طريقة النص المستقل لتحديد هوية المتحدث باستخدام صوته

الدكتور المهندس: حسان محمد أحمد
كلية هندسة الحاسوب والمعلوماتية والاتصالات – الجامعة السورية الخاصة

ملخص

تم في هذا البحث دراسة طريقة النص المستقل (Text-independent) لتحديد هوية الشخص باستخدام صوته (Voice Identification) والمبنية على أساس استخراج الميزات/السمات (Features) الخاصة من الإشارة الصوتية، والتي تميز التنبؤ الخطي (Linear Prediction) لسلوك دالة الترابط الذاتي (Autocorrelation Function) لسبستروم (Cepstrum) الإشارة الصوتية. بني النموذج الصوتي للشخص على أساس متجه الميزات (Features Vector) الخاصة بشكل خليط أكثر معقولية من وحدات غاوص (Maximally Plausible Mixture of Gaussians) التي تعرّف متجه الميزات. نفذت عملية إثبات هوية الشخص بواسطة الصوت بطريقة اختيار النموذج ذات أعظم احتمالية لاحقة (Maximum a Posteriori Probability) لاستعادته بناءً على إشارة الدخل الصوتية.

تعرض الطريقة المدروسة والمقترحة الدقة العالية و الكافية لتحديد شخصية المتحدث بواسطة الصوت و بنص مستقل، بالمقارنة مع النتائج الحاصلة على المستوى العالمي في مثل هذه النظم.

الكلمات المفتاحية: بصمة الصوت، نبرة الصوت، تحديد هوية الشخص، التحقق من هوية الشخص، سمات الصوت، سبستروم الإشارة الصوتية، النموذج الصوتي، النموذج الخلفي العام، نموذج خليط غاوس.

Abstract

In this paper, the text-independent method of person voice identification based on the features extraction from speech signal that characterize the linear prediction of the behavior of the autocorrelation function of the voice signal cepstrum are considered and developed. On the basis of a features vector the person voice model is constructed in the maximum-plausible Gaussian mixture form that describe the feature vector. The Voice identification is executed by selecting model having the maximum of a posteriori probability of its restoration by the input voice signal.

The studied and proposed method demonstrates the higher and sufficient accuracy for speaker personal identification by his voice using text independently, compared with results taking place at the global level in such systems.

Keywords: voice identification, verification, voice features, voice signal cepstrum, Gaussian mixture model.

1- مقدمة

تعتبر اليوم معالجة الصوت وتقنيات الكلام من أهم الاتجاهات الرائجة في الأبحاث العلّمية. يعود الاهتمام الزائد في هذا المجال إلى الطلب الكبير على نتائج تطوير نظم تحليل الكلام، والتي لها أكبر مجال تطبيقات تبدأ من علم كشف الجريمة (Criminalistics)، وتوفير الأمن وتصل حتى المنتجات البسيطة الخاصة بالاستخدام اليومي.

ففي مجال معرفة المتحدث، شاع استخدام البيانات الصوتية في أمرين: الأول، تحديد هوية المتحدث (Speaker Identification)، وغالباً مايستخدم في حالات الجرائم إذ يقارن صوت مسبق التسجيل مع صوت المتهم للتحقق فيما إذا كان الصوت المسجل صادراً عن الشخص نفسه. والثاني، التحقق من هوية المتحدث (Speaker Verification)، وتستخدم هذه التقنية لمطابقة هوية الشخص مع صوته، ومن تطبيقاته الدخول على الحسابات المصرفية وفتح الأبواب والحساب الآلي لساعات عمل الموظفين.

يعتبر استخدام هذه التقنيات جنباً إلى جنب مع غيرها من طرائق معالجة المعلومات الكلامية أمراً ممكناً، على سبيل المثال، لأجل حل مسألة أتمتة مراكز خدمة الهاتف (Call-center) (جرد المكالمات وربطها مع قاعدة بيانات الزبائن، التحليل التلقائي وإحصاء الطلبات).

نقدم في هذا البحث دراسة لعمل طريقة النص المستقل في تحديد هوية الشخص باستخدام صوته ونعرض تحليل دقة عمل هذه الطريقة، والتي أجريت على عدد كبير من البيانات الصوتية الحقيقية المختلفة عن بعضهعا البعض باللغات، وبشروط تسجيل الإشارات الصوتية وبجنس الشخصية (Gender)، ذكر أم أنثى.

