# تمييز أصوات الأرقـام العربيـة

**شهلة عبد الوهاب عبد القادر**

*قسم أنظمة الحاسبات، المعهد التقني الموصل، الموصل، العراق*

**الملخص**

في هذا البحث ، تم استخدام خوارزمية انحراف الوقت الديناميكي ( Dynamic Time Warping) في عملية تمييز أصوات الأرقام العربية التي تعتبر إحدى تطبيقات البرمجة الديناميكية ( Dynamic Programming) باستخدام طريقة مطابقة الأنماط القادرة على تمييز أصوات الأرقام من خلال البحث عن أفضل تطابق وحيد بين مجموعة مفردات قاعدة البيانات الصوتية للأرقام العربية .تم استخدام طريقتين لتحليل الإشارة الصوتية وهما :

MFCC) Mel- Frequency Cepstrum Coefficients
Linear Frequency Cepstrum Coefficients ( LFCC )

والمقارنة بين الطريقتين وبيان مدى تأثيرهما على عملية التمييز النهائية .

يسمح البرنامج باختيار أصوات أرقام مختلفة مع محددات تحليل ومحددات لخوارزمية انحراف الوقت الديناميكي ليكون إدخالاً ويكون الإخراج هو صوت الرقم المطابق للإدخال مع تحديد المتكلم الذي لفظ ذلك الصوت .تم إجراء تجارب كثيرة واختبارات لتحسين أداء البرنامج.وشملت مجموعة التدريب ثلاثة متكلمين (طفل،وإمرة ،ورجل) لمفردات قاعدة البيانات الصوتية مع أصوات مختلفة للتدريب.أثبتت النتائج العملية نجاح الطريقة المستخدمة في التمييز.

**المقدمة**

أنظمة تمييز الكلام تعتمد بالأساس على أشكال الموجات الصوتية التي يتم من خلالها تشخيص الكلمة المستلمة من المتكلم ، فقبل ٤٠ سنة من الدراسات والأبحاث ، اقترحت خوارزميات عديدة لتمييز الكلام الآلي . تعد مطابقة الأنماط (Pattern-Matching) واحدة من هذه الدراسات ، ففيها يخزن البرنامج النماذج التي يطلق عليها مصادر النظام(References) لكل صوت من مجموعة الأصوات لتشكل قاعدة البيانات، ثم تقارن إشارة الصوت القادمة (Test) بكل الأصوات لإيجاد أحسن تطابق ناتجاً للتمييز [2].

**أساسيات التمييز**

اعتمد في البحث على مرحلتين أساسيتين تتألف منها كل أنظمة التمييز في أعلى المستويات وهي :

١- استخلاص الصفات( Feature Extraction):-أو ما تسمى بعملية التحليل ، تقوم ببناء المصادر لكل صوت رقمي من مجموعة المفردات لتشكيل قاعدة البيانات الصوتية. هذه الرقم عبارة عن إشارة صوتية للرقم الملفوظ يطلق عليها إشارة المصدر، بحيث تقسم كل إشارة إلى قوالب بعينات متساوية الطول من البداية إلى النهاية ، ثم تحول ولكل قالب إلى متجه الصفات الذي يستخلص صفات الإشارة في ذلك القالب. تضم تلك المتجهات في مجاميع يطلق عليها متجهات الصفات ( Feature Vectors). تكرر هذه المعالجة لكل صوت رقمي في مجموعة المفردات .

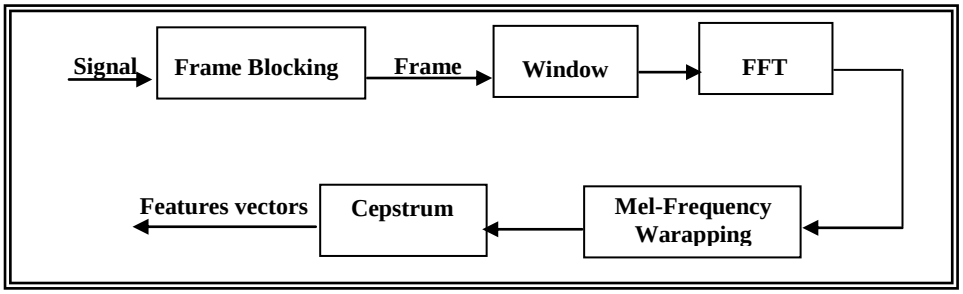
٢- مطابقة الصفات (Feature Matching) :- أو ما تسمى بعملية التمييز،تحول فيها الإشارة القادمة المراد تمييزها والتي يطلق عليها إشارة الاختبار إلى سلسلة من متجهات الصفات باستخدام تحويل البداية-النهاية نفسها في معالجة استخلاص الصفات وهذه الصفات تقارن مع كل الاحتمالات الموجودة في قاعدة البيانات باستخدام طريقة مطابقة الأنماط. يعطى قرار التمييز من جودة التطابق بواسطة دالة المسافة الأقليدية (Euclidean Distance) بين سلسلتين من متجهات الصفات،إحداها تمثل متجهات صفات إشارة المصدر والأخرى متجهات صفات إشارة الاختبار[3] .

