نظام آلي لتمييز اللكنات

إسراء جاسم حرفش الرحماني(*) عبدالكريم عبدالوهاب حسن(**)

الملخص

تناولنا في هذا البحث بناء نظام تمييز لكنات آلي(Automatic Accents Recognition)على مجموعة مغلقة من المتكلمين، إن النظام يقوم بتمييز مجموعة من اللكنات الإنكليزية من أشخاص من أصل عربي وأميركي وبريطاني.إن اللكنات هنا تتكون من ستين مقطع صوتي ،وتكون جميع المقاطع الصوتية مقسمة إلى مجموعتين مجموعة التدريب ومجموعة الاختبار حيث يقوم النظام بتمييز لكنة المقطع الصوتي الذي يختاره المستخدم من مجموعة الاختبار بالاعتماد على الصفات المميزة التي يستخرجها منه ويقوم بمقارنتها مع الصفات المميزة التي قام باستخراجها وتخزينها على شكل قاعدة بيانات من بيانات التدريب.

في مرحلة استخراج الصفات المميزة لكل إشارة صوتية استخدمنا لهذا الغرض معامل الطيف الترددي، (Mel frequency cepstral coefficients)والذي يعتبر من المعاملات الشائعة الاستعمال في هذا المجال .

أما في مرحلة التمييز أي تمييز اللكنه المدخلة، فهنا أجرينا تجارب على ثلاثة نماذج من طرق التمييز طبقناها كلا على حدة منها إحصائية وهي التحليل التمييزي Linear discriminant analysis ومسافة الارتباط Correlation distance ومنها ما يعتمد قياس المسافة وهي إزاحة الوقت الديناميكية Dynamic Time Wrapping ، ولكل طريقة من هذه الطرق الثلاث تعتمد قدرتها في التمييز على مجموعة من الأمور تم ملاحظتها خلال العمل منها ما يعتمد على كمية بيانات التدريب ومنها ما يعتمد بكثرة على مدى ارتباط البيانات في نفس الصنف .

الكلمات المفتاحيه : اللكنة، تمييز اللكنة، التحليل التمييزي، مسافة أو بعد الارتباط.

المقدمـــــة

إن التحدي الأساسي للبحوث الحديثة في علم تكنولوجيا الكلام هو في فهم ونمذجة التغير في لغة الفرد المنطوقة إذ أن الأفراد لهم أساليب في الكلام (اعتمادًا على عدة عوامل) والتي تتمثل باللغة المحلية (العامية)، لهجة المتكلم،الخلفية الاجتماعية للمتكلم، والأقليات العرقية ومتغيرات سياق الكلام. لذا فإن الأنماط المحددة في لفظ المتكلمين تدخل كمميز لأصواتهم وبالتالي للتغيرات في الكلام[٢].

أن اللكنة فهي تتعلق بطريقة إخراج الحروف والأصوات وتظهر في كلام كل منّا بأي لغة وأي لهجة يحكى بها بغض النظر عن تواجده في أي قطعة من العالم، وقليل من يستطيع التخلص منها بشكل تام، فمثلاً، يقال أن فلان يتحدث العربية بلكنة أمريكية وليس صحيحاً أن يقال أن فلان يتحدث العربية بلهجة أمريكية. و اللكنة تخص الفرد فقط أي أن لكل فرد لكنة خاصة به.

إذا فإن تمييز اللكنه هو تصنيف إلى المتكلم إلى أي مجموعة ينتمي إليها، هذه المجاميع تعرف وفق: المناطق الجغرافية، أو وفق تصنيف اجتماعي ـ اقتصادي ، أو وفق الانتماء العرقي أووفق اللغة الثانوية للمتكلم. إن المستمع المتحسس للكنة يمكن أن يخبر إلى أي مجموعة ينتمي إليها المتكلم . لذا ولتكون عملية التمييز ممكنة هنا فإنه يوضع مقطع كلامي يتكرر بين المتكلمين. ومن الجدير بالذكر أن تكنولوجيا معالجة الكلام تتعامل بشكل كافي مع تغيرات التلفظ ضمن أصناف اللكنات[٤][١]

(*) كلية العلوم، جامعة البصرة، العراق.
(**) كلية العلوم، جامعة البصرة، العراق.