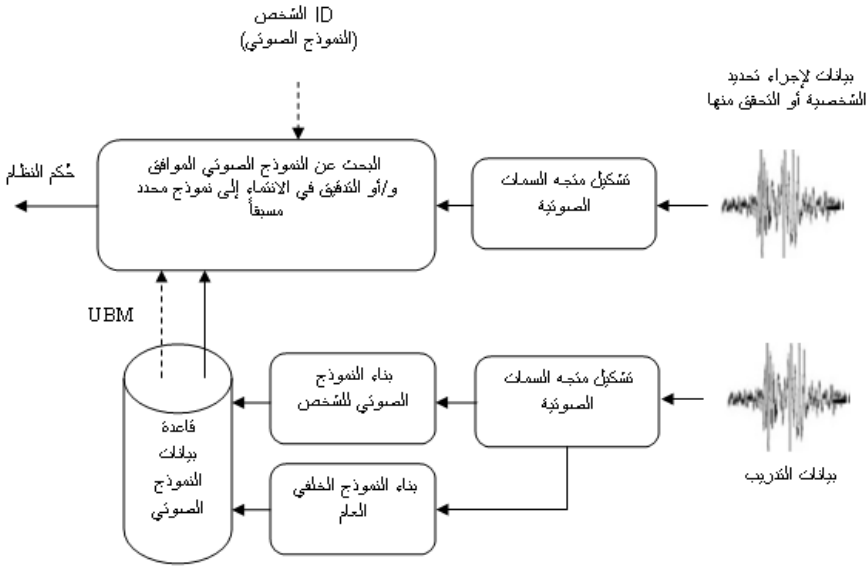
تتمتع هذه الطريقة بالمقارنة مع التقنيات المماثلة الأخرى والمتوفرة حالياً بالسمات المميزة التالية:

- متطلبات منخفضة لنوعية/جودة الإشارة الصوتية.
- التبعية المعتدلة لدقة تحديد الشخصية بشروط تسجيل الإشارة الصوتية، والتي تتراوح قيمتها في حدود ±5 % في حال تغيرات واسعة لشروط التسجيل.

وللمقارنة، فإن دقة الكثير من تقنيات تحديد الشخصية بوساطة الصوت والمتوفرة حالياً تتراوح مابين (1 – 5)% وبنفس تغيرات شروط تسجيل الإشارة الصوتية [14].
يشير مفهوم شروط التسجيل إلى مجموعة أدوات تسجيل الإشارة والوسط السمعي المحيط وصيغة تخزين الإشارة الصوتية.

2- دراسة مرجعية

تعمل نظم تحديد الشخصية والتحقق منها الحالية بوساطة الصوت وفق طورين [14]:
طور التدريب: يتم تحديد السمات المميزة لصوت الشخص والتي على أساسها يُشكل النموذج الصوتي (Voice Model) للشخص، أي بصمة الصوت (Voiceprint)، ومن ثم يتم تخزين هذا النموذج في قاعدة البيانات.
طور العمل: يتم تحديد السمات المميزة لإشارة صوت الشخص وينفذ البحث في قاعدة البيانات عن النموذج الصوتي، الموافق لهذه السمات (تحديد الشخصية)، أو التدقيق في انتماء هذه السمات إلى نموذج صوتي محدد (التحقق من الشخصية). يظهر الشكل (1) المخطط الوظيفي لهذه النظم.



الشكل (1): المخطط الوظيفي لنظام التعرف على الصوت

إضافة إلى ذلك، في طور التدريب يتم أيضاً إعداد ما يسمى النموذج الخلفي العام (Universal Background Model, UBM) الذي يصف السمات الصوتية المتوسطة لجميع المتحدثين الموجودة في قاعدة البيانات.

يعتبر النموذج الخلفي العام إطار عمل فعالاً، وقد وجد نجاحاً كبيراً في التعرف على المتحدث. و هو عبارة عن خليط كبير من وحدات غاوص التي تغطي كل الكلام، وفي سياق التعرف على المتحدث، ويتكيّف هذا النموذج مع كل متحدث باستخدام مخطط أعظم احتمالية لاحقة [12، 13، 15، 17].

وفي طور العمل أيضاً، يتم على أساس النموذج الخلفي العام حساب درجة تميز الإشارة الصوتية، التي تسمح بالحكم عن حقيقة تحديد الشخصية / التحقق منها، والتي تعتبر أيضاً جزءاً من آلية اتخاذ القرار النهائي.

تحظى طريقة النص المستقل لتحديد الشخصية باستخدام الصوت الاهتمام الأكبر في الأبحاث. تعتبر الطريقة مستقلة النص إذا لم تحصل خلال عملها على أية معلومات عن أية جملة أو كلمة محددة سيقوم بلفظها المتحدث.

