

توليف النص العربي الى كلام منطوق عالي الجودة

سندس محمد الشريف
جامعة طرابلس، كلية العلوم، قسم علوم الحاسوب
طرابلس، ليبيا
sharif.sondos@gmail.com

إبراهيم علي المرهق
جامعة طرابلس، كلية تقنية المعلومات، قسم الشبكات
طرابلس، ليبيا
i.almerhag@uot.edu.ly

ان عملية بناء النظام النطقي مقيدة بالمدخلات (الحروف او النصوص) التي انطلق منها البرنامج، ولقد اعتمدنا هنا اسلوب تسجيل وتخزين الاحرف العربية بصوت بشري في ملفات صوتية. بحيث يكون لكل حرف مرسوم مقابله حرف منطوق (صوتياً) مع عدم إهمال تشكيل الحرف أي (اَ اِ اُ اِوْ اِوْ اِوْ) ولكامل الحروف العربية.

ان نظرة سريعة الى الحركات في اللغة العربية توجي بوجود 3 حركات هي الفتحة والضمة والكسرة. إلا أن الحقيقة هي أنه هناك عدد أكبر من ذلك نتيجة تعاقب الحركات مع حروف العلة (الألف والواو والياء) سواء كانت قبلها او بعدها وكذلك وجود الشدة على حرف العلة [2].

فالکسرة قبل الواو المفتوحة (مثل سوى) هي غير الكسرة قبل الياء المفتوحة (مثل علياً)، وتجدر الإشارة كذلك إلى أن بعض الحروف العربية لها أكثر من صوتيم/فونيم (Phoneme) كاللام المرققة والمفخمة والراء المرققة والمفخمة كما أن الياء والواو هما حرفان صامتان وهما حركات في الوقت نفسه. أما الهمزة فهي حرف صامت بينما الألف فهو حركة فقط. ان المهمة الاولى التي تواجه نظم توليف النصوص Text To Speech (TTS) هي تحويل المدخلات النصية الى تمثيل لغوي، وعادة ما يسمى هذا بالتحويل الصوتي Text To phonetic (TTP) أو Grapheme To (GTP) Phoneme وهو ما يشكل تعقيداً بالنسبة للغة العربية والانجليزية واللغات الأخرى ذلك لوجود عدة قواعد مختلفة لإنتاج النطق ومن ثم توليف الكلام [2].

ويمكن تقسيم عملية التحويل الى ثلاثة أطوار رئيسية وهي: (1) معالجة النص، (2) تصحيح النطق و (3) تحليل الصوتيات.

تعتبر معالجة النص المكتوب عملية معقدة، بما في ذلك معالجة الارقام والفواصل فمثلاً عند احتواء النص على "567" سنتطق "خمسائة وسبعة وستون"، الكسور والتواريخ ايضا مشكلة أخرى مثلا الرقم 3/2 هنا يحتمل ان يفسر على انه الكسر "ثلثان" والاحتمال الاخر انه تاريخ "الثاني من مارس". وتتسع المشكلة بشكل أكبر لتشمل كلمات كاملة مثل "كجم" وهي اختصار كيلو جرام، وايضا "د." وهي اختصار دكتور، وربما كلمتان مثل "الخ" وتعني الى اخره. الا انه في بعض الاحوال تحل المشكلة بالاستعانة بالرموز الى جانب الكلمة مثل "100\$" هنا تنطق مائة دولار استنادا الى الرمز بجانبها.

ان استخدام القواعد اللغوية او بعضها على الاقل في كثير من الاحيان يكون ضروريا لتحقيق النطق الصحيح، مثلا الصوتيم "س" في نطق كلمة "صراط" وكذلك الصوتيم "ص" في كلمة "مُسْتَقِيم"، هذه قد تشكل أحد المشاكل الأخرى عند بناء تطبيق توليف النصوص [3].

وعلى الرغم من ثراء التراث العربي بكتب قواعد نطق اللغة العربية بغرض تجويد القرآن الكريم، إلا انه لا توجد دراسات حديثة تقوم بتعريف وتحديد الصوتيات (فونولوجي) للغة العربية الفصحى الحديثة الدارجة الآن، سواء لتحديد مخارج الحروف أو لتقنين قواعد نطقها لهذا اعتمدنا هنا على قواعد اللغة العربية الفصحى الكلاسيكية.

ان من أهم الوحدات في مراحل تحويل النص إلى كلام هي وحدة التحويل من الحرف إلى صوت حيث تعتمد على التحديد التلقائي للترميز

المخلص— في هذه الورقة نستعرض نموذج توليف الخطاب العربي من نص مكتوب ضمن إطار يرتكز على نتائج صوتية عالية الجودة من ناحية طبيعتها، حيث تم البحث في نظام النطق عند الإنسان وقواعد النطق في اللغة العربية ومن ثم تطوير وتطبيق خوارزمية تعتمد على المقاربة للخوارزمية النطقية عند البشر أي محاكاة النطق البشري، مما سمح بإنتاج مختلف الأصوات اعتماداً على مكونات صوت ذات جودة عالية لأنها اعتمدت على مدخلات صوت بشري طبيعي، كذلك تم تطبيق بعض القواعد النطقية للغة العربية.

