التعرف الآلي على الكلام العربي: التطورات والتحديات

أ.د. مصطفى الشافعي

ملخص:

شهدت الآونة الأخيرة اهتمامات وتطورات سريعة في تكنولوجية التعرف على الكلام العربي لما لها من أهمية كبيرة في البحث والتصنيف الصوتي وتحويل الكلام إلى نص مكتوب والترجمة الفورية والتعليم والتطبيقات التي تتطلب تفاعل المستخدم مع الحاسب أو الآلة. ومن المتوقع ان يكون التخاطب بالكلام الطبيعي التلقائي هو التكنولوجيا الرئيسية للتعامل مع أجهزة الحاسب والأجهزة المحمولة الذكية. وتقدم هذه الورقة أولا عرض مختصر للتطبيقات الحديثة لنظم التعرف على اللغة. ثم تستعرض المكونات الرئيسية لهذه النظم وخاصة قواعد المعلومات المطلوبة لتدريب النظام على اللغة العربية. ثم نستعرض تسلسليا التطورات الرئيسية في هذا المجال حتى الآن. وتقدم الورقة أيضا عرضا للبحوث التي قامت بها المجموعة البحثية في جامعة الملك فهد للبترول والمعادن بمساعدة مدينة الملك عبد العزيز للعلوم والتقنية وكذلك آخر الاتجاهات البحثية حتى الآن. وتختم الورقة بعرض عدد من المشاكل والتحديات التي تواجه الباحثين وتتطلب تعاون ودعم لتحتل اللغة العربية المكانة التي تليق بمكانتها التي حباها الله بها.

١- مقدمة:

مما لا شك فيه أن التقدم السريع في تكنولوجيا المعلومات والاتصالات يعيد تشكيل مجتمعنا ويزيد إلى حد بعيد قدراتنا على التعلم بشكل أسرع، وكفاءتنا في الإنجاز، ومقدرتنا على تواصل أفضل وأوسع وتعاون أكثر فعالية وإنتاجية. وأحد أهم الوسائل التكنولوجية هي واجهة المستخدم، وهي كيف يمكن جعل تعامل الانسان مع الآلة بيئة أكثر طبيعية وذات فاعلية باستعمال وسائل التواصل الطبيعية بين البشر. ولهذا عكف الباحثون على اختراع وتطوير وسائل للاتصال بين الإنسان والآلة لتحاكي قدرة الأنسان على الكلام والرؤية. ومن الإنصاف أن نقول إنه على الرغم من العديد من النجاحات التكنولوجية المذهلة التي تحققت في تطوير قدرات الجهاز حتى الآن فهي لا تزال بدائية مقارنة مع نظرائها في الإنسان، وبالتالي فإن الجهود البحثية سوف تستمر لا هوادة فيها لتحقيق هذا الهدف بعيد المدى.

يعتبر التعرف الآلي على الكلام من أهم مكونات تقنية المعلومات والاتصالات التي تجتاح العالم الآن. ولذلك فإن تطوير نظم التعرف الآلي على الكلام بما يتمشى مع لغة كل بلد أصبح ضرورة استراتيجية هامة لتوسيع مجالات الاستفادة من تقنية المعلومات، وبالتالي إحداث التطوير الاقتصادي المرجو.

فالخطاب هو الشكل الطبيعي للتواصل بين البشر. إلا أن انتاج الكلام والتعرف عليه عملية معقدة تتأثر بكثير من العوامل مثل العمر والجنس ومعدل الكلام، واللهجات المختلفة واللهجات الإقليمية، والحالة النفسية. هذا بالإضافة إلى الصعوبات المتعلقة بالضوضاء والأصوات الخلفية، والتداخل من المتحدثين الآخرين، والخصائص الصوتية للغرفة، ومعدات التسجيل وخصائص قناة الاتصال.

مع دخول الحاسب الآلي في الكثير من التطبيقات المعاصرة ومع زيادة الاهتمام في الأبحاث التي تعنى بالمعالجات الصوتية، كان لتقنية التعرف الآلي على الكلام نشاط واضح في عدد من اللغات العالمية وخصوصا اللغة الإنجليزية. وفي مقالة نشرت عام ٢٠٠٧ [١] احصى المؤلفون بعض من تطبيقات التعرف على الكلام، نذكر منها على سبيل المثال لا الحصر:

- تحويل النص المكتوب الى كلام منطوق وبالعكس (الاملاء الآلي)
- الترجمة الآلية الفورية
- البحث والتصنيف الصوتي
- التعرف على المواضيع

- تعلم اللغة العربية وتجويد القرآن
- الخدمات الهاتفية الاستعلام الآلي
- التعليم بالوسائط المتعددة
- الأنظمة التفاعلية مع الحاسب والأجهزة المنزلية
- خدمات ذوي الاحتياجات الخاصة
- انظمة التحكم

ومع أن البحث العلمي في مجال التعرف الآلي على الكلام العربي قد بدأ متأخرا، إلا أن هناك العديد من التطبيقات والابحاث التي ظهرت في الفترة الأخيرة لخدمة هذا الحقل العلمي الهام. وتجدر الإشارة إلى أن هذا التأخر في الانتاج العلمي والبحثي لا ينفي أصالة علم الصوتيات عند العلماء العرب لقرون عديدة خلت. فقد اهتم العلماء العرب ومنذ القرن الثاني الهجري بالصوتيات نظراً للحاجة إلى معرفة الوجوه الصوتية التي اشتملت عليها القراءات القرآنية والتي أدت إلى ظهور علم التجويد، وهو العلم الذي يحدد الطريقة المثلى والصحيحة لتلاوة القرآن الكريم.

أما في العصر الحاضر فقد توسع هذا الاهتمام بحيث شمل توظيف تقنية التعرف على الكلام وتأليفه في علوم القراءات القرآنية. فقد طورت البرمجيات المختلفة لقياس مدى مطابقة التلاوة لأحكام التجويد [٢] كما طورت برامج أُخرى لتدريب المتعلمين على القراءة حسب قارئ معين. واهتم الباحثون مؤخراً بتطوير البرمجيات الصوتية الخاصة بالبحث عن موضوع معين في كم هائل من الدروس والخطب الصوتية [٣]. كأن يقوم المستخدم بالاستعلام صوتياً عن كلمة (الزكاة) مثلاً ليقوم النظام بالبحث في جميع الدروس والخطب التي تتحدث عن هذا الموضوع وذلك حسب البيانات المتوفرة في قاعدة البيانات الصوتية. فلم يعد المستخدم مقيداً بالبحث في النصوص مع توفر إمكانية إجراء الاستعلامات صوتيا داخل ملفات صوتية.

ولعبت أنظمة التعلم والتعليم الإلكترونية دوراً هاما في الاهتمام بالتعرف الآلي على الكلام، حيث يساعد نظام التعرف على الكلام المدرسين والمتعلمين على حد سواء في الدراسة الذاتية بحيث يقوم النظام بالاستماع لحديث المتكلم وتقييم دقة كلامه ودرجة صحته. كما تقوم هذه الأنظمة بإعطاء مصادر الخطأ للمتعلمين وإرشادات عن كيفية تجاوزها. إن نظاماً من هذا النوع يسمح لدارسي اللغات الأجنبية بالاستعانة بتقنية التعرف على الكلام من أجل إثراء معرفتهم وقدرتهم على نطق الكلام بالشكل الصحيح بصورة ذاتيه. فاللغة العربية أحوج ما تكون لمثل هذه الأنظمة في ظل الإقبال الكبير لتعلم هذه اللغة في الوقت الحاضر. ومن الأمثلة على استخدام تقنية التعرف على الكلام في أنظمة التعلم والتعليم الإلكتروني نظام حفص (عبدو ومن معه، ٢٠٠٦) [٢] حيث يقوم النظام بالاستماع إلى المستخدم وتقييم جودة قراءته للقرآن الكريم كما يستطيع النظام أن يحدد مكان الخطأ وإعطاء إرشادات لتجنب الخطأ. إن مثل هذا النظام يعتبر ذو أهمية كبيرة خصوصاً في البلاد التي لا يتوفر بها مدرسين لعلم التجويد والقراءات القرآنية. وقس على ذلك أهمية توفر مثل هذه التقنية في البلاد التي لا يتوفر بها مدرسين للغة معينة كالعربية مثلاً. إن الحصول على جودة عالية في أنظمة التعلم والتعليم الإلكتروني بحاجة إلى مراعاة عدد من الأمور منها: قدرة النظام على التعامل مع الألفاظ المتحدث بها من أشخاص غريبين على اللغة [٤] Non-native speakers بحيث يتوجب على النظام أن يكون قادراً على التعرف على كلامهم. وبرغم الكم الهائل من الناطقين بهذه اللغة علاوة على اللغات التي تستخدم حروف العربية كالفارسية والأوردو، إلا أن الاستفادة من التقنية الحديثة بقيت بعيدة نسبياً عن اللغة العربية إذا ما قورنت بلغات أخرى كالإنجليزية والصينية وغيرها من اللغات العالمية.

ومن المتوقع ان يكون التخاطب بالكلام الطبيعي التلقائي هو التكنولوجيا الرئيسية للتعامل مع الحاسبات والأجهزة المنزلية وأن تكون هي التكنولوجيا المهيمنة على واجهة الإنسان والآلة في المستقبل القريب. ومع ذلك، على الرغم من العديد من الإنجازات الرائعة في مجال التعرف على الكلام، والوصول إلى مستوى الأداء البشري لا يزال هدفا بعيد المنال.

وسنستعرض في الباب التالي مكونات نظام التعرف على الكلام. ويقدم الباب الثالث عرضا مختصرا عن الجهود البحثية المبذولة مؤخرا. أما الباب الرابع فيقدم البحوث التي انجزتها المجموعة البحثية في جامعة الملك فهد للبترول والمعادن والتحديات التي تواجه

الباحثين في هذا المجال.