نظام تمييز اللكنة الآلي المقترح

فيما يلي توضح لمسيرة العمل التي اتبعت في بناء النظام:

تحظير البيانات

بالنسبة أن البيانات المستعمله هنا تم تحميلها من قاعدة بيانات The Speech Accent Archive، إذ أن معدل النمذجة هنا هو ٤٤١٠٠ H، ومعدل التفاصيل ١٦ bit، وتثبيت القناة Mono.حيث كما قلنا استخدمت في النظام لتمييز مجموعة من اللكنات لأشخاص من جنسيات عدة (العربية، والبريطانية، والأمريكية) وهم يتحدثون نصًا ثابتًا باللغة الإنكليزية أي أنه معتمد على النص (text-dependent) ، والنص المستخدم هو:

((Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station)).

تنقية الإشارة باستخدام الترشيح Filtering

استعملت هنا دالة Butterworth في Matlab يمكن استخدامها كمرشحات مرور منخفض واطئ Low pass Filters، أو مرشحات مرور عالHigh pass، أو مرشحات مرورBand pass؛وذلك حسب المعلمات المدخلة إلى الدالة، وقد استخدمناها في نظامنا كـ Band pass لإزالة الترددات التي هي أقل أو أعلى من ترددات الكلام حيث أن الترددات بين ٠ـ ٢٠٠، والترددات بين ٦٤٠٠ ـ ٨٠٠٠ لا تتم إزالتها بعملية التعيين Sampling.

حذف مناطق السكون

إن إشارة الكلام تحتوي على مناطق كثيرة ساكنه ليس لها فائدة في عملية استخراج الصفات،بل إنها قد تقلل من كفاءة هذه المرحلة وتزيد من الحسابات والوقت في مرحلة استخراج الصفات ومرحلة التمييز؛ لذا من المفروض إزالة هذه المناطق من الإشارة، وهنالك تقنيات عدة تستخدم لهذا الغرض منها ما يعتمد على طاقة الإشارة، أو التردد، أو معدل التقاطع الصفري.

استخدمنا في هذا البحث معيار معدل الطاقة لكل إطار Frame حيث أننا وبالتجربة قمنا بتحديد عتبة Threshold للتمييز بين السكوت والكلام وكالآتي:

ـ بعد تقسيم الإشارة الصوتية إلى إطارات، يتم حساب معدل الطاقة لكل إطار

$$AVG_K = (\sum_{i=1}^{n} abs(x_i))/n \quad \ldots(1)$$

حيث n: طول الإطار ، x_i هي الإشارة في الوقت i

ـ تقارن AVG_K بالعتبة التي استخرجت بالتجربة وهي ٠,٠٠٢٣ فإذا كان معدل طاقة الإطار أقل أو مساوي لمقدار العتبة يتم حذف الإطار، وبخلافة يحتفظ بالإطار.

استخلاص الصفات المميزة Features Extraction

استخدمنا معامل الطيف الترددي Mel Frequency Cepstral Coefficients في عملية استخلاص الصفات المميزة من إشارة الكلام بسبب قدرتها المميزة في تمييز اللهجات واللكنات، وفيما يلي سنوجز الخطوات التي قمنا بعملها للحصول على MFCC [٧][٦].

ـ تطبيع الإشارة لإزالة القيم الشاذة منها وحسب المعادلة الآتية:

$$x(i)=x(i) - 0.95 * x(i1-) \quad \ldots (2)$$

حيث i هو تسلسل العينة في إشارة الكلام X

ـ ثم إن الإشارة يجب أن تكون مقسمه إلى إطارات لتسهيل عمليات المعالجة والحصول على إشارة مستقرة نسبياً.