اختبر النظام على نطق المقاطع الصوتية المنفردة والمقاطع الصوتية المركبة بمعنى آخر اختبر على نطق الكلمات ذات المقطع الواحد والكلمات متعددة المقاطع الصوتية حيث سجلت نتائج جيدة في نطق الكلمات ذات المقطع الصوتي الواحد ونتائج تحتاج الي تحسين في الكلمات متعددة المقاطع الصوتية.

الكلمات المفتاحية— نظم توليف النصوص، معالجة اللغة العربية، تحويل النصوص الى كلام منطوق، تمثيل الصوت العربي آليا

1. المقدمة

بالرجوع الي القدرة البشرية على النطق ولو افترضنا تدريب شخص لديه مشاكل بالنطق فإن التدريبات تبدأ بنطق كل حرف على حده ومن ثم الحرف بالتشكيل وهكذا حتى ينطق كل الحروف بالحركة ومن ثم يتم الانتقال الي الكلمات البسيطة وبعدها الي الكلمات الأكثر تعقيداً وهكذا حتى يصل الإنسان الي النطق السليم مع تحليل تطورات القدرة في كل مرحلة وتحديد معيقات النطق السليم مالم يكن هناك اعاقه حيوية لإنتاج الكلام.

ان الهدف من هذه الدراسة هو تكوين مقارنة بين آلية النطق البشري وامكانية تمثيلها بخوارزميات بحيث يتم انتاج خطاب صوتي أقرب ما يكون للصوت البشري ووضع التحديات الحاسوبية في مشكلة تمثيل الصوت العربي آليا وتداعيات تطبيق القواعد النطقية للغة العربية، والتي تمثلت في الفية ابن مالك، وهي ما يقارب الألف (1000) قاعدة نطقية.

وبالرغم من الكم الكبير للقواعد النطقية الا ان ما يميز اللغة العربية هو عدم وجود شواذ، حيث أن كل ما هو منطوق في العربية له قاعدة وقانون وبالتالي هناك إمكانية لتنفيذه برمجياً.

ان نظام تحويل النص الى كلام منطوق هو نظام مستقل وقائم على قواعد نطقية، ونظرا لأهمية اللغة العربية والاحتياجات المختلفة لدعمها حاسوبياً عكفت دراساتنا الحالية على البحث في تطوير نظام النطق العربي. أي تطوير وسيلة مبسطة للفظ الكلمات العربية في حالة الحرف المجرد أو الحرف المشكل أو الحرف ضمن مقطع صوتي واحد او متعدد المقاطع. بحيث يمكن المستخدم من سماع النطق الصحيح للحرف في اشكاله المختلفة وسماع نطق الكلمات الناتجة من كتابة نص بحروف مشكّلة، ويكون مزودا بواجهة واضحة وسهلة الاستخدام حتى بالنسبة للمبتدئين في استخدام الحاسوب.

لقد عملنا على تحليل تركيب الكلام بالعربية وفهم خوارزمية النطق العربية والطبيعية عند الإنسان وبالتالي قمنا بتصميم نظام تحويل النص الي كلام ومن ثم تنفيذه برمجياً من خلال وضع الية لتوليف خطاب عربي بصوت عالي الجودة، حتى نتمكن من تحديد المشاكل والتحديات الجديدة التي تواجه انظمة تحويل النص العربي الي كلام [1].

يعتبر العثور على الطول الصحيح للوحدة الصوتية التي ستكون حلقة في تسلسل الكلام المولد أحد النقاط المهمة في توليد الكلام بالتسلسل. وغالبا ما يقع الخيار على طول وسط بين الطويلة والقصيرة فالوحدة الطويلة تكون أقرب للطبيعية وتقلل من نقاط حلقات التسلسل وتعطي تحكماً أكثر في النطق الانتقالية إلا أن عدد الوحدات يكون أكثر ومن ثم تحتاج إلى ذاكرة أكبر وبالمقابل فإن الوحدات القصيرة تحتاج إلى ذاكرة أقل، ولكن جمع عيناتها وتسميتها تكون أصعب وأعد، أي أننا يجب أن نراعي عدة عوامل منها سعة الذاكرة وجودة الصوت ونوفق بينهما. وفي وحدات النظم الحالية عادة ما تستخدم الكلمات أو المقاطع أو أشباه المقاطع أو الصوتيمات المفردة أو الثنائيات (الدايفونات) وأحيانا الصوتيمات الثلاثية.

ويعتبر استخدام كامل الكلمة كوحدة صوتية أقرب الطرق للنطق الطبيعي ذلك لكونها متكاملة، فهي تحتوي على النطق الانتقالية داخلها لأنها لا تضم وحدات أصغر منها تؤثر على جودة الصوت، ومما يميز هذا النوع من الوحدات أنه من السهل سلسلة كلماتها. إلا أن هناك فرق كبير بين الكلمات التي تنطق منفردة وتلك المكونة لجملة مما يجعل العبارات الناتجة عنها بعيدة إلى حد ما عن الكلام الطبيعي.