٢-نظام التعرف الآلي على الكلام

والطريقة السائدة في بناء نظم التعرف على الكلام الحديثة تعتمد على الطرق الاحصائية في بناء نماذج للوحدات الصوتية وللغة. ومن أهم النماذج المستعملة في بناء الوحدات الصوتية هي نماذج ماركوف المخفية التي اقترحها باكر [٥] (١٩٧٥) وربينر و جوانج [٦] (١٩٩٣) وجلينك [٧] (١٩٩٨). وكان أول تطبيق ناجح لمورجان وبورلاند [٨] (١٩٩٥) ويانج [٩] (١٩٩٦).

وعادة تتكون مراحل التعرف على الكلام من ثلاثة مراحل [١٠]. في المرحلة الأولى يتم استقبال الكلام وتحويلة الى صورة رقمية ثم تقسيمه إلى اطارات من ١٠ مللي ثانية واستخلاص السمات المميزة له. وفي الخطوة الثانية تستخدم السمات المميزة للصوت في التعرف على الحروف الصوتية. وينتج عن هذه المرحلة عدة سلاسل محتملة من الحروف الصوتية. وفي المرحلة الثالثة يتم تحويل السلاسل الفونيمية إلى كلمات واختيار تتابع الكلمات الأكثر احتمالية.

ويمكن تقسيم نظم التعرف على الكلام إلى نوعين. النوع الأول وهو يستعمل في التحكم والاستعلامات dialog-based حيث يتم التحدث مع الحاسوب طبقا لخطة أو أسئلة أو اختيارات محددة [١١]. وفي هذه الحالة يتطلب التعرف على عدد محدود نسبيا من الكلمات والجمل. وحققت هذه التطبيقات نجاحا باهرا ودقة عالية. والنوع الثاني large vocabulary speech recognition هو الأكثر صعوبة وهو يتطلب التعرف على الكلام من عدد كبير من الكلمات والجمل المحتملة.

نتناول في هذه الدراسة تقنية التعرف الآلي على الكلام المتصل [٨] حيث عمل الباحثون لعشرات السنين لتعزيز هذه التقنية ورفع أدائها لما له من فوائد كبيرة. وهناك نظامين رئيسيين للتعرف على الكلام يستخدمها الباحثون، وهما نظام سفينكس [١٢] [١٣] [١٤] من جامعة كارنيجي ميلون في الولايات المتحدة، ونظامHTK [١٥] [١٦] من جامعة كامبريدج في بريطانيا. ويستخدم كلا النظامين نماذج ماركوف المخفية في عملية النموذجة الصوتية [١٧] [١٨]. وتعتبر نماذج ماركوف المخفية الطريقة المثلى حتى اليوم في النماذج الصوتية (Acoustic Models) وذلك لقدرتها على التعامل مع الأطوال المختلفة للوحدات الصوتية، إذ أن الفونيم الصوتي يمكن أن يكون بأطوال مختلفة لنفس المتحدث. إن هذه الإمكانية لنماذج ماركوف المخفية غير موجودة في الطرق الأخرى للتعرف على الكلام مثل طريقة الشبكات العصبية. ولهذا السبب فإن أغلب الأعمال البحثية في هذا المجال قد استخدمت نماذج ماركوف المخفية كما هو موضح في المسح الأدبي في الباب الثالث.

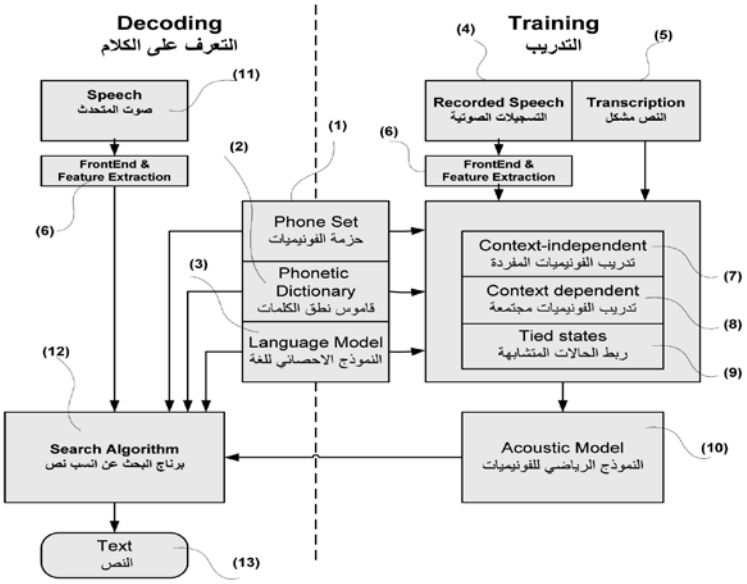
يقسم الكلام لأغراض التعرف، إلى ثلاثة أنواع: الكلام المتصل، والكلام المنفصل (كلمات أو أوامر معينة)، والأرقام. كما أن النظام قد يصمم للتعامل مع مستخدم واحد فقط (speaker-dependent) وقد يكون لأكثر من مستخدم [١٩].

يوضح الشكل رقم ١ الاجزاء الرئيسية في نظام التعرف على الكلام. وينقسم النظام إلى مجموعة البرمجيات وقواعد البيانات المستعملة في تدريب النظام، ومجموعة البرمجيات وقواعد البيانات الخاصة بالتعرف على الكلام. وسنبدأ أولا بالأجزاء الخاصة باللغة المطلوبة وهي مطلوبة في كلا المرحلتين: التدريب والتعرف على الكلام.

(١) الحزمة الفونية: وهي قائمة بالحروف الصوتية التي تكون نطق جميع كلمات اللغة. يُعرَّف الحرف الصوتي لكلمة معينة على أنه أصغر وحدة صوتية يتم من خلاله الحصول على كلمة مختلفة عند استبداله بحرف آخر. ويُعتبر الحرف الصوتي الوحدة الأساسية المستخدمة في عملية التعرف على الكلام. حيث يوضح الجدول ١ الحروف الصوتية المستخدمة في النظام، وقد استعملنا حزمة من ٤٥ وحدة صوتية تغطي الاصوات العربية وبعض التغييرات الشائعة في نطق أحرف عربية وكذلك اصوات من اللغة الإنجليزية حتى يمكن التعامل مع الأسماء الأجنبية. وقد اختيرت هذه المجموعة من الفونيميات بناء على خبرتنا السابقة في تحويل النصوص العربية إلى كلام وقد حصلنا على هذا التصنيف من [٢٠] [٢١] [٢٢].

الجدول ١. الحروف الصوتية العربية كاملة مع ترميزها المستخدم في النظام

الحرف الصوتي	الترميز المستخدم في النظام	الحرف الصوتي	الترميز المستخدم في النظام
ر (Reh)	/R/	بَ◄ـَ Fatha	/AE/
ز (Zain)	/Z/	بَا◄باب	/AE:/
س (Seen)	/S/	فتحة مفخمة بَ◄خَ	/AA/
ش (Sheen)	/SH/	فتحة نصف مفخمة بَ◄دَ	/AH/
ص (Sad)	/SS/	بُ◄ـُ Damma	/UH/
ض (Dad)	/DD/	بُو◄دُون	/UW/
ذ (Thal)	/DH/	بُ◄غُصْن	/UX/
ر (Reh)	/R/	بِ◄ـِ Kasra	/IH/
ط (Tah)	/TT/	بِي◄فِيل	/IY/
ظ (Thah)	/DH2/	بِ◄صِنف	/IX/
ع (Ain)	/Al/	بُو◄لُوم	/AW/
غ (Ghain)	/GH/	بَي◄ضَيف	/AY/
ف (Feh)	/F/	ء (Hamza)	/E/
ق (Qaf)	/Q/	ب (Beh)	/B/
ك (Kaf)	/K/	ت (The)	/T/
ل (Lam)	/L/	ث (Theh)	/TH/
م (Meem)	/M/	جيم فصحة (Jeem)	/JH/
ن (Noon)	/N/	جيم مصرية	/G/
ه (Heh)	/H/	ح (Hah)	/HH/
و (Waw)	/W/	خ (Khah)	/KH/
ي (Yeh)	/Y/	د (Dal)	/D/
		ذ (Thal)	/DH/



شكل ١: مكونات نظام التعرف على الكلام.

(٢) القاموس الفونيمي: وهو قاموس يحتوي على الكلمات التي ستستخدم في التطبيق مع طريقة نطقها (كتابتها) بالحروف الصوتية. يعتبر القاموس الصوتي أحد أركان انظمة التعرف الآلي على الكلام، وذلك لأنه يمثل حلقة الوصل بين الكلمات الموجودة في المدونة (بيانات التدريب والفحص) وبين الحروف الصوتية الخاصة بكل كلمة على انفراد. وهناك ثلاثة طرق لتكوين هذا القاموس: الطريقة اليدوية بواسطة خبراء في الصوتيات العربية، وطريقة قواعد صرف اللغة العربية بتكوين الكلمات ونطقها ابتداء من الجزور العربية وإضافة السوابق واللواحق وغيرها من الكلمات المشتقة [٢٣] [٢٤] ولكن أحد عيوب هذه الطريقة أنها تضيف كلمات كثيرة غير مستعملة. والطريق الثالثة تستعمل قواعد نطق اللغة العربية لتكوين النطق الفونيمي للمفردات الموجودة بالفعل في النصوص المتوقعة للتطبيق. وقد قمنا في [٢٥] بتطوير برنامج يقوم بإنتاج القاموس الصوتي باستعمال قواعد نطق اللغة ونشير هنا الى ان هذا البرنامج ينتج القاموس مع الاخذ بعين الاعتبار عدد من الاختلافات التي قد تظهر في نطق الكلمة. ومن الأمثلة على ذلك تطبيق قاعدة الإقلاب إذا حصل داخل الكلمة، وليس بين كلمات متجاورة. بالإضافة الى ذلك فانه يأخذ بعين الاعتبار همزة الوصل في بداية الكلمة والتاء المربوطة في نهايتها. ومن الامثلة على ذلك ما هو موضح في الشكل ٢. وقد قمنا بعمل قاموس فونيمي لأكثر من ١٢٠ ألف كلمة عربية.