يعدّ، في الوقت الحالي، بناء النماذج الصوتية على أساس نماذج خليط غاوص (Gaussian Mixture Model, GMM) المنهج الأكثر فعالية لحل المسائل المتعلقة بتحديد الشخصية بطريقة النص المستقل [2، 3، 14]. تُبنى النماذج ذاتها، وكما تم الإشارة إليه سابقاً، على أساس حزمة من السمات الصوتية، والتي يمثل تشكيلها الصعوبة الأساسية. الطريقة الأكثر انتشاراً لبناء السمات الصوتية تتمثل في تشكيل متجه معاملات MFCC (Mel–Frequency Cepstral Coefficient,) من التسجيل الصوتي [1، 2، 9، 10].

مع ذلك، وعلى الرغم من النتائج الجيدة للعمل ضمن الشروط المخبرية، فإن منهجية GMM–MFCC لا يمكن أن تكون مستخدمة لبناء نظم حقيقية للتحديد والتحقق من الشخصية بوساطة الصوت [9، 13، 14].

تعود عدم إمكانية استخدام منهجية GMM–MFCC إلى الأسباب التالية:

1) المتطلبات العالية لنوعية الصوت.

2) تبعية النتائج وبشكلٍ شديد إلى نوع المواد المستخدمة في التدريب (على أساس النوع تشكل قاعدة بيانات للنموذج الصوتي وللنموذج الخلفي العام).

3) تبعية النتائج وبشكلٍ شديد لشروط تسجيل الإشارة الصوتية.

ومن العيوب أيضاً، أن الوقت المطلوب لتشكيل متجه السمات الصوتية يكون كبيراً نسبياً [1، 4].

بالتالي، تظهر الآن الحاجة الملحة لطريقة نوعية لتحديد السمات الصوتية للشخص، والقادرة على العمل مع مواد صوتية ذات نوعية متوسطة (على سبيل المثال، تسجيل مكالمة هاتفية) وتكون أقل حساسية لتغيرات شروط تسجيل الإشارة الصوتية.

3- الطريقة المقترحة

يتجلى جوهر طريقة تحديد والتحقق من الشخصية باستخدام الصوت في استخدام الأساليب، التي تم التوصل إليها لتحديد متجه السمات، أو الميزات الصوتية، وعلى أساسه بناء نماذج صوت الشخص. يتمثل متجه السمات الصوتية بمتجه من الاثني عشر معاملاً (12 معاملاً) الأولى من التنبؤ الخطي لسلوك دالة الترابط الذاتي لسبسترووم الإشارة الصوتية.

يسبق عملية حساب السبسترووم عملية ترشيح نوعي (Filtration) للإشارة الصوتية ضمن مجال نبرة الصوت (Pitch) (عادة تتم عملية ترشيح الصوت في المجال الترددي، [16]) والذي يسمح بتحجيم عناصر التحلل الترددي (Frequency Decomposition) المؤثرة بشكل ضعيف على الميزات الصوتية، و على العكس، بتأكيد المناطق التي تحتوي على المعلومات الأكثر أهمية والتي تميز الخصائص الصوتية الفردية للمتكلم.

تُبنى النماذج الصوتية على أساس متجهات السمات الناتجة بطريقة اختيار نموذج GMM الأكثر معقولية ذي 1024 مكونة، وكذلك نموذج UBM ذي 1024 مكونة.

يتم إجراء عملية تحديد الشخصية (اختيار النموذج الصوتي الأكثر توافقاً مع الإشارة الصوتية المحددة) بطريقة تعظيم الاحتمال اللاحق. عملية التحقق من الشخصية تبدو كمسألة تصنيف ثنائي (Binary Classification) [18، 19، 20]، وتنفذ بطريقة

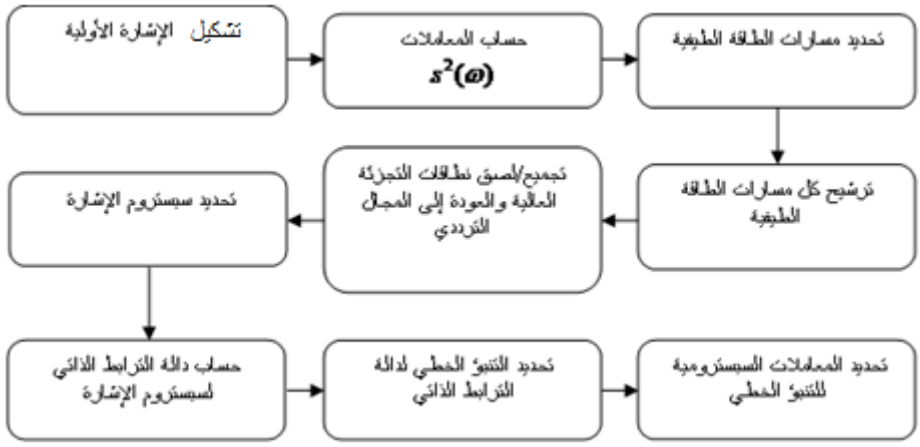
التحقق وبآنٍ واحد من فرضية انتماء الإشارة الصوتية للنموذج الصوتي المحدد مسبقاً ومن فرضية عدم وجود انتمائه للنموذج الخلفي العام.