ولأن تعداد الكلمات وأسماء الأعلام في أية لغة يتجاوز مئات الآلاف فإن الكلمات لا تصلح أن تكون وحدات لنظام ناطق آلي لنصوص مفتوحة الكلمات. وعلى الرغم من أن عدد المقاطع المختلفة في أية لغة أقل بكثير من عدد كلماتها، فعلى سبيل المثال هناك ما يقرب من 100 ألف مقطع في الإنجليزية فإن عدد الوحدات هذا يظل كبيراً بالنسبة لنظام ناطق آلي. وعكس ما هو موجود في الكلمات فإن تأثير النطق الانتقالية لا يكون موجوداً بين المقاطع المخزنة عند تسلسلها لتوليد الكلام، لذا فإن استخدام المقاطع كوحدات في نظم النطق الآلي ليست مناسبة، كما أنه ليس من الممكن التحكم في تكوين العبارة. ففي الوقت الحاضر لا يوجد نظام ناطق آلي كامل قائم على كلمات أو مقاطع. بل إن النظم الحالية تقوم أساساً إما على الصوتيمات أو الثنائيات أو أشباه المقاطع أو أي شكل من أشكال الجمع بينها [3].

1. أشباه المقاطع: يمثل شبه المقطع بداية ونهاية المقطع. وتستخدم أشباه المقاطع بدلا من الصوتيمات لأن أحد مميزاتها أن ألفا منها فقط يمكن أن تبني عشرة آلاف من مقاطع الإنجليزية. إلا أن النظام الذي يعتمد كلياً على أشباه المقاطع لا يمكن أن يولد بشكل طبيعي جميع الكلمات الممكنة في اللغة. بينما النظام الذي يستخدم المقاطع وأشباه المقاطع يمكن أن يكون ناجحاً إذا استعمل وحدات وزوائد كلمات متعددة الطول.

2. الصوتيمات: هي أكثر الوحدات شيوعاً في الاستخدام في نظم النطق الآلي نظراً لكونها تمثل الوحدات الطبيعية اللغوية التي تستخدم أثناء حديث الإنسان. ويكون مجموعها في الغالب بين أربعين وخمسين وحدة. وتعطي الصوتيمات أعلى مرونة ممكنة للنظم القائمة على القوانين.

3. الثنائيات: وتعرف بأنها المنطقة الممتدة من منتصف الحالة المستقرة لصوت ما إلى الحالة المستقرة للصوت الذي يليه. لذا فإنها تحمل معها الحالة الانتقالية بين الصوتين. مما يعطيها ميزة أخرى كونها لا تحتاج لوضع قوانين للنطق الانتقالية. نضيف هنا أن عدد الثنائيات يساوي مربع عدد الصوتيمات مضافاً إليه عدد المقاطع (الألفونات). ففي نظم النطق الآلية يندر استخدام الوحدات الطويلة، كالألفونات الثلاثية والرابعة. فالألفونات الثلاثية تشبه الألفونات الثنائية إلا أنها تحتوي على صوتيم واحد في الوسط.

ففي نظم التوليد بالتسلسل يتم إعداد قائمة بالوحدات الصوتية على ثلاثة مراحل هي:

- تسجيل الكلام الطبيعي بحيث يحتوي على جميع الوحدات التي ستستخدم صوتيمات في جميع صيغ الفونات.
- تسمية الوحدات واستخلاصها من المعطيات.
- يتم اختيار أكثر الوحدات ملائمة.

ومن الجدير بالذكر أن نظم توليد الكلام بالتسلسل تعاني من عدة مشاكل عند مقارنتها بغيرها من الطرق، من أهمها:

- تشويه انسياب الصوت في نقاط التسلسل، وهذه يمكن تداركها باستعمال الثنائيات أو استخدام طرق معالجة الإشارة الصوتية لتحسين سلاستها.

الصوتي للنص المدخل، أي أن وظيفتها تعتمد أساساً على المعجم والقواعد بالرغم من وجود عمليات أخرى وسيطة.

2. الدراسات السابقة

نظراً للتعقيدات والصعوبات التي واجهها الباحثون للخروج بنطاق آلي ذي جودة عالية فإنهم نهجوا طرقاً عديدة ومتباينة لتحقيق هذا الهدف، ويمكن تصنيف هذه الطرق في المجمل إلى صنفين: تخليق الكلام وتسلسل الكلام [3]. وكذلك تحتوي عملية توليف الكلام على خطوتين: تحليل النص وتركيب الكلام [4].

أ. الصنف الأول: تخليق الكلام

في هذا الصنف (تخليق الكلام أو التوليد البارامتري) يتم توليد الكلام بشكل آلي دون اللجوء إلى أصوات بشرية، ومن هذا الصنف النماذج التالية:

1. طرق التنبؤ الخطي: وهي عبارة عن نموذج لمرشحات تثار بضجيج صوتي ومصدر متسلسل لنبضات منتظمة.
2. تشبيه مخارج الأصوات: وهي محاولة لعمل نموذج للجهاز الصوتي أو محاكاة الجهاز الصوتي عند الإنسان ومن ثم إخراج الأصوات اللغوية بطريقة مشابهة لما يقوم به الجهاز الصوتي الطبيعي.
3. توليد النطق الرنينية: وهي عبارة عن نموذج لتسلسل ترددات الإشارة الصوتية للكلام أو الاعتماد على نموذج مرشحات المصدر لنقل وظائف الجهاز الصوتي.
4. أنموذج المنحنى الجيبي: وهو تمثيل الإشارات الصوتية بموجات جيبيية (الدالة جا وجتا).
5. أنموذج التوافقية: وهو إمكانية تمثيل كل جزء من الموجة الصوتية بتوافقية النغمة (harmonic) والضجيج (noise).

