Arabic Text-To-Speech: Speech Units

Mansour Al-Ghamdi, Moustafa Elshafei and Husni-Al-Muhtaseb

Supported by King Abdulaziz City for Science and Technology
Project Number AT-18-12
Correspondence: husni@ccse.kfupm.edu.sa

Abstract: A growing Attention has been paid to text-to-speech for many languages in different methods. One of the widely used methods nowadays is concatenated speech method. It relies on prerecorded speech units that are utilized by the system to generate concatenated speech. This paper shows how 387 speech units were extracted, digitized and stored to be used in concatenated speech systems. Arabic language specific is taken into account to extract units that can be used in a concatenated speech system to generate high quality speech. The speech units cover Arabic sound distribution, although they are relatively a few in number and small in size (1226 kilobytes).

وحدات صوتية لتوليد الكلام العربي آليا

منصور بن محمد الغامدي، مصطفى أحمد الشافعي، حسني عبد الغني المحتسب

بدعم من مدينة الملك عبدالعزيز للعلوم والتقنية
مشروع رقم أت-18-12

للمتابعة: حسني المحتسب – قسم علوم الحاسب الآلي والمعلومات
جامعة الملك فهد للبترول والمعادن – الظهران 31261
husni@ccse.kfupm.edu.sa

الملخص: هناك اهتمام بارز لتوليد الكلام آليا لجميع اللغات الحية بطرق شتى. ومن أكثر الطرق شيوعا في هذه الفترة توليد الكلام بالتسلسل. وتعتمد هذه الطريقة على وحدات صوتية مسبقة التسجيل ويقوم النظام في توظيفها لإنتاج كلام متسلسل. وتشرح هذه الورقة كيف تم تكوين 387 وحدة صوتية يمكن استخدامها في نظم توليد الكلام بالتسلسل. أُخذ في الحسبان عند عمل هذه الوحدات الخصائص اللسانية للعربية للخروج بكلام عال الوضوح. ومما يميز هذه الوحدات قلة عددها وصغر حجمها الذي لا يتجاوز 1226 ألف بايت.

وحدات صوتية لتوليد الكلام العربي آليا

منصور بن محمد الغامدي، مصطفى أحمد الشافعي، حسني عبد الغني المحتسب

بدعم من مدينة الملك عبدالعزيز للعلوم والتقنية – مشروع رقم أت-18-12

للمتابعة: حسني المحتسب – قسم علوم الحاسب الآلي والمعلومات
جامعة الملك فهد للبترول والمعادن – الظهران 31261
Husni@ccse.kfupm.edu.sa

الملخص: هناك اهتمام بارز لتوليد الكلام آليا لجميع اللغات الحية بطرق شتى. ومن أكثر الطرق شيوعا في هذه الفترة توليد الكلام بالتسلسل. وتعتمد هذه الطريقة على وحدات صوتية مسبقة التسجيل ويقوم النظام في توظيفها لإنتاج كلام متسلسل. وتشرح هذه الورقة كيف تم تكوين 387 وحدة صوتية يمكن استخدامها في نظم توليد الكلام بالتسلسل. أُخذ في الحسبان عند عمل هذه الوحدات الخصائص اللسانية للعربية للخروج بكلام عال الوضوح. ومما يميز هذه الوحدات قلة عددها وصغر حجمها الذي لا يتجاوز 1226 ألف بايت.

تقديم

من أبرز التطورات التقنية في هذا العصر ظهور نظم مختلفة للنطق الآلي التي تتمتع باستخدامات عديدة وأن هذه الاستخدامات في تنام مطرد مع مرور الزمن والتطور المستمر في الإلكترونيات والحاسبات لكونها أصبحت جزءاً من الاستخدامات اليومية في حياة الناس. ويدخل ضمن هذه الاستخدامات: الأجهزة المعِينة للمعاقين والمكفوفين، والأجهزة المساعدة للتعليم، والوسائط المتعددة، ونظم الاتصالات، والعاب الأطفال ووسائل الترفيه وغيرها. هذا ما دفع مراكز البحث والتطوير في الجامعات والشركات المصنعة إلى استقطاب عدد من الباحثين واعتماد المبالغ للخروج بنظم للنطق الآلي.

للإحاطة بالمصطلحات المستخدمة في هذا البحث يرجى الرجوع لملحق رقم 1.

نظم توليد الكلام

نظراً للتعقيدات والصعوبات التي يواجهها الباحثون للخروج بناطق آلي ذي جودة عالية فإنهم نهجو طرقاً متباينة لتحقيق هذه الغاية. ويمكن تصنيف هذه الطرق إلى صنفين:

الأول: توليد الكلام آليا دون اللجوء إلى أصوات بشرية (إنشاء الكلام) Synthesized Speech. ويقع تحت هذا الصنف ثلاثة أنواع:

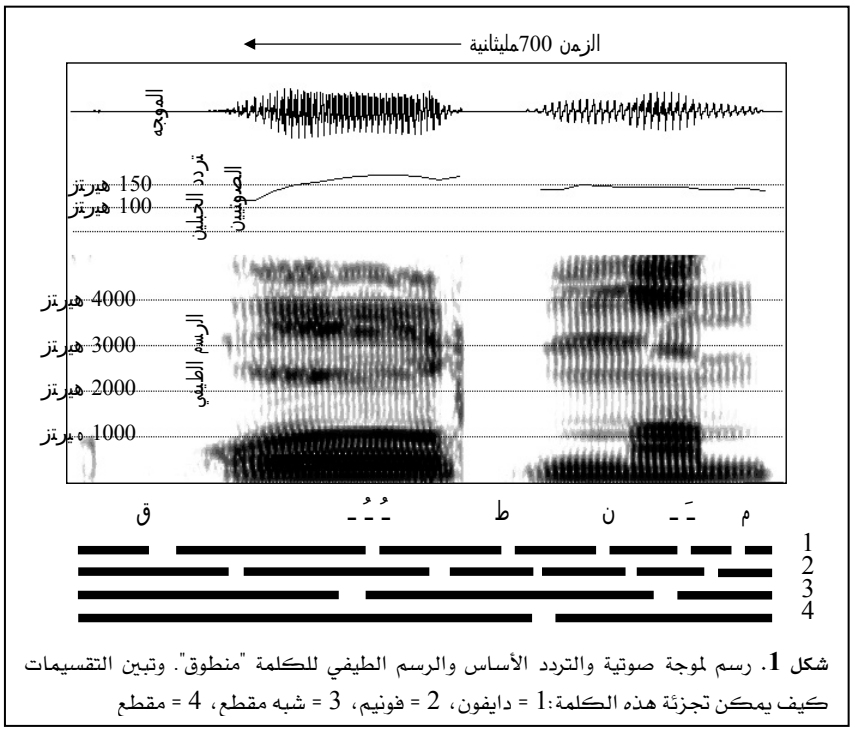
1. التنبؤ الخطي Linear Prediction وهو عبارة عن أنموذج لمرشحات تثار بضجيج صوتي ومصدر متسلسل لنبض منتظم.
2. توليد النطق الرنينية Formant Synthesis وهي أنموذج لتسلسل ترددات الإشارة الصوتية للكلام أو الاعتماد على أنموذج مرشحات المصدر لنقل وظائف الجهاز الصوتي.
3. تشبيه مخارج الأصوات Articulatory Synthesis وهي محاولة لعمل أنموذج يحاكي الجهاز الصوتي عند الإنسان ومن ثم إخراج الأصوات اللغوية بطريقة مشابهة لما يقوم به الجهاز الصوتي الطبيعي.