إن الكلمة الاولى في الشكل ٢ تشير الى اللفظ الافتراضي لكلمة "أدنبرة"، اما الكلمات المرقمة التالية من ١ إلى ٣ فهي تشير الى الالفاظ المختلفة التي يمكن حدوثها اثناء نطقها. وللتعرف على الكلام التلقائي يضاف أيضا قائمة تمثل الأصوات غير المكتوبة مثل "الضوضاء"، "التنفس" و "آه" وغيرها من الأصوات غير الإرادية أو اصوات خلفية مثل فتح الباب وغلقه أو ضجيج السيارات إلى آخره.

مثال لتعدد نطق كلمة واحدة	أمثلة من القاموس الفونيمي
أدِنبرَة E AE D IH M B R AA H	آبَارِ E AE: B AE: R IX N
(١)أدِنبرَة E AE D IH M B R AA T	آخَر E AE: KH AA R
(٢)أدِنبرَة E AE D IH N B R AA H	آخَرَ E AE: KH AA R AA
(٣)أدِنبرَة E AE D IH N B R AA T	آخَرُوْنَ E AE: KH AA R UW N AE
	آخَرِيْنَ E AE: KH AA R IX: N AE

شكل ٢: أمثلة من القاموس الفونيمي.

(٣) النموذج الإحصائي للغة. ويعرف أيضا بالنموذج اللغوي. ولبناء هذا النموذج لابد من وجود مدونة للنصوص المشكلة التي سيتم تدريب النظام على التعرف عليها. ويقوم برنامج خاص بالتعامل مع قاعدة النصوص واستخراج المفردات وحساب احتمالاتها وتكوين جداول للتتابع الثنائي والثلاثي للكلمات وحساب احتمالات كل منها [٢٦]. يستخدم النموذج اللغوي في كثير من تطبيقات معالجة اللغات الطبيعية وأنظمة التعرف على الكلام وأنظمة الترجمة الفورية وأنظمة التشكيل الآلي. ويعتبر أحد المكونات الاساسية في هذه الانظمة. حيث يتم بناء النموذج اللغوي بإسناد قيم الاحتمالات التالية (بالاعتماد على المدونة):

١) إيجاد احتمالية الكلمات المستعملة

٢) إيجاد احتمالية كلمتين متتابعتين

٣) إيجاد احتمالية تتابع ثلاث كلمات

ان الهدف الرئيس من وجود النموذج اللغوي هو توقع حدوث كلمة معينة بعد كلمة او كلمتين متتاليتين، وبذلك امكانية تشكيل الجملة التي يراد التعرف عليها. فعند تحديد المقطع الصوتي للتعرف عليه، يبدأ النظام بالتعرف على الكلمة بالاستعانة بالنموذج الصوتي والقاموس الصوتي. بعد تحديد الكلمة الأقرب في القاموس الصوتي والحصول على كلمة جديدة مجاورة، يتم استخدام النموذج اللغوي لربط الكلمتين، ثم لربط ثلاث كلمات، وهكذا تستمر العملية حتى يتم تشكيل الجملة التي احتوى عليها المقطع الصوتي.

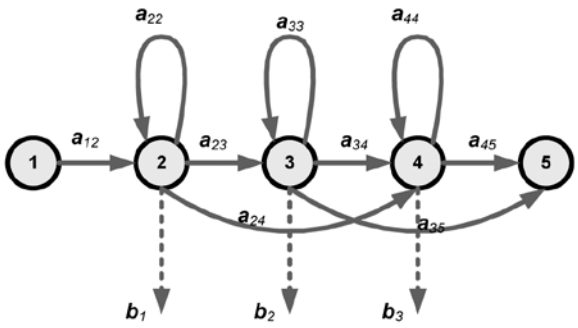
(٤) التسجيلات الصوتية. وهي ضرورية لتدريب النظام على سماع الأصوات وتكوين نماذج للفونيميات. ولتدريب النظام للتعرف على

الكلام من أي مستخدم يجب ان تكون التسجيلات الصوتية من عدد كبير من الناس. ولتقليل نسبة الخطأ في التعرف على الكلمات يجب تدريب النظام على اعداد كبيرة من التسجيلات. وفي النظام الذي تم تطويره في جامعة الملك فهد للبترول والمعادن [٢٢] لتدوين الأخبار العربية قمنا بعمل قاعدة صوتية من التسجيلات الصوتية للغة العربية الحديثة بلغت ٢٤ ساعة من عدد محدود من المحطات الفضائية. وبلغت التكلفة المباشرة لتسجيل كل ساعة وكتابتها أكثر من ٢٠٠٠ دولار. ووصلت نسبة الخطأ في الكلمات حوالي ١٤٪. وجدير بالذكر أن النظام الذي طورته IBM لكتابة الأخبار الإنجليزية من أجهزة الراديو والتلفاز تطلب استخدام ١٩٧ ساعة من التسجيلات الصوتية [٢٧]. بجانب نصوص مكتوبة من الصحب تعادل أكثر ٤٠٠ مليون كلمة كان معدل الخطأ في الكلمات ٢١٪. فوجود قاعدة صوتية مدونة في غاية الأهمية لبناء التطبيقات المختلفة مثل تعليم اللغة العربية وتعليم تجويد القرآن والتعرف الآلي على الشعر العربي [٢٨] وغيرها من التطبيقات الحديثة.

(٥) النص المشكل. وهو النص الذي يقابل كل تسجيل صوتي والنصوص التي تغطي الجمل المتوقعة في موضوع المحادثة. ومن الأفضل أن تكون كل النصوص مشكلة ومحققة.

(٦) المعالجة الأمامية واستخراج الخصائص الصوتية. وفي هذه المرحلة يتم تحويل الصوت إلى صورة رقمية بمعدل ١٦ ألف عينة في الثانية وتقسيم الصوت إلى اطارات بطول ١٠ مللي ثانية. يتم في هذه المرحلة التخلص من الضوضاء المحيطة بالإشارات الصوتية الرقمية بحيث يتم استخلاص السمات المميزة للصوت بوضوح. أيضا يتم توحيد زمن تردد الأصوات ومعالجتها بعدة خطوات لتكوين قائمة من ١١ قيمة تمثل كل ١٠ مللي ثانية من الصوت. وهناك طريقتين MFCC and PLP وإن كانت طريقة MFCC هي الأكثر انتشارا في النظم الضخمة للتعرف على الكلام [١٩].

(٧) تدريب الفونيميات المنفصلة. وفي هذه المرحلة يتم بناء نموذج رياضي لكل حرف صوتي منفصل. وتعتمد الطرق الحديثة على استعمال نماذج ماركوف الخفية (Hidden Markov Models) وهي سلسلة من الحالات ويتم الانتقال من حالة الى حالة أخرى عن طريق مصفوفة الاحتمالات. يمكن معرفة الناتج من الدالة العشوائية وعلاقتها مع الحالة ولكن لا يمكن معرفة مراحل الإنتقال من حالة الى أخرى ولا معرفة تسلسل الملاحظات، لذلك سُميت بنماذج ماركوف الخفية. والنموذج المستعمل للحرف الصوتي يتكون عادة من ثلاثة حالات فعالة وحالتي الدخول والخروج من النموذج كما هو مبين في شكل ٣. وهذه الحالات تمثل تقريبا بداية الوحدة الصوتية ووسطها وآخرها.



شكل ٣ نموذج ماركوف المخفي.

(٨) تدريب الأحرف الصوتية مجتمعة. وفي هذه المرحلة يتم بناء نماذج لكل الوحدات الصوتية مع الأخذ في الاعتبار تأثير الاحرف الصوتية السابقة والتالية. وتسمي هذه النماذج تراي فون أي الأحرف الصوتية الثلاثية. فمثلا يتم بناء نموذج لحرف م إذا كان قبله ن وبعده واو (ن–م–و) وبالمثل (ش–ن–ط) إلى آخره. كما يتم بناء نموذج رياضي لاحتمال كل مرحلة من نموذج ماركوف المخفي مع

أي سمات صوتية ملحوظة.

(٩) ربط الحالات المشابهة. ونظرا للأعداد الكبيرة من الترايفون وعدد المعاملات اللازمة لوصف كل نموذج تركز هذه المرحلة على اكتشاف المتشابه من الترايفون وربطها ببعض أو جعلها تشارك في دوال التوزيع الاحتمالي لحالات نماذج ماركوف الخفي [١٩]. ويؤدي ذلك إلى تقليل عدد النماذج الرياضية وعدد المعاملات وحجم الذاكرة المطلوبة للنظام وتسريع عملية التعرف على الكلام.

(١٠) النموذج الرياضي للفونيميات. وبعد انتهاء تدريب النظام يتم تخزين النماذج الرياضية للترا يفون وتكون جاهزة لمرحلة التعرف على الكلام.

(١١) المتحدث. وتبدأ عملية التعرف على الكلام من الميكرفون الذي يحول الاصوات إلى اشارات كهربية ويتم تحويلها إلى قراءات رقمية بالمعدل المطلوب وتعالج بواسطة خوارزميات لإزالة الضوضاء وتقطع إلى إطارات وتكوين السمات الخاصة بكل إطار كما ذكر في الخطوة (٦).