**عملية تحليل الإشارة الصوتية**

تم في تحليل الإشارة اختيار أحسن الطرق لاستخلاص الصفات بدقة للحصول على معلومات الإشارة الصوتية ؛ لأنها من أهم العوامل في تحقيق تمييز عالٍ.الهدف هو إلغاء المعلومات غير ملائمة لبيانات الإشارة المحللة،وتحسين جميع النواحي فيها والتي تساعد على اكتشاف الأهمية لتلك الإشارة وتعد طريقة Mel- Frequency Cepstrum Coefficients وطريقة Linear Frequency Cepstrum Coefficientsمن التقنيات ذات المواصفات العالية التي تزودنا بمستوى أعظم للإدراك الحسي بتمثيل دقيق للنظام الصوتي ، وطرق جيدة في عملية تحليل الإشارة الصوتية [5] وكالاتي:-

**تحليل الإشارة الصوتية باستخدام (MFCC)** Mel- Frequency Cepstrum Coefficients

تعد MFCC طريقة لتحليل الإشارة الصوتية ذات مواصفات عالية لتثمين مواصفات الكلام مثل طبقة الصوت ونغمته . تعتمد بالأساس على ترشيح الترددات الواطئة خطيا ، والعالية لوغاريتما لاستخلاص المواصفات المهمة للصوت. الحالة المثالية للإشارة يجب أن تسجل بمعدل عينات أعلى من 10KHZ ، وذلك لاستخلاص كل الترددات الأعلى من 5KHZ ؛لأنها تغطي غالبية الطاقة الصوتية المتولدة من قبل الإنسان [1] ،يوضح الشكل(١) المخطط الكتلي للمعالجة MFCC وتشمل المراحل الآتية:




**شكل ١ : المخطط الكتلي للمعالجة MFCC.**

١- **Frame Blocking** : يستخدم التحليل الطيفي لجزء قليل من الإشارة لان مواصفاتها ثابتة. تختار قوالب الوقت هذه لتصف أهمية الأحداث الصوتية في وحدة الزمن. من هذا المبدأ يتم تقسيم إشارة الرقم الواحد إلى عدة قوالب متداخلة وبمعدل عينات N لكل قالب ، والقالب الثاني يبدأ عادة بعد M من العينات من بداية القالب الأول.القيم المختارة في هذا البحث لـ NوM هي 256 ،125 على الترتيب ، أي بحدود 30msec من الإشارة.

٢- **Windowing**: عبارة عن دالة رياضية تشبه في وظيفتها عمل النافذة الاعتيادية بإدخال كمية الضوء. تقوم هذه النافذة بتنعيم الإشارة وإزالة التعرجات الحادة منها ولكل قالب في الإشارة. وتكون نتيجة النافذة ممثلة بإشارة المعادلة الآتية:

$$x_i(n) = x_i(n).w(n), \qquad 0 \le n \le N-1$$

حيث:

$x_i$ : عينات القالب الواحد .

N : حجم القالب أي عدد عينات القالب.

وهناك أنواع عديدة منها نافذة Tukey-Hamming Window المستخدمة بصورة خاصة مع إشارات الكلام والتي تعرف رياضياً كما يأتي -:[6]

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N-1}\right) \qquad 0 \le n \le N-1$$

٣- **Fast Fourier Transform أو FFT** : يتم فيه تحليل الطاقة التي تمتلكها الإشارة عند ترددات معينة.إن الفكرة المستخلصة من التحليل الطيفي هي أن المنطلق التكراري يعطينا معلومات مهمة لا نستطيع أن نحصل عليها في المنطلق الزمني .لذلك يطبق هذا التحليل على كل قالب كما في المعادلة الآتية:

$$X_n = \frac{1}{N}\sum_{k=0}^{N-1}\left(X_k * e^{\frac{-2\pi jkn}{N}}\right) \qquad n=0,1,2,\ldots\ldots,N-1$$

حيث أن : N : تمثل حجم القالب .

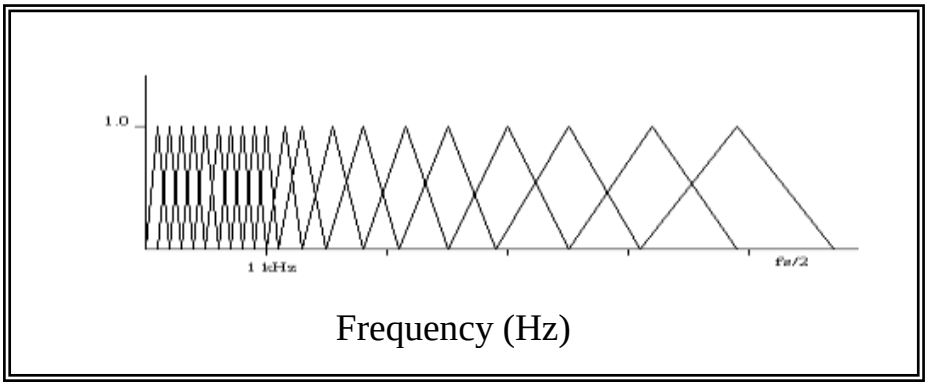
$X_k$ : قيمة العينة قبل FFT

$X_n$ : قيمة العينة بعد FFT

j : تمثل جذراً خيالياً .