نستخدم في عملية تنفيذ الطريقة المذكورة إشارات صوتية بتردد أخذ العينات (Sampling Frequency) $f_S = 8000\,\text{Hz}$ (بالمقارنة مع نوعية التسجيل الصوتي بواسطة الجهاز الخليوي) ولمدة زمنية عظمى $t = 20\text{s}$.

3-1- تشكيل متجه السمات

كما ذكرنا سابقاً، يتكون متجه السمات من المعاملات الـ 12 الأولى للتنبؤ الخطي لسلوك دالة الترابط الذاتي لسبستروم الإشارة الصوتية. يسبق حساب السبستروم ترشيح الإشارة الصوتية ضمن مجال ارتفاع نغمة الصوت.

يتم تشكيل متجه السمات الصوتية وفق الخوارزمية الآتية والمبينة في الشكل (2):



الشكل (2): خوارزمية تشكيل متجه السمات الصوتية

1. تُحدد الإشارة الصوتية بمدة زمنية $t = 20\text{s}$ ويحدد تردد أخذ العينات $f_S = 8000\,\text{Hz}$.

2. يُجرى تحويل فوريير السريع (Fast Fourier Transformation, FFT) على إشارة الدخل (الإشارة الأولية) وتحسب القيمة التربيعية للمعاملات الطيفية (Spectral Coefficient) $s^2(\omega)$.

3. يتم تجزئة المجال الترددي $[0,\ 0.5f_S]$ إلى 14 نطاقاً حرجاً لإدراك الصوت (Perception Critical Band)، والتي تتوافق مع التجزئة المتساوية لمجال نغمة الصوت (z، bark) الحاصلة من مقياس التردد (ω، Hz) (Frequency Scale) وفق العلاقة التالية [17، 18، 19، 20]:

$$z = 6\log\left(\frac{\omega}{600} + \sqrt{\left(\frac{\omega}{600}\right)^2 + 1}\right) \qquad (1)$$

ومن ثم تحدد مسارات الطاقة الطيفية (Spectrum Energy Paths) $\ln s^2(z)$ في كافة النطاقات الحرجة.

4. تُنفذ عملية ترشيح المسارات $\ln s^2(z)$ لقطع الطريق أمام المكونات الطيفية، والتي تختلف سرعة تغيرها عن سرعة تغيرات مكونات الكلام الموافقة، وتنفذ أيضاً عملية تمدد مطال (Amplitude Spreading) المعاملات الطيفية والتي تحتوي على السمات الصوتية الأكثر وضوحاً. يتميز المرشح الذي تم استخدامه بدالة تحويل متقطعة من الشكل:

$$\Phi(z) = 0.1z^4\,\frac{1+z^{-1}-3z^{-3}-2z^{-4}}{1-0.9z^{-1}} \qquad (2)$$

5. يتم تجميع طيف الطاقة $\ln s^2(z)$ من 14 نطاقاً حرجاً، ومن ثم إعادة صياغته في مقياس التردد الخطي $\ln s^2(\omega)$.

6. يُنفذ تحويل فوريير السريع العكسي (Inverse Fast Fourier Transform, IFFT) من طيف الطاقة، والذي بنتيجته نحصل على السبسترום $C_s(q)$ المميز لسمات الطاقة الترددية للإشارة الأولية في حيز الزمن السبسترومي q (المرتبط بالتردد).

7. تحسب دالة الترابط الذاتي $R_c(k)$ من السبسترום $C_s(q)$:

$$R_c(k) = \sum_q M\left[C_s(q)\cdot C_s(q-k)\right] \quad (3)$$

حيث $M[\cdot]$ – عملية حساب التوقع الرياضي (Mathematical Expectation).

على اعتبار أن الطريقة تفترض حساب 12 معاملاً سبسترومياً، فإن حساب دالة الترابط الذاتي تكون ممكنة فقط من أجل $k = 1\ldots13$.