الثاني: تسلسل الكلام آلياً Concatenative Synthesis بالاعتماد على أصوات طبيعية سبق تخزينها وإضافتها إلى النظام ويستخدم هذا النوع عينات مختلفة الطول من الأصوات اللغوية التي سبق تسجيلها لأحد المتكلمين [2].

ومن أكثر النظم شيوعاً في الوقت الحاضر النوع الثاني من الصنف الأول، والصنف الثاني. وقد هيمن أسلوب توليد النطق الرنينية لفترة طويلة إلا أن تسلسل الكلام أصبح الآن يستحوذ على اهتمام الباحثين أكثر من أي نظام سواه. ومن النظم الواعدة في المستقبل إنشاء مخارج الأصوات إلا أنه لا يزال معقد التطبيق للخروج بنظام ناطق آلي ذي جودة عالية. وحيث أن هذه الورقة تركز على وحدات تسلسل الكلام فإننا سنعطيها شيء من التفصيل.

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 محدودا بمتحدث واحد وصوت واحد كما أنه يحتاج إلى ذاكرة أكبر من تلك في النظم الأخرى.



شكل 1. رسم لموجة صوتية والتردد الأساس والرسم الطيفي للكلمة "منطوق". وتبين التقسيمات كيف يمكن تجزئة هذه الكلمة:1 = دايفون، 2 = فونيم، 3 = شبه مقطع، 4 = مقطع

أحد النقاط المهمة في توليد الكلام بالتسلسل هي العثور على الطول الصحيح للوحدة. وغالبا ما يقع الخيار على طول وسط بين الطويل والقصير. والوحدة الطويلة تكون أقرب للطبيعية وتقلل من نقاط التسلسل وتعطي تحكم أكثر في النطق الانتقالية formant transition، إلا أن عدد الوحدات يكون أكثر وتحتاج إلى ذاكرة أكبر. وبالمقابل، فإن الوحدات القصيرة تحتاج إلى ذاكرة أقل، ولكن جمع عيناتها وتسميتها تكون أصعب وأعقد. وفي وحدات النظم الحالية، عادة ما تستخدم الكلمة word أو المقطع syllable أو نصف المقطع demisyllable أو الفونيم phoneme أو الفون المزدوج diphone وأحيانا الفون الثلاثي triphone.

والكلمات هي أكثر الوحدات طبيعية للنصوص والرسائل المكتوبة ذات المفردات المحدودة. ومن السهل عمل تسلسل للكلمات كما أنها تحتوي على النطق الانتقالية داخلها. إلا أن هناك فرق كبير بين الكلمات التي تنطق منفردة وتلك المكونة لجملة، وهذا يجعل الكلام المتصل بعيدا عن [2] . ولأن هناك مئات الآلاف من الكلمات وأسماء الأعلام في كل لغة فإن الكلمات لا تصلح أن تكون وحدات لنظام نطق غير محدود الكلمات.

إن عدد المقاطع المختلفة في أية لغة أقل بكثير من عدد كلماتها، إلا أن عدد الوحدات في قاعدة البيانات يظل كبيرا بالنسبة لنظام الناطق الآلي. فعلى سبيل المثال، هناك ما يقرب من 100 ألف مقطع في الإنجليزية. وعكس ما هو موجود في الكلمات، فإن تأثير النطق الانتقالية لا يكون موجودا بين المقاطع المخزنة، لذا فإن استخدام المقاطع كوحدات في نظم النطق الآلي لا تكون مناسبة. كما أنه ليس من الممكن التحكم في تطريز الجملة. وفي الوقت الحاضر، فإنه لا يوجد نظام ناطق آلي كامل قائم على كلمات أو مقاطع. إن النظم الحالية تقوم أساسا إما على الفونيمات أو الدايفونات أو أشباه المقاطع أو أي شكل من أشكال الجمع بينها (الشكل 1).

أشباه المقاطع

يمثل شبه المقطع بداية ونهاية المقطع. وأحد مميزات أشباه مقاطع الإنجليزية أنه يمكن أن نبني من ألف منها عشرة آلاف من المقاطع [3]. واستخدام أشباه المقاطع بدلا من الفونيمات، على سبيل المثال، لا يتطلب إلا قليلا من نقاط التسلسل. كما أنها تحتوي على جل النطق الانتقالية ومن ثم تشمل مناطق النطق المزدوج coarticulation، كما تحتوي على الاختلافات الألفونية نتيجة لعزل صوامت بداية ونهاية المقاطع. إلا أن الذاكرة المطلوبة تبقى كبيرة مع أنها ممكنة. وعند مقارنتها بالفونيمات والدايفونات فإن عددها الدقيق لا يمكن تحديده في أية لغة. كما أن النظام الذي يعتمد كليا على أشباه المقاطع لا يمكن أن يولد بشكل طبيعي جميع الكلمات الممكنة في اللغة. إلا أن النظام الذي يستخدم المقاطع وأشباه المقاطع يمكن أن يكون ناجحا إذا استعمل وحدات و زوائد كلمات متعددة الطول، كما في النظام HADIFIX [4]

الفونيمات

هي أكثر الوحدات شيوعا في الاستخدام في نظم النطق الآلي نظرا لكونها تمثل الوحدات الطبيعية اللغوية التي تستخدم أثناء حديث الإنسان. ويكون مجموعها في الغالب بين أربعين وخمسين وحدة، وهذا يوضح أنها أقل بكثير من الوحدات الأخرى[2] . وتعطي الفونيمات أعلى مرونة ممكنة في النظم القائمة على القوانين. إلا أن بعض الفونيمات التي ليس لها وضع مستقر كنقطة هدف، مثل الفونيمات الشديدة، من الصعب توليدها. كما أن مخرج الأصوات يحتاج أن يعمل كقانون. وتستخدم الفونيمات أحيانا كمدخلات للناطق الآلي لتوليد، على سبيل المثال، الناطق الآلي القائم على الدايفونات. ونظراً لكونها لا تحتوي على النطق الانتقالية فإنه لا يمكن الاعتماد عليها كليا في نظم النطق بالتسلسل.

