum_x\sum_y (x-\bar{x})^p(y-\bar{y})^q f(x,y) \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

عند حساب العزوم المركزية للمرتبة الثالثة تكون كما يأتي:

$$\mu_{ij} = \sum_x\sum_y (x-\bar{x})^i(y-\bar{y})^j f(x,y) = m_{ij} - \frac{m_{ij}m_{ji}}{m_{jj}},$$

بإيجاز

$$\mu_{00} = m_{00} \quad ,$$
$$\mu_{10} = 0 \quad ,$$
$$\mu_{01} = 0 \quad ,$$
$$\mu_{20} = m_{20} - \bar{x}m_{10},$$
$$\mu_{02} = m_{02} - \bar{y}m_{01},$$
$$\mu_{11} = m_{11} - \bar{y}m_{10},$$

$\mu_{30}=m_{30}-3\bar{x}m_{20}+2m_{10}\bar{x}^2$,

$\mu_{12}=m_{12}-2\bar{y}m_{11}-2\bar{x}m_{02}+2\bar{y}^2m_{10}$,

$\mu_{21}=m_{21}-2\bar{x}m_{11}-\bar{y}m_{20}+2\bar{x}^2m_{01}$,

$\mu_{03}=m_{03}-3\bar{y}m_{02}+2\bar{y}^2m_{01}$,

وتعرف العزوم المركزية المقيسة مشاراً إليها بـ η_{pq} كما يأتي:

$$\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad \ldots\ldots (4)$$

إذ

$\gamma = 1/2(p+q)+1$ ……………………………………(5)

$p+q = 2,3,\ldots$

يمكن اشتقاق سبعة عزوم غير متغيرة Invariant Moment من العزم الثاني والثالث تعطى كما يأتي:

$\phi_1 = \eta_{20}+\eta_{02}$, ……………………………………………(6)

$\phi_2 = (\eta_{20}-\eta_{02})^2 + 4\,\eta_{11}^2$ ……………………………………(7)

$\phi_3 = (\eta_{30}-3\eta_{12})^2 + (3\eta_{21}-\eta_{03})^2$, …………………………(8)

$\phi_4 = (\eta_{30}+\eta_{12})^2 + (\eta_{21}+\eta_{03})^2$, …………………………(9)

$\phi_5 = (\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})((\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2)$,
$+ (3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})(3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)$,….
(10)

$\phi_6 = (\eta_{20}-\eta_{02})((\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)$,
$+ 4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})$,…………………………….(11)

$\phi_7 = (3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})((\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2)$,
$+ (3\eta_{12}-\eta_{30})(\eta_{21}+\eta_{03})(3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)$……(12)

لقد برهن (Hu, 1962) أن هذه العزوم غير متغيرة بالنسبة للتزحيف Translation والدوران Rotation وتغير المقياس Scaling (Gowell, 2001) (Bozinovic, 1989) (البدراني، ٢٠٠١).

٦- الخوارزمية المعتمدة في البحث:

تم في هذا البحث اعتماد خوارزمية ذات مفهوم جديد للتعامل مع إشارة الصوت وتضمنت المراحل الآتية:

المرحلة الأولى:

الخطوة الأولى: تم في هذا البحث اعتماد أسلوب جديد في تحويل إشارة الكلام الداخل (ذو بعد واحد) إلى مصفوفة ثنائية مربعة ذات بعدين إذ يتم حساب بعد المصفوفة كما يأتي:

أ- حساب الجذر التربيعي لطول إشارة الكلام مقربة إلى أصغر رقم صحيح (لأجل الحصول على مصفوفة ثنائية مربعة).

ب- يتم إملاء المصفوفة الثنائية (سطر بعد سطر) من عناصر مصفوفة الكلام ذات البعد الواحد على أن يتم إهمال العناصر الأخيرة من مصفوفة الكلام.

الخطوة الثانية: يتم تحويل المصفوفة ذات البعدين إلى تحويل فورير (لأجل الاستفادة من مواصفات فورير للتخلص من الترددات العالية (غير المسموعة)). أي إن المصفوفة الثنائية الجديدة تكون غنية بالمعلومات (أي الترددات الواطئة).

الخطوة الثالثة: حساب العزوم السبعة للمصفوفة الثنائية الجديدة.

