التعرف على جنس المتكلم بالاعتماد على إشارة الصوت*

زيد جعفر فاضل

مدرس مساعد، كلية علوم الحاسوب والرياضيات، جامعة الموصل

المستخلص

في هذا البحث تم اقتراح خوارزمية تعمل على تمييز جنس المتكلم بالاعتماد على اشارة صوته. الخوارزمية المقترحة اعتمدت استخلاص خواص إشارة الكلام لأجل الحصول على تمثيل لتلك الإشارة ومن ثم الاستفادة من تلك الخواص في عملية بناء قاعدة معلومات تتضمن الهوية التعريفية للمتكلم فضلاً عن الخواص المستخلصة وجنسه.

يقدم هذا البحث خوارزمية للتحقق من جنس المتكلم، وتم اعتماد تحويل فورير من أجل استخلاص بعض المعاملات المعتمدة على عوامل تحويل فورير إذ تم موائمة هذه المعاملات للحصول على معادلة من الدرجة الخامسة عشر المقاربة لاشارة المتكلم بعد تحويلها إلى تحويلات فورير.

يظهر التطبيق العملي بأن مقدار التقارب للمعادلة المستخرجة غير متاثر بدرجة تقارب المعادلة بعد الدرجة الخامسة عشرة. تم تطبيق الخوارزمية المقترحة على عينة تتألف من أكثر من خمسين متكلماً وفئات عمرية مختلفة ولكلا الجنسين إذ تم اكتساب ثلاث كلمات لكل متكلم، وتم تحويلها إلى تحويلات فورير إذ تم احتساب المعدل بعد موائمة الاشارة إلى متعدد الحدود من الدرجة الخامسة عشرة.

يتم خزن هذه القيم المستخلصة بعد ترتيبها تنازلياً لتكون خواص لتمييز جنس المتكلم. التطبيق العملي للخوارزمية المقترحة ادى إلى الحصول على نسبة تمييز تصل إلى ٧٢٪ من ٢٠ نموذج. معامل التقارب (Correlation Factor) تم توظيفه لأجل اتخاذ اقرار حول نسبة التمييز.

(*) تم استلام البحث في ٢٠١٣/٥/٢٣ وحصل على قبول النشر في ٢٠١٣/١٠/٢٣.

Gender Identification Based on Speech Signal

Zaid Jafer Fadil
Asst. Lect., College of Computer Science and Math, Mosul University

Abstract

In this paper algorithm has been proposed to identify the speaker's gender based on the signal of her sound. This algorithm relied on deriving the features of speaker's signal in order to gain a representation of this signal and to get benefit from those features in establishing data base that includes the identity of the speaker as well as the drawn features and her gender.

This paper present an algorithm confirm the speaker’s gender. The Fourier transformation is adopted here in order to derive some of the factors that depend upon the Fourier transformation factors where those factors have been adjusted to gain an equality of the 15^{th} degree that is too close to the speaker's signal after transforming it to Fourier transformation.

The practical application reveals that the amount of the approach of the derived equality is not affected by the equality degree of the polynomial after the 15^{th} degree. This algorithm has been applied on a sample of more than 50 speakers of different groups of age of the two genders where three words of each speaker have been gained then those words are transformed to Fourier transformation where the rate is calculated after adjusting the signal into polynomial of 15^{th} degree.

The derived values will be stored after arranging them in decreasing way to be features by which the speaker's gender can be recognized. The practical application of the proposed algorithm resulted in getting on distinction ratio approaches to 72% out of 20 samples. The approaching factor is employed to determine has to do with the distinction ratio.

المقدمة:

إن تحليل العلاقة بين السمات الصوتية، السمعية للكلام وجنس المتكلم لها تطبيقات كثيرة. إذ إن إشارة الكلام تحوي الكثير من المعلومات فضلاً عن الرسالة التي يتم نقلها على لسان المتكلم، فهي تعطي للمستمع فكرة عن جنس المتكلم، فضلاً عن حالته العاطفية. إن اختلاف نمط الكلام بين الإناث والذكور يأتي من الاختلافات في الخصائص اللغوية وغير اللغوية (Paralinguistic) لإشارة الكلام. فمثلاً، في حالة الإناث فإن الخصائص اللغوية تتمثل في وجود الترددات العالية، على العكس من ذلك بالنسبة للذكور [1].

كما أن التغييرات الفسلجية تؤثر في صوت المتكلم، إذ تمر الحنجرة بعملية نضوج التي تبدأ في وقت مبكر من العمر وتستمر بالتغييرات خلال الحياة. إذ إن أكثر التغييرات إثارة تحدث خلال سن البلوغ، ومع مرور الوقت فإن غضاريف الحنجرة تبدأ بالتكلس وتصبح متصلبة أكثر بقليل. كما أن المفاصل التي تسمح لحركة الطيات الصوتية ثلاثية الأبعاد بأن تصبح أصلب وحجم عضلة الطيات الصوتية بأن يقل [2].