الدايفونات

تُعرف الدايفونات بأنها المنطقة الممتدة من منتصف الحالة المستقرة لصوت إلى الحالة المستقرة للصوت الذي يليه. لذا فإنها تحمل معها الحالة الانتقالية بين الصوتين. هذا يعني أن التسلسل يقع بين أكثر الأماكن استقرارا للإشارة الصوتية. ميزة أخرى للدايفونات أنه لا حاجة لوضع قوانين للنطق الانتقالية. ومن حيث المبدأ ، فإن عدد الدايفونات يساوي مربع عدد الفونيمات مضافا إليه عدد الألفونات، إلا أنه ليس جميع التركيبات الفونيمية ضرورية. وعلى أية حال، فإن عدد الوحدات قابل للتطبيق خاصة أن الدايفونات مناسبة جدا كوحدات لنظام النطق الآلي القائم على العينات. إن عدد وحدات الدايفونات يمكن أن ينخفض إذا ما تم عكس الوحدات مثل /سَ/ لتصبح /سَ/.

يندر استخدام الوحدات الطويلة، كالفونات الثلاثية والرباعية. فالفونات الثلاثية شبيه بالفونات الثنائية (الدايفونات) إلا أنها تحتوي على فونيم آخر في الوسط (فهي تتكون من: نصف فونيم + فونيم كامل + نصف فونيم). وهناك أكثر من 10.000 فونيما في الإنجليزية [5].

يتم إعداد قائمة الوحدات على ثلاث مراحل [6]. الأولى، يجب تسجيل الكلام الطبيعي بحيث يحتوي على جميع الوحدات التي ستستخدم (فونيمات) في جميع سيق الكلام (ألفونات). الثانية، تسمية الوحدات واستخلاصها من المعطيات. الثالثة، يتم اختيار أكثر الوحدات ملائمة. يستغرق جمع العينات من كلام طبيعي وقتا طويلا، إلا أنه يمكن عمل ذلك آليا باختيار النص المدخل للتحليل بطريقة مناسبة. كما أن تطبيق قوانين اختيار العينات الصحيحة كوحدات يجب عملها بحذر.

الدايفون هو الوحدة الصوتية الممتدة من منتصف الحالة المستقرة لصوت لغوي إلى منتصف الحالة المستقرة للصوت الذي يليه [7]، [8]. ومما يميز هذا النوع من الوحدات الصوتية احتوائها على النطق الانتقالية بين الأصوات اللغوية التي تجعل الكلام المولد باستخدام هذا النوع من الوحدات أكثر وضوحا من الوحدات الألفونية. إلا أنه يؤخذ عليها كبر حجم المعجم المكون لها وكثرة عددها.

الخصائص الأصواتية العربية

المعرفة الدقيقة بالخصائص الأصواتية العربية هي من العوامل المهمة التي ينبغي أخذها في الحسبان عند التعامل الحاسوبي مع الأصوات اللغوية، خاصة في مجالي: التعرف الآلي على الكلام، وتوليد الكلام آليا. ومن الخطأ الاعتماد

الكلي على الدراسات والنتائج القائمة على اللغات الأخرى. فلكل لغة خصائصها الأصواتية التي تنفرد بها. ومن هذه الخصائص النظام الصوتي phonological system ونظام التسلسل الأصواتي phonetactics والتراكيب المقطعية syllabic structure. وتتباين اللغات الطبيعية natural languages تبايناً واضحا في هذه الخصائص سواء من حيث الكم أو الكيف، مما يحتم الدراية الكاملة بخصائص اللغة تحت الدراسة.

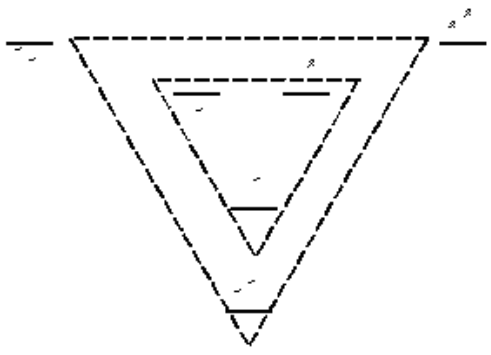
فمما يميز العربية نظامها الصوتي الذي يحتوي على 28 صامت consonant. (جدول 1 [9]) و6 صوائت vowel (الشكل 2). أي أن نسبة الصوائت إلى الصوامت هي 21٪. وتقل هذه النسبة إذا ما علمنا أن نصف هذه الصوائت يقابل النصف الآخر في الكيفية quality ويختلف فقط في الكمية quantity. أي أن الفتحة القصيرة /ـَ/ في كلمة "سَد" يقابلها الفتحة الطويلة /ـَـ/ في كلمة "سَاد"، والضمة القصيرة /ـُ/ في كلمة "عُد" يقابلها الضمة الطويلة /ـُـ/ في كلمة "عُود"، والكسرة القصيرة /ـِ/ في كلمة "سِن" يقابلها الكسرة الطويلة /ـِـ/ في كلمة "سِين". هذا عكس لغات أخرى كالسويدية والإنجليزية والألمانية التي تزيد فيها نسبة الصوائت إلى الصوامت على 50٪.