(١٢) البحث والتمييز. في هذه المرحلة يتم البحث في النماذج الفونيمية عن أحسن تتابع للحروف الصوتية يمثل الصوت المنطوق. ويستعمل القاموس لتحديد انسب تتابع من الكلمات يمثل سلسلة الحروف الصوتية ويدخل في الاعتبار هنا النموذج اللغوي لإيجاد أحس احتمالات لتتابع الكلمات.

(١٣) النص. وهنا يتم كتابة النص الذي تم التعرف عليه وهو عبارة عن سلسلة الكلمات التي لها أعلى احتمال تراكمي.

٣- المسح الأدبي:

نعرض في هذا الباب بعض من جهود الباحثين في مجال التعرف على الكلام العربي. فقد قدم الشافعي في [٢٠] (١٩٩١) وصفا لنظام تجريبي لتحويل النص العربي إلى أصوات حيث تناول المؤلف تقسيم أ صوات اللغة العربي وتأثرها بموقعها في المقطع الصوتي وقواعد الإدغام والإقلاب والإخفاء وتتابع الأصوات ومخارجها واستعمال الحركات القصيرة والطويلة. وقام [٣٠] (١٩٩٦) باستخدام نموذج ماركوف المخفي لعمل نماذج صوتية للفونيميات العربية. ونشر الغامدي [٣١] (٢٠٠٠) كتابا باللغة العربية تناول فيها الصوتيات العربية بمزيد من التفصيل. واستخدم باهي و ساليمي [٣٢] (٢٠٠٣) نظاما مدمجا مكونا من الشبكات العصبية وأنظمة ماركوف المخفي للتعرف على الكلام العربي. وقدم الشوباصي ومن معه [٣٣] (٢٠٠٣) نظاما للتعرف على الكلام العربي المنفصل باستخدام الشبكات العصبية الراجعية. وعقدت ورشة عمل في جامعة جون هوبكنز لمناقشة مسألة التعرف على الكلام العربي والعقبات التي تواجه هذه التقنية [٢٣] (٢٠٠٣). وقدم شعيب ومن معه [٣٤] (٢٠٠٣) نظاما للتعرف على الكلام العربي باستخدام خواص (intensity contours) و (formant frequency). وأنجزت مدينة الملك عبد العزيز للعلوم والتقنية في بحث للغامدي [٣٥] (٣٢٠٠) قاعدة بيانات صوتية لمتحدثين سعوديين حيث ضمت تفاصيل دقيقة عن أصوات اللغة العربية من حيث مخارج الحروف وكيفية اصدارها.

واشار فيجري [٣٦] (٤٢٠٠) إلى أن توظيف علم الصرف في بناء نظام التعرف على الكلام يساعد على تحسن الأداء. بينما قدم نوفل ومن معه [١١] (٢٠٠٤) نظاما للتعرف على أوامر التحكم في اللغة العربية باستخدام نماذج ماركوف المخفية. وقدمنا في [٣٧] (٢٠٠٤) دراسة عن القواعد الفونولوجية العربية. أما كيرشوف وفيرجيري [٣٤] (٢٠٠٤) فقد قدموا نظاما مدمجا قادرا على التعرف على الكلام لأكثر من لهجة للغة العربية المعاصرة بما في ذلك اللهجة المصرية. وقدم العتيبي [٣٩] (٢٠٠٤) نظام للتعرف على الأرقام العربية باستعمال الخلاية العصبية وقد حقق النظام دقة بلغت ٩٩,٥٪ لنظام متعدد المتحدثين. وقد قدم بوروبا ومن معه [٤٠] (٢٠٠٦) نظاما يجمع بين نماذج ماركوف المخفية وخلاية عصبية تعرف للتمييز تعرف بـ support vector machine. وعرض الشمسي وجوسيوم [٤١] (٢٠٠٦) طريقة لوسم كلمات الجمل العربية باستخدام نموذج ماركوف المخفية. وعرض شوتير ومن معه [٤٢] (٢٠٠٦) نظاما صوتيا لتعليم اللغة العربية يعتمد على المورفيم (المقاطع الصوتية) كأساس لإنتاج النموذج اللغوي. وقد طور عبدو ومن معه [٢] (٢٠٠٦) نظاما صوتيا لتعليم اللغة العربية لغير الناطقين بها. واقترح كسيانج ومن معه [٤٣] (٢٠٠٦) طريقة لتعزيز الأداء في أنظمة التعرف على الكلام العربي وذلك بتقسم الكلمات صرفيا إلى أجزاء مختلفة ومن ثم تمثيل هذه الأجزاء في النماذج الصوتية واللغوية. وقدم سولتاو

ومن معه [٤٤] (٢٠٠٧) نظام للتعرف على الكلام GALE (Global autonomous language exploitation)

حيث احتوى النظام على ١٨٠٠ ساعة صوتية مشكلة وغير مشكلة. واستخدم ساتوري ومن معه [٤٥] (٢٠٠٧) نظام سفنكس ٤ للتعرف على الأرقام العربية. كما استخدم أمامي و مانجو [٤٦] (٢٠٠٧) الشبكات العصبية لإنتاج النموذج اللغوي في انظمة التعرف على الكلام العربي. وقدم طه ومن معه [٤٧] (٢٠٠٧) نظاما للتعرف على الكلام العربي المنفصل باستخدام الشبكات العصبية.

وقدم عزمي ومن معه [٤٨] (٢٠٠٧) نتائج تجاربه لنظام للتعرف على الأرقام تم بنائه بواسطة مدونة صوتي بواسطة ٤٤ شخصا. وكانت نسبة الخطأ بين ٠,٥٪ إلى ٢,٥٪. وقدم عزمي وتولبا [٤٩] (٢٠٠٨) نظاما للتعرف على الكلام العربي باستخدام نظام جامعة كمبردج HTK وقارن الباحثان بين استخدام الفونيميات والمورفيميات Syllables)) وقد وجدوا أن الأداء باستخدام المورفيميات يفوق أداء الفونيميات. وفي عام (٢٠٠٨) قدم ستالارد ومن معه نظام للترجمة الصوتية من اللهجة العراقية إلى الإنجليزية [٥٠]. وقدم شوي ومن معه أيضا في نفس العام [٥١] (٢٠٠٨) نظاما للترجمة الصوتية بين اللهجة العراقية واللغة الانجليزية. واستخدم بارك ومن معه [٤٨] (٢٠٠٩) خلاية عصبية في بناء نظام للتعرف الآلي على الكلام العربي. وقدم المهدي ومن معه [٥٢] (٢٠٠٩) نظاما للتعرف على الارقام العربية باللهجة المصرية باستخدام مدونة خاصة باللغة العربية المعاصرة. وقدمنا في عام ٢٠٠٩ أداة برمجية لإنتاج القواميس الصوتية للغة العربية المعاصرة مع الأخذ في الاعتبار المتغيرات الصوتية للكلمات[٢٥]. كما قدمنا في ٢٠٠٩ نظاما للتعرف الآلي على الكلام العربي المتصل وكتابة النص المقابل لمحطات الأخبار [١٠]. وركزت التسجيلات الصوتية على الأخبار الاقتصادية والرياضية. وحققت نسبة نجاح ٨٨٪ للنص المشكل ونسبة صواب ٩٣٪ إذا تم اهمال الخطأ في التشكيل. وقدم فرغلي وشعلان [٥٣] (٢٠٠٩) دراسة شاملة عن التحديات التي تواجه أنظمة معالجة اللغات العربية والحلول المقترحة. وقدم لاميل ومن معه [٥٤] (٢٠٠٩) نظاما للتعرف على الكلام العربي المتصل مع الأخذ في الاعتبار التغيرات الصوتية للهجات المختلفة. واقترح كوه ومن معه [٥٥] (٢٠١٠) استخدام الطرق التركيبية الصرفية لتحسين الأداء في انظمة التعرف الآلي على الكلام العربي. وقدم سلواني والعتيبي [٥٦] (٢٠١١) خوارزمية لتعديل النماذج الصوتية بحيث تكون قادرة على التعامل لمتحدث من غير أهل اللغة حيث تم العمل على اللغة العربية المعاصرة.

شهدت العشر سنوات الماضية اهتماما كبيرا باستعمال نموذج صوتي للوحدات الصوتية باستعمال الخلاية العصبية الصناعية واستعماله للتعرف على اللغات المختلفة لتطبيقات السحب والأجهزة المحمولة.

واقترح قام ليو [٥٧] (٢٠١١) نموذج يجمع بين نموذج ماركوف و SVM (support vector machine) وبين أن النموذج المقترح يحقق نتائج أفضل بكثير من نموذج ماركوف التقليدي.

وفي [٥٨] (٢٠١٢) استعملت الشبكات العصبية العميقة في التعرف على الكلام لما لها من قدرة على تمييز الفونيميات ونمذجة التأثيرات من الفونيميات القريبة والتغير نتيجة اللهجات. وكمثال لآخر التطورات في نظم التعرف على الكلام نذكر هنا آخر تقييم لنظام للتحدث مع الحاسوب طرته ميكروسوفت [٧٢] (٢٠١٧). فقد استعملوا الشبكات العصبي لبناء النموذج الصوتي و كذلك لبناء النموذج اللغوي. واستعملوا الشبكات العصبية المتراجعة recurrent neural networks في عمل النموذج اللغوي. وكان معدل الخطأ في الكلمات ٦,٢٪.