٤- **Mel-Frequency Wrapping** : إن محتويات الترددات لأصوات الإشارات لأتتبع مقياساً خطياً ؛ لهذا فكل نغمة في الترددات الطبيعية ($f$) ، تقاس بالهرتز (Hz) ، أما طبقة الصوت ؛ فتقاس بمقياس Mel . مقياس الترددات Mel يعطينا ترددات خطية بالمسافات التي يكون ترددها أقل من 1000Hz ويعطي ترددات لوغارتيمية في المسافات الأعلى من 1000Hz ؛ لهذا يتم استخدام الصيغة التقريبية لحساب الـ Mel لتردد معين إذا أعطينا $f$ بـ ( Hz ):$\text{Mel}(f) = 2595 * \log_{10}(1+f/700)$

إن الاستفادة من مقياس الترددات Mel هو عمل بنك الترشيح ( Filter Bank ) الذي يحوي مرشحاً واحداً لكل مكون من مكونات الـ Mel المطلوبة ويتم بموجبه تجميع مرشحات تمرير الحزمة مثلثة الشكل (Triangular Bandpass Filter) ، يحدد عرض المرشح بوساطة ثابت التردد Mel التي في النهاية تمثل بنك الترشيح. يتم بوساطة هذه الطريقة تغطية كل الترددات بدون أي فقد للمعلومات بحيث أنه عند إدخال الإشارة الناتجة FFT على المرشح تظهر مواقع القوة لتلك الإشارة ،والحالة المثالية لعدد المرشحات في هذه الطريقة تساوي 20 والتي سوف تعتبر عدد معاملات معادلة الـ MFCC . ويمثل الشكل(٢) بنك الترشيح الذي يطبق على الإشارة [4] [7].




**شكل ٢ : بنك الترشيح-Mel**

**٥-Cepstrum:** في الخطوة الأخيرة ، تحول نتيجة بنك الترشيح الى المنطلق الزمني وهي MFCC والتي تجهزنا بتمثيل جيد لمواصفات الطيف لجزء من إشارة الرقم الصوتي للقالب المحلل،وتستخدم المعادلة الآتية لحساب MFCC :-

$$MFCC_i = \sum_{k=1}^{20} X_k COS \frac{\pi i(k-0.5)}{20} \qquad i=1,2,\ldots..,P$$

حيث أن: 20 : عدد المرشحات في بنك الترشيح.

$X_k$ : هي إشارة القالب الواحد Mel Spectrum .

P : عدد معاملات المتجه الواحد للإشارة المحللة [4][8].

**تحليل الإشارة الصوتية باستخدام (LFCC)** Linear Frequency Cepstrum Coefficients

تعد طريقة ثانية لتحليل الإشارة الصوتية للرقم الملفوظ حيث تجهز بنموذج جيد وتقريبي لما يصدره جهاز النطق من أصوات ،وإزالة كل الترددات عديمة الأهمية ، وتحسين نواحي الإشارة جميعها التي تساعد على دقة التمييز. تكون المراحل الموضحة في الشكل (٢) مماثلة لمراحل هذه الطريقة عدا الإخراج للتحليل الطيفي لا يدخل إلى مرحلة (Mel-Frequency Wrapping ) بل يدخل مباشرة إلى مرحلة (Cepstrum) ، وبحسب المعادلة الآتية:

$$LFCC_i = \sum_{k=0}^{NYQ-1} X_k COS \frac{\pi ik}{NYQ} \qquad i=1,2,\ldots..,P$$

حيث أن:

NYQ : نصف عينات القالب الواحد .

$X_k$ : هي إشارة القالب الواحد (Spectrum ).

P: عدد معاملات المتجه الواحد للإشارة المحللة.

بواسطة تطبيق طريقتي التحليل أعلاه ،من بداية إشارة الرقم إلى نهايته وكل قالب يكون بحدود 30msec بتداخل ، ستحسب مجموعة من معاملات MFCC و LFCC ولكل رقم، تلك المجموعات من المعاملات هي متجهات الصفات. في النهاية إشارات الأرقام المحللة تشكل قاعدة البيانات بهذه الطرق ، التي يستفاد منها في عملية التمييز [8].