	Bilabial شفتاني	Labiodental شفوي أسناني	Interental بين أسناني	Alveodental لثوي أسناني	Alveopalatal غاري لثوي	Palatal غاري	Velar طبقي	Lab-velar شفوي طبقي	Uvular لهوي	Pharyngeal حلقي	Glottal حنجري
Nasal غني	م m			ن n							
Stop شديد	ب b			ت t د d			ك k		ق q		ء ʔ
Emphatic Stop*				ط tˤ ض dˤ							
Fricative رخو		ف f	ث θ ذ ð	س s ز z	ش ʃ				غ ɣ خ χ	ع ʕ ح ħ	هـ h
Emphatic fricative**			ظ ðˤ	صˤ sˤ							
Affricate مزجي					ج dʒ						
Glide لزلاقي						ي j		و w			
Lateral جانبي				ل l							
Trill تكراري				ر r							

جدول 1 أصوات العربية الفصحى المعاصرة، الأصوات المهموسة تحتها خط بينما المجهورة ليس تحتها خط

* تعني مفخم شديد و** تعني مفخم رخو

وتتميز العربية بخصائص أصواتية تسهل التعامل معها حاسوبيا، فنظام التسلسل الأصواتي phonetactics بسيط إذا ما قورن بذلك في اللغات الأخرى كالإنجليزية مثلا. فالنظام لا يسمح بورود أكثر من صامتين متتاليين وهو ما يعرف عند اللغويين العرب بعدم التقاء الساكِنَين.



شكل 2. الصوائت العربية الستة: ثلاثة منها قصيرة والأخرى طويلة. وما يقابلها في الرموز الدولية: /ـَ/ = /a/، /ـِ/= /i/، /ـُ/ = /u/، /ـًـ/ = /a:/، /ـِـ/= /i:/، /ـٌـ/ = /u:/.

هذا يعني أنه لابد لكل صامت من صائت إما سابق له أو لاحق به وغالبا ما يقع الصامت بين صائتين. فالصوائت منتشرة بكثرة في الكلام العربي رغم عددها القليل. وتكمن ميزة وجود الصوائت بكثرة احتوائها على المشعرات الاكوستية للصوامت المجاورة. فالفرق الأساس بين /ط/ ، /ت/ هو فرق في الصائت المجاور لهما لا في الصوت نفسه فكلاهما ينطق بلا ترددات صوتية حيث أنهما صوتان شديدان ومهموسان. وحيث أن أعلى شدة في الموجات الصوتية موجودة في الصوائت فإن المشعرات التي تحملها غالبا ما تكون واضحة للسامع بعكس تلك الموجودة في الصوامت. فلو كان الصامت بين صامتين فإن

المشعرات الخاصة به تكون في الصوامت المجاورة ويجد السامع صعوبة لحد ما في التعرف عليه، مما يجعل أيضا من الصعوبة بمكان توليد الصوت المناسب في نظام النطق الآلي.

والتراكيب المقطعية محدودة وذات سمة مميزة، فكلها تبدأ بصامت متبوع بصائت. وهذا عكس ما هو قائم في لغة كالإنجليزية مثلا التي يمكن أن يبدأ فيها المقطع أو ينتهي بصامت أو صامتين أو ثلاثة صوامت أو بلا صامت، هذا يعني أن هناك ما يقرب من عشرين نوع من المقاطع. أما في العربية فإن التراكيب البسيطة في المقطع العربي تسهل من عملية توليد الأصوات أو التعرف الآلي عليها، فمن المقطع تتكون الكلمات وكلما كانت تراكيب المقاطع محدودة كلما كان من السهولة توليد الكلام آليا أو التعرف عليه. وتتكون المقاطع العربية من:

1- صامت متبوع بصائت، هذا الصائت إما أن يكون:

أ. قصيراً كما في /كَـ/ في كلمة "كَتَبَ"،

ب. طويلاً كما في /سَـ/ في كلمة "سَاروا".

2- صامت متبوع بصائت متبوع بصامت، ويكون الصائت إما:

أ. قصيراً كما في /فَـكْ/ في كلمة "فَكَّر"،

ب. طويلاً كما في /فَـلْ/ في كلمة "فَال".

3- صامت متبوع بصائت قصير ثم صامت فصامت كما في /بَـدْ ر/ في كلمة "بَدْر".

هذه الخصائص الأصواتية مجتمعة تسهل من عملية إنشاء نظام حاسوبي لتوليد الكلام العربي. وفي حالة الإدراك الآلي للكلام العربي فإنها تُعِين في التعرف على أماكن الأصوات اللغوية ووضع القوانين الدالة على أماكن تواجدها وتسلسلها.

تسجيل واستخلاص الوحدات

تم وضع الفونيمات المستهدفة في نص حامل باستخدام النص الحامل carrying token الذي اُستخدم في قاعدة بيانات الصوتيات العربية الذي أنتجته مدينة الملك عبد العزيز للعلوم والتقنية وهو /زVCVز/ حيث يرمز V للصائت و C للصامت المستهدف.

وكان النظام المستخدم في تسجيل الوحدات واستخلاصها هو CSL 4300B ، ومما يميز هذا النظام نقاء الصوت المسجل عن طريقه فهو نظام مصنع خصيصا لتسجيل وتحليل الأصوات اللغوية. هذا النظام يحفظ ملفاته على هيئة NSP format التي لا يمكن للبرامج الأخرى فتحها. ولتحويلها إلى ملفات يمكن الاستفادة منها من قبل نظم وبرامج أخرى أكثر شيوعا فقد تم تحويلها إلى WAV format باستخدام النظام Multi-speech 3700، وكلا النظامين من إنتاج شركة Kay Elemetrics.

الترددات الواقعة في نطاق 3 كيلو هيرتز تكفي لفهم الكلام المنطوق. وهذا هو القائم في كثير من أنظمة الاتصالات الهاتفية [10]. إلا أن بعض الصوامت الرخوة fricatives تقع ترددتها فوق 3 كيلو هيرتز، وللحصول على وحدات تحمل الخصائص الأكوستية الضرورية للسامع العربي فقد استخدمنا نسبة تمثيل تعادل 10000 عينة / الثانية التي تغطي ترددات من الصفر إلى 5 كيلو هيرتز. ونرى بأن هذه النسبة مناسبة لتسجيل دايفونات واضحة كما أنها لا تحمل ترددات غير ضرورية قد تعيق عملية التخزين والنقل لكبر حجمها (المرجع السابق).

وتم اختيار عدة وحدات صوتية موزعة كالتالي:

1. كلمة كاملة:

لفظ الجلالة لاحتوائه على اللام المفخمة التي لا توجد إلا فيه وتأتي بهذه الصفة في جميع الحالات إلا عندما تكون مسبوقة بالصائتين /ـِ/ ، /ـِ/: كما في "ما شاءَ اللّه" ، "اللّهُ العلي القدير"، "رحمةُ اللّهِ واسعة"، فهي هنا مفخمة. وتكون لام عادية كما في: "باللّه عليك".