ب. الصنف الثاني: تسلسل الكلام

تسلسل الكلام ألياً بالاعتماد على أصوات طبيعية سبق تخزينها وإضافتها إلى النظام ويستخدم هذا النوع وحدات مختلفة الطول من الأصوات اللغوية التي سبق تسجيلها لأحد المتكلمين.

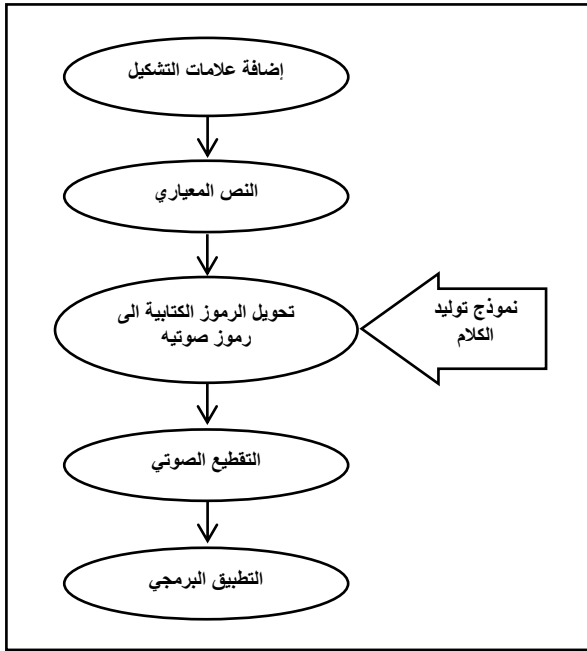
لقد هيمن أسلوب توليد النطق الرنينية على نظم النطق الآلي لفترة طويلة. إلا أن أكثر النظم شيوعاً في الوقت الحاضر هو النوع الثاني من الصنف الأول (تشبيه مخارج الأصوات) بالإضافة إلى الصنف الثاني (تسلسل الكلام). أما الآن فقد استحوذ أسلوب تسلسل الكلام على اهتمام أكثر الباحثين من أي نظام سواه. ومن النظم الواعدة في المستقبل نظام تشبيه مخارج الأصوات إلا أن الخروج بنظام ناطق آلي ذي جودة عالية يطبق هذا الأسلوب لا يزال معقداً. وفيما يلي سنتحدث بشيء من التفصيل عن طريقة توليد الكلام بالتسلسل [5].

ج. توليد الكلام بالتسلسل

قد تكون طريقة توصيل الكلام المسبق التسجيل مع بعضه من أسهل الطرق لإنتاج كلام آلي مفهوم وقريب من الطبيعي. إلا أن التوليد بالتسلسل عادة ما يكون محدوداً بمتحدث واحد وصوت واحد لكل مفردة أو مقطع وبالتالي فهو يحتاج عادة إلى ذاكرة أكبر من تلك التي تحتاجها النظم الأخرى ويقارن (جدول 1) بين متطلبات التوليد بالتسلسل والتوليد البارامتري [3].

جدول 1. مقارنة بين توليد الكلام بالتسلسل والتوليد البارامتري.

توليد بالتسلسل	توليد بارامتري	الوحدات الأساسية
أجزاء صوت بشري	خوارزمية	الجودة
طبيعي أكثر	اصطناعي أكثر	علم العروض
منخفض	مرتفع	متطلبات الذاكرة
مرتفع	منخفض	الخوارزمية
لصق الصوتيمات	المسار الصوتي	الجهد اللازم لإنتاج صوت جديد
مرتفع	منخفض	



شكل 1. مراحل بناء النظام

ج. تحويل الرموز الكتابية إلى رموز صوتية

ويجري فيها تحويل النص من شكله الإملائي أو الرموز الكتابية Grapheme إلى مكافئه الصوتي أو المنطوق Phoneme وذلك باستخدام خوارزمية متعلقة باللغة، حيث تم تخزين كل حرف مرسوم ومقابلته الصوتي في قاعدة البيانات التي تم بنائها لهذا الغرض [7].

د. عملية التقطيع الصوتي

يجري تمثيل الرموز الصوتية الناتجة في المرحلة السابقة بشكل سلسلة من الصوامت والصوائت والتي يطبق عليها مجموعة من القواعد المطردة لتحديد البنية المقطعية للجملة.