**عملية تمييز الإشارة الصوتية**

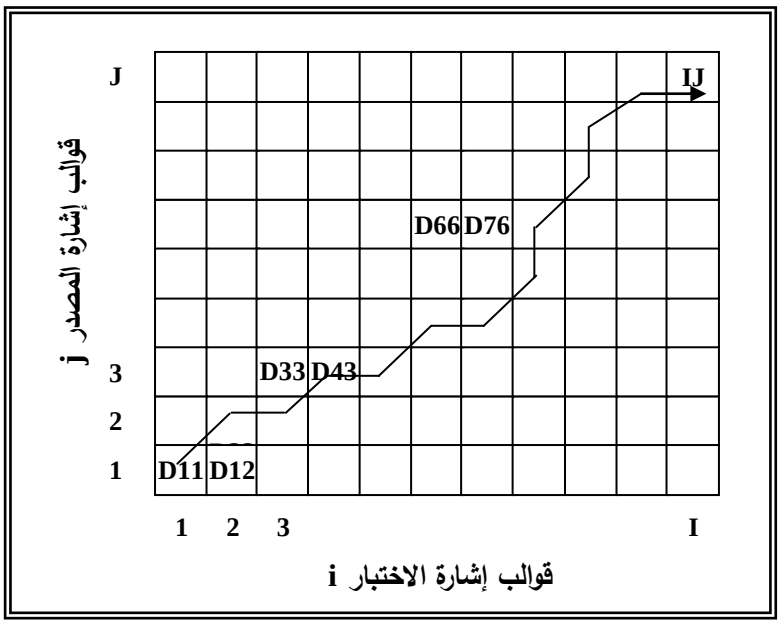
استخدمت طريقة مطابقة الأنماط لتمييز أصوات الأرقام .تستخدم هذه الطريقة مقياس المسافة لإيجاد التطابق غير الخطي بين إشارة الاختبار وكل إشارات المصادر،ثم اختيار المصدر الأقل مسافة والرقم الذي يمثله ذلك المصدر يشير إلى نتيجة التمييز،ذلك التطبيق في البرمجة الديناميكية يسمى انحراف الوقت الديناميكي DTW .المطابقة غير الخطية تنجز عادة بمخطط شبكي كما في الشكل (٣). تُمثل إشارةالإختبار T على طول المحور السيني، وإشارة المصدر R على طول المحور الصادي حيث أن T و R هي متجهات الصفات لكل رقم وكالآتي [10] :-

$R=\{R_1 , R_2 , R_3 ,\ldots\ldots, R_J\}$

$T=\{T_1 , T_2 , T_3 ,\ldots\ldots, T_I\}$

حيث أن I و J تمثل عدد القوالب المحللة في T و R على الترتيب. من الجدير بالذكر أنه ليس بالضرورة ( I=J) ،كل تقاطع في ذلك المخطط يعرف بالعقدة، حيث أن العقدة (i,j) تعني تطابق القالب i من الإشارةT مع القالب j من الإشارة R. تعرف الكلفة المتعلقة على أنها مسافة مقطعية $d_f(i ,j)$ بين متجهات الصفات T(i)، R(j) ويعرف المسار بأزواج من العقد$(i_k , j_k)$ $(i_{k-1} , j_{k-1})$ التي تعني الامتداد من العقدة $(i_{k-1} , j_{k-1})$ إلى العقدة $(i_k , j_k)$ .تشير $i_k$ إلى الإشارة T في الوقت k ، كذلك $j_k$ تشير إلى الإشارة R في الوقت k . يبدأ مسار البحث من العقدة(1,1) وينتهي بالعقدة$(i_k , j_k)$، يمتلك كلفة كلية تمثل الكلفة المتراكمة من نقطة البداية لهذا المسار إلى حين الوصول إلى النقطة النهائية $(i_k , j_k)$ التي هي(I,J) . نستعمل $D(i_k , j_k)$ للإشارة إلى هذا النوع من الكلفة،ويعرف بالمعادلة الآتية [9] :-

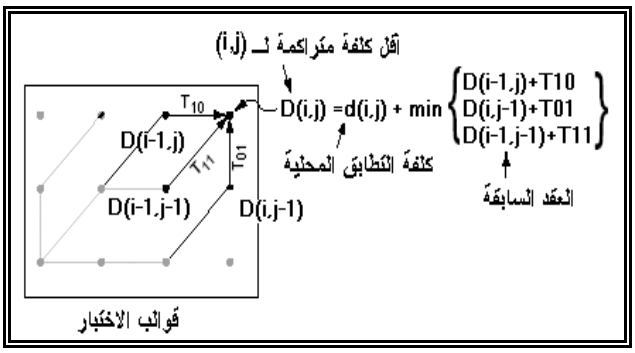
$$D(i_k , j_k) = Min\left[D(i_{k-1}, j_{k-1})\right] + d_f(i_k , j_k)$$




**شكل ٣ : المخطط الشبكي**

تطبق قيود خاصة في هذه الطريقة استناداً إلى معلومات بديهية معروفة عن إشارات أصوات الأرقام الملفوظة ، تلك القيود تعرف كيف يتغير مقياس الوقت فجأة لإشارة واحدة نسبة إلى الإشارة الثانية ،والمسارات القانونية في فضاء البحث تؤخذ بنظر الاعتبار للتطابق وغير القانونية تستبعد وتم اعتماد القيود الآتية لتطبيق هذه الطريقة:

١- قيد المسار المحلي Local Path Constraint: تحدد آلية الانتقال،أي مسار البحث القانوني في منطقة محلية بين عقد الشبكة والقيد الموضح في الشكل(٤) هو القيد المستخدم في هذا البحث. تقول الفرضية في هذا القيد أن هناك مساراً واحداً فقط بأقل كلفة هو الصحيح للوصول إلى العقدة (i,j).على سبيل المثال ،العقدة (i,j) هناك ثلاثة مسارات محتملة للعقد السابقة وهي :(i,j-1) أو (i-1,j) أو ( i-1,j-1) ممكن أن تصل إليها ، وترجح العقدة السابقة ذات الأقل كلفة. تستمر عملية تركيم الكلفة الأقل لإيجاد المسار النهائي (I,J ) [5].