٣. التطورات والتحديات

نستعرض أولا النظام الذي طورناه في جامعة الملك فهد للبترول والمعادن [٥]. وهو نظام لنسخ الأخبار من القنوات الفضائي. وقد تم بناء قاعدة صوتية من حوالي ٧,٥ ساعة (KACSTv١,٩) لتطوير وتدريب نظام التعرف على الكلام العربي. واقتصرت على الموضوعات الاقتصادية والرياضية. وتم تسجيل الملفات الصوتية من عدد محدود من القنوات التلفزيونية الإخبارية العربية. واحتوت التسجيلات على ٢٣٤ موضوع. وقصمت إلى ٦١٤٦ مقطع صوتي. وكان متوسط طول المقاطع الصوتية ٤,٥ ثانية. وجدير بالذكر أن نظام نسخ الأخبار الفرنسي قد احتاج في المرحلة الأولى إلى ٤٤ ساعة [٥٩]. بينما نظام اللغة الانجليزية تطلب بناء قاعدة صوتية من ١٩٧ ساعة [٦٠]

وحقق نسبة خطأ في الكلمات WER ١٧,٥٪. وبالرغم أن القاعدة الصوتية التي بنيناها تعتبر محدودة إذا ما قورنت بالنظم الأخرة فقد حققت حوالي ١٤٪ على النصوص المشكلة. ويمكننا أن نحقق مزيدا من التحسن إذا ما توفر مزيدا من الدعم. واقتصرنا على اللغة العربية المعاصرة (MSA)، حيث أنها تستخدم على نطاق واسع وقبلت على المنطقة برمتها. ولكنها تحتاج إلى تحليل وتقنين ووصف علمي للفونيميات الخاصة بها وقواعد نطقها ومفرداتها. وهي تختلف اختلافا جوهريا في كتابتها ونطقها عن القرآن الكريم. وتحتاج إلى إدخال أصوات جديدة مثل G، P، V . فعلى سبيل المثل نجد أن كلمة شائعة مثل google تكتب غوغل أو قوقل. وهناك العديد من القضاية تحتاج إلى تقنين متفق عليه، مثال لذلك هل اتباع قواعد الإدغام والإقلاب ملزمة أو غير ملزمة؟ هل تفخيم بعض الحروف مثل ب في "بصره" والميم في "مطار" مقبول أو غير مقبول. فوجود تقنين standard متفق عليه ضرورة ملحة حتى تقتحم اللغة العربية مجالات التقنية الحديثة.

واقترح أبو شعيرة ومن معه [٦١] [٦٢] استعمال نصوص غنية بالحروف الصوتية للغة العربية المعاصرة لتدريب النظام الصوتي بكمية قليلة من التسجيلات وحققت نجاحا كبيرا.

النسخ:

وقد نسخت كل التسجيلات الصوتية في نظام (KACSTv١,٩) بنصها المشكل (diacritized). وهي تتكون من ٥٢٧١٤ كلمة وتتضمن قائمة المفردات ١٧٢٣٤ كلمة. تشكيل الكلمات يلعب دوراً أساسيا في عملية التعرف على الكلام. فمن وجهة نظر أنظمة التعرف على الكلام فإن التشكيل يساعد في تقليل الغموض أثناء عملية التعرف على الكلمة. فعلى سبيل المثال تحتمل كلمة (كتب) الألفاظ التالية (كُتِبَ، كَتَبَ، كُتُبٌ) التي يجب أن تأخذ بعين الاعتبار عند تحديد كيفية نطق الكلمات في القاموس الصوتي، لأنها ألفاظٌ مختلفة يجب أن يفرق بينها في الكتابة الصوتية للفونيميات في القاموس الصوتي. وبالتالي فإن قضية التشكيل تعتبر أساسية في أنظمة التعرف على الكلام العربي.

ومن المشاكل التي واجهتنا بجانب الأخطاء العديدة في التشكيل (حروف بدون تشكيل-حروف لها أكثر من تشكيل-تشكيل بدون حروف-وجود شدة بدون تشكيل) عدم وجود طريقة موحدة في التشكيل فمثلا الحرف الساكن قد يترك بدون تشكيل أو يكتب معه سكون. آداة التعريف قد تشكل كالآتي (ال-الّ -أل-أَل-أَلْ) والتنوين (سَرِيعًا-سَرِيعًا-سَرِيعًا-سَرِيعٌ). وهي تختلف أيضا اختلافا جوهريا عن تشكيل النص القرآني. فلابد من وجود قواعد متفق عليها لتسهل على الباحثين بناء مكانز للغة وتبادل نتائج ابحاثهم.

وقد قمنا مؤخرا ببناء قاعدة صوتية من ٢٤ ساعة. ولكن النسخ الأصلي مازال به العديد من الأخطاء التي تتطلب تصحيحات دقيقة ومراجعات قبل استعمالها ووضعها لخدمة الباحثين.

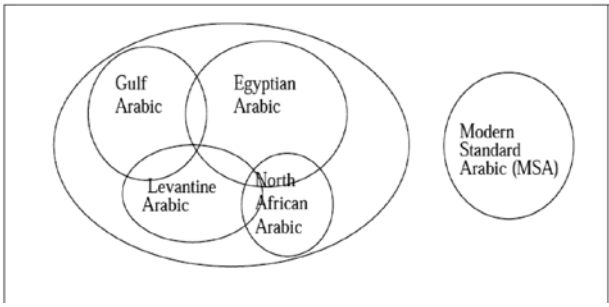
وقد واجهتنا هذه المشاكل أيضا في تطوير نظام للتشكيل الآلي [[٦٢] [٦٣] [٦٤] وما زالت هي العقبة الرئيسية لتحسين أداء المشكلات الآلية. وفي مجال التشكيل الآلي قمنا بدراسة عدة اتجاهات منها مثلا استعمال نموذج ماركوف المخفي على مستوى تتابع الكلمات [٦٢] [٦٣]، والعلاقات بين التشكيل وتتابع الحروف باستعمال النموذج الإحصائي للغة واستعمال الشبكات العصبية [٦٤].

التعرف على كلام اللهجات العربية:

بالإضافة إلى عوائق المدونة والتشكيل، فإن اللهجات المختلفة تشكل مسألة ثالثة يجب أن يتم مراعاتها عند تطوير أنظمة التعرف على الكلام العربي. فاللهجات العربية تفتقر إلى الدراسات الوصفية والتحليلية والدراسات الفونيمية. ولذلك فإن البحث العلمي في أنظمة التعرف على الكلام قد انصب على اللغة العربية المعاصرة. إلا أن هناك أعمالا تمت لخدمة لهجات خاصة كاللهجة العراقية واللهجة المصرية واللهجة السعودية وغيرها. وشكل رقم ٤ يبين تقسيم اللهجات العربية يمكن تقسيم اللهجات العربية، كما يمكن أن يحتوي بلداً ما على أكثر من لهجة.

ومن أنجح النظم للغة العربية الدارجة الآن نظام جوجل. حيث تم تجميع تسجيلات من اللهجات العربية [٥٩] كما هو موضح في الشكل رقم ٥. وهذه القاعدة من الاصوات تحتوي على ٦١٧ ألف كلمة وتمثل ٢ مليون تسجيل صوتي. ووجد أن نسبة الخطأ في الكلمات

وصلت إلى ١٨,٣٪ عند استعمال ١٩٦٧ تسجيل صوتي لتدريب النظام، وكانت نسبة الخطأ ١٩,٨٪ عندما استعمل ٥٣٨ ألف تسجيل لتدريب النظام.



Dialect	Set	# Speakers	# Queries	Duration
EG	train	604	245K	223 hrs
	test	29	15K	12.4 hrs
JO	train	848	260K	224 hrs
	test	21	15K	10.7 hrs
SA	train	745	299K	226 hrs
	test	29	15K	10.6 hrs
AE	train	587	235K	193 hrs
	test	29	15K	10.6 hrs
LB	train	795	264K	219 hrs
	test	29	15K	13.8 hrs

شكل ٤ اللهجات العربية [٥٩] شكل ٥ حجم التسجيلات الصوتية لتدريب نظام التعرف على الكلام [٥٨]

للتغلب على مشكلة تعدد اللهجات العربية وعدم وجود دراسات تحليلية وتقنين اقترحنا بناء طرق تقوم على نظم الحديثة للذكاء الصناعي للتعرف على الوحدات الصوتية لأي لغة. ففي [٦٦] [٦٧] قمنا باستعمال شبكة عصبي تعرف بـLVQ لتقسم السمات الصوتية ثم تجميعها في وحدات صوتية. واستعملنا أدوات سفنكس لتقسيم الكلام إلى كلمات ومن ثم عمل القاموس الفونيمي طبقا للوحدات الصوتية المستنبطة. وفي دراسة أخري تم إيجاد تقسيم للأصوات طبق لمعدل قطع الصفر ومستوى الطاقة ومستوي الأنتروبي. ويلي ذلك تحديد الحروف الصوتية للغة، ثم توليد القاموس الفونيمي للغة.

هناك خمس طرق نقترحها لتعزيز الأداء وهي دمج الكلمات الصغيرة، ودمج الكلمات حسب حالتها الإعرابية، ودمج الكلمات حسب القواعد الفونولوجية العربية، واستخراج المتغيرات الصوتية من بيانات التدريب، وإعادة تقييم الفرضيات باستخدام القواعد التركيبة. نوضح هذه الطرق في النقاط التالية.

طريقة دمج الكلمات الصغيرة [٦٨]

تصنف طريقة دمج الكلمات الصغيرة على أنها طريقة تعتمد على بيانات التدريب (data-driven method) حيث لا يتم استخدام أية معلومات لغوية مسبقة، إنما يتم استخدام كلمات المدونة النصية فقط. تقوم الطريقة على دمج الكلمات حسب مجموع طولهما حيث أن طول الكلمتين بعد دمجهما يعتبر المعيار للدمج من عدمه. أُجريت عدد من التجارب لدمج الكلمات الصغيرة، وقد تبين أن أفضل أداء يتحقق عند دمج الكلمات الصغيرة التي يكون مجموع أطوالهما ١٠ حروف. فقد ارتفع الأداء في هذه الحالة بنسبة ٢,١٦٪ بالنسبة للنظام المرجعي. والكلمات المدموجة تضاف إلى القاموس الفونيمي وتضاف الجملة التي تحتويها إلى المدونة ليتم إدخالها ضمن النموذج اللغوي.