كذلك تتضمن العوامل المتعلقة بالمتكلم الخصائص الطبيعية (تشريحية وفسلجية) مثل الجنس، والوزن، والصحة، والتدخين، والحالة الفسلجية. فمثلاً في حالة الجنس إذ إن معظم الخصائص الصوتية تختلف بين الجنسين بسبب اختلافات تشريحية مثل حجم الحنجرة الأكبر والمنطقة الصوتية (Vocal Tract) الأطول قليلاً كمعدل في حالة الذكور. إذ إن حجم المنطقة الصوتية في حالق الإناث هي ١٥٪ أقصر من الذكور [1].

إذ إن تمييز الكلام آلياً تقنية واعدة أصبحت حقيقة تؤكد على استعمال السمع والكلام للتحاور من خلال واجهة اللغة الطبيعية مع الأجهزة والمعدات (الشبيهة بالبشر من ناحية الإدراك). عليه أصبحت أنظمة تمييز الكلام بواسطة الآلة هي واحدة من أكثر الطرق كفاءة لتفاعل الإنسان مع الآلة [2].

وممكن تعريف الصوت انه عبارة عن تذبذب في ضغط الهواء أو الوسط الناقل سببه جسم مهتز [3]، وان كان الصوت كلاماً فإن الجسم المهتز الذي سبب الكلام هو الأوتار الصوتية للإنسان [4]، أما أصوات الموسيقى فسببها الاهتزاز الحاصل في الأدوات الموسيقية [5][6].

يشار عادة لهذه التغييرات المتذبذبة في ضغط الهواء بمصطلح موجة الصوت (Sound Wave)، ولكي تكون هذه التغييرات مسموعة يجب أن تحصل بتردد مناسب لتكون محسوسة وتسبب تأرجح طبلة الإذن، فالترددات التي تتمكن الإذن البشرية من تمييزها يتراوح مداها ما بين (20Hz - 20KHz)، ومع تقدم العمر يقل هذا المدى [4] فقد يصل إلى (20Hz – 17KHz) [3][7].

وتجدر الإشارة إلى أن الترددات التي تكون أقل من هذا المدى تدعى بـ (Infrasound)، والتي تكون أعلى منه تدعى بـ (Ultrasound) [7].

١- الأعمال السابقة:

في عام ٢٠٠٠ قدم (Childers) حقيبة ادوات حاسوبية لأجل تحليل ومعالجة الكلام [8].

وفي عام ٢٠٠٣ حلل العالمان (Bruckl and Sendlmeier) عينات من كلام مقروء لاحرف العلة وكلام تلقائي ايضا لعينة تزيد عن ٥٦ متكلماً من الاناث اللاتي تتراوح اعمارهن من ٢٠ إلى ٨٧ سنة في بحثهما (Aging Female Voices: an Acoustic and Perceptive Analysis) [9]. وفي العام نفسه قدم الباحث مصطفى نظام قاعدة بيانات لتشخيص المتكلمين إذ استخدمت عدة تقنيات في كل من المنطق الزمني والمنطق الترددي لغرض استخلاص السمات الصوتية وتطبيق طريقة ترميز التنبؤ الخطي لمعالجة الاصوات [10].

وقد تضمنت كثير من أعمال الباحثين في مجال معالجة الإشارة الصوتية اتجاهات كثيرة منها ما دار حول تخمين عمر المتكلم من خلال استخلاص خواص وصفات لإشارة الكلام للمتكلمين إذ قدمت الباحثة Schötz حقيبة متكاملة حول تخمين العمر من خلال

اطروحة الدكتوراه في عام ٢٠٠٦ [11]. وفي العام نفسه نشر (Harb & et. Al.) تحديد نوع الجنس باستخدام الصوت العام [13].

وفي عام ٢٠٠٧ قدم الباحثان Schötz & Müller أيضاً دراسة حول مدى التقارب الصوتي بين المتكلمين المتقاربين في العمر إذ استخلصت ست خواص صوتية وبشكل آلي باستخدام برنامج Praat ومن ست كلمات منتجة من ٥٢٧ متكلماً ومن كلا الجنسين [12]. أما الباحثان (Matthew Blackshaw, Eric Steinlauf) فقد قدما بحثهما الموسوم (Gender and Age Determination of a Speaker) في عام ٢٠٠٧ أيضاً والذي تضمن أسلوباً إحصائياً يعتمد على الترددات وذلك بعدّ الوسيط والانحراف المعياري للترددات في صوت المتكلم في حروف العلة له ارتباط بعمر المتكلم [14].

وقدم الباحث (Janda) في عام ٢٠٠٩ بحثه الموسوم "معاملات كلام الاطفال لتحديد العمر" إذ عمل على دراسة العوامل الرئيسة في كلام الاطفال والتي يمكن اعتمادها لأجل تحديد العمر [15].