**شكل ٤ : قيد المسار المحلي**

٢- قيد المسار العام Global Path Constraint: تختبر العقد وتقيد بمدى معين من الوقت هذا يعني ان البحث يحدد بمنطقة قانونية كما في الشكل (٥) ، ثم اعتماد فرضية لتحديد تلك المنطقة وذلك لتقييد حد الانضغاط أو التمدد في إشارات الرقم لمدى معين من الوقت وحدود القيم العليا للمؤشر j تحدد بوساطة القيمة i من خلال المعادلتين الآتيتين :

$$\max J = \min\begin{pmatrix} s(i-1)+eps+1 \\ \frac{1}{s}(i-1)-\frac{1}{s}(I-eps-1)+J \end{pmatrix}$$

وحدود القيم الدنيا تحدد بالمعادلتين :

$$\min J = \max\begin{pmatrix} \frac{1}{s}(i-1-eps)+1 \\ s*i-(s*I-J)-eps \end{pmatrix}$$

حيث أن S : هو الميل الذي يعرف القيد .

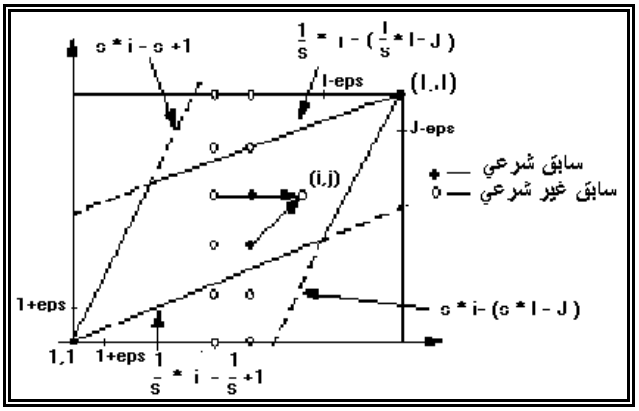
eps: هي حد الانضغاط أو التمدد في الوقت .

٣-دالة المسافة الاقليدية : تعد من التقنيات الشائعة في تمييز الأنماط التي تعتمد مبدأ المسافة بين متجهين من خلال البعد بين نقاطهما وقد تم اعتماد المعادلة التالية :

$$D[i][j] = \frac{1}{P}\sqrt{\sum_{k=1}^{P}\left(ref(k,j)-test(k,i)\right)^2}$$

حيث إن: ref : قيمة متجه صفات إشارة المصدر المحللة .

test : قيمة متجه صفات إشارة الاختبار المحللة [3][10] .




**شكل ٥ : قيد المسار العام**

**خوارزمية التمييز باستخدام DTW: (الجانب التطبيقي)**

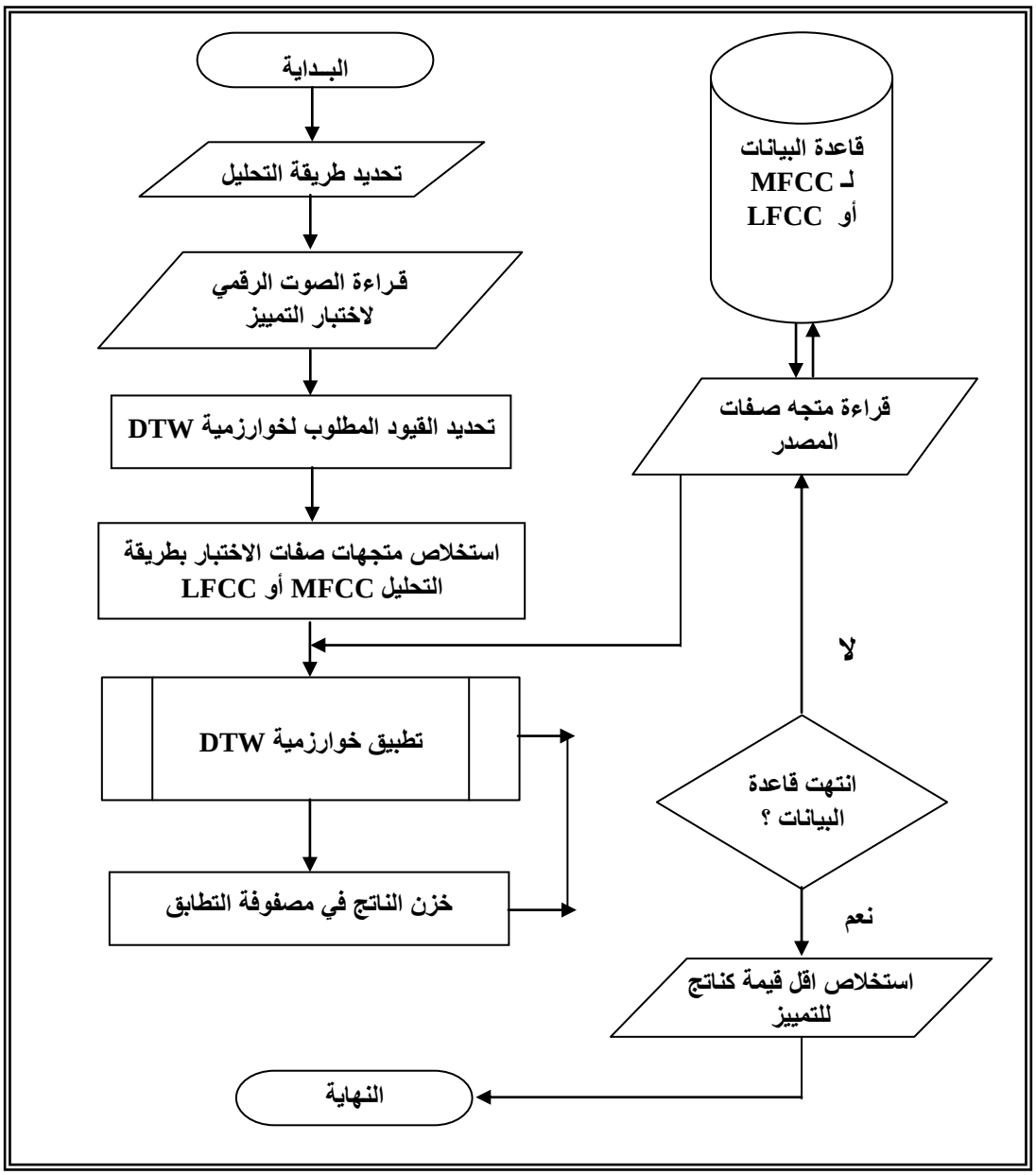
تتعلق هذه الفقرة بالخوارزمية المستخدمة في البرنامج ويوضحها الشكل(٦) وكالآتي :