يتم تكرار الخطوة الأولى إلى الثالثة على مجموعة العينات التي تم تسجيلها والتي تضمنت (٥٠ عينة لصوت ذكور و٥٠ عينة لصوت إناث). ليتم بعدها إيجاد المعدل للعزوم السبعة لكل عينة.عليه يتم حساب حد العتبة للعزوم السبعة لفئة الذكور وحد العتبة للعزوم لفئة الإناث.

في الجدول (١) نموذج لبعض تلك العزوم علما أن النماذج المعتمدة لا تتضمن حالات الغضب والمرض أو التعب.

جدول (١)

قيم العزوم السبعة لخمسين أنثى وخمسين ذكراً مع حد العتبة لكلا الجنسين

العزوم السبعة لعينة صوت انثى١	٠.٣٦١٥	٠.١٨٦١	٥.٢٧٤٢	٦.٣٠٥٠	١٢.٢٤٧٤	٦.٤٥٤٢	١٢.٣٩٤٩
العزوم السبعة لعينة صوت انثى٢	٠.٤٩٥٨	٠.٣٠٨١	٥.٢٢٠٥	٥.٥٣٦٢	١١.٥٣٥١	٥.٥٤٢٠	١١.٠٧٩٩
العزوم السبعة لعينة صوت انثى٣	٠.٥٣١٣	٠.٣٥٥٧	٤.٠٦٣٧	٤.٨٥٧٧	٩.٥٠٢٤	٥.٠١٥٦	٩.٣٧٤٢
			.				
			.				
العزوم السبعة لعينة صوت ذكر١	٠.١٥٠٣	٠.٥٠٥٩	٧.٤١٣٠	٨.٥٠٩٩	١٦.٦٨١١	٩.٠٠٨٧	١٧.١٢٤٥
العزوم السبعة لعينة صوت ذكر٢	٠.٠٤٨٠	٠.٨٥٥١	٧.٠١٧٤	٧.٩٠٨٣	١٥.٥٧٣٤	٨.٣٨٧٠	١٦.٢٤٦٤
العزوم السبعة لعينة صوت ذكر٣	٠.٠٩٨٩	٠.٦٥٤٠	٦.٥٣٣٦	٧.٦٢٠٠	١٥.٠٩١٨	٨.٥٢٩٨	١٥.٠٠٢٣
			.				
			.				
حد العتبة للإناث	٠.٤٤٤٩	٠.٦٦٠٢	٥.٥٠١٦	٦.٣٠٢٢	١٢.٥٠٣٢	٦.٨١٧٦	١٢.٥٠٤٦
حد العتبة للذكور	٠.١٥٩٣	٠.٧٩٢٤	٦.٨٤٧٢	٧.٥٩٢٨	١٥.٠٩٨٥	٨.١٣٦٥	١٥.٢٤٧٢

المرحلة الثانية:

في هذه المرحلة يتم تدقيق صوت متكلم لجملة الاختبار نفسها والتي تم اعتمادها في المرحلة الاولى والتي اعتمدت لحساب حد العتبة للفئتين (الذكور / الإناث) ويمكن تلخيص خطوات هذه المرحلة كما يأتي:

الخطوة الأولى: تضمنت هذه الخطوة الأسلوب ذاته الذي تم اعتماده في المرحلة الأولى لتحويل إشارة الكلام ذات البعد الواحد إلى مصفوفة ثنائية ليتم بعدها تحويلها إلى المجال الترددي باعتماد تحويل فورير يتبعه حساب العزوم السبعة للمصفوفة الثنائية.

الخطوة الثانية: في هذه الخطوة يتم اعتماد معامل الاقتراب من خلال المشاهدة البصرية للمنحنيات وذلك بعرض الأشكال التي تبين منحني العزوم السبعة لإشارة

المتكلم ومنحنيي العزوم السبعة للذكور والإناث لاكتشاف مدى تقارب العزوم السبعة للإشارة الداخلة مع حد العتبة لفئة الذكور وحد العتبة لفئة الإناث.