2. الدايفونات:

وقد استخلصت من النصوص الحاملة التالية:

/زَـ C ـَز/،

/زَّـ C ـَّز/،

/زُـ C ـُز/،

/زُّـ C ـُّز/،

/زِـ C ـِز/،

/زِّـ C ـِّز/،

وتحتوي على التراكيب الصوتية الأكثر شيوعا في العربية وهي الصوامت وما جاورها من الصوائت وتتكون من:

- 168 دايفون تحتوي على جميع الصوامت متبوعة بجميع الصوائت (6×28).
- 168 دايفون تحتوي على جميع الصوائت متبوعة بجميع الصوامت (6×28).

وتم استخلاص الدايفونات من منتصف الوضع المستقر للصامت إلى منتصف الوضع المستقر للصائت المجاور.

3. الفونيمات:

وتتكون كل واحدة منها من 100مليثانية من الفترة المستقرة لجميع الصوامت عدا الشديد منها: (ب ، ت ، ط ، د ، ض ، ك ، ق ، ء)، حيث استخدمت فترة صمت لمدة 100 مليثانية وهذا ما يحدث في حقيقة الأمر عند نطق هذه الأصوات بالتضعيف. وتفيد هذه الوحدات عند نطق الحروف المشددة كما في الكلمات: "عَدَّاد" ، "فَكَّر" ، "مَهَّد".

وقد استخلصت من النص الحامل: //زَـ CC ـَز/، حيث ترمز CC للصامت المستهدف في حالة التشديد.

4. الفونات:

وتتكون من 28 فونا لجميع الصوامت ، حيث تمتد كل وحدة من منتصف الفترة المستقرة للصامت إلى نهايتها. وتستخدم هذه الوحدات لنطق الحروف في نهاية الكلام، كما في الحرف /د/ في الجملة: "يوفي عليٌ بالعهْد". واُستخلصت هذه الوحدات من النص الحامل /زَـ C/ حيث C هي الصامت في نهاية الحاملة.

التحليل الأكوستي والإحصائي

اُستخدم نظام CSL 4003B لتحليل الموجات الصوتية للدايفونات وقياس أمد كل منها وتردده الأساس ، وتم قياس التردد الأساس للدايفونات من منتصف الوضع المستقر للتردد الأساس أثناء نطق الصائت أما في الفونيمات والفونات فإن القياس كان في منتصف الوضع المستمر للتردد الأساس أثناء نطق الصوامت المجهورة. واستخدم برنامج Excel لمعرفة المعدل والانحراف المعياري وإجمالي أمد الدايفونات.

الترميز

يمكن تسمية ملفات الوحدات بأي اسم يراه الباحث مناسباً لخطة بحثه، إلا أنه نظرا لكثرة ملفات الوحدات الصوتية فإنه من الأفضل أن يتم ترميزها بطريقة تجعل من السهل على أي مستخدم آخر الاستفادة منها. فتم أولا ترميز الصوائت والصوامت العربية. فكانت كما هو موضح في الجدول 2. والترميز وضع هنا على أساس ما يقابل الحرف العربي في الإنجليزية، فإذا كان له شبيه فإننا استخدمنا الشبيه مضافا له حرف S كما هي الحالة في /ب/ حيث رمزها BS، وإذا لم يكن هناك شبيه له فإننا نستخدم الأقرب من قائمة الرموز الدولية للصوتيات كما في حالة /خ/ XS، /ق/ QS،

وفي حالة الأصوات المتشابهة فتم استخدام C في الخانة الثانية للأصوات الخاصة بالعربية أو الأصوات المجهورة كما في /ص/ SC، /ط/ TC، /غ/ XC.

الرمز	الحرف	الرمز	الحرف	الرمز	الحرف	الرمز	الحرف
YS	ي	XC	غ	RS	ر	HZ	ء
AS	ـَ	FS	ف	ZS	ز	BS	ب
US	ـُ	QS	ق	SS	س	TS	ت
IS	ـِ	KS	ك	JS	ش	VS	ث
AC	ـً	LS	ل	SC	ص	JC	ج
UC	ـٌ	MS	م	DC	ض	HC	ح
IC	ـٍ	NS	ن	TC	ط	XS	خ
		HS	هـ	ZC	ظ	DS	د
		WS	و	CS	ع	VC	ذ

جدول 2. نظام الترميز لاصوات العربية المستخدم في الوحدات الصوتية.

واُستخدمت 6 خانات لترميز كل ملف من ملفات الوحدات الصوتية (الملحق 2). فكانت الخانتان الأولى من اليسار ترمزان للصائت المكون للدايفونات التي تبدأ بصائت. أما الخانتان اللتان في المنتصف فترمزان للصامت المكون للدايفون، كما أنهما يرمزان للفونيم. والخانتان الأخيرة ترمزان للصائت المكون لدايفون يبدأ بصامت، كما أنهما ترمزان للفون في نهاية الكلمة. ويفسر جدول رقم 3 أمثلة على ذلك.

الشرح	الرمز
دايفون يبدأ بالصامت /ب/ متبوع بالصائت /ـَ/	00BSAS
دايفون يبدأ بالصائت /ـِ/ متبوع بالصامت /ح/	ICHC00
الفونيم /س/ منفردا	00SS00
الصامت /ت/ منفردا في نهاية الحاملة	0000TS

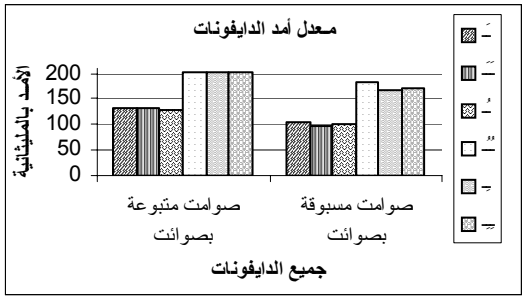
جدول 3. أمثلة على ترميز الملفات.