إن عملية تحديد جنس المتكلم من الصوت بدأت منذ زمن بعيد إذ ظهرت تقنيات ممكننة لأجل التعرف على جنس المتكلم من خلال استخلاص خواص لإشارة الكلام آلياً، إذ ظهرت تقنيات حاسوبية واسعة غطت مجالات كبيرة من الامكانيات التي توفرها الحاسبات الالكترونية.

في عام ٢٠٠٩ غطى المؤلف Sedaaghi، دراسة مقارنة لنوع الجنس والسن في تصنيف إشارات الكلام [16].

في عام ٢٠١١ قدمت الباحثة هبة رسالة ماجستير بعنوان "معاملات المواصفات الصوتية لتحديد العمر ذاتياً" إذ درست صفات صوتية معتمدة على مواءمة شكل موجة الصوت على وفق أعمار بفئات مختلفة [2].

٢- **الصوت:**

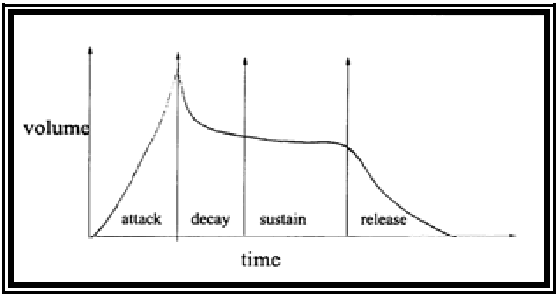
الإشارات الصوتية يمكن أن تكون لصوت مسموع أو غير مسموع معتمدٍ على الترددات التي يتكون منها ذلك الصوت. ومن الممكن تصنيف الصوت إلى أصوات الآلات الموسيقية والكلام البشري أو أصوات الحيوانات (جميعها تشترك بخاصية كونها مسموعة أم لا بالاعتماد على الترددات التي تحتويها).

٣- **خصائص الصوت** [17]:

هناك صفات كثيرة يوصف بها الصوت منها:

- **التردد (Frequency):** ويتعلق بهذا المصطلح الفيزيائي مصطلح آخر خاص بالإدراك السمعي للإنسان (Pitch)، إلا أن هذا المصطلح تسيطر عليه محددات سمع الإنسان وهما:
  - ✓ محدودية المدى الطبيعي للأصوات التي يمكن للإنسان أن يدركها.
  - ✓ مع الموجات المعقدة فان هذه الترددات لا يمكن إدراكها بصورة مستقلة لأنها تصبح ذات (Pitch) واحدة.
- **سعة الموجة (Amplitude):** غالباً مايستخدم هذا المصطلح للإشارة إلى قمة الاتساع أو ذروته أي أكبر تغيير حاصل في ضغط الهواء والذي سببته موجة الصوت. وترتبط أيضاً مع هذه الخاصية خاصية محددة للصوت وهي الصخب (Loudness) ولكنها لا تطابقها تماماً.
- **النبرة (Timbre):** تعد النبرة ثالث المكونات الأساسية للطبيعة المدركة للصوت ومن خلال هذه الخاصية يمكن التمييز بين شخص إلى شخص أخر ومعرفة صوت آلة عن صوت آلة أخرى، ولا توجد صفة فيزيائية متعلقة بها لأنها عبارة عن ناتج لأشياء عديدة.
- **الغلاف (Envelope) [18]:** يعبر الغلاف غالباً عن شكل التغييرات الحاصلة في جهاز الصوت (Volume)، ويمكن تقسيم الغلاف إلى أربعة أقسام:

- ✓ **الهجوم (Attack): هي المدة الزمنية التي تبدأ مع بداية الصوت إلى الوصول إلى القمة أو ذروة سعة الموجة.**
- ✓ **الاضمحلال (Decay): يمثل الاضمحلال المدة الزمنية التي يتناقص فيها حجم الصوت أو جهارته إلى مستوى آخر أدنى منه.**
- ✓ **الثبات (Sustain): هو مدة استمرار جهارة الصوت على مستوى واحد.**
- ✓ **الإطلاق (Release): مدة تقليل جهارة الصوت من الثبات إلى حالة الصفر أي انتهاء موجة الصوت بصورة كاملة. والشكل (١) يوضح أجزاء الغلاف في موجة صوتية.**



شكل (١)

غلاف لموجة صوتية

٤- تمييز جنس المتكلم:

إن القانون التنفيذي وسلطات الأمن العسكرية كانت من بين الأوائل التي استخدمت تقنية تمييز المتكلم آلياً. إذ كان النوع الأول لتمييز المتكلم آلياً هو باستخدام الطيف (Spectrograms) **لأصوات المتكلمين، والذي يدعى ببصمة الصوت** (Voiceprint) **أو الكلام المرئي الذي بدا عام ١٩٦٠. وهذا المصطلح** (Voiceprint) **مشتق من المصطلح** (Fingerprint) **الأكثر شيوعاً** [19].

إن تحليل بصمة الصوت هي عملية شبه آلية إذ يتم أولاً تكزين تمثيل بياني لكل أصوات المتكلمين، ثم يقرر الخبراء يدوياً فيما إذا كان هناك اثنان من الرسوم البيانية تمثل تعبيرات للشخص نفسه.