الخطوة الثالثة: في حالة التوصل إلى نتائج مقاربة لحقيقة جنس المتكلم يكون هناك خيار لإضافة العزوم السبعة التي تم حسابها للإشارة الجديدة إلى مجموعة العزوم السبعة للفئة التي تنتمي إليها الإشارة الجديدة على وفق المعادلة (١) الآتية:

$$(x*z)+(y/z+1 \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

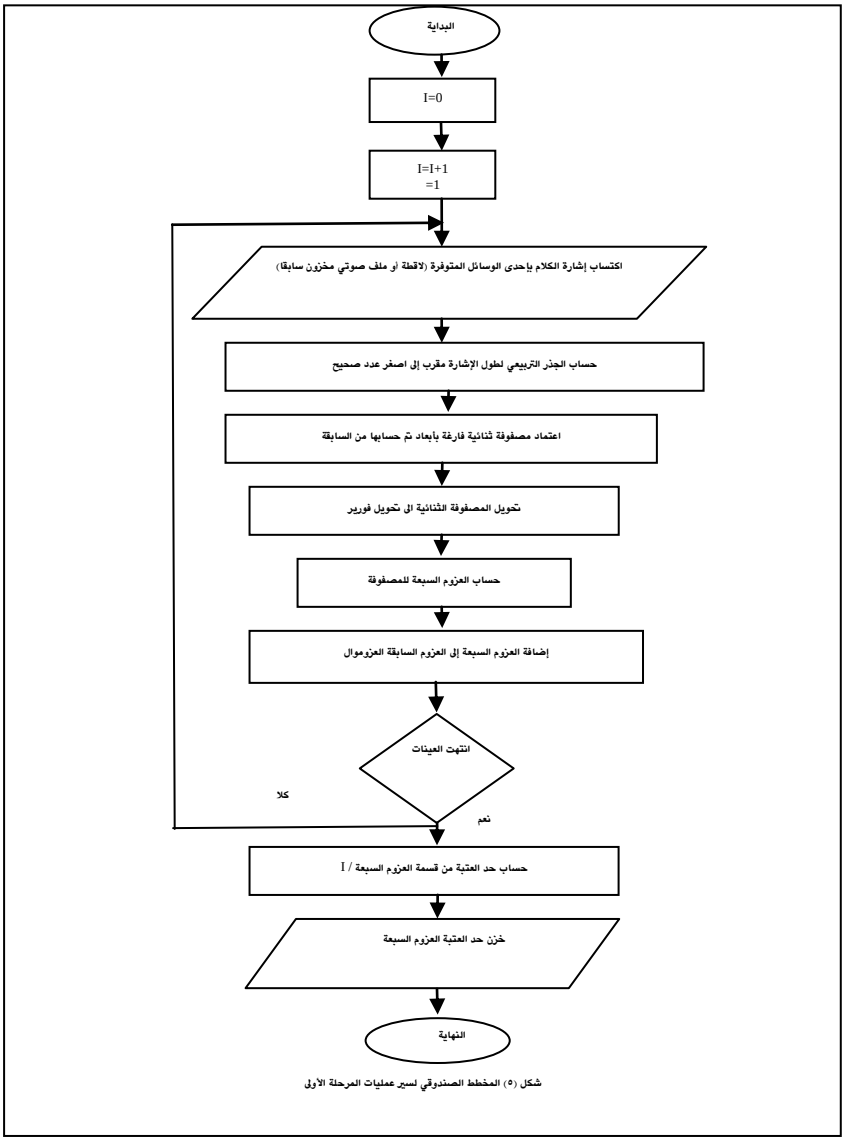
X: العزوم السبعة السابقة

Y: العزوم السبعة للإشارة الجديدة

Z: العدد الكلي للعينات

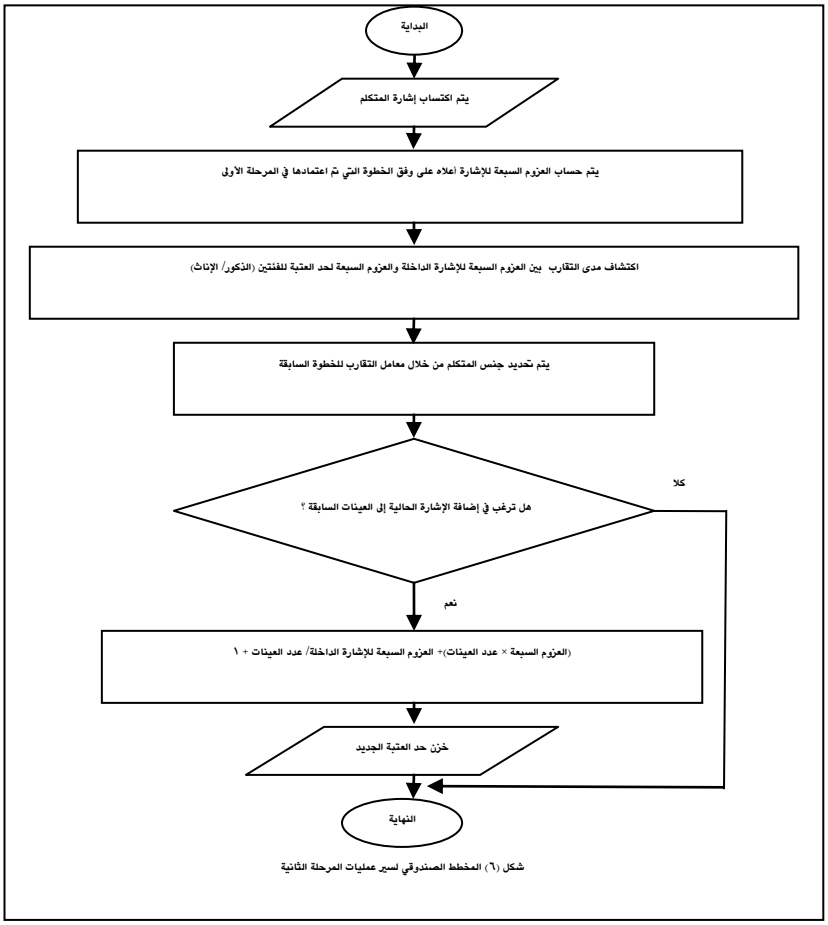
إذ إن العلاقة أعلاه تتضمن تغيراً خطياً على حد العتبة المعتمد سابقاً.

ويوضح الشكل (٥) المخطط الانسيابي لسير عمليات المرحلة الأولى للخوارزمية المقترحة:



شكل (٥) المخطط الصندوقي لسير عمليات المرحلة الأولى

إما الشكل (٦) فيمثل المخطط الانسيابي لسير عمليات المرحلة الثانية للخوارزمية المقترحة:



شكل (٦) المخطط الصندوقي لسير عمليات المرحلة الثانية

٧- مثال تطبيقي ومناقشه النتائج:

فيما يأتي مثال تطبيقي على نطق جملة البسملة كاملةً بفترة زمنية ٥ ثوانٍ وبمعدل تعيين (Sample Rate= 8000) ليصبح طول الإشارة (40000 Sample). ومن ثم تحويلها إلى مصفوفة مربعة بأبعاد (200×200) ليتم بعد ذلك تحويلها إلى Fourier Transform, إذ يتم حساب العزوم السبعة للمصفوفة

الجديدة ليتم بعد ذلك مقارنتها مع حد العتبة لكلا الجنسين لتحديد جنس المتكلم.

وفيما يأتي قيم المصفوفات للإشارة الداخلة و العزوم السبعة:

صيغة لمصفوفة مربعة لإشارة الكلام لأجل اعتمادها في حساب العزوم السبعة وتكون كالآتي:

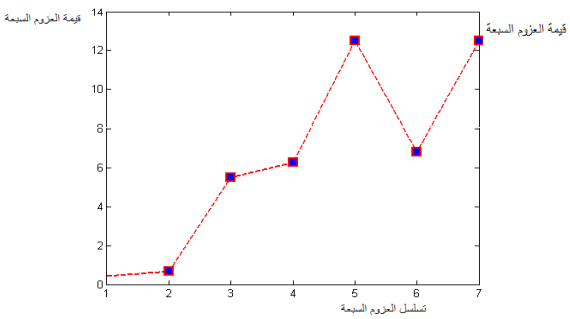
المصفوفة الأصلية	مصفوفة فورير
1.0e-004 * 0	0
0	0
-0.3052	0
0	1
0	1
0	1
0	1
-0.3052	2
0	2
0	2

المصفوفة المربعة الناتجة من المصفوفة الأحادية

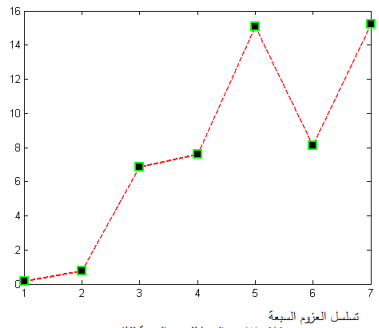
(0	0	0	1	1	1	1	2	2	2
1	1	0	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0)

قيم العزوم السبعة لعينة صوت الذكر في الشكل (١٠)