8. إذا عبَّرنا عن قيم دالة الترابط الذاتي $R_c(1)\ldots R_c(11)$ بشكل مصفوفة توبلتز (Toeplitz Matrix):

$$T = \begin{bmatrix} R_c(1) & R_c(2) & \ldots & R_c(11) \\ R_c(2) & R_c(1) & \ldots & R_c(10) \\ \vdots & \ddots & \ddots & \vdots \\ R_c(11) & \ldots & R_c(2) & R_c(1) \end{bmatrix} \quad (4)$$

وعن مسألة حساب التنبؤ الخطي بالشكل:

$$\begin{bmatrix} R_c(1) & R_c(2) & \ldots & R_c(12) \\ R_c(2) & R_c(1) & \ldots & R_c(11) \\ \vdots & \ddots & \ddots & \vdots \\ R_c(12) & \ldots & R_c(2) & R_c(1) \end{bmatrix} \begin{bmatrix} a_2 \\ a_3 \\ \vdots \\ a_{13} \end{bmatrix} = \begin{bmatrix} -R_c(2) \\ -R_c(3) \\ \vdots \\ -R_c(13) \end{bmatrix} \quad (5)$$

فإنه بإمكاننا تحديد معاملات التنبؤ الخطي لسلوك دالة الترابط الذاتي باستخدام خوارزمية ليفينسون– دوربين التكرارية (Levinson–Durbin Recursion Algorithm) المجدية من الناحية الحسابية [5].

9. يجرى حساب المعاملات السبسترومية للتنبؤ الخطي من خلال العلاقات المتكررة

$$c_1 = -a_2; \quad c_{n+1} = -a_n - \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}, \qquad n = 1,\ldots,11 \quad (6)$$

بالتالي، نحصل على المتجه النهائي للسمات الصوتية $X = \{c_1,\ldots,c_{12}\}$ ، والذي يصف بشكل جيد السمات الصوتية الفردية للشخص المشروطة/المرتبطة بطبيعته الفيزيولوجية وقناته الصوتية (Vocal Tract)، وغير المرتبطة بالمعلومات الكلامية الملفوظة من قبله.

3-2- تشكيل النماذج الصوتية

لبناء النماذج الصوتية على أساس متجهات السمات الناتجة يُستخدم نموذج GMM المركب الأكثر معقولية ذي 1024 مكونة. تكمن الفكرة الأساسية لتركيبة GMM في تمثيل كثافة توزيع متجه السمات الصوتية X بشكل المجموع الموزون لكثافات غاوص للتوزيع (Gaussian density of distribution) [14]:

$$p(X)=\sum_{m=1}^{M}\alpha_m p_m\left(X,\mu_m,D_m\right) \tag{7}$$

حيث: α_m – وزن مكونة الخليط ذات المؤشر m؛ $p_m\left(X,\mu_m,D_m\right)$ – كثافة غاوس للتوزيع ذات توقع رياضي μ و مصفوفة التباين D (covariance matrix)، وتتمثل هذه الكثافة بالشكل:

$$p_m\left(X,\mu,D\right)=\frac{1}{\sqrt{2\pi\det D}}\exp\left(-0.5(X-\mu)^T D^{-1}(X-\mu)\right) \tag{8}$$

في الحقيقة، تمثيل الكثافة $p(X)$ بشكل مجموع M غاوسات يتوافق مع تجزئة مجموعة المحددات (parameters) الصوتية إلى M من الأصناف الفرعية $(M=1024)$.

من الجدير بالذكر أيضاً، أنه ليس مهماً بالنسبة لنماذج غاوس ترتيب تعاقب الإشارات الصوتية المحددة الواحدة تلوى الأخرى، وذلك بسبب أن هذه التركيبة تعمل مع الإحصائيات المتراكمة للمحددات.

تتمثل مسألة التأكد من المستخدم بواسطة الصوت بالتصنيف الثنائي، وشكلياً بالتحقق من فرضيتان:

الأولى: H_0 – عبارة Y لفظها الشخص S؛

الثانية: H_1 – عبارة Y ليس الشخص S من لفظها.

تعدّ نسبة الأرجحية (Likelihood Ratio) التدقيق المثالي لاختيار واحدة من الفرضيتين الاثنتين. بذلك، تبدو عملية اتخاذ القرار على النحو التالي:

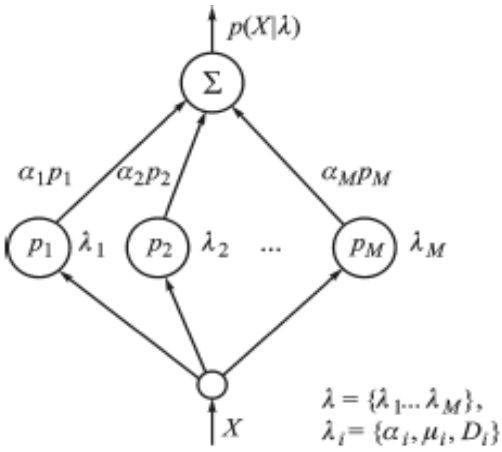
$$\frac{p\left(Y\mid H_0\right)}{p\left(Y\mid H_1\right)}\begin{cases}\geq\theta\rightarrow \text{ accept } H_0\\ <\theta\rightarrow \text{ accept } H_1\end{cases} \tag{9}$$

حيث: $p(Y|H_0)$ – دالة كثافة احتمال الفرضية ؛ والمقيّمة على المقطع الكلامي Y؛ θ – عتبة اتخاذ القرار. رياضياً، يمكن أن تكون الفرضية H معرّفة بالنموذج λ، الذي يميز المتكلم 3 في حيز السمات.