هـ. تنفيذ برنامج النطق بالكلام

في هذه المرحلة يتم تطوير برنامج حاسوبي ينطق بالكلمات التي جرى إدخالها، وتمت معالجتها في المراحل سالفة الذكر. تتميز في هذه المرحلة منهجين رئيسيين متبعين هما: المركب بالضم أو السلسلة. بحيث يتم الاعتماد على سلسلة (stringing together) من خطاب مسجل، وينتج خطاب ذو صدق طبيعي عند عملية التوليف. ومع ذلك فإن الاختلافات بين التعبيرات الطبيعية في الكلام وطبيعة التقنيات الآلية لتقسيم الطول الموجي يؤدي في بعض الأحيان إلى مواطن خلل مسموعة في الإخراج.

وبالاعتماد على ما سبق تم تصميم وبناء قاعدة بيانات خزنت بها اغلب اشكال الحروف العربية المكتوبة وما يقابلها بالحرف الصوتي كما هو منطوق. وقد تم استخدام ملفات من نوع MP3 لحفظ الملفات الصوتية من نطق بشري سليم.

وفي تنفيذ البرنامج استخدمنا لغة البايثون لما لها من مميزات تساهم في تصغير الملفات التنفيذية كما انها تعمل على عدة منصات (الويندوز واللينكس والماكنتوش والبونكس) بدون تغيير الكود مما يؤدي الي تصغير الملف التنفيذي لنظامنا هذا وايضا يزيد مرونته في العمل على المنصات المذكورة سابقا.

4. النتائج والمناقشة

ركزنا في اثناء بناء التطبيق على الخطوات المنطقية لنقل المعرفة المتمثل في قدرات النطق وتحويلها الي نظم حاسوبية (شكل 2) بحيث تم بناء نظام حاسوبي يهدف الي تمثيل القدرات البشرية في النطق وحلول مشاكل القراءة لنصوص مختلفة. الا ان هذا النظام لايزال يحتاج الي جهود علماء اللغة لتوضيح كامل عمليات النطق والقراءة عند الناطقين باللغة العربية.

- الذاكرة المطلوبة غالبا ما تكون عالية، خاصة إذا ما كانت الوحدات طويلة عند استخدام المقاطع أو الكلمات [6].
- تستغرق عملية تسمية الوحدات وجمعها وقتا طويلا ونظريا، فان الوحدات لا بد أن تشمل على جميع الألفونات الممكنة، ولكنه يمكن الموازنة الي حد ما بين عدد الوحدات وجودة الكلام.

3. وصف النظام المقترح

يُعمد في نظم تحليل الكلام وتركيبه لإنتاج كلام غير محدد، أو غير مسجل من قبل، إلى اختيار مجموعة وحدات صوتية تمكن من تركيب أي جملة كلامية، وبذلك يجب فقط تخزين هذه الوحدات أو موسطات مناسبة لإنتاجها. لا يمكن أن تكون هذه الوحدات كلمات، لأن عدد كلمات اللغة هائل ويمكن أن يزداد مع مرور الزمن بإضافة كلمات جديدة لم تكن معروفة من قبل. إن أصغر الوحدات الصوتية هي المقاطع الصوتية أو الصوتيمات، وهي محدودة العدد في كل لغة، ويتراوح عددها ما بين 35 و 60 مقطعا في اللغات الأوروبية والسامية.

لذا يمكن بسهولة تخزين هذه الصوتيمات وتركيب أي كلمة منها. ولكن استعمال هذه الوحدات وحدها لا يحل مشكلة التفاضل بين صوتيين، فالتفصل بين صوتيين أهم بكثير من الصوتيمات نفسها، على سبيل المثال، إذا اقتطعنا الصوتيم /ب/ من كلمة والصوتيم /أ/ من كلمة أخرى وجمعناهما معاً فلن نسمع بالضرورة المقطع /با/. وهذا ما قاد إلى اعتماد الثنائيات الصوتية كأصغر وحدات صوتية، إذ تتضمن كل ثنائية النقلة بين الجزء المستقر من الصوتيم الأول والجزء المستقر بين الصوتيمين الأول والثاني. عدد هذه الثنائيات للغات السابقة يتراوح بين 1500 و 2000 ثنائية ويمكن تشكيل أي كلمة منها وهناك مركبات كلامية كثيرة تعتمد هذا المبدأ. كما توجد أيضاً مركبات تعتمد المقاطع (syllables) أو أنصاف المقاطع كوحدة صوتية وهي تعطي نتائج أجود ولكن على حساب ازدياد عدد هذه الوحدات، الذي يفوق 30000.

ان تركيب الكلام من نص يتطلب تحليل هذا النص قبل تركيبه، وتتضمن هذه المرحلة تحويل الكتابة إلى رموز قراءة صوتية، وإضافة معلومات تطريزية (prosody) بالاستعانة بأدوات التقييم والتحليل الصرفي والنحوي والدلالي للنص. كطريقة النطق، مدة النطق، شدة الكلام، النبرة. وتتعلق هذه المعلومات بكون الجمل إخبارية أو استفهامية أو تعجبية أو أمرية. ونظراً لتعدد طرق لفظ الكلمات غير المشكلة في اللغة العربية، يتطلب الأمر إضافة مرحلة لتشكيل النص لإزالة اللبس، وهذه المرحلة تستفيد من عمليات التحليل الصرفي والنحوي والدلالي أيضاً.