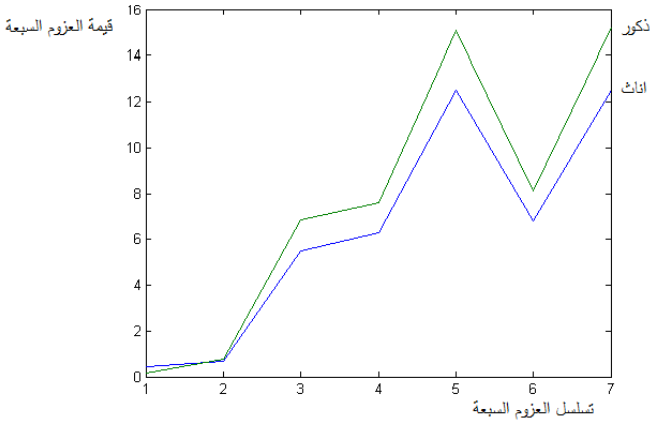
0.504 2.145 9.318 9.892 19.415 11.053 19. 642



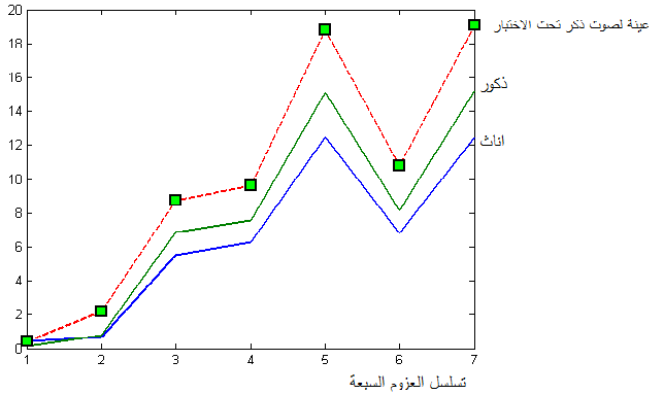
شكل (٧) حد العتبة للعزوم السبعة للإناث



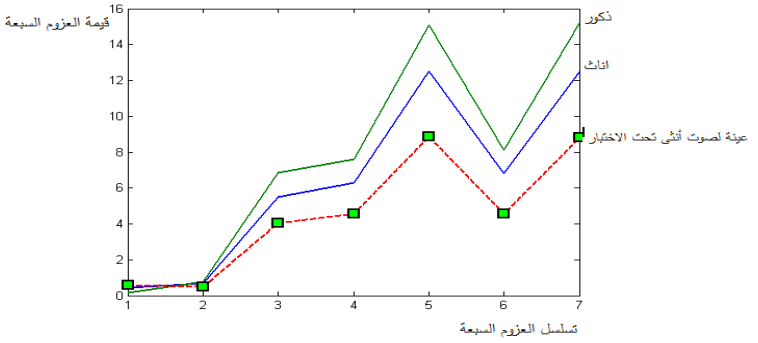
شكل (٨) حد العتبة للعزوم السبعة للذكور



(٩) حد العتبة للعزوم السبعة لكلا الجنسين



شكل (١٠) حد العتبة للعزوم السبعة لكلا الجنسين مع العزوم السبعة لعتبة صوت ذكر قيد الاختبار



شكل (11) حد العتبة للعزوم السبعة لكلا الجنسين مع العزوم السبعة لعتبة صوت أنثى قيد الاختبار

في الأشكال (٧-١١) تبين الخطوات مدى التقارب لصوت ذكر في الشكل (١٠) ومدى التقارب في الشكل (١١) من خلال دراسة نماذج كثيرة طبقت على الخوارزمية المقترحة تبين هناك نسبة خطأ في تمييز جنس المتكلم كما في الجدول (٢).

جدول (٢)

نسبة الخطأ

الجنس	عدد مرات الاختبار	عدد العينات ذات النتائج غير الصحيحة	نسبة الخطأ
ذكور	١٥	٣	٪٢٠
إناث	٢٠	٣	٪١٥

من الجدول (١) تبين أن الخوارزمية المعتمدة قد أعطت نسبة اكتشاف لجنس المتكلم من الإناث أعلى بكثير من جنس المتكلم من الذكور ويمكن تفسير ذلك إلى أن هناك إمكانية في احتواء صوت الإناث على ترددات واطئة (صوت خشن) أكثر بكثير من إمكانية احتواء صوت الذكور على ترددات عالية (صوت ناعم).