النتائج

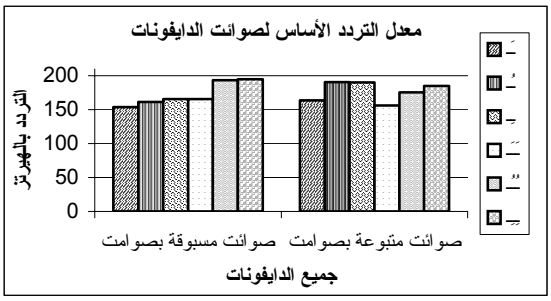
كانت نتيجة جميع هذه الوحدات الصوتية هو 387 وحدة: واحدة منها لفظ الجلالة، ومنها 336 دايفونا نصفها يشمل الدايفونات المكونة من صوامت متبوعة بصوائت، والنصف الآخر يشمل صوامت مسبوقة بصوائت، و22 فونيما لجميع الصوامت عدى 7 منها تتصف بخاصية الشدة فيحل محلها فترة صمت تمتد لـ 100مليثانية. كما تشمل 28 فونا.

وتحتل جميع الوحدات مساحة تساوي 1226 ألف بايت، بينما الزمن لا يتجاوز الدقيقة الواحدة. وهي مساحة صغيرة نسبيا إذا ما قورنت بما عُمل على الفرنسية على سبيل المثال التي بلغت عدد الدايفونات فيها 1200 دايفون، وبلغ الزمن 3 دقائق [11].

تبين نتائج التحليل الإحصائي المعروضة في الملاحق 1-4 والمختصرة في الشكل 3 و4 أن المعدل الكلي أمد الدايفونات المكونة من صوامت متبوعة بصوائت أطول من نظائرها المكونة من صوامت مسبوقة بصوائت إذ يبلغ معدل الفرق 30 مليثانية. بينما نجد المعد الكلي للتردد الأساس نفسه تقريبا بالنسبة للنوعين من الدايفونات (175 هيرتز تقريبا). إلا أن الفروق تكون أكثر بروزا عند مقارنة مكونات كل نوع من الدايفونات. فمعدل أمد الدايفونات المكونة من صوائت طويلة أطول من تلك المكونة من صوائت قصيرة (72 مليثانية)، وهذا طبيعي لكون الصوائت الطويلة أطول من تلك القصيرة بما مقداره 80 مليثانية تقريبا [12].



شكل 3. الأمد



شكل 4. التردد

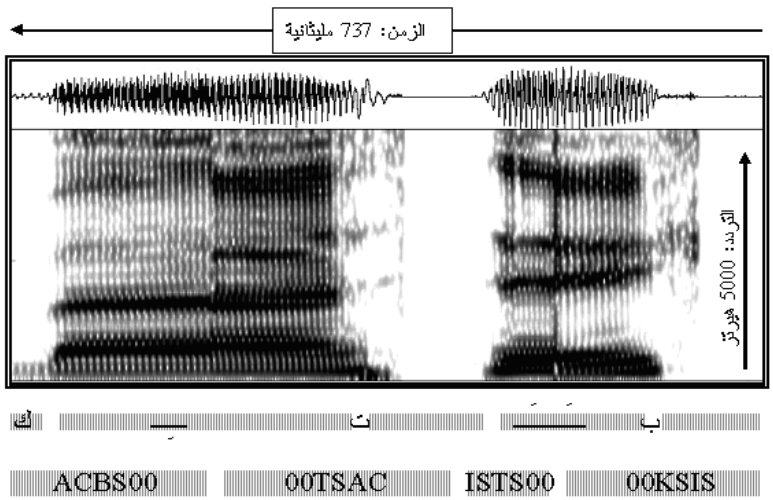
وفي حالة توليد الكلام فإنه من الممكن التحكم في درجة التردد الأساس والأمد وذلك بعمل البرامج الحاسوبية المناسبة لتحقيق هذه الغاية. هذا التحكم ضروري للخروج بنظام يمكن أن ينتج كلام بدرجات مختلفة في الأمد والشدة والتردد الأساس، هذه العوامل الثلاث ذات أهمية كبيرة في حالتي التطريز prosody والتأكيد stress. وتكمن أهمية هذه العوامل في علاقتها القوية بطبيعة الكلام. إذ أنها ليست دائما ثابتة في أصوات اللغة وإنما تتغير بناء على موقعها في الكلام وصوت المتحدث وأسلوبه.

الكلمة	التسلسل
عُلُوم	00CSUS USLS00 00LSUC UCMS00 0000MS
مَكْتَب	00MSAS ASKS00 000100 00TSAS ASBS00 0000BS
قَوم	00QSAS ASWS00 00UCMS 0000MS
النَّاس	00HZAS ASNS00 00NS00 00NSAC ACSS00 0000SS
عِبء	00CSIS ISBS00 0000HZ

جدول4 . أمثلة على تسلسل الوحدات.

يبين الجدول رقم 4 بعض الأمثلة على تسلسل الوحدات المذكورة في هذه الورقة لنطق بعض الكلمات.

ويلاحظ أن التسلسل في الجدول أعلاه يبدأ من اليسار. كما يلاحظ أيضا أن الرمز 000100 في "مكتب" يدل على أمد 100 مليثانية، سبق وأن ذكرنا أن هذه الوحدة تستخدم في تضعيف الصوامت الشديدة، ونظرا لأنه في حالة عنقود الصوامت consonant cluster وهي الصوامت التي ليس بينها صائت، فإننا استخدمناها هنا للحاجة إلى أمد أطول للصامت الأول.



شكل 5. رسم للموجة الصوتية ورسم طبقي للكلمة "كِتَاب" ويظهر أسفل منه تقطيع الموجة غلأى فونيمات ودايفونات.

ويبين الشكل 5. الموجة الصوتية والرسم الطيفي للكلمة "كِتَاب" الناتجة عن تسلسل أربع وحدات صوتية. وتبين الحدود بين هذه الوحدات الاختلاف بين نهاية وحدة وبداية الأخرى في الشدة والتردد، وهذا ما يجعل مبرمجي نظم النطق الآلي القائم على التسلسل يقومون بعمل برنامج يتولى التوفيق بين حدود الوحدات الصوتية ليجعل الكلام سلس smooth.

ونختم بالقول أن الوحدات المستخلصة والمجربة على عدد كبير من الكلمات العربية تكفي لتوليد الكلام بوضوح كبير للسامع العربي. ومن نافلة القول أن العمل في هذا المجال لا يمكن الحكم عليه بالكمال فهو يتطلب المزيد من التطوير والتحسين المستمرين خاصة في الأصوات المحسنة كالترقيق والتفخيم لبعض الصوامت العربية كالراء على سبيل المثال.

شكر

نشكر مدينة الملك عبد العزيز للعلوم والتقنية على دعم المشروع رقم أت -18 – 12. كما نشكر جامعة الملك فهد للبترول والمعادن.