وبالاعتماد على التحليل السابق فان بناء النظام موضوع هذا البحث مر بعدة خطوات كما هو موضح في (شكل 1) وتتمثل في المراحل التالية:

أ. إضافة علامات التشكيل

وتتضمن إضافة علامات التشكيل لضمان النطق السليم للنص ولإزالة أي لبس قد يحدث اثناء توليف النص.

ب. انتاج النص المعياري

وهي عملية تعرف بتطبيع النص Text Normalization أو Tokenization أي إعادة كتابة النص ليبدو بالشكل القانوني/النظامي وفيها يجري ما يلي:

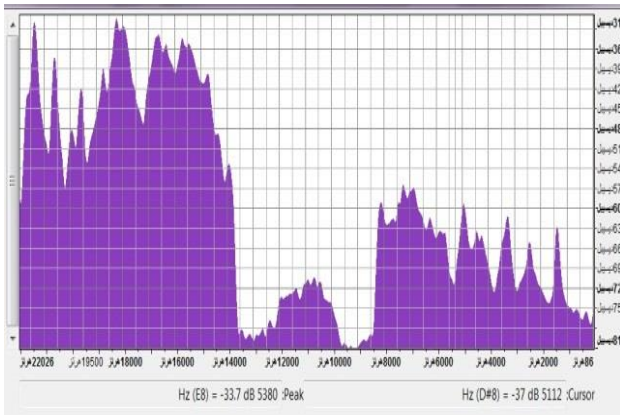
- تأخذ هذه المرحلة النص كما هو مكتوب بشكله الأولي (مع التشكيل) وتعيد كتابته بعد تحويل كل الرموز والاختصارات والأرقام الموجودة فيه إلى كلمات.
- يجري رد حروف الكتابة إلى أصلها ويقصد بذلك رد أشكال الحروف كما ترد داخل الكلمات إلى صورتها الأساسية (على سبيل المثال يجري رد كل أشكال حرف العين: ع ع ع إلى ع).
- فك الحروف العربية ذات الطابع الثنائي إلى عناصرها المفردة فمثلاً: تُرد الألف الممدودة إلى حرف الهمزة يتبعها فتحة طويلة.
- رد أشكال كتابة الهمزة (أ، و، ئ) إلى ما يناظرها من صامت وصائت.

ولغرض اختبار الخوارزمية والتطبيق الذي تم تطويره تم إجراء ثلاثة اختبارات مختلفة هذه الاختبارات تقيس جودة التوليف للنصوص الناتجة على ثلاثة مستويات وهي قياس نطق: الحرف المنفرد، الكلمة ذات المقطع الصوتي الواحد والكلمة ذات المقاطع الصوتية المتعددة. وفيما يلي النتائج المتحصل عليها من تلك الاختبارات.

أ. الاختبار الأول

وهو اختبار نطق الحرف المستقل أو الحرف المنفرد مثل "س" وكذلك الحرف "ن" وأشكاله المتعددة المنفردة. ويوضح (شكل 5) الرسم البياني لطيف نطق الحرف س. في هذا الاختبار تم تسجيل نماذج من الخطاب الناتج من التطبيق ثم عرضت على سمع عدد من الأشخاص لديهم القدرة على السمع الطبيعي فتجاوزت نسبة مدى مفهومية الحرف ووضوحه %84.

هذه النتيجة تدعم ما توصلت اليه دراسة سابقة [8] حيث تؤكد ان النماذج الصوتية لا ترتبط مباشرة بالشكل الموجي المادي بل ترتبط بالتمثيل الرمزي وزمن النطق.



شكل 5. الرسم البياني لطيف نطق الحرف س.

ب. الاختبار الثاني

وهو اختبار الكلمة ذات المقطع الصوتي الواحد Monosyllabic word واخترنا كلمة "نور" لهذا الاختبار وهي صوتيا من النوع ساكن - متحرك - ساكن consonant-vowel-consonant (CVC). ويبين (شكل 6) طيف نطق الكلمة.

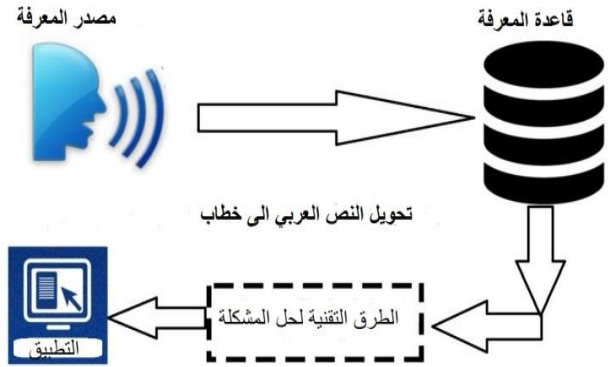
في اللغة العربية هناك قاعدة نطقية وهي متعلقة بالقواعد النطقية للحرف المدود وهي خاصة بالضممة التي يليها حرف الواو والتي تنص على عدم نطق الحرف التالي لحرف النون المضمومة وهو في هذه الحالة حرف الواو.