١- تعرف طريقة التحليل والصوت الرقمي المراد تمييزه والمتغيرات وفضاء البحث للتمييز.

٢-استخلاص متجهات صفات الاختبار للصوت الرقمي بطريقة التحليل المختارة.

٣-تأخذ خوارزمية الـDTW سلسلة متجهات الصفات لإشارة المصدر من قاعدة البيانات الصوتية للأرقام وإشارة الاختبار من عملية التحليل كإدخال ويمثل الإخراج أقل قيم التطابق .

٤-خزن ناتج الخوارزمية في مصفوفة التطابق النهائية .




**شكل ٦ : المخطط الانسيابي لعملية التمييز**

٥- تستمر الخطوات (٣،٤) إلى نهاية أصوات الأرقام المحللة بطريقة التحليل المختارة.

٦- استخلاص اقل قيمة من مصفوفة التطابق النهائية.

٧- استخراج ناتج التمييز الذي يمثل الصوت الرقمي المقابل للقيمة النهائية.

**تهيئة الملفات الصوتية ومكونات البرنامج**

لغرض اختبار طريقة التمييز المستخدمة في هذا البحث، في البداية تم تسجيل أصوات عشرة أرقام عربية من الصفر إلى العشرة لثلاثة متكلمين مختلفين( رجل ،وامرأة وطفل) وخزنت البيانات بـ 16-Bit وبمعدل عينات يساوي 11025Hz في ملفات صوتية مفصولة بصيغة (.Wav)، أي وصلت العينات المسجلة إلى ٣٠ رقم صوتي تمثل بمجموعها قاعدة البيانات الصوتية للأرقام العربية للبرنامج .

استخدمت لغة MATLAB Version 6 التي تحتوي على دوال خاصة لتحويل الإشارة الصوتية إلى عيناتها وإجراء عمليات المعالجة عليها داخل البرنامج . يتكون البرنامج من جزأين أساسيين وهما : التحليل باستخدام (MFCC) أو (LFCC) حيث شكلت قاعدة البيانات الصوتية للأرقام المحللة بالطريقتين كل على إنفراد،والتمييز باستخدامDTW والملحق رقم (١) يبين البرامج الفرعية المعدة لغرض عملية التحليل والتمييز

**تجربة التمييز**

أختبر التمييز بوساطة محددات متعددة ، p عدد معاملات المتجه الواحد لكل قالب من قوالب متجهات الصفات للإشارة الواحدة. Eps تمثل احتمالية عدد قوالب الإشارة الصوتية للرقم التي يتم البحث بها من البداية الى النهاية ، أي قيمة التمدد الأقصى في الوقت المحدد من النطق والقيمة الدنيا للانضغاط وبقيمة ميل S لتحديد المنطقة القانونية للبحث. تم تطبيق ستة حالات مختلفة بحسب الجدول (١) كانت دقة نتائج التمييز

كبيرة في معظم الحالات ووصلت أكثر من 90%. كما لوحظ أن إنقاص P من 10 إلى 8 لا يؤثر على نتيجة التمييز.لكن عند إنقاص eps إلى 0 خفضت دقة التمييز ووصلت إلى 88% في MFCC، 85 % في LFCC. كذلك لوحظ أن تعديل المحددات الشاملة S مربحة لدقة النتائج في تجهات متوقعة، وذلك بتحديد أقصى منطقة بحث تساوي 3 حداً أعلى و 0.333 حداً أدنى بوساطة السماح لمنطقة البحث أن تتضمن أكثر احتماليان المسار وتعطي نتائج دقيقة. وبينت النتائج أن هناك فرقاً في التمييز بين طريقتي التحليل حيث أثبتت النتائج أداءً أفضل ودقة أعلى لـMFCC إذا قارناها مع LFCC من خلال استخلاص الصفات لكل الحالات . عملياً تم إثبات أن أفضل اختيار للمعاملات في الطريقتين هو p=10 .