المراجع

[1] Ahhmed, M. Elshafei (1991) Toward an Arabic Text-to Speech System, *The Arabian Journal for Science and Engineering*, Vol 16, 4B, 565-583.

[2] Allen J., Hunnicutt S., Klatt D. (1987). From Text to Speech: The MITalk System. Cambridge University Press, Inc.

[3] Donovan R. (1996). Trainable Speech Synthesis. PhD. Thesis. Cambridge University Engineering Department, England.

[4] Dettweiler H., Hess W. (1985). Concatenation Rules for Demisyllable Speech Synthesis. Proceedings of ICASSP 85 (2): 752-755.

[5] Huang X., Acero A., Hon H., Ju Y., Liu J., Mederith S., Plumpe M. (1997). Recent Improvements on Microsoft's Trainable Text-to-Speech System - Whistler. Proceedings of ICASSP97 (2): 959-934.

[6] Hon H., Acero A., Huang X., Liu J., Plumpe M. (1998). Automatic Generation of Synthesis Units for Trainable Text-to-Speech Systems. Proceedings of ICASSP 98 (CD-ROM).

[7] Al-Muhtaseb, Husni, Moustafa Elshafei and Mansour Al-ghamdi (2000) Techniques for High Quality Arabic Speech Synthesis, The Third KFUPM Workshop on Information & Computer Science, 73-82.

[8] Elshafei, Moustafa, Husni Al-Muhtaseb and Mansour Al-ghamdi, Techniques for High Quality Arabic Speech Synthesis, accepted for publication in a special issue on Software Engineering: Systems and Tools, Information Sciences vol. 140/3-4.

[9] الغامدي، الدكتور منصور بن محمد (1421هـ) الصوتيات العربية، مكتبة التوبة، الرياض، المملكة العربية السعودية.

[10] O'Saughnessy D. (1987). Speech Communication - Human and Machine, Addison-Wesley, United States of America.

[11] Dutoit, Thierry (1997) An Interoduction to Text-to-Speech Synthesis, Kluwer Academic Publishers, Dordrecht.

[12] Hussain, A. A. (1985) An Experimental Investigation of Some Aspects of the Sound System of the Gulf Arabic Dialect with Special Reference to Duration. (Unpublished Ph. D. thesis, Essex).

ملحق رقم 1: قائمة المصطلحات

إنجليزي/عربي

acoustics	اكوستية (العلم المتعلق بالموجات الصوتية)
allophone	ألوفون (الفونيم عندما يظهر بشكل مختلف بناء على موقعه في الكلام)
articulatory synthesis	تشبيه مخارج الأصوات
carrying token	نص حامل
coarticulation	نُطْق مزدوج
concatenated speech	تسلسل الكلام
consonant	صامت
consonant cluster	عنقود صوامت
demisyllable	نصف مقطع
diphone	فون مزدوج
formant	نُطُق رنينية
formant synthesis	توليد النُطُق الرنينية
formant transition	نُطُق انتقالية
format	بُنْية
linear prediction	تنبؤ خطي
Modern Standard Arabic	العربية الفصحى المعاصرة
natural language	لغة طبيعية (كالعربية والإنجليزية...)
phone	فون (اقصر وحدة صوتية متجانسة)
phoneme	فونيم (اصغر وحدة صوتية تغير معنى الكلمة)
phonotactics	نمط تتابع صوتي
prosody	تطريز
quality	كيفية
quantity	كمية
smooth	سلس
speech synthesis	إنشاء الكلام
stress	تأكيد
syllable	مقطع
triphone	فون ثلاثي
vowel	صائت
word	كلمة

عربي/إنجليزي

ألوفون (الفونيم عندما يظهر بشكل مختلف بناء على موقعه في الكلام)	allophone
إنشاء الكلام	speech synthesis
اكوستية (العلم المتعلق بالموجات الصوتية)	acoustics
بُنْية	format
تأكيد	stress
تسلسل الكلام	concatenated speech
تشبيه مخارج الأصوات	articulatory synthesis
تطريز	prosody
تنبؤ خطي	linear prediction
توليد النُطُق الرنينية	formant synthesis
سلس	smooth
صائت	vowel
صامت	consonant
العربية الفصحى المعاصرة	Modern Standard Arabic
عنقود صوامت	consonant cluster
فون (اقصر وحدة صوتية متجانسة)	phone
فون ثلاثي	triphone
فون مزدوج	diphone
فونيم (اصغر وحدة صوتية تغير معنى الكلمة)	phoneme
كلمة	word
كمية	quantity
كيفية	quality
لغة طبيعية (كالعربية والإنجليزية...)	natural language
مقطع	syllable
نص حامل	carrying token
نصف مقطع	demisyllable
نُطُق انتقالية	formant transition
نُطُق رنينية	formant
نُطْق مزدوج	coarticulation
نمط تتابع صوتي	phonotactics