إن التمثيل البياني يأخذ أحد الشكلين، إما كطيف للكلام (Speech Spectrogram)، أو مخطط بصمة الصوت (Voiceprint Contour)، والصيغة الأكثر شيوعاً هي تمثيل التعبير أو الكلام المنطوق بحيث يكون الزمن على المحور الأفقي والتردد على المحور العمودي والطاقة الطيفية تمثل الظلام عند نقطة معطية [20].

٥- التمييز الآلي والسمعي للمتكلم:

يقوم الإنسان والحاسبات بتمييز أنماط الكلام بطرائق مختلفة. فالبشر يستخدمون الفهم الأذني السمعي واللغوي وأنواع أخرى من المعرفة لتمييز الآخرين من أصواتهم. على العكس فان أنظمة التمييز الآلي للمتكلم تعتمد على استخدام الخصائص السمعية التي تستخلصها الحاسبات آلياً.

إن أنظمة التمييز الآلي للمتكلم أيضاً يمكن أن تتدرب لتتعلم خصائص الصوت لشخص محدد. وكما هو الحال بالنسبة للبشر عندما تكون بيانات التدريب أكثر فإن تمييز المتكلم تكون أفضل. ويمكن أن يعزى السبب إلى القدرة الذكائية التي يتميز بها الإنسان والذي يعطيه خاصية مختلفة عما تتمتع به الآلة [21].

٥-١ بعض الصعوبات في عملية تمييز المتكلم:

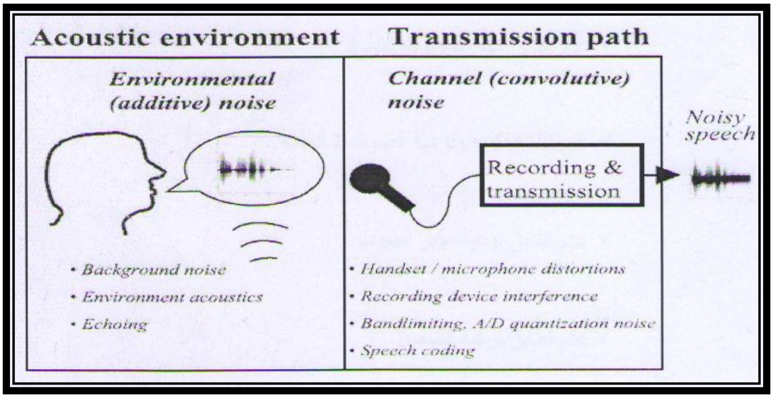
هناك أسباب مختلفة لصعوبة تمييز المتكلم، بعض منها يعتمد على المتكم نفسه والبعض الآخر لأسباب تقنية وأهمها:

١. تغييرات داخل الفرد (Intra-Individual Variation): من المعلوم ان الحالة الصحية والنفسية تؤثر على صوت المتكلم. كما أن الادوية، والمنبهات والمخدرات تؤثر في الصوت أيضاً، فمثلاً الاشخاص المدخنون تكون اصواتهم أكثر خشونة من

الاشخاص غير المدخنين. وأيضاً عندما يكون المتكلم في حالة إجهاد فالمعالم السمعية تختلف مما لو كان المتكلم في حالة راحة.

٢. **تنكر ومحاكاة الصوت** (Voice Disguise and Mimicry): إن عملية تنكر الصوت يعني تغيير الصوت بتعمد بحيث لا يتطابق مع العينة الاخرى المنتجة لدى المتكلم نفسه. وهو أمر شائع في الحالات القضائية عندما يقوم المجرم بتغيير صوته أثناء تحقيقات الشرطة.

٣. **مصادر تقنية للخطأ** (Sources Technical Error): هناك العديد من المصادر التقنية للخطأ والتي تقلل من أداء تمييز المتكلم (السمعي والآلي). ويمكن توضيحها بالشكل (٢) [22].



شكل (٢)

مصادر الخطأ التقنية في تمييز المتكلم

إذ يتم أولاً تسجيل الصوت باستخدام مكبر الصوت أو جهاز التلفون، ثم تتم إضافة الضوضاء البيئية (محرك السيارة، نقرات لوحة المفاتيح، ضوضاء المرور والموسيقى) إلى إشارة الكلام، كما إن الصدى يضيف نسخة متأخرة من الإشارة الأصلية إلى الإشارة المسجلة. كما أن مكبرات الصوت رديئة النوعية تضيف تشويشاً لا خطياً إلى طيف إشارة الكلام. كما أن محول الإشارة التناظرية إلى رقمية (A/ D Converter)

يضيف التشويش الخاص به وإذا تم نقل إشارة الكلام عبر شبكة الهاتف يتم كبسها باستخدام تقنيات الكبس والذي من المحتمل أن تضيف الضوضاء إلى إشارة الكلام.