**جدول ١ : حالات خوارزمية التمييز DTW**

| | P | S | 1 / S | Eps |
|---|---|---|---|---|
| Case1 | 10 | 2 | 0.5 | 3 |
| Case2 | 8 | 2 | 0.5 | 3 |
| Case3 | 12 | 2 | 0.5 | 3 |
| Case4 | 10 | 3 | 0.333 | 3 |
| Case5 | 10 | 2 | 0.5 | 0 |
| Case6 | 10 | 2 | 0.5 | 5 |

**الاستنتاجات**

يحدد أداء أنظمة التمييز بوساطة مصادر الأصوات المتضمنة مواصفات صوت المتكلم ،ومعدل إصدار الأصوات ،ومصادر إيصالها ، ووسائط التسجيل ، ودقة التمييز تأتي من سلوك المتكلم في إصدار هذه الأصوات . وقد أثبتت نتائج هذا البحث ما يأتي :

- إن ( DTW) هي طريقة رياضية مرنة وتعطي نتائج ذات دقة عالية وتبين أنه من الممكن تحسين الأداء بوساطة اختيار المصادر بعناية ؛إذ لها دور كبير في التأثير في دقة التمييز .
- استعمال القيود الشاملة تقلل منطقة البحث من ثم تنقص الحسابات الضرورية لإيجاد التطابق مما يؤدي إلى عدم إشغالها مساحات كبيرة جدا أثناء التنفيذ.
- على الرغم من أن المسافة الاقليدية تعد أسلوباً بسيطاً في حساب الفرق بين متجهين إلا أنه أسلوب فعال في هذا المجال.
- إن معالجة الإشارة باستخدام طريقتي التحليل تزودنا بمعاملات الطيف التي تمثل النظام الصوتي الدقيق للأرقام .على الرغم من أن (LFCC) تتجاهل التناقض بين الترددات وملائمة للوصول إلى أهداف التمييز إلا أن (MFCC) تزودنا بمستوى عالٍ من الإدراك الحسي لأصوات الإنسان حيث تعمل على إزالة كل المعلومات عديمة الأهمية ، ثم إعطاء أفضل تمثيل للإشارة التي تقود إلى دقة أعلى في أداء التمييز باستخدام هذه المعالجة.

من ذلك كله يمكن استنتاج حقيقة مهمة جدا أن معاملات الطيف ذات المواصفات العالية تظهر أهميتها بالاعتماد على المتكلم نفسه وطريقة إنتاج الأصوات وبالدرجة الأساس أسلوب اللفظ الصوتي والذي يمكن الاستفادة منها في تطبيقات عديدة مثل الأنظمة الأمنية حيث أن أصوات الأشخاص تختلف كما تختلف بصمات الأصابع .

## References


1. Brandstein M. ,2001. Speech Recognition of English Digits, Harvard University.
2. Charles A. , 2002. Speech Processing , Electrical and Computer Engineering, Purdue University, January 28.
3. Land J., 2000. Isolated Word Speech Recognition of English Digits, http://www.gerc.eng.ufl.edu/defult.htm.
4. Minh N., 2000, "An Automatic Speaker Recognition System", http://svr-www.eng.cam.ac.uk, Swiss Federal Institute of Technology, Lausanne , Switzerland.
5. Picone J., 1996. Fundamentals of Speech Recognition, I. S. I. P. , Mississippi State University, May, pp 5-84.
6. Rabiner L. and Juang , B. H.,1993 . Fundamentals of Speech Recognition, Prentice - Hall, Englewood Cliffs, New Jersey, US.
7. Robinson T.,1996.Speech analysis, University of Leeds, http://svr-www.eng.cam.ac.uk, September 30.
8. Shaffer J. , Hamaker and Picone J. ,1998. The Visualization of Signal rocessing Concepts , IEEE Transactions on Acoustics, Speech ,and Signal Processing, May, pp 1853-1856.
9. Silverman H. F. and Morgan D. P. ,1990. The Application of Dynamic Programming to Connected Speech Recognition , IEEE ASSP Magazine, July, pp 7-25.
10. Zeng Y. ,2000. Dynamic Time Warping Digit Recognition, I. S.I.P, Mississippi State University ,November, pp 1-51, http://www.isip.msstate.edu /projects/speech.

# Arabic Numbers Sound Recognition

**shahla Abd Al-Wahab**
*Department Computer System, Technical Institute in Mosul, Mosul, Iraq*

## Abstract

In this research, dynamic time warping algorithm was used in the form of recognition Arabic sound numbers which is regarded one of the dynamic programming applications using the method of superimposition of the patterns capable of recognition the sounds of Arabic numbers through the search for best unique superimposition between the group elements of audio database of Arabic numbers . Two methods were used to analyze the voice signal. These are Mel-Frequency Cepstrum Coefficient (MFCC) and Linear Frequency Cepstrum Coefficient (LFCC) analyses. These two methods of analysis were compared and their effect on the final distinction process is shown.