طريقة دمج الكلمات حسب نوعها[٧١] [٦٩]

بالإضافة إلى دمج الكلمات الصغيرة فإن دمج الكلمات حسب نوعها (حالتها الإعرابية) يمكن أن يعزز الأداء في أنظمة التعرف الآلي على الكلام العربي. وتستخدم برمجيات تحديد أقسام الكلام (Part of Speech Tagging) لوسم الحالة الإعرابية للكلمات. فعلى سبيل المثال فقد وُجد أن المتحدث العربي عادةً ما يدمج الأسماء المتبوعة بالصفات مثل (منافسة شديدة)، وكذلك حروف الجر والكلمات التي بعدها مثل (في القدس). ويتم إنتاج كلمة مندمجة جديدة بنفس طريقة دمج الكلمات الصغيرة ليتم فيما بعد تمثيلها في جملة جديدة. وقد أدت هذه الطريق إلى تحسن في الأداء بلغ ٢,٢٨٪.

طريقة دمج الكلمات حسب القواعد الفونولوجية [٧٠]

تعتبر طريقة دمج الكلمات حسب القواعد الفونولوجية العربية واحدة من الطرق المستخدمة لتحسين الأداء في أنظمة التعرف الآلي على الكلام العربي، إذ يمكن استخدام قواعد النطق وأشهرها قواعد الإدغام والإقلاب لإنتاج كلمات جديدة تكون أكثر قرباً لطريقة نطقها الفعلي. وتشمل قواعد الإدغام النون الساكنة أو التنوين التي تأتي قبل ستة حروف مجموعة في كلمة (يرملون). ومثال على الإدغام (من رفعها تنطق مرّفعها). وكذلك فإن الإقلاب هو أن تأتي النون الساكنة أو التنوين قبل كلمة مبتدئة بحرف الباء بحيث تقلب النون الساكنة أو التنوين ميماً ساكنة مثل (من بعد تنطق ممّبعد). وقد تم التعامل مع الكلمات الناتجة من الاندماج بنفس طريقة التعامل مع دمج الكلمات الصغيرة. بعد تحديد المتغيرات الصوتية الجديدة للكلمات تُمثل في الجمل الخاصة بها ليتم استخدامها في بناء القاموس الصوتي والنموذج اللغوي. أظهر تمثيل القواعد الفونولوجية العربية في نظام التعرف على الكلام العربي تحسناً ملحوظاً في الأداء. فقد تحسن الأداء بنسبة ٢,٣٪.

طريقة اعادة تقييم الفرضيات [٧١]

باستخدام القواعد التركيبة إن الطرق التي تم ذكرها في النقاط السابقة كانت تتعلق بنمذجة بعض الظواهر الصوتية لمرة واحدة فقط بحيث يتم تمثيل المتغيرات في القاموس الصوتي والنموذج اللغوي. وتختلف هذه الطريقة عن الطرق السابقة في كونها تنفذ في كل مرة يتم فيها التعرف على الكلام. فبمجرد انتهاء عملية التعرف تخضع النتائج لعملية إعادة تقييم لاختيار الفرضية الأفضل من الناحية التركيبية. ويعاد تقييم الفرضيات من خلال عدد من قواعد اللغة تستخلص من بيانات المدونة النصية بعد وسم (Tagging) كلماتها. وعلى سبيل المثال، تحتوي الجملة التالية على الكلمات وأوسامها حيث تم استخدام أداة الوسم (Stanford tagger) لوسم الكلمات.

إلا أن الطريقة المقترحة على استخراج عدد من القواعد التركيبية باستخدام برنامج (ويكا) weka .

برنامج (ويكا) [٧٣] فهو عبارة عن أداة برمجية تستخدم في استخلاص العلاقات الأشهر بين كمية كبيرة من البيانات وهو أحد التطبيقات في برامج التنقيب عن البيانات (Data-Mining). إن المدخلات لبرنامج (ويكا) هي أوسام الكلمات فقط لكل الجمل. أما المخرجات فتتمثل في مجموعة القواعد الأشهر التي تم استنباطها من مجموعة أوسام الجمل المدخلة. استخدمت هذه القواعد لإعادة تقييم نتائج التعرف على الكلام من خلال إنتاج عدد من الفرضيات لكل جملة يتم فحصها (ميزة إنتاج عدد من الفرضيات متاحة في أنظمة التعرف على الكلام ومنها سفينكس). ويعاد تقييم الفرضيات من خلال إيجاد الفرضية التي تتناسب أكثر مع قواعد اللغة التي تم استخلاصها. فمن الممكن أن تكون الفرضية الثالثة متوافقة أكثر مع القواعد المستخلصة بشكل أكبر من الفرضية الأولى (الأعلى احتمالاً حسب ما هو مستخدم في أنظمة التعرف على الكلام) ففي هذه الحالة يتم اختيارها (أي الفرضية الثالثة) كأفضل نتيجة ممكنة ويتم بالتالي إرجاعها إلى المستخدم كنتيجة نهائية. ولكن لم نتمكن من الحصول على نتائج يعتد بها بسبب وجود أخطاء في الوسم الناتج من نظام ستنفورد [٧٤]. ومن ناحية أخرى كانت الفرضيات المستخرجة كنصوص مشكلة تختلف اختلافا طفيفا في التشكيل.

٤. التوصيات والنتائج

ومما لا شك فيه أن التأخر العلمي العربي قد انعكس سلباً على جودة التطبيقات اللغوية العربية التي توظف ثورة تكنولوجيا المعلومات لخدمة الإنسان. فمن المعلوم أن الحاجة الملحة لإنتاج تطبيقات صوتية علمية أو تجارية على حد سواء قد أثرى علم الصوتيات في الدول المتقدمة بشكل كبير، الأمر الذي لم يتحقق في البلاد العربية. فندرة التطبيقات العلمية وقلة التطبيقات التجارية التي تعتمد على تقنية التعرف على الكلام أدت إلى ضعف التطور في هذا المجال. وبذلك فإنه يمكن القول إن البحث العلمي في مجال الصوتيات العربية لا زال محدوداً مقارنة مع اللغات الأخرى. ولعل أسباب ذلك تعود إلى الضعف العام الذي تعاني منه الجامعات ومراكز البحوث العربية بالإضافة إلى قلة التمويل للبحث العلمي بشكل عام. فلا عجب أن نرى الكثير من الأبحاث التي تخص اللغويات العربية قد مولت من قبل مراكز بحوث

أجنبية. إن أنظمة التعرف على الكلام العربي التي طورتها جوجل لم تستخدم على حد علمنا المعرفة اللغوية كمصدر من مصادر المعرفة، فعمل هذه الأنظمة لا يعدو أكثر من مقارنة حروف صوتية تم إدخالها من قبل المستخدم مع ما هو مخزن في النظام من حروف صوتية تم تدريب النظام عليها مسبقاً بطرق تعتمد على التحليل الإحصائي لقاعدة ضخمة من البيانات. وستظل هذه الطرق عاجزة عن تحسين الأداء مالم تدعم بالمعرفة اللغوية. وللأسف فإن معرفتنا اللغوية وقفت عند الأعمال التي قام بها علماء العرب في قرون الإسلام الأولي.

وهناك عدة دراسات وجهود محفزة سيكون لها أبلغ الأثر في دعم توظيف التكنولوجيا لنشر اللغة العربية. ومنها وضع توحيد قياسي للغة العربية الحديثة ويشمل ذلك الأصوات والمفردات والمصطلحات. ونحتاج إلى طريقة موحدة لكتابة التشكيل وانشاء مكانز ضخمة للنصوص المشكلة تكون مفتوحة للباحثين. وتطوير نظم تحليل النصوص والتعرف على اقسام الكلام. وقضية التسجيلات الصوتية وكتابتها وهي مكلفة وتحتاج لجهد كبير متصل نظرا لأهميتها لنظم نسخ الصوت والترجمة الفورية ونظم التعليم وتعلم القرآن (مخارج الحروف، قواعد التجويد، التحفيظ، المراجعة) والتعرف على أوزان وعروض الشعر المنطوق وغيرها من التطبيقات التي تخدم اللغة العربية. وهناك أيضا قضية تعدد اللهجات العربية فهي تحتاج إلى دراسات وصفية وتحليلية وصوتية من قبل المتخصصين، وتسجيلات صوتية حتى تصل تكنولوجية التعامل بالصوت إلى الجيل القادم من الجوالات الذكية والحاسبات الذكية والأجهزة المنزلية الذكية.

شكر وتقدير: يقدم المؤلف وافر الشكر والتقدير إلى مدينة الملك عبد العزيز للعلوم والتقنية وإلى جامعة الملك فهد للبترول والمعادن على دعمها للبحوث وللباحثين وللطلاب لخدمة اللغة العربية.

المراجع

[١] Baker، J.، Deng،L.، Glass J.، Khudanpur S.، Lee، C. and Morgan، N.، Historical Development and Future Directions in Speech Recognition and Understanding، MINDS report، (٢٠٠٧).http://www-nlpir.nist.gov/MINDS/FINAL/speech.web.pdf

[٢] Abdou SM، Hamid SE، Rashwan M، Samir A، Abd-Elhamid O، Shahin M، Naz W (٢٠٠٦) Computer aided pronunciation learning system using speech recognition techniques، NTERSPEECH ٢٠٠٦، ICSLP، pp ٢٥٢–٢٤٩

[٣] Billa، J.، Noamany، M.، Srivastava، A.، Liu، D.، Stone، R.، Xu، J.، Makhoul، J.، & Kubala، F. (٢٠٠٢). Audio indexing of Arabic broadcast news. In Proceedings (ICASSP '٠٢). IEEE international conference on acoustics، speech، and signal processing (Vol. ١، pp. I٥-–I٨-).