ـ لحساب تحويل فورير المتقطع تم استعمال تحويل فورير السريع (Fast Fourier Transformation) أو (FFT):

$$X[K] = \sum_{n=0}^{N-1} X[n] e^{-j\frac{2\pi}{N}kn} \quad \ldots(3)$$

ـ تحويل القيم التي حصلنا عليها إلى مقياس الـ MEL وحسب المعادلة :

$$Mel(f) = 2595\, log_{10}\left(1 + \frac{f}{700}\right) \quad \ldots(4)$$

ـ أخذ اللوغاريتم لكل قيم MEL التي حصلنا عليها.

ـ المرحلة الأخيرة في عملية الحصول على معامل الطيف الترددي هو تحويل جيب التمام المتقطع Discrete Cosine Transformation الذي نجريه على القيم التي حصلنا عليها من الخطوة السابقة.

ومن الجدير بالذكر هنا أننا وجدنا بالتجربة أن معاملات MFCC بطول ١١ معاملا ملائمة لإيجاد نسب مطابقة جيدة.

مرحلة التمييز Classification stage

هناك العديد من طرق التمييز المستخدمة لمعالجة تمييز الإشارة الصوتية منها طرق تعتمد على قياس المسافة Distance Measurement،ومنها طرق إحصائية وغيرها، وفيما يلي سنعرض بعضا من هذه الطرق التي استخدمت في التمييز الآلي في هذا البحث. ومن الجدير بالذكر هنا أننا في كل طريقة من طرق التمييز أدناه أجرينا التجارب الآتية:

١ ـ التمييز بين أشخاص سعوديين الأصل وأشخاص من المملكة المتحدة.

٢ ـ التمييز بين أشخاص سعوديين الأصل وأشخاص من الولايات المتحدة.

٣ ـ التمييز بين أشخاص من الولايات المتحدة وأشخاص من المملكة المتحدة.

٤ ـ التمييز بين اللكنات الثلاث.

التحليل التمييزي أو التصنيفي Discriminant Analysis

إن التحليل التمييزي أو التصنيفي يستخدم عندما يكون لدينا مشاهدات من مجاميع محددة مسبقاً، وتحتوي هذه المجاميع على إثنين أو أكثر من المتغيرات predictors حيث تقوم هذه التقنية بتوليد تركيبة خطية من المتغيرات التي تعظم احتمالية تنسيب المشاهدة إلى مجموعتها المحددة مسبقاً، أو التي بإمكانها تصنيف مشاهدة جديدة إلى أحد المجاميع. كما أنه يمكننا من معرفة أي المتغيرات لها التأثير الأكبر في التمييز بين المجاميع [٣].

إن المشكلة الإحصائية هنا تكمن في كيفية ايجاد دالة تمييزية Discriminant Function على وفق المعايير أو القياسات التي يمكن الحصول عليها من العينات، والتي بوساطتها يمكن تصنيف، أو تمييز العينات الجدد (المجهولة الانتماء) إلى المجموعة الصحيحة.

إن عدد دوال التمييز يعتمد على عدد المجاميع ويكون دائماً «قل منها بواحد، كما» ن عدد المتغيرات وطريقة اختيارها تحدد دقة التمييز. وفيما يلي خطوات الحصول على الدالة التمييزية [١٢]:

١ ـ ايجاد متوسط كل متغير في كل مجموعة، ثم ايجاد الفرق بين متوسطي كل متغير في المجموعتين (على فرض ان المسألة تحتوي على مجموعتين).

$$M = (\sum_{i=1}^{n} X1(i))/n \qquad ..(5)$$

حيث M: متوسط المتغير X1

X1 : هو المتغير الاول في المجموعة الاولى.

n:عدد العناصر في المتغير X1.