The program allows the testing of different sound numbers with analysis and dynamic time warping algorithm determinants to be inputs and the output is the number consistent with input with the determination of the speaker who spelled the number. Many experiments were held and also many tests to enhance the program performance. Training group included three speakers (a child, a female and a male) for audio database with different voices for training. The practical results proved the success of the method used in recognition.

## الملحق (١)

**- برنامج تحليل الاشارة باستخدام MFCC :**

```
function ccep=mfcc(y,M,N,P);
NYQ=N/2;
% Triangle Filter Defs
seqlen=size(y,1);
m=0; startpt=1;endpt=N; m=1;
while endpt<=seqlen
winseq=hamming(N).*y(startpt:endpt);
magspec=abs(fft(winseq));
for i=1:20 % Calc triangle filter outputs
for j=fstart(i):fcent(i)
filtmag(j)=(j-fstart(i))/(fcent(i)-fstart(i));  end;
for j=fcent(i)+1:fstop(i)
filtmag(j)=1-(j-fcent(i))/(fstop(i)-fcent(i));  end;
Y(i)=sum(magspec(fstart(i):fstop(i)).*filtmag(fstart(i):fstop(i))');  end;
Y=log(Y.^2);
for i=1:P
coefwin=cos((pi/20)*i*(linspace(1,20,20)-0.5))';
ccep(i,m)=sum(coefwin.*Y');  end;
m=m+1;  startpt=1+(m-1)*M;
endpt=startpt+N-1;  end;
```

**- برنامج تحليل الاشارة باستخدام LFCC :**

```
function ccep=lfcc(y,M,N,P);
NYQ=N/2;  seqlen=size(y,1);  m=0; startpt=1;  endpt=N;m=1;
while endpt<=seqlen
winseq=hamming(N).*y(startpt:endpt);
logspec=log(abs(fft(winseq)));
for i=1:P
coefwin=cos(pi*i*linspace(0,NYQ-1,NYQ)/NYQ)';
ccep(i,m)=sum(coefwin.*logspec(1:NYQ));  end;  m=m+1;
startpt=1+(m-1)*M;  endpt=startpt+N-1; end;
```

**- برنامج DTW :**

```
function Dmin_Avg=dtw1(test,ref,eps,delmax,delmin)
% Dynamic Time Warping Algorithm
%eps=3; %delmax=2.0;  %delmin=0.5;
I=size(test,2);  J=size(ref,2);  Dmin=1e6*ones(I,J);  C_t=zeros(I,J);
for i=1:1+eps
   s=0;
  for hh=1:10
```

```
      u=ref(hh,1)-test(hh,i);          s=s+u*u;
  end
  Dmin(i,1)=sqrt(s)/10;    % L2norm(i,1,test,ref);     C_t(i,1)=0;    end;
for j=1:1+eps
       s=0;
       for hh=1:10
          u=ref(hh,j)-test(hh,1);        s=s+u*u;      end
     Dmin(1,j)=sqrt(s)/10;     C_t(1,j)=0;    end;
Dmin_old=Dmin;     C_told=C_t;
for i=2:I
jmax1=delmax*(i-1)+eps+1; % Global Constraints
jmax2=delmin*(i-1)+J-delmin*(I-eps-1);
jmin1=1+delmin*(i-1-eps);
jmin2=1+J-1-eps+delmax*(i-I);
jmax=min(J,floor(min(jmax1,jmax2)));
jmin=max(1,ceil(max(jmin1,jmin2)));
for j=jmin:jmax
  s=0;
      for hh=1:10
         u=ref(hh,j)-test(hh,i);      s=s+u*u;     end
  C_node=sqrt(s)/10;%RMS Error
Prev_Min=1e6*ones(3,1);
for k=1:3
idiff=1; jdiff=1;
if k==1 jdiff=0; end;
if k==3 idiff=0; end;
if i-idiff>=1 & j-jdiff>=1
  Prev_Min(k)=Dmin(i-idiff,j-jdiff);end;end;
[Dmin(i,j),k_ind]=min([Prev_Min;Dmin_old(i,j)]);
Dmin(i,j)=Dmin(i,j)+C_node;      idiff=1;      jdiff=1;
if k_ind==1 jdiff=0; end;
if k_ind==3 idiff=0; end;
if i-idiff>0 & j-jdiff>0
if i~=1 & j~=1 & i-idiff~=I & j-jdiff~=J
C_t(i,j)=1+C_t(i-idiff,j-jdiff);
else
C_t(i,j)=C_t(i-idiff,j-jdiff);    end;  end;
if k_ind==4 C_t(i,j)=C_told(i,j); end;
T_Cost=0;
if C_t(i,j)>0 T_Cost=1; end;
Dmin(i,j)=Dmin(i,j)+T_Cost;
Dmin_old=Dmin;
C_told=C_t;    end;    end;
Dmin_F=1e4;     C_t_F=1;
for i=I-eps:I
if Dmin(i,J)<Dmin_F
Dmin_F=Dmin(i,J);
C_t_F=C_t(i,J);    end;     end;
for j=J-eps:J-1
if Dmin(I,j)<Dmin_F
Dmin_F=Dmin(I,j);
C_t_F=C_t(I,j);    end;    end;
Dmin_Avg=Dmin_F/C_t_F;
```