كما أن عدم تطابق الظروف يعد من أكثر مصادر الأخطاء لتمييز المتكلم، والتي تعني أن ظروف كل من مرحلتي التدريب والتمييز تكون مختلفة فضلاً عن التغييرات التي تحصل في الفرد كعدم مطابقة درجة الصوت وإن عدم التطابق سببه عامل واحد من العوامل الآتية أو أكثر [22]:

- ✓ **عدم تطابق علم الصوتيات البيئي.**
- ✓ **عدم تطابق كمية الضوضاء الخلفية ونوعيتها.**
- ✓ **عدم تطابق نوعية مكبر الصوت.**
- ✓ **عدم تطابق نوعية التسجيل.**

٦- تحويل فورير (Fourier Transform):

يعرف تحويل فورير Fourier Transform **بأنه عبارة عن تحليل الإشارة إلى مركباتها التكرارية. ويعدّ تحويل فورير من الطرق الرئيسة من التحليل الطيفي التي تلعب دوراً مهماً فيه. لذلك يمكن أن نعدّ أن تحويل فورير مرادفاً للتحليل الطيفي. لتكن** f(x) **تمثل إشارة مستلمة من مصدر ما فإن تحويل فورير لها، يرمز له** ($\Im\{f(x)\}$) **ويعرف رياضياً كالآتي** [23]:

$$\Im\{f(x)\} = F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)e^{-2j\pi ux/N} \qquad (1)$$

إذ إن u = 0، 1، 2،......، N-1 **و** j **تمثل** $\sqrt{-1}$ **وهي الرقم العقدي** Complex Number**، لذلك فإن تحويل فورير يعطينا أرقاما عقدية، و**N **تمثل عدد النقاط (حجم العينة).**

ويمكن تمثيل الدالة الآسية لأس عقدي بدلالة الجيب والجيب تمام كالآتي:

$$e^{jt} = \cos(t) + j\sin(t), \quad j = \sqrt{-1} \qquad (2)$$

وبتعويض (٢) في (١) نحصل على الآتي:

$$F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)[\cos(2\pi ux/N) - j\sin(2\pi ux/N)] \qquad (3)$$

وفي بعض الأحيان يجري تغيير على الإشارة في المنطلق التكراري ثم يتم إرجاعها إلى المنطلق الزمني كما يحدث في ترشيح الإشارة. للحصول على الإشارة في المنطلق الزمني من المنطلق التكراري نستخدم تحويل فورير العكسي Inverse Fourier Transform (نرمز له $\Im^{-1}\{f(x)\}$) ويعرف كالآتي:

$$\Im^{-1}\{F(u)\} = f(x) = \sum^{N-1} F(u)e^{2j\pi ux/N} \qquad (4)$$

٧- تحويل فورير السريع Fast Fourier Transform (FFT):

لتطبيق المعادلة (١) نحتاج إلى N*N من عمليات الجمع والضرب. إن تحويل فورير السريع يتضمن تطبيق المعادلة (١) في عدد أقل من عمليات الجمع والضرب يساوي $N*\log_2 N$، إذ إن تقليل العمليات من N*N إلى $N*\log_2 N$ يزيد من سرعة التحويل. يتبين مدى تفوق تحويل فورير السريع على التحويل المباشر (والذي نرمز له DFT) في توفير الوقت عند زيادة النقاط. وعلى سبيل المثال فإن تحويل فورير المباشر عندما N=8192 على الحاسب IBM 7094 يتطلب ثلاثة أرباع الساعة. بينما العملية نفسها يمكن أن تطبق على هذا الحاسب خلال خمس ثوانٍ باستخدام تحويل فورير السريع [23].

ولتوضيح الخوارزمية لحساب تحويل فورير السريع يمكننا صياغة المعادلة (١) بشكل آخر كالآتي [23]:

$$F(u) = \sum_{x=0}^{N-1} f(x) W_N^{ux}, \qquad W_N = e^{-j2\pi/N} \qquad (5)$$

كما يمكن صياغة المعادلة (٥) كالآتي:

$$F(u) = \sum_{x=0}^{N/2-1} f(2x)\, W_N^{2xu} + \sum_{x=0}^{N/2-1} f(2x+1) W_N^{(2x+1)N}$$
$$= \sum_{x=0}^{N/2-1} f(2x)(W_N^2)^{Ux} + W_N^u \sum_{x=0}^{N/2-1} f(2x+1(W_N^2)^{Ux}) \qquad (6)$$

ونلاحظ أن:

$$W_N^2 = e^{-2j2\pi/N} = W_{N/2} \qquad (7)$$

وعند تعويض (٧) في (٦) نحصل على ما يأتي:

$$F(u) = \sum_{x=0}^{N/2-1} f(2x)\, W_{N/2}^{ux} + W_N^u \sum_{x=0}^{N/2-1} f(2x+1) W_{N/2}^{ux}$$
$$F(u) = G(u) + W_{N/2}^u H(u) \qquad (8)$$

مما سبق نلاحظ إمكانية تمثيل تحويل فورير بحجم N من النقاط بواسطة تحويليين بحجم N/2 من النقاط هما G(u) وH(u). إن G(u) يمثل تحويل فورير للنقاط الزوجية، في حين إن H(u) يمثل تحويل فورير للنقاط الفردية.