[٤] Selouani S-A، Alotaibi YA (٢٠١١) Adaptation of foreign accented speakers in native Arabic ASR systems، Appl Comput Informat ١٠–١:(١)٩

[٥] Baker، J. K. (١٩٧٥). Stochastic modeling for automatic speech understanding. In R. Reddy (Ed.). Speech recognition (pp. ٥٤٢–٥٢١). New York: Academic Press.

[٦] Rabiner ، L. ، Juang ، B.، Fundamentals of speech recognition، Prentice-Hall، Inc.، Upper Saddle River، NJ، ١٩٩٣.

[٧] Jelinek، F. (١٩٩٨). Statistical methods for speech recognition. Cambridge: MIT Press.

[٨] Morgan ، N. and Bourlard ،H.، "Continuous speech recognition"، IEEE signal processing magazine، vol. ١٢، no. ٣، pp. ٤٢-٢٥، May ١٩٩٥.

[٩] Young، S. (١٩٩٦)، "A review of large-vocabulary continuous-speech recognition"، IEEE Signal Processing Magazine، pages ٥٧-٤٥.

[١٠] Alghamdi M.، Elshafei M.، Almuhtasib H.، "Arabic broadcast news transcription system"، International journal of speech and technology، ٢٠٠٩ ،١٩٥–١٨٣ :١٠.

[١١] Nofal M. Abdel Reheem E et al (٢٠٠٤) The development of acoustic models for command and control Arabic speech recognition system. ٢٠٠٤ international conference on electrical, electronic and computer engineering, ٢٠٠٤. ICEEC'٠٤

[١٢] Lee, K. F. ; Hon H. W.; and Reddy R.; "An overview of the SPHINX speech recognition system," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ٣٨,no. ١, pp. ٤٥–٣٥, Jan. ١٩٩٠.

[١٣] Huang, X.; Alleva, F.; Hon, H. W.; Hwang, M. Y.; & Rosenfeld, R. (١٩٩٣). The SPHINX-II speech recognition system: an overview. Computer Speech and Language, ١٤٨–١٣٧ ,(٢)٧.

[١٤] CMU Sphinx Downloads (٢٠١١). http://cmusphinx.sourceforge.net/wiki/download. Accessed ١ September ٢٠١١.

[١٥] Young ,S. J.; Evermann , G.; Gales , M. J. F.; Hain , T.; Kershaw ,D.; Moore , G. L.; Odell ,J. J.; Ollason , D.; Povey D.; Valtchev V. & Woodland P. C. (٢٠٠٤). The HTK Book.

[١٦] HTK (٢٠١١). http://htk.eng.cam.ac.uk/. Accessed January ٢٠١٧ ١.

[١٧] Rabiner, L. R. (١٩٨٩). "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE ٢٨٦-٢٥٧ :(٢)٧٧.

[١٨] Rabiner , L. , Juang , B.; Fundamentals of speech recognition, Prentice-Hall, Inc.; Upper Saddle River, NJ. ١٩٩٣.

[١٩] Huang , X.; Acero ,A.; and Hon, H.; Spoken Language Processing, Prentice Hall PTR, ٢٠٠١.

[٢٠] Elshafei-Ahmed, M. (١٩٩١). Toward an Arabic text-to-speech system. The Arabian Journal of Science and Engineering, ٤)١٦B), ٥٨٣–٥٦٥.

[٢١] Elshafei, M.; Almuhtasib, H.; & Alghamdi, M. (٢٠٠٢). Techniques for high quality text-to-speech. Information Science, ٢٦٧–٢٥٥ ,(٤–٣) ١٤٠.

[٢٢] Alghamdi, M.; Elshafei, M.; & Almuhtasib, H. (٢٠٠٢). Speech units for Arabic text-to-speech. In The fourth workshop on computer and information sciences (pp. ٢١٢–١٩٩).

[٢٣] Kirchhofl, K.; Bilmes, J.; Das, S.; Duta, N.; Egan, M.; Ji, G.; He, F.; Henderson, J.; Liu, D.; Noamany, M.; Schoner, P.; Schwartz, R.; & Vergyri, D. (٢٠٠٣). Novel approaches to Arabic speech recognition: report from the ٢٠٠٢ John-Hopkins summer workshop. In ICASSP ٢٠٠٣ (pp. I٣٤٤--I٣٤٧-).

[٢٤] Kuo HJ, Mangu L et al (٢٠١٠) Morphological and syntactic features for Arabic speech recognition. ٢٠١٠ IEEE international conference on acoustics speech and signal processing (ICASSP)

[٢٥] Ali, M.; Elshafei, M.; Alghamdi, M.; Al-Muhtaseb, H.; & Al-Najjar, A. (٢٠٠٨). Generation of Arabic phonetic dictionaries for speech recognition. In The ٥th international conference on innovations in information technology, United Arab Emirates, December ٢٠٠٨.

[٢٦] Clarkson, P.; & Rosenfeld, R. (١٩٩٧). Statistical language modeling using the CMU-Cambridge toolkit. In Proceedings of the ٥th European conference on speech communication and technology, Rhodes, Greece, Sept. ١٩٩٧.

[٢٧ Hagen, S. (٢٠٠٧). The IBM ٢٠٠٦ GALE Arabic ASR system. In ICASSP, ٢٠٠٧.

[٢٨] Abdulkareem Al-Zahrani and Moustafa Elshafei , "Arabic Poetry Meter Identification And Method", US Patent # ٨,٢١٩,٣٨٦ , July, ٢٠١٢.

[٢٩] Hwang ,M. Y. and Huang ,X." Shared-distribution hidden Markov models for speech recognition". IEEE Transactions on Speech and Audio Processing, ٤٢٠--٤١٤:(٤)١, October ١٩٩٣.

[٣٠] Mokhtar MA، El-Abddin AZ (١٩٩٦) A model for the acoustic phonetic structure of Arabic language using a single ergodic hidden Markov model. In: Proceedings of the fourth international conference on spoken language، ١٩٩٦. ICSLP ٩٦.

[٣١] Alghamdi M (٢٠٠٠) Arabic phonetics. Attaoobah، Riyadh

[٣٢] Bahi، H.، & Sellami، M. (٢٠٠٣). A hybrid approach for Arabic speech recognition. In ACS/IEEE international conference on computer systems and applications، ١٨–١٤ July ٢٠٠٣.

[٣٣] El Choubassi، M. M.، El Khoury، H. E.، Alagha، C. E. J.، Skaf، J. A.، & Al-Alaoui، M. A. (٢٠٠٣). Arabic speech recognition using recurrent neural networks. In Proceedings of the ٣rd IEEE international symposium on signal processing and information technology (ISSPIT) (pp. ٥٤٧–٥٤٣)، Dec. ٢٠٠٣.

[٣٤] Shoaib M، Rasheed F، Akhtar J، Awais M، Masud S، Shamail S (٢٠٠٣) A novel approach to increase the robustness of speaker independent Arabic speech recognition. ٧th international multi topic conference، ٢٠٠٣. INMIC ٩–٨ .٢٠٠٣ Dec ٢٠٠٣، pp ٣٧٦–٣٧١

[٣٥] Algamdi، M.، KACST Arabic Phonetics Database، The Fifteenth International Congress of Phonetics Science، Barcelona، ٢٠٠٣، ٣١١٢-٣١٠٩.

[٣٦] Vergyri D، Kirchhoff K، Duh K، Stolcke A (٢٠٠٤) Morphology-based language modeling for Arabic speech recognition. International conference on speech and language processing. Jeju Island، pp ١٢٥٥–١٢٥٢

[٣٧] Alghamdi M.، Almuhtasib H.، Elshafei M.، Arabic Phonological Rules، King Saud University Journal: Computer Sciences and Information. Vol. ١٦، pp. ٢٠٠٤ ،٢٥-١. networks، ٢٠٠٤

[٣٨] Kirchhoff، K.; Vergyri، D.; ، Cross-dialectal acoustic data sharing for Arabic speech recognition، Acoustics، Speech، and Signal Processing، ٢٠٠٤. Proceedings. (ICASSP '٠٤). IEEE International Conference on ، vol.١، no.، pp. I- ٨-٧٦٥ vol.،١ ٢١-١٧ May ٢٠٠٤ doi: ١٠ ,١١٠٩/ICASSP.٢٠٠٤ , ١٣٢٦٠٩٨

[٣٩] Alotaibi، Y. A. (٢٠٠٤). Spoken Arabic digits recognizer using recurrent neural networks. In Proceedings of the fourth IEEE international symposium on signal processing and information technology (pp. ٢١–١٨٠(١٩٩–١٩٥ Dec. ٢٠٠٤.

[٤٠] Bourouba، H.، R. Djemili، et al. (٢٠٠٦). New Hybrid System (Supervised Classifier/HMM) for Isolated Arabic Speech Recognition. Information and Communication Technologies، ٢٠٠٦. ICTTA '٢ .٠٦nd.

[٤١] Al-Shamsi F. and Guessoum A.، "A Hidden Markov Model –Based PS Tagger for Arabic"،٢٠٠٦، CiteSeerX.