يتم إنشاء نموذج صوتي لكل متكلم على أساس التسجيل الحاصل لكلامه. لأجل الفرضية H_1 يُبنى نموذج خلفي عام (UBM) والذي يميز كل ما يمكن من الناس المتكلمين في كافة السياقات الممكنة. يتم تدريب هذا النموذج على عدد كبير من البيانات الصوتية المتوازنة وفق نوع الجنس (مثال، 450 ذكور + 750 إناث)، وعلى الأجهزة المستخدمة و شروط تسجيل الإشارة الصوتية.

بالتالي، نماذج GMM يجب أن تكون مُدَرَبة بشكل مستقل لأجل كل شخص، أي أنه لكل شخص يجب أن يكون قد وجد مجموعة من المحددات $\lambda = \{\alpha_i, \mu_i, D_i\},\ \ i = 1 \ldots M$، الشكل (3).

تعدّ متجهات السمات الصوتية $X = \{c_1, \ldots, c_{12}\}$ المعطيات الأولية للتدريب. إضافة إلى ذلك، تتم عملية تدريب نماذج GMM وفق خوارزمية الأرجحية العظمى المبنية على أساس خوارزمية التوقع – التعظيم (Expectation–Maximization, EM) [8].



الشكل (3): خوارزمية تدريب نموذج GMM لكل شخص

يمكن أن تقيّم أرجحية النموذج λ على متتالية متجهات التدريب $T=\{X_1,\ldots,X_T\}$ وفق العلاقة

$$p(T\mid\lambda)=\prod_{t=1}^{T}p(X_t\mid\lambda) \qquad (10)$$

تنحصر فكرة خوارزمية الأرجحية العظمى في التغير المتتالي لمحددات النموذج $\lambda_n\rightarrow\lambda_{n+1}$ بحيث $p(T\mid\lambda_{n+1})\geq p(T\mid\lambda_n)$ إلى حين، حتى الوصول إلى عتبة التقارب (convergence) أو حتى تتوقف الخوارزمية. يجري تقييم الأرجحية العظمى في المنهجية المدروسة باستخدام خوارزمية باوم- ويلش (Baum–Welsh) والتي تستخدم بشكل اعتيادي لإيجاد محددات نماذج غاوس (GMM) [6]، الشكل (3).

بشكلٍ مماثل، يتم تشكيل النموذج الخلفي العام λ_{UBM} ماعدا أن متتالية متجهات التدريب T يتم تشكيلها من جميع متجهات السمات الصوتية الممكنة X.

3-3- تحديد الشخصية والتحقق منها

لتكن $G=\{S_1,\ldots S_k\}$ مجموعة من الأشخاص والتي تتمثل ببصمات صوتهم $\Psi=\{\lambda_1,\ldots,\lambda_k\}$ في نظام تحديد الشخصية في قاعدة بيانات GMM. يجري تحديد أياً من النماذج في القاعدة Ψ أكثر توافقاً مع متجه ما من السمات X، بطريقة اختيار النموذج λ_m الذي يكون لديه أعظم احتمالية لاحقة:

$$\hat{S}=\arg\max_{1\leq m\leq k}\Pr(\lambda_m\mid X)=\arg\max_{1\leq m\leq k}\frac{p(X\mid\lambda_m)\Pr(\lambda_m)}{p(X)} \qquad (11)$$

أو الأخذ بعين الاعتبار الاحتمال المتساوي لظهور كل شخص من قاعدة النماذج الصوتية

$$\hat{S}=\arg\max_{1\leq m\leq k}p(X\mid\lambda_m) \qquad (12)$$

بعد اختيار النموذج الصوتي الأكثر توافقاً λ_m يتم إجراء عملية التحقق:

$$\frac{p(X \mid \lambda_m)}{p(X \mid \lambda_{UBM})} \begin{cases} \geq \theta \rightarrow X \text{ be up to } \lambda_m \\ < \theta \rightarrow X \text{ be up to NOT } \lambda_m \end{cases} \quad (13)$$

حيث تم اختيار العتبة $\theta = 1.65$ تجريبياً كعتبة مثالية. من الجدير بالتنويه إلى أن قيمة هذه العتبة تغيرت بشكل غير ملحوظ حسب الطرق المختلفة لإجراء الاختبارات.