لاحظنا ان عدم تطبيق هذه القاعدة يؤدي الى نطق الكلمة بشكل غير صحيح وبالتالي اختلاف المعنى، اما في حال تطبيق القاعدة النطقية او كتابة الكلمة عند الإدخال بشكل اخر مثل "نور" فهذا نتج النطق الصحيح لكلمة "نور". ومن هنا تبين الزامية تطبيق كل القواعد النطقية التي تتضمنها اللغة العربية لضمان النطق السليم. وعند عرضها على المستخدمين لاختبار مدى الوضوح والفهم من قبلهم كانت ردود الفعل تتلخص بوضوح الكلمة الا ان سرعة النطق ابطأ من النطق الطبيعي لدي الانسان خصوصا عند الانتقال بين حروف الكلمة.

ج. الاختبار الثالث

اختبار الكلمة متعددة المقاطع الصوتية Multisyllabic word وتم تجربة اكثر من كلمة: مَحْمَدٌ \ أَحْمَدٌ وكانت النتائج الصوتية غير واضحة المعالم بالنسبة للكلمة ككل حيث تتكون هذه الكلمات من مقاطع متعددة وبالتالي تحتاج الى تطوير عملية الانتقال بين المقاطع الصوتية وتنفيذها بسلاسة لتبدو طبيعية أكثر وهذا ما يمثل أحد اهم التحديات الجديدة.

وعليه أثبت نجاح هذا الاسلوب في الحالتين الاولى والثانية بشرط تنفيذ كل القواعد النطقية للغة العربية.



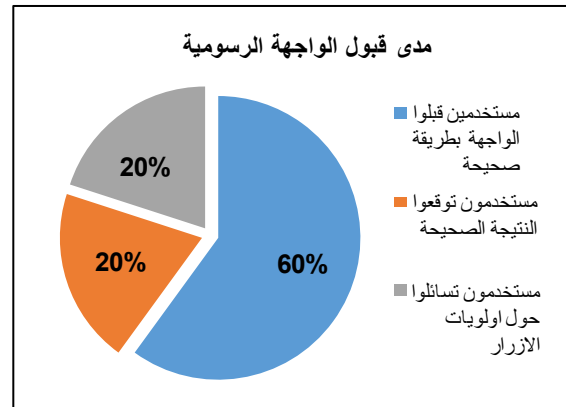
شكل 2. تمثيل عملية نقل المعرفة البشرية الي نظم حاسوبية.

طُور التطبيق باستخدام لغة البايثون وقاعدة البيانات اكسس كما استخدمنا واجهة مستخدم رسومية Graphical User Interface وذلك لتعزيز مفاهيم التفاعل بين الحاسوب والبشر. لقد راعينا البساطة في تصميم الواجهة (شكل 3) المستخدمة حاليا وذلك لجعل التعامل مع التطبيق من قبل المبتدئين أسهل.



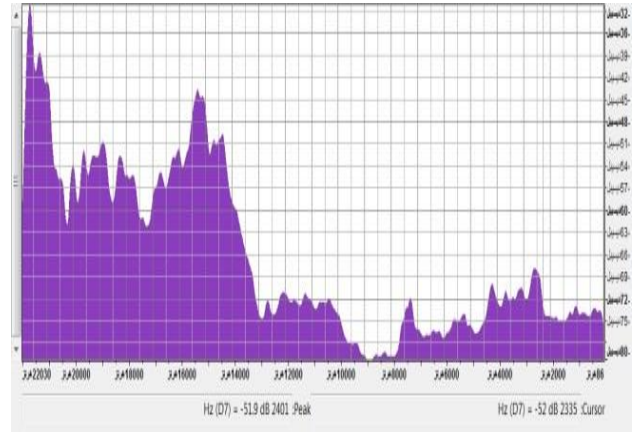
شكل 3. واجهة المستخدم الرسومية.

استطلعت آراء المستخدمين لاختبار مدى قبول المستخدمين متوسطي المهارة للواجهة، وقد تمكن ما نسبته 60% منهم من استخدام البرنامج بالشكل الصحيح ويوضح (شكل 4) تفاوت المستخدمين في قبول الواجهة وتوقعهم لوظيفه البرنامج.



شكل 4. مدى قبول المستخدمين لواجهة البرنامج.

- [4] Maheswari, Uma, and Rajeswari K C. "Prosody Modeling Techniques for Text-to-Speech Synthesis Systems - A Survey." International Journal of Computer Applications 39.13 (2012): 8-11.
- [5] Mazin Hamad, Mustafa Hussain, "Arabic Text-To-Speech Synthesizer". Electrical and Electronics Engineering Department, University of Khartoum, Khartoum, Sudan. In the proceeding of the IEEE Student Conference on Research and Development, 2011.
- [6] M. Z. Rashad, Hazem M. El-Bakry, Islam R. Isma'ili. "Diphone Speech Synthesis System for Arabic Using MARY TTS". International journal of computer science & information Technology (IJCSIT) Vol.2, No.4, August 2010.
- [7] Zouhir ZEMIRLI. "ARAB_TTS: An Arabic Text To Speech Synthesis". Institut National d'Informatique, LMCS, BP 35M – 16309 - Oued-smar Algiers ALGERIA, 2006.
- [8] Mnasri, Zied. "F 0 Contour Modeling for Arabic Text-to-Speech Synthesis Using Fujisaki Parameters and Neural Networks." Signal Processing: An International Journal (SPIJ) 4.6 (n.d.): 352-69.