لو فرضنا أن $N = 2^n$ إذ n عدد صحيح، فإن N/2 هو عدد زوجي أيضاً، وبالإمكان تقسيم البيانات إلى مجموعتين وبالتالي نحصل على أربع مجاميع من البيانات التي عددها الكلي N. تستمر عملية التجزئة إلى أن نحصل على مجموعة تتكون من نقطتين يمكن إجراء تحويل فورير المباشر عليها بسهولة. إن عملية تقسيم النقاط إلى مجاميع صغيرة تقلل من عمليات الجمع والضرب وبالتالي تسرع من عملية التحويل، هذا ما نطلق عليه (بالاختزال العشري الأساس ٢ في الزمن) (decimation radix 2 in time) أو تحويل فورير السريع FFT.

٨- الخوارزمية المقترحة:

يتم استخلاص الخواص التي اعتمدت في هذا البحث من خلال حساب معاملات تحويل فورير ومبدأ تنعيم المنحنيات (Curve Fitting) لأجل الحصول على تمييز جنس المتكلم من خلال كلمات منطوقة يتم خزنها لأجل بناء قاعدة معلومات المعتمدة لاستخلاص معاملات تمييز للذكور والاناث.

تضمنت الخوارزمية المعتمدة مرحلتين لأجل التمييز:

١- **مرحلة بناء الثوابت المشتركة لكلا الجنسين:** في هذه المرحلة يتم عملية إدخال إشارات صوتية يتم معالجتها لأجل استخلاص الخواص لكل منها وكما يلي:

أ- تجميع العينات لمجموعة الذكور ومجموعة الاناث: إذ تتمكن الخوارزمية المقترحة من استقبال إشارة صوتية من إدخالات متعددة (على سبيل المثال من خلال لاقطة الصوت أو من خلال ملف مخزون مسبقاً يحتوي على إشارة صوتية).

ب- لأجل معالجة الاشارة الصوتية تظهر الحاجة إلى تهيئة الاشارة بشكل يسهل عملية التعامل معها من خلال اجراء معالجات اولية وذلك بإمرارها بمرشح إمراراً واطئاً لامتصاص كل الترددات غير المسموعة والتي قد تشكل ضوضاء تعرقل عمليات اكتشاف خصائص الاشارة المطلوبة.

ت- تتم في هذه المرحلة اعتماد الاشارة الصوتية الداخلة في المجال الترددي من خلال تطبيق تحويلات فورير عليها.

ث- لكون معاملات تحويلات فورير تكون عالية جداً ومعتمدة على عدد العينات التي يمتلكها الصوت عليه يتم إجراء مواءمة للمنحني الخاص بإشارة الصوت في المجال الترددي باستخدام تطبيق المواءمة (Curve Fitting) (Polynomial Fitting Order 15).

ج- لأجل بناء قاعدة بيانات لمتكلمين سيتم تحديدهم لغرض إيجاد قاسم مشترك لكل الصفات الست عشرة لكلا الجنسين (إذ ينتج ستة عشر معاملاً ليتم تحديد جنس أي متكلم لاحقاً من خلال مدى تقاربه من الصفات التي تم الاعتماد عليها أعلاه).

٢- **مرحلة التمييز:** في مرحلة التمييز يتم اكتساب صوت لأي متكلم بإحدى الوسائل المتاحة من قبل الخوارزمية المعتمدة لأجل تطبيق مشابه لخطوات إيجاد معاملات كلا الجنسين اعلاه، والتي ستؤدي للحصول على ستة عشر معاملاً تمثل الصفات الخاصة لاشارة المتكلم.

يليها إجراء مقارنة مع الصفات الست عشرة لكلا الجنسين لتحديد مدى تقاربها من أي منهما، وذلك من خلال حساب معامل الاقتراب (Correlation Factor) والذي اعتمد كمعيار لتحديد جنس المتكلم. في حالة كون معامل الاقتراب بعيداً عند حد العتبة الذي اعتمد لتحديد مدى الاقتراب يتم إضافته إلى قاعدة البيانات لأجل استخلاص مجموعة جديدة تمثل معاملات إشارة الكلام لكلا الجنسين.

٩- مثال تطبيقي:

عند تطبيق الخوارزمية المقترحة على مجموعة من المتكلمين تجاوزت العينة المعتمدة خمسون شخصاً. الجدول (١) يبين نموذج تسجيل معاملات بعض منهم.