4- النتائج التجريبية

نفذّت وطبقّت الطريقة المدروسة لتحديد و التحقق من الشخصية بواسطة الصوت بشكل كامل في وسط بيئة برمجيات ماتلاب (Matlab) [11]. تم استخدام معطيات تقييم نظم التعرف على المتكلمين NIST SRE data للأعوام 2004، 2006، 2008، وذلك كمواد للتدريب والاختبار [7، 20]، والتي منها تم اختيار الأصوات المسجلة (phonogram) للمتكلمين الذين لديهم 6-10 تسجيلات لخطابهم الصوتي وبطول زمني من 16 ثانية لكل تسجيل، الجدول (1).

تحتوي الأصوات المسجلة على عدد كبير من العبارات المتنوعة، والتي لفظت بلغات مختلفة وفي ظروف أجواء سمعية مختلفة أيضاً (مبنى، شارع...الخ).

الجدول (1) قاعدة الأصوات المسجلة المستخدمة

نوع الجنس	العدد الكلي للمشاركين وأصواتهم المسجلة	القنوات		
		هاتف – هاتف	مكرفون – مكرفون	هاتف – مكرفون
ذكور	متكلمون	475	95	92 هاتفاً+95 مكرفوناً
	أصوات مسجلة	3930	910	1375
إناث	متكلمون	625	125	120 هاتفاً+125 مكرفوناً
	أصوات مسجلة	5150	1175	1830

اختبرت كل الأساليب الممكنة لتدريب النموذج الخلفي العام (UBM) و شكلت النماذج الصوتية لكل المتكلمين واختبر نظام تحديد الشخصية على هذه المعطيات. جرى اختيار

الأصوات المسجلة لأجل التدريب والاختبار عشوائياً من المتوفرة، بحيث الأصوات المستخدمة في بناء النماذج الصوتية لا تشترك في عملية الاختبار.

لأجل تدريب النموذج الخلفي العام (UBM) تم اختيار أصوات إضافية من قاعدة بيانات NIST SRE المشار إليها سابقاً، والتي لم تستخدم في تشكيل النماذج الصوتية ولا في عملية الاختبار، الجدول (2).

الجدول (2) عدد الأصوات لتدريب نموذج UBM

نوع الجنس	**القنوات**		
	هاتف – هاتف	مكرفون – مكرفون	هاتف – مكرفون
ذكور	645	870	605
إناث	735	1340	595

بنتيجة التجربة، تم تحديد احتمالية دقة تحديد شخصية المتكلم بوساطة التسجيل الصوتي وفق المعطيات المختلفة لتدريب نموذج UBM و معطيات الاختبار. الجداول (3)، (4) و (5) تبين النتائج الحاصلة.

الجدول (3) دقة تحديد الشخصية في القناة هاتف–هاتف

معطيات تدريب UBM	بيانات الاختبار		
	ذكور	إناث	ذكور + إناث
ذكور	%96	–	–
إناث	–	%95.6	–
ذكور+ إناث	%95.3	%94.8	**%93.8**

الجدول (4) دقة تحديد الشخصية في القناة مكرفون – مكرفون

معطيات تدريب UBM	بيانات الاختبار		
	ذكور	إناث	ذكور + إناث
ذكور	%97.1	–	–
إناث	–	%97.8	–
ذكور+ إناث	%96.2	%96.7	**%94.4**

الجدول (5) دقة تحديد الشخصية في القناة هاتف– مكرفون

معطيات تدريب UBM	بيانات الاختبار		
	ذكور	إناث	ذكور + إناث
ذكور	%92.1	–	–
إناث	–	%92.8	–
ذكور+ إناث	%91.5	%92.1	**%90.3**

من الجداول (3) و(4) و(5) نستنتج التالي:

- الطريقة المدروسة والمقترحة تعرض الدقة العالية و الكافية لتحديد شخصية المتحدث بواسطة الصوت ذي النص المستقل، بالمقارنة مع النتائج الحاصلة على المستوى العالمي في مثل هذه النظم [1، 4، 14].
- يلاحظ تبعية معقولة لدقة التحديد من الشروط التي تم وفقها تسجيل الصوت لأجل عمليات التدريب وتشكيل النماذج الصوتية، وكذلك لأجل الاختبار، والتبعية لنوع الجنس في تركيبة قاعدة النماذج الصوتية.

يبين تحليل النتائج النهائية والمتوسطة أن عدداً ملحوظاً من الأخطاء يحصل بنتيجة الاختيار الخاطئ للنموذج الصوتي λ_m الموافق لمتجه السمات X. تبرر أهمية هذه الملاحظة بأن النظام يمكن أن يعطي إشارة خاطئة حتى في حالة بناء النماذج الصوتية الفعلية على حساب فقط الجهاز غير الكامل وحده لاختيار النموذج الصوتي المحدد λ_m (أو الإشعار إلى عدم وجود هكذا نموذج) الموافق لمتجه السمات X.