٢ ـ تكرر الخطوة (١)على المتغيرات جميعها في المجموعتين الأولى والثانية. ومن ثم يحسب d الآتي:

$$d_{ii} = \sum x_i^2 - \frac{(\sum x_i)}{n} \qquad ....(6)$$

$$d_{ii} = \sum x_i x_j - \frac{(\sum x_i)(\sum x_j)}{n} \qquad ....(7)$$

٢ ـ ايجاد مصفوفة التباين والتباين المشترك المدمج (داخل المجاميع):

$$V_{ii} = \frac{d_{ii}(1) + d(2)}{n1 + n2 - 2} \qquad ...(8)$$

$$V_{ij} = \frac{d_{ij}(1) + d_{ij}(2)}{n1 + n2 - 2} \qquad ...(9)$$

حيث V_{ii} تشير إلى المجموعة الأولى، و V_{ij} إلى المجموعة الثانية.

وأخيرًا فإن دالة التمييز L هي توليفه خطية من المتغيرات تكتب كالآتي:

$$L = \alpha 1\, X1 + \alpha 2\, X2 + \alpha 3\, X3 \ldots \alpha n Xn \qquad ...(10)$$

حيث

L: هي دالة التمييز

Xn...X2,X1 : هي المتغيرات Predictors المستخدمة في عملية التمييز وأن $\alpha_1, \alpha_2, \ldots, \alpha_n$ التي تختار بحيث تعطي أعلى تمييز بين المجموعتين.فإذا رمزنا لنسبة الاختلافات بين المجموعتين إلى الاختلافات داخل المجموعتين بالرمز λ مثلاً أي أن:

$$\lambda = \frac{Between-Groupvariation}{Within-Groupvariation} \qquad ...(11)$$

فأننا نختـــار $\alpha_1, \alpha_2, \ldots, \alpha_n$ بحيث تكون λ أكبر ما يمكن. حيث أن المعادلة الطبيعية لإيجاد α هي:

$$V\,\alpha = d \qquad (12)$$

وبالتعويض يمكننـــا الحصول على قيم α، حيث أن قيم d و V أصبحتا معلومة، هذا ويمكننا حساب أهمية كل متغير αi* في دالة التمييز كالآتي:

$$\alpha_i^* = \alpha i\sqrt{Vii} \qquad (13)$$

في الجدول (١) توضيح لنتائج تطبيق LDA .

الجدول (١) نتائج تطبيق النظام على اللكنات الإنكليزية بطريقة LDA

LDA – English Accents		
No.	Experiment	Success Percentage
1	KSA-UK	82.15%
2	KSA-USA	85%
3	USA-UK	54.29%
4	KSA-USA-UK	50%

إزاحة الوقت الديناميكية Dynamic Time Warping

إزاحـــة الوقـــت الديناميكية (DTW) هـــي خوارزمية معروفة تطورت وزاد اســـتخدامها بعد استخدامها في أنظمة تمييز الكلام.Animation.وتستخدم خوارزمية DTW لإيجاد التشابه بين متسلسلتين زمنيتين Time Series ممكن أن تكونان مختلفتين في الزمن، أو السرعة.[٥][١٢].

ولمحـــاذاة المتسلســـلتين (x1)X=(x1,x2,...,xn و (y1)Y=(y1,y2,...,ym يجب :

١ ـ علينـــا في البداية ايجاد مصفوفة الكلفة (Cost Matrix d (qi,cj لهما حيـــث أن كل عنصر من عناصرها يحتوي على المســـافة بين النقطتين qi، والتي يمكن حسابها إما بطريقة اقليدس العادية المستمدة من نظرية فيثاغورس [٨][١٠]:

$$d_{x,y}^2 = (x1 - y1)^2 + (x2 - y2)^2 \qquad (14)$$

$$d\,x,y = \sqrt{(x1 - y1)2 + (x2 - y2)2} \qquad (15)$$

سنقوم بكتابة المعادلة بالشكل الآتي:

$$d(qi,cj) = (qi - cj)^2 \qquad (16)$$

أو باستخدام طريقة مانهاتن

$$d(qi,cj) = (qi - cj) \qquad (17)$$

٢ ـ بنـــاء أفضل طريق (أقل كلفـــة) (pl,...,p2,p= p1) من مصفوفة الكلفة، وحسب الدالة الآتية:

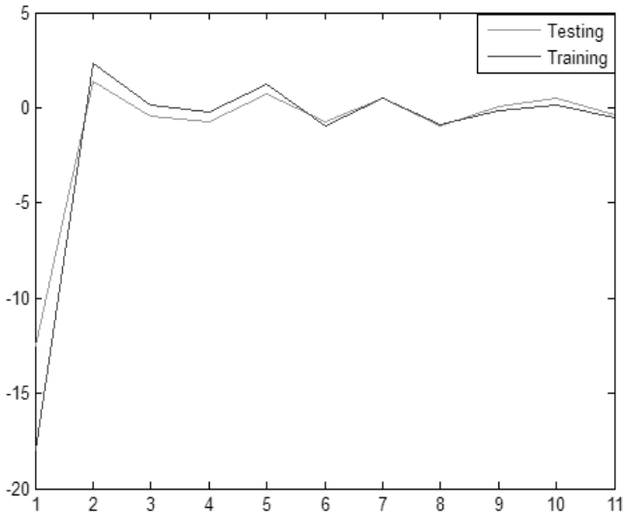
$$c_p(X,Y) := \sum_{l=1}^{L} c(x_{nl}, y_{ml}). \qquad (18)$$

٣ ـ إن أمثل مســـار بين X و Y هو المسار P* اللذي يحمل أقل كلفة كلية من بين كل المســـارات المحتملة؛ لذا فإن مسافة DTW بين X و Y تعرف بالكلفة الكلية لـ P*:

$$DTW(X,y) = c_p*(X,Y)$$
$$= \min\{c_p(X,Y) | p \text{ is an } (N,M) - warping\ path\}$$

ومما يجدر الذكر هنا أننا طبقنا هذه الطريقة على مصفوفة الصفات لبيانات التدريـــب وبيانات الاختبار وهي بحجمها الأصلي ٤٧٦×١١ لكل صوت أي بدون أخذ معدل الأسطر، أو الأعمدة، وأيضًا طبقناها على هذه البيانات (صفات إشارة الاختبار، وصفات إشارات التدريب) بعد أخذ معدل الأسطر لها.

الشكل (١) يوضح شكل متجه الصفات لإحد إشارات الاختبار، ومتجه الصفات لإشـــارة التدريـــب الأقرب إليه. والجدول (٢) يوضح نتائج هذه الاختبارات.



الشكل رقم (١) متجهي الصفات لإشارة الاختبار وإشارة التدريب الأقرب إليه

الجدول رقم (٢) نتائج تطبيق النظام على اللكنات الإنكليزية بطريقة DTW

DTW – English Accents		
No.	Experiment	Success Percentage
1	KSA-UK	96.43%
2	KSA-USA	95%
3	USA-UK	58.57%
4	KSA-USA-UK	61.11%

مسافة أو بعد الارتباط Distance Correlation

إن بعد الارتبـــاط Distance Correlation هو مقياس احصائي للاعتمادية بين متغيرين أو متجهين عشـــوائيين ليس بالضروري أن يكونا متســـاويين في الأبعاد. هذا المقياس، أن معادلة مسافة الارتباط [١١][٩]:

$$dCor(X,Y) = \frac{dCov(X,Y)}{\sqrt{dVar(X)\, dVar(Y)}} \qquad (19)$$

لحساب بعد التباين المشـــترك Distance covariance نفرض أن

$a_{j,k} = (X_j - X_k)$, $\quad j,k = 1,2,\ldots,n$

$b_{j,k} = (Y_j - Y_k)$, $\quad j,k = 1,2,\ldots,n$

ثم تحسب

$$A_{j,k} = a_{j,k} - \bar{a}_j. - \bar{a}_k + \bar{a}..,\quad B_{j,k} := b_{j,k} - \bar{b}j. - \bar{b}.k + \bar{b}.. \qquad (20)$$