جدول (١)

معاملات تحويل فورير لاشارة الكلام المكتسبة

	1	2	…	999	1000
1	-٤.33615 +0.00000i	-٨1.80083 -1٤.3147٦i	…	1٢.70481 +٣٨.20070i	-٧1.80083 +15.31479 i
2	٤١.47595 + 0.00000i	٩٠.35607 - 68.09299i	…	-1٣.76408 -3٢٩٨٦8i	٨١.35607 +٥٩.09١٢٩i
. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
9	1.72239+02 +0.00000i	6.04608 +13.29194i	…	-60.09569 +14.29315i	7.94608 -13.91979i
10	73.27938 +0.00000i	16.95429 -21.04095i	…	59.91412 - 0.49875i	16.25429 +22.01095i

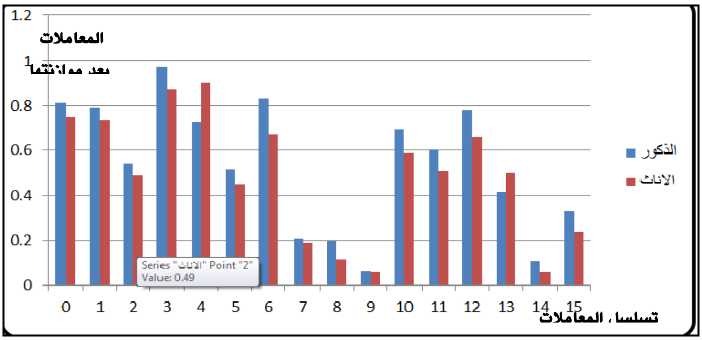
من الجدول (١) والشكل (٣) تبين ان اعتماد معاملات تحويل فورير إلى الحد العاشر وإجراء مواءمة من الدرجة الخامسة عشرة لمتعدد الحدود يقدم وصفاً قريباً جداً لصفة اشارة الكلام للمتكلم. إن الشكل المذكور يعمل على الحصول على تقارب عالٍ بين إشارة المتكلم ومقدار حد العتبة لكل من المعاملات ولكلا الجنسين. تبين أيضاً وجود تقارب في المعاملات البعيدة أي بعد التسلسل الحادي عشر مما يجعل تلك المعاملات ذات تأثير قليل بحيث يتم إهماله.

عند حساب مدى التقارب بين الاشارة التي تم مواءمتها مع الإشارة الاصل كان مدى التقارب إلى حد (٤٥٪) تقريباً عند اعتماد الـ (Correlation Factor) **كمقياس لمدى التقارب مما يجعله نسبة جيدة في حالة تمييز جنس متكلم من خلال إشارة صوته، أما حساب** MSE **فقد اظهر هناك فرقاً ليس عالياً بسبب موازنة مقدار الاشارة** (Amplitude) **لكلا الإشارتين.**

جدول (٢)

قيم حد العتبة لإشارة الكلام لكلا الجنسين

تسلسل متعدد الحدود	الذكور	الاناث	تسلسل متعدد الحدود	الذكور	الاناث
٠	٠,٨١٠٣	٠,٧٥١	٨	٠,٢٠٠١	٠,١١٩
١	٠,٧٩٣	٠,٧٣٣	٩	٠,٠٦٣	٠,٠٥٩
٢	٠,٥٤	٠,٤٩	١٠	٠,٦٩٧	٠,٥٨٩
٣	٠,٩٧٣	٠,٨٧٢	١١	٠,٦٠٢	٠,٥١٢
٤	٠,٧٣٢	٠,٨٩٨	١٢	٠,٧٨١	٠,٦٦١
٥	٠,٥١٥	٠,٤٤٩	١٣	٠,٤١٩	٠,٥٠٢
٦	٠,٨٣١	٠,٦٧١	١٤	٠,١٠٩	٠,٠٥٩
٧	٠,٢١	٠,١٩٣	١٥	٠,٣٣١	٠,٢٣٩



شكل (٣)

المدرج التكراري لقيم حد العتبة لاشارة الكلام لكلا الجنسين

١٠- مناقشة النتائج:

من خلال الإجراءات التي تم تطبيقها على الخوارزمية المقترحة تبين أن اعتماد معاملات فورير لأجل تمييز جنس المتكلم من خلال إضافة فكرة المواءمة عليها تؤدي

إلى نتائج تعمل على اختزال الوقت من خلال اعتماد معاملات متعدد الحدود والتي تكون قليلة جداً عند مقارنتها بمعاملات فورير، فضلاً عن إمكانية الوصول إلى علاقة رياضية تتحكم بوصف كامل لإشارة صوت المتكلم. من النتائج التي تم الحصول عليها من خلال التطبيق العملي للخوارزمية تبين أن اعتماد فكرة معامل الاقتراب (Correlation Factor) تعطي نتائج اقرب إلى الواقع الحقيقي لجنس المتكلم من خلال تمييز معاملات اشارته مقارنة مع معاملات أخرى تم الاعتماد عليها من قبل خوارزميات سابقة مثل البعد الديكارتي أو معدل مربع الخطأ التراكمي (MSE).