5- الخلاصة

في هذا البحث، تم تطوير طريقة النص المستقل لتحديد هوية الشخص باستخدام صوته والتي دقة عملها مماثلة لدقة عمل النظم الصوتية العالمية الرائدة.

تتميز هذه الطريقة عن غيرها بخاصية التبعية المعقولة لشروط تسجيل الإشارات الصوتية (أدوات التسجيل، الوسط المحيط، قنوات نقل الإشارة).

يمكن اعتماد هذه الطريقة في قاعدة عمل نظم تحديد والتحقق من هوية الشخص بواسطة صوته كتطبيق تجاري، وكذلك في نظم التحكم بالدخول الحقيقي إلى أماكن محددة والتي تتطلب الحماية العالية.

المراجع

1. REYNOLDS, D, 1994 Experimental evaluation of features for robust speaker identification. IEEE Trans. On Speech and Audio Processing. Vol. 2. No. 4, 639–643.
2. BIMBOT, F, A, 2004 tutorial on text-independent speaker verification. EURASIP J. on Applied Signal Processing. No. 4, 430–451.
3. REYNOLDS, D; ROSE, R, 1995 Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. On Speech and Audio Processing. No. 3, 72–83.
4. HERMANSKY, H; MORGAN, N, 1994 RASTA processing of speech. IEEE Trans. On Speech and Audio Processing. Vol. 2. No. 6, 578–589.
5. MUSICUS, B, 1998 Levinson and fast Choleski algorithms for Toeplitz and Almost Toeplitz Matrices. RLE TR, MIT. No. 538.
6. WELCH, L, 2003 - Hidden Markov Models and the Baum-Welch algorithm. IEEE Information Theory Society Newsletter,.
7. NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, NIST, 2004, 2006, 2008, Speaker Recognition Evaluation. <http://www.itl.nist.gov/iad/mig/tests/sre/>
8. GEOFFREY, J.; McLACHLAN; THRIYAMBAKAM KRISHNAN, 1997 - The EM Algorithm and Extensions. A Wiley-Interscience Publication. JOHN WILEY & SONS, INC.
9. BAGUL, S, G; SHASTRI, R, K, 2012, Text Independent Speaker Recognition System using GMM. International Journal of Scientific and Research Publications. Vol. 2, Issue 10.
10. ALFREDO MAESA; FABIO GARZIA and others, 2012 Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models. Journal of Information Security. No.3, 335-340.
11. MIKE BROOKES. VOICEBOX: Speech Processing Toolbox for MATLAB. Department of Electrical & Electronic Engineering, Imperial College, Exhibition Road, London SW7

2BT, UK. <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
12. DOUGLAS, R, Universal Background Models. MIT Lincoln Laboratory 244 Wood St., Lexington, MA 02140, USA.
13. REYNOLDS, D; QUATIERI, T; DUNN, R, 2000 Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing10 (1), 19–41.
14. SELVA NIDHYANANTHAN, S; SHANTHA SELVA KUMARI, R, 2013 Language and Text-Independent Speaker Identification System Using GMM. Issue 4, Volume 9.
15. TOMI KINUNEN; HAIZHOU LI, 2010 An overview of text-independent speaker recognition: From features to supervectors. International journal of Speech Communication Speech Communication 52 , 12- 40.
16. BASHAR, M A,; TOFAEL AHMED, M; SYDUZZAMAN, M; PRITAM JYOTI RAY; TOUHIDUL ISLAM, A Z M, 2014 Text Independent speaker identification system using average pitch and formant analysis. International Journal on Information Theory (IJIT), Vol.3, No.3.
17. MOHAMED KAMAL OMAR and JASON PELECANOS, 2010 Training Universal Background Models for Speaker Recognition. The Speaker and Language Recognition Workshop, 28 June – 1 July, Brno, Czech Republic.
18. DOMINIQUE GENOUD, MIGUEL MOREIRA, EDDY MAYORAZ, 1998 Text dependent speaker verification using binary classifiers. IDIAP Research Report 97-08. Proceedings of the International Conference on Automatic Speech and Signal Processing, ICASSP'.
19. HUNG-SHIN LEE; YU TSO; YUN-FAN CHANG; HSIN-MIN WANG, 2014 Speaker verification using kernel-based binary classifiers with binary operation derived features. Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on.
20. JOSEPH KESHET; SAMY BENGIO, 2009 - Automatic Speech and Speaker Recognition. Large Margin and Kernel Methods. John Wiley & Sons Ltd.