٨- الاستنتاجات والأعمال المستقبلية:

إن الخوارزمية المعتمدة في البحث تعطي نتائج جيدة ويمكن الاستفادة منها في إمكانية بناء منظومة مبسطة لاختبار جنس المتكلم وهي ذات حساسية عالية مع أصوات الإناث. يمكن تطوير الخوارزمية أعلاه من خلال إضافة شبكة عصبية لها القابلية على إيجاد مدى التقارب بين النموذج المطلوب اختباره وحد العتبة لكل الجنسين. وكذلك يمكن اعتماد النتائج التي تم الحصول عليها في هذا البحث لأجل تمييز جنس المتكلم بشكل أسرع من خلال اعتماد التقارب المعتمد على شكل العينة وحد العتبة. وأيضاً بالامكان الاستفادة من أفكار الخوارزمية المعتمدة في البحث لأجل تصنيف الفئات العمرية للمتكلم من خلال جمع عينات عالية (كثيرة) لفئات عمرية متباينة لأجل حساب حد العتبة لكل فئة عمرية.

المصادر

أولاً: المصادر العربية

أ. الرسائل والاطاريح

١. البدراني، هدية صالح عبدا لله (٢٠٠١)، تمييز الحروف العربية المكتوبة بنظام (word 7)، **رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.**

٢. حليم، علياء موفق عبد المجيد (٢٠٠٣)، تشفير إشارة الكلام بطريقة البعثرة، **رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.**

٣. عبد القادر، إسراء عبد السلام (٢٠٠١)، كبس الصوت عند الزمن الحقيقي، **رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.**

٤. قدو، سجى جاسم محمد (٢٠٠٤)، كبس إشارة الكلام بواسطة استخلاص الخواص، **رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.**

٥. الوكيل، هناء محسن احمد (٢٠٠٦)، تصميم نظام مسيطر حركة مضبب لروبوت متنقل ذاتي الحركة، **اطروحة دكتوراه، قسم علوم الحاسبات، الجامعة التكنولجية.**

ب. الكتب

١. عبد الحميد، محمد محيي الدين (١٩٩٦)، **التحفة السنية بشرح المقدمة الأجرومية**، المكتبة العصرية للطباعة والنشر، بيروت، صيدا.

ثانياً: المصادر الأجنبية

1. Abdul-Majeed, Elaf Osamh (2000), Voice Recognition Using Neural Networks, **M.Sc. Thesis, computer science, college of Computer and Mathematics, Mosul University.**

2. Kivimaki, Jukka (2000), Very low bit rate speech coding using speech recognition, analysis and synthesis, **M.Sc. Thesis, Tampere university of technology, Department of information Technology.**

3. Brooks, David W.& Carr, Adam & Edkins, Keith & et. al. (2004), **"data Compression",** Wikipedia the Free Encyclopedia, GNU Free Documentation License, Boston, U.S.A. **http:// en.wikipedia.org/wiki/**

4. Dobrian, Christopher (1997), **"Digital Audio",** from MSP: The Documentation Cycling '47 and IRCAM, Dec. 1997.

5. Elseom-Cook, Mark (2001), **Principles of Interactive Multimedia,** the McGraw. Hill Companies.

6. Goebel, Greg (2003), **"Introduction / Lossless Data Compression",** v1.1.0, The Public Domain. **http://www.vectorsite.net/ttdcmp1.html**

7. Bozinovic R. M. & Srihari S. N. (January 1989), **"Off-line cursive script word recognition"**, IEEE Transactions on pattern analysis and machine intelligence, Vol.11, No.1.

8. Gowell J. & Hussain F. (2000), **"The confusion matrix-identifying conflicts in Arabic and Latin character recognition",** Eurographic 2001.

9. Kientzle, Tim (1997), **"programmer's Guide to Sound",** Addison_Wesley Developers Press, ISBN 0-201-41972-6.

10. McEnary, John, (2001), **"Computer in Music/ What is Sound?"** Lectures/ the Classroom Instructional

Materials/Orange Coast College in Costa Mesa, California. http:// www.occ.cccd.edu/faculty/jmcenary/sound/sound.htm

11. Minasi, Mark (1996), **the Complete PC Upgrade and Maintenance Guide**, 4th Edition, Sybex-Inc.
12. Saeed, V. (2002), **"Speed Processing Digital filter ……."** http://www.brunel.ac.uk/depts/ee/com/home speed vaseghi