١١- الاستنتاجات والإعمال المستقبلية:

إن الخوارزمية المعتمدة في البحث أعطت نتائج صحيحة وجيدة وبذلك يمكن الاستفادة منها في إمكانية بناء منظومة مبسطة لاختبار جنس المتكلم. يمكن تطوير الخوارزمية أعلاه من خلال إضافة إحدى الشبكات العصبية التي لها القابلية على إيجاد مدى التقارب للنموذج المطلوب اختباره مع النماذج الموجودة في قاعدة البيانات لكلا الجنسين. وأيضاً بالإمكان الاستفادة من أفكار الخوارزمية المعتمدة في البحث لأجل تصنيف الفئات العمرية للمتكلم من خلال جمع عينات كثيرة لفئات عمرية متباينة. إن ما تم الحصول عليه من تمييز لجنس المتكلم يصل إلى حدود مقاربة لما اجرته الباحثة هبة من تمييز لعمر المتكلم وجنسه من خلال بحثها [2]، فضلاً عن أنه تم حساب الوقت في بحثها والذي كان غير محسوس مما يجعله قريباً من الزمن الحقيقي.

قائمة المصادر

أولاً: المصادر باللغة العربية

١. الحيالي، شيماء قيدار عبد الجبار (٢٠٠٥)، كبس الصوت في الزمن الحقيقي، **رسالة ماجستير، علوم الحاسبات، كلية علوم الحاسبات والرياضيات، جامعة الموصل.**

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

٣. الوكيل، هناء محسن احمد (٢٠٠٦)، تصميم نظام مسيطر حركة مضبب لروبوت متنقل ذاتي الحركة، رسالة دكتوراه، قسم علوم الحاسبات، الجامعة التكنولوجية.

٤. يونس، هبة ادريس (٢٠١١)، معاملات المواصفات الصوتية لتحديد العمر ذاتياً، رسالة ماجستير، علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل.

ثانياً: المصادر باللغة الأجنبية

1. Elseom – Cook, Mark, 2001, “Principles of Interactive Multimedia”, The McGraw, Hill companies.
2. Saeed, V., 2002, “Speed Processing Digital Filter”, M.Sc. Thesis. Burnel University.
3. http://www.brunel.ac.uk/depts/ee/com/home speed vaseghi.
4. Mc. Enary, John, 2001, “Computer in Music / What is Sound?”, Lectures / the Classroom Instructional Materials / Orange Coast College in Costa Mesa, California.
5. http://www.occ.cccd.edu/faculty/jmcenary/sound/sound.html.
6. Minasi, Mark, 1996, “The Complete PC Upgrade and Maintenance Guide”, 4th Edition, Sybex-Inc.
7. 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.
8. Childers, D. G. (2000), “Speech Processing and Synthesis Toolboxes”, John Wiley & Sons, Inc.
9. Br”uckl, M. Sendlmeier, W., 2003, “Aging Female Voices: an Acoustic and Perceptive Analysis”, Proc of ISCA workshop Voqual ’03, pp. 163-168.

10. Mustafa, Bassam Ali, 2003, "A Data-Base System for Speaker Identification", Ph.D. Thesis, College of Computers Sciences and Mathematics, University of Mosul.
11. Schötz, S., 2006, "Perception، Analysis and Synthesis of Speaker Age", Ph.D. thesis, Department of Linguistics and Phonetics, Center for Languages and Literature, Lund University.
12. Schötz, S., Müller C., 2007, "A Study of Acoustic Correlates of Speaker Age". In Müller C. (Ed.) Speaker Classification II, Lecture Notes in Computer Science, Springer, pp. 1-9.
13. Harb, H. and Chen, L. 2006, "Gender Identification using a general Audio Classifier". Multimedia Tools and Applications, volume 34, No. 3, 375-395.
14. Blackshaw, M., Steinlauf, E., 2007, "Gender and Age Determination of a Speaker".
15. http://mattblackshaw.com/project/agegender/paper.pdf.
16. Janda J., 2009، "Age Dependence of Children's Speech Parameters', Acta Polytechnica, Czech Technical University Publishing House, Vol. 49, No. 2-3, pp. 40-43.
17. Sedaaghi, M. H. 2009, "A Comparative Study of Gender and Age Classification in Speech Signals". Iranian Journal of Electrical & Engineering. Vol. 5, No. 1.
18. Jean-Francois Bonastre, Frederic Bimbot, et. al., 2003", Person Authentication by Voice: A Need for Caution", eurospeech 2003-Geneva.
19. Baken, R., J., Orlikoff, R. F. , 2000, "Clinical Measurement of Speech and Voice", Edition2, Singular Publishing Group, ISBN 1-56593-869-0، San Diego.
20. Mireia Farr'usi Cabeceran, 2008, "Fusing Prosodic and Acoustic Information for Speaker Recognition", PhD. Thesis, Department of Signal Theory and Communications, Universitat

Polit'ecnica de Catalunya, TALP Research Center, Speech Processing Group.

21. Kinnunen, T., 2003, "Spectral Features for Automatic Text-Independent Speaker Recognition", Licentiate's Thesis, Department of Computer Science, University of Joensuu.
22. Gonzalez, R. C., and wintz, P. (1977) “Digital Signal Processing” Addison-Wesley Publishing Conpany.