شكل 6. رسم بياني لطيف نطق كلمة نُر

5. الاستنتاجات والتوصيات

في بحثنا هنا اجتزنا العمليات الحيوية والفيزيائية للقدرة البشرية واعتمدنا على تخزين الحروف العربية ضمن قاعدة بيانات وبالتالي بدأت التجارب لجعل البرنامج ينطق بالحروف المجردة و ثم الحرف بالحركة علمنا على تحديد المعينات الحاسوبية للنطق السليم في كل مرحلة وذلك بعد اجراء الاختبارات الثلاثة السابقة الذكر وتم تحديد المعينات النطقية و علمنا على تحسينها ما امكن، ولكن عملية الوصول الي برنامج متكامل باللغة العربية تحتاج الي المزيد من الجهد من قبل علماء اللغة Linguistics والمتخصصين بالنحو والصرف والتقنيين وذلك لنتمكن من عملية نقل المعرفة البشرية والقدرة علي النطق حسب قوانين اللغة الي العمليات الحاسوبية لنتمكن من برمجتها.

مما يعزز ثقتنا بنجاح تلك العمليات هي طبيعة تكوين اللغة العربية كلغة خالية من الشواذ أي ان كل ما يكتب ينطق استناداً على قاعدة، أي انها لغة خوارزمية والنطق فيها عبارة عن مجموعة متسلسلة تسلسل منطقي ثابت لا يتغير.

كذلك يجدر بنا الإشارة هنا الي اهمية اجراء المزيد من الاختبارات من النوعين الاول والثاني وذلك لتعزيز نتائج دراستنا هذه بحيث يتم اجراء اختبار الاحرف المتقاربة النطق مثل (ث ، س) وكذلك (ذ ، ز).

بالإضافة الي دراسة موسعة وتحليل عملية النطق البشري للكلمة متعددة المقاطع الصوتية وعملية الانتقال الصوتية بين الحروف داخل الكلمة متعددة المقاطع الصوتية.

تحويل كامل القواعد النطقية للغة العربية المتمثلة في الفية ابن مالك وترجمتها خوارزمية ليتم تمثيلها وتنفيذها برمجياً كما يجدر الإشارة هنا الي اهمية البحث والتطوير للقواعد الخاصة بالأرقام والتواريخ والعملات.

المراجع

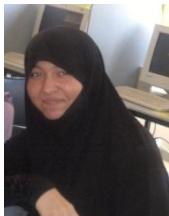
- [1] Rudi Studer, V. Richard Benjamins, and Dieter Fensel. "Knowledge Engineering-Principles and Methods". University of Karlsruhe, Germany. (IIIA), Spanish Council for Scientific Research (CSIC), Spain. Social Science Informatics (SWI), 1997.
- [2] El-Imam, Yousif A. "Phonetization of Arabic: Rules and Algorithms." Computer Speech & Language 18.4 (2004): 339-73.
- [3] Mohammed, Ahmed Qasim, ALjayousi." Arabic TEXT-TO-SPEECH Synthesizer. Thesis. Faculty of computer science and information technology university of Malaya -KUALA LUMPUR, 2007.

السير الذاتية للمؤلفين



إبراهيم علي المرهق حاصل على درجة البكالوريوس في الهندسة الكهربائية والإلكترونيات من جامعة بنغازي / ليبيا في عام 1986. وفي عام 1997، منح درجة الماجستير في هندسة الالكترونيات وهندسة الكمبيوتر من جامعة وارسو للتكنولوجيا / بولندا. كما أنه حاصل على ماجستير في إدارة الأعمال من جامعة برادفورد / المملكة المتحدة في عام 2002. وعلاوة على ذلك، حصل

على درجة الدكتوراه في المعلوماتية من جامعة برادفورد / المملكة المتحدة في عام 2006. وهو حالياً يعمل بقسم شبكات الحاسوب في جامعة طرابلس / ليبيا كأستاذ مشارك. له اهتمامات في هذه المجالات البحثية: جودة الخدمات، التوجيه والتحويل، أمن الشبكات وأنظمة الحاسوب، الأنظمة الخبيرة والجودة في التعليم العالي.



سندس محمد الشريف حصلت على درجة البكالوريوس في علوم الحاسب الآلي جامعة طرابلس / ليبيا في عام 2007. التحقت بالدراسات العليا كلية العلوم قسم الحاسب الآلي سنة 2010، تحصلت على دبلوم إدارة مشاريع من الأكاديمية الأمريكية لإدارة المشاريع الفرع الليبي في عام 2014، حاصلة على منحة دبلوم حوكمة الانترنت

من منظمة هيغوس الدولية في عام 2015، وهي تعمل حالياً كمدير ويب وتسويق الكتروني بشركة تاسيلي للاستشارات والتدريب طرابلس / ليبيا. لها اهتمامات في هذه المجالات البحثية: برمجة وتطوير الويب، الذكاء الاصطناعي، هندسة البرمجيات وإدارة المشاريع.