حيث

$\bar{a}_j$ المعدل للصف j

$\bar{a}_k$ المعدل للعمود k

$\bar{a}$ معدل مصفوفة المسافة

في النهاية، فأن مربع بعد التباين المشترك يعطى بالمعادلة الآتية:

$$dCov2n(X,Y) := \left(\frac{1}{n2}\right) \sum_{j,k=1}^{n} A_{j,k} B_{j,k} \qquad (21)$$

ولحساب Distance variance بعد التباين:

$$dVar^2n(X) := dCov^2n(X,X) = (1/n2) \sum_{j,k} A_{j,k}^2 \qquad (22)$$

ثم تحسب مسافة الانحراف المعياري بأخذ الجذر التربيعي إلى $dVar^2_n$.

$$dCov2n(X,Y) := \left(\frac{1}{n2}\right) \sum_{j,k=1}^{n} A_{j,k} B_{j,k} \qquad (23)$$

في الجدول(٣) الآتي نتائج تطبيق النظام على اللكنات.

الجدول رقم (٣)

نتائج تطبيق النظام على اللكنات بطريقة Correlation

Correlation – English Accents		
No.	Experiment	Success Percentage
1	KSA-USA	76.67%
2	KSA-UK	84.52%
3	USA-UK	47.86%
4	KSA-USA-UK	56.35%

٣ ـ والاستنتاجات:

١ ـ لاحظنا من التجارب أن أفضل نسبة حصلنا عليها في النظام بهذه الطريقة هي للأشخاص العرب لأن أغلبهم يتكلمون لغة واحدة فقط.

٢ ـ نلاحظ أن أفضل نسبة نجاح حصلنا عليها في تمييز اللكنات بطريقة التمييز التصنيفي هي أيضًا بين الأصوات السعودية والأمريكية وهي ٨٥٪ حيث كانت نسبة نجاح الأصوات السعودية هي مئة.

٣ ـ إن زيادة زمن الإشـــارة قد يعطي نتائج أفضل ولكن على حساب زمن المعالجة كما أن هناك تطبيقات يجب أن يكون فيها طول الإشـــارة الصوتية قليل مثل تطبيقات المساعدة عن طريق الهاتف لذلك فإن طول الإشارة الصوتية يحدد حسب نوع التطبيق.

المصـــــــادر

Behravan H. ,[2012] "Dialect and Accent Recognition" ,Master thesis ,university of Eastern Finland, School of Computing.

Biadsy F. ,[2011], "Automatic Dialect and Accent Recognition and its Application to Speech Recognition", Ph. thesis,COLUMBIA UNIVERSITY, New York USA.

Burns R. B. and Burns R. A. [2008] "Business Research Methods and Statistics Using SPSS ", SAGE publication

Englund C. ,[2004]," Speech recognition in the JAS 39 Gripen aircraft adaptation to speech at di_erent G-loads", Department of Speech, Music and Hearing Royal Institute of Technology, Stockholm.

Gangonda S. and Mukherji P. [2012] " **Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features**", **International Journal of Engineering Research and Applications (IJERA) .**

Kondoz A. M. [2004] "Digital Speech Coding for Low Bit Rate Communication Systems", Wiley Series in Communication and Distributed Systems, second edition.

Logan B. [2000]"Mel Frequency Cepstral Coefficients for Music Modeling ", Cambridge Research Laboratory, Compaq Computer Corporation,one Cambridge Center, Cambridge MA 02142.

Marie M. ,Deza E. [2009] " Encyclopedia of Distances", Springer Berlin Heidelberg.

Meinard M. [2007] " information retrieval for Music and Motion", Springer

PavelSenin [2008] " Dynamic Time Warping Algorithm Review ", Information and Computer Science Department University of Hawaii at Manoa Honolulu,Honolulu, USA.

SZÉKELY G.,RIZZO M. AND BAKIROV N. [2007] " MEASURING AND TESTING DEPENDENCE BY CORRELATION OF DISTANCES", Bowling Green State University, Bowling Green State University and USC Russian Academy of Sciences.

المدخل إلى تحليل الانحدار ١٩٨٧ د.خاشـــع محمود الراوي جامعة الموصل.