[٤٢] Choueiter G، Povey D et al (٢٠٠٦) Morpheme-based language modeling for Arabic LVCSR. ٢٠٠٦ IEEE international conference on acoustics، speech and signal processing. ICASSP ٢٠٠٦ proceedings

[٤٣] Xiang، B.، Nguyen، K.، Nguyen، L.، Schwartz، R.، and Makhoul، J. ٢٠٠٦. Morphological ecomposition for Arabic Broadcast News Transcription. In Proceedings of ICASSP. Vol. I. Toulouse، ١٠٩٢–١٠٨٩.

[٤٤] Soltau H، Saon G et al (٢٠٠٧) The IBM ٢٠٠٦ Gale Arabic ASR system. IEEE international conference on acoustics، speech and signal processing، ٢٠٠٧. ICASSP ٢٠٠٧

[٤٥] Satori H، Harti M، Chenfour N (٢٠٠٧) Introduction to Arabic speech recognition using CMU Sphinx system. Information and communication technologies international symposium proceeding ICTIS٢٠٠٧ ،٠٧

[٤٦] Emami A، Mangu L (٢٠٠٧) Empirical study of neural network language models for Arabic speech recognition. IEEE workshop on automatic speech recognition and understanding، ٢٠٠٧. ASRU

[٤٧] Taha M، Helmy T et al (٢٠٠٧) Multi-agent based Arabic speech recognition. ٢٠٠٧ IEEE/WIC/ ACM international conferences on web intelligence and intelligent agent technology workshops.

[٤٨] Azmi M، Tolba H،Mahdy S، FashalM(٢٠٠٨) Syllable-based automatic Arabic speech recognition in noisy-telephone channel. In: WSEAS transactions on signal processing proceedings، World Scientific and Engineering Academy and Society (WSEAS)، vol ٤، issue ٤، pp ٢٢٠-٢١١.

[٤٩] Azmi، M. M. and H. Tolba (٢٠٠٨). Syllable-based automatic Arabic speech recognition in different conditions of noise. Signal Processing، ٢٠٠٨. ICSP ٩ .٢٠٠٨th International Conference on

[٥٠] Stallard، D.; Chia-lin Kao; Krstovski، K.; Daben Liu; Natarajan، P.; Prasad، R.; Saleem، S.; Subramanian، K.; ، Recent improvements and performance analysis of ASR and MT in a speech-to-speech translation system، Acoustics، Speech and Signal Processing، ٢٠٠٨. ICASSP ٢٠٠٨. IEEE International Conference on ، vol.، no.، pp.٤٩٧٦-٤٩٧٣، March ٣١ -٢٠٠٨April ٢٠٠٨ ٤

[٥١] Choi F، Tsakalidis S et al (٢٠٠٨) Recent improvements in BBN's English/Iraqi speech-to-speech translation system. IEEE Spoken language technology workshop، ٢٠٠٨. SLT ٢٠٠٨.

[٥٢] El-Hadj Y.، Al-Sughayeir I. and Al-Ansari A.، Arabic Part-Of-Speech Tagging using the Sentence Structure، Proceedings of the Second International Conference on Arabic Language Resources and Tools، The MEDAR Consortium،٢٠٠٩.

[٥٣] Farghaly A، Shaalan K (٢٠٠٩) Arabic natural language processing: challenges and solutions. ACM Trans Asian Lang Inform Process ٢٢-١:(٤)٨.

[٥٤] Lamel L، Messaoudi A et al (٢٠٠٩) Automatic speech-to-text transcription in Arabic. ACM Trans Asian Lang Inform Process ٢ ١٨٢٢-١:(٤)٨ Arabic Speech Recognition Systems

[٥٥] Kuo HJ، Mangu L et al (٢٠١٠) Morphological and syntactic features for Arabic speech recognition. ٢٠١٠ IEEE international conference on acoustics speech and signal processing (ICASSP)

[٥٦] Selouani S-A، Alotaibi YA (٢٠١١) Adaptation of foreign accented speakers in native Arabic ASR systems. Appl Comput Informat ١٠-١:(١)٩.

[٥٧] Luo، X. (٢٠١١). Chinese Speech Recognition Based on a Hybrid SVM and HMM Architecture Advances in Neural Networks – ISNN ٢٠١١. D. Liu، H. Zhang، M. Polycarpou، C. Alippi and H. He، Springer Berlin / Heidelberg. ٦٣٥-٦٢٩ :٦٦٧٧.

[٥٨] Geoffrey Hinton، Li Deng، Dong Yu، George Dahl، Abdel-rahman Mohamed، Navdeep Jaitly، Andrew Senior، Vincent Vanhoucke، Patrick Nguyen، Tara Sainath، and Brian Kingsbury، "Deep Neural Networks for Acoustic Modeling in Speech Recognition"، IEEE Signal Processing Magazine (Volume: ٢٩، Issue: ٦، Nov. ٢٠١٢).

[٥٩] Irina Illina، Dominique Fohr، Odile Mella، Christophe Cerisara، "The Automatic News Transcription System : ANTS some Real Time experiments"، ٨th International Conference on Spoken Language Processing – ICSLP'٢٠٠٤ ،٢٠٠٤، Jeju، Corée du Sud، ٤ p، ٢٠٠٤.

[٦٠] S. Chen، M.J.F. Gales، P.S.Gopalakrishnan، R.A.Gopinath، H. Printz، D. Kanevsky، P. Olsen، and L. Polymenakos، "IBM's Lvcsr System For Transcription Of Broadcast News Used In The ١٩٩٧ Hub٤ English Evaluation"، In: DARPA Broadcast News Transcription and Understanding Workshop، ٨-٢-١٩٩٨ to ١١-٢-١٩٩٨، Lansdowne، VA، US.

[٦١] Mohammad A. M. Abushariah، Raja Noor Ainon، Roziati Zainuddin، Moustafa Elshafei، Othman Omran Khalifa:

Phonetically rich and balanced text and speech corpora for Arabic language. Language Resources and Evaluation :(٤)٤٦ ٢٠١٢) ٦٣٤-٦٠١)

[٦٢] Mohammad Abd-Alrahman Mahmoud Abushariah، Raja Noor Ainon، Roziati Zainuddin، Moustafa Elshafei، Othman Omran Khalifa: Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus. Int. Arab J. Inf. Technol. ٢٠١٢) ٩٣-٨٤ :(١)٩)

[٦٣] Elshafei، M.، Al-Muhtaseb، H.، & Alghamdi، M. (٢٠٠٦a). Statistical methods for automatic diacritization of Arabic text. In Proceedings ١٨th national computer conference NCC'١٨، Riyadh، March ٢٠٠٦، ٢٩-٢٦.

[٦٤] Elshafei، M.، Al-Muhtaseb، H.، & Alghamdi، M. (٢٠٠٦b). Machine generation of Arabic diacritical marks. In Proceedings of the ٢٠٠٦ international conference on machine learning; models، technologies، and applications (MLMTA'٠٦)، June ٢٠٠٦، USA.

[٦٥] Mansour Alghamdi، Muhammad Khursheed، Mustafa Elshafei، Fayz Alhargan، Muhammed Alkanhal، Abu Aus Alshamsan، Saad Alqahtani، Syed Zeeshan Muza ffar، Yasser Altowim، Adnan Yusuf، and Husni Almuhtasib، "Automatic Arabic Text Diacritizer"، KACST Project final report، July، ٢٠٠٦.

[٦٦] Khalid M. O. Nahar، Mohammed Abu Shquier، Wasfi G. Al-Khatib، Husni Al-Muhtaseb، Moustafa Elshafei، "Arabic phonemes recognition using hybrid LVQ/HMM model for continuous speech recognition"، International Journal of Speech Technology، Volume ١٩، Issue ٣، pp ٥٠٨-٤٩٥، September ٢٠١٦.

[٦٧] Khalid Nahar، Husni Al-Muhtaseb، Wasfi Al-Khatib، Moustafa Elshafei and Mansour Alghamdi، "Arabic Phonemes Transcription Using Data Driven Approach"، The International Arab Journal of Information Technology (IAJIT)، International Arab Journal of Information Technology ٣)١٢) • May ٢٠١٥.

[٦٨] Dia AbuZeina and Moustafa Elshafei ، Cross-Word Modeling for Arabic Speech Recognition، (Springer Briefs in Electrical and Computer Engineering / Springer Briefs in Speech Technology، ٢٠١١.

[٦٩] AbuZeina D.، Al-Khatib W.، Elshafei M.، Al-Muhtaseb H.، "Cross-word Arabic pronunciation variation modeling for speech recognition" ، International Journal of Speech Technology، Volume ١٤ Issue ٣، September ٢٠١١.

[٧٠] AbuZeina D.، Al-Khatib W.، Elshafei M.، "Modeling of Cross-word Pronunciation Variation for Arabic ASRs: A Knowledge-Based Approach"، Journal of Communications and Computer Engineering، Vol ٢، No ٢٠١٢) ١)،pp ٣٠-٢٤.

[٧١] Dia AbuZeina، Wasfi Al-Khatib، Moustafa Elshafei، Husni Al-Muhtaseb، "Toward enhanced Arabic speech recognition using part of speech tagging"، International Journal of Speech Technology ،Volume ١٤، Issue ٤، pp ٤٢٦-٤١٩، December ٢٠١١،

[٧٢] W. Xiong، J. Droppo، X. Huang، F. Seide، M. Seltzer، A. Stolcke، D. Yu، G. Zweig، "The Microsoft ٢٠١٦ Conversational Speech Recognition System"، ICASSP ٢٠١٧، https://arxiv.org/abs/١٦٠٩,٠٣٥٢٨; last visited January ٢٠١٧،٣٠.

[٧٣] Weka، Machine Learning Group at University of Waikato، ٢٠١١.http://www.cs.waikato.ac.nz/ml/weka/

[٧٤] Stanford Log-linear Part-Of-Speech Tagger، ٢٠١١، http://nlp.stanford.edu/software/tagger.shtml