ملحق رقم 2: Statistics-1

	Consonants + long high front vowel			Consonants + short high front vowel			Consonants + long high back vowel			Consonants + short high back vowel			Consonants + long low vowel		
Script Arabic	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0
ء	00hzic	234	185	00hzis	130	164	00hzuc	206	200	00hzus	107	172	00hzac	196	165
ب	00bsic	172	182	00bsis	88	156	00bsuc	206	164	00bsus	126	169	00bsac	212	161
ت	00tsic	210	205	00tsis	128	154	00tsuc	206	182	00tsus	145	167	00tsac	226	162
د	00dsic	158	179	00dsis	101	153	00dsuc	177	185	00dsus	111	159	00dsac	202	167
ط	00tcic	203	204	00tcis	142	162	00tcuc	186	186	00tcus	136	169	00tcac	203	160
ث	00vsic	175	164	00vsis	164	170	00vsuc	220	193	00vsus	143	149	00vsac	217	166
ذ	00vcic	173	185	00vcis	113	165	00vcuc	184	176	00vcus	117	161	00vcac	205	154
ج	00jcic	213	185	00jcis	37	139	00jcuc	202	189	00jcus	135	164	00jcac	220	166
ش	00jsic	224	194	00jsis	158	152	00jsuc	221	190	00jsus	153	149	00jsac	216	170
ح	00hcic	246	188	00hcis	163	158	00hcuc	231	170	00hcus	137	167	00hcac	216	167
ع	00csic	186	185	00csis	120	182	00csuc	193	185	00csus	124	169	00csac	200	164
خ	00xsic	214	196	00xsis	152	167	00xsuc	214	196	00xsus	141	166	00xsac	220	174
غ	00xcic	171	213	00xcis	115	178	00xcuc	203	217	00xcus	126	147	00xcac	190	167
ر	00rsic	187	217	00rsis	106	152	00rsuc	181	190	00rsus	115	145	00rsac	180	162
س	00ssic	211	188	00ssis	147	156	00ssuc	208	193	00ssus	123	133	00ssac	212	161
ز	00zsic	201	203	00zsis	53	152	00zsuc	213	196	00zsus	130	169	00zsac	216	170
ص	00scic	208	196	00scis	148	164	00scuc	190	208	00scus	170	159	00scac	221	161
ض	00dcic	209	228	00dcis	144	192	00dcuc	202	179	00dcus	145	173	00dcac	202	
ظ	00zcic	220	209	00zcis	175	178	00zcuc	202	200	00zcus	156	175	00zcac	205	161
ف	00fsic	215	208	00fsis	154	204	00fsuc	192	213	00fsus	166	204	00fsac	208	164
ق	00qsic	209	223	00qsis	136	170	00qsuc	214	213	00qsus	139	147	00qsac	208	172
ك	00ksic	227	192	00ksis	161	154	00ksuc	246	196	00ksus	171	149	00ksac	237	164
ل	00lsic	188	200	00lsis	96	159	00lsuc	192	173	00lsus	105	143	00lsac	160	170
م	00msic	173	212	00msis	103	161	00msuc	162	200	00msus	112	159	00msac	186	170
ن	00nsic	187	204	00nsis	112	176	00nsuc	190	205	00nsus	95	162	00nsac	156	170
و	00wsic	211	204	00wsis	167	156	00wsuc	200	209	00wsus	140	170	00wsac	191	173
ي	00ysic	223	112	00ysis	146	200	00ysuc	212	218	00ysus	137	159	00ysac	182	164
ه	00hsic	196	200	00hsis	146	164	00hsuc	200	190	00hsus	115	159	00hsac	208	161
Mean		202	195		129	166		202	193		133	161		203	165
SD		22	22		34	15		17	14		20	14		18	5
Total Duration		5644			3605			5653			3720			5695	

ملحق رقم 2: Statistics-2

	Consonants + short low vowel			Consonant steadystates			Short high front vowel + consonants			Long high front vowel + consonants			Short low vowel + consonants		
Script Arabic	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0
ء	00hzas	123	152	00hz00	100		ishz00	85	196	ichz00	175	169	ashz00	82	167
ب	00bsas	78	147	00bs00	100		isbs00	87	175	icbs00	145	175	asbs00	83	161
ت	00tsas	120	154	00ts00	100		ists00	105	192	icts00	150	172	asts00	102	167
د	00dsas	112	149	00ds00	100		isds00	82	176	icds00	134	173	asds00	80	156
ط	00tcas	151	152	00tc00	100		istc00	99	193	ictc00	154	175	astc00	134	164
ث	00vsas	124	159	00vs00	100		isvs00	114	200	icvs00	205	182	asvs00	117	172
ذ	00vcas	117	167	00vc00	100	125	isvc00	95	172	icvc00	155	166	asvc00	85	164
ج	00jcas	109	166	00jc00	100	115	isjc00	119	149	icjc00	185	175	asjc00	107	179
ش	00jsas	174	164	00js00	100		isjs00	113	185	icjs00	169	189	asjs00	120	164
ح	00hcas	125	150	00hc00	100		ishc00	102	185	ichc00	193	197	ashc00	114	173
ع	00csas	130	156	00cs00	100	122	iscs00	87	182	iccs00	193	175	ascs00	46	130
خ	00xsas	135	158	00xs00	100		isxs00	94	182	icxs00	166	175	asxs00	112	162
غ	00xcas	169	154	00xc00	100	141	isxc00	86	200	icxc00	162	189	asxc00	119	170
ر	00rsas	114	146	00rs00	130	154	isrs00	74	196	icrs00	153	192	asrs00	108	156
س	00ssas	168	156	00ss00	100		isss00	136	182	icss00	173	181	asss00	138	167
ز	00zsas	151	147	00zs00	100	123	iszs00	130	185	iczs00	173	185	aszs00	122	159
ص	00scas	209	150	00sc00	100		issc00	119	192	icsc00	186	182	assc00	142	157
ض	00dcas	200	161	00dc00	100		isdc00	84	218	icdc00	143	201	asdc00		
ظ	00zcas	177	151	00zc00	100	139	iszc00	91	209	iczc00	178	185	aszc00	128	167
ف	00fsas	122	154	00fs00	100		isfs00	115	214	icfs00	165	218	asfs00	105	164
ق	00qsas	154	161	00qs00	100		isqs00	103	208	icqs00	181	196	asqs00	113	169
ك	00ksas	142	164	00ks00	100		isks00	104	193	icks00	171	182	asks00	107	178
ل	00lsas	91	164	00ls00	100	170	isls00	69	185	icls00	162	182	asls00	71	175
م	00msas	133	173	00ms00	100	169	isms00	91	193	icms00	163	185	asms00	105	176
ن	00nsas	110	159	00ns00	100	167	isns00	81	190	icns00	145	181	asns00	80	175
و	00wsas	113	130	00ws00	100	164	isws00	106	200	icws00	198	205	asws00	111	150
ي	00ysas	97	132	00ys00	100	169	isys00	126	182	icys00	204	205	asys00	100	147
هـ	00hsas	129	141	00hs00	100		ishs00	99	192	ichs00	174	182	ashs00	116	154
Mean		135	154		101	147		100	190		170	185		105	164
SD		31	10		6	21		17	14		19	12		22	11
Total Duration		3777			2830			2796			4755			2847	

ملحق رقم 2: Statistics-3

	Long low vowel + consonants			Short high back vowel + consonants			Long high back vowel + consonants			Final Consonants		
Script Arabic	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0	Code	Duration	F0
ء	achz00	138	159	ushz00	78	208	uchz00	193	179	0000hz	254	
ب	acbs00	200	152	usbs00	105	196	ucbs00	173	154	0000bs	237	130
ت	acts00	172		usts00	96	197	ucts00	141	164	0000ts	271	
د	acds00	189	154	usds00	85	178	ucds00	142	181	0000ds	177	169
ط	actc00	189	159	ustc00	108	186	uctc00	176	172	0000tc	132	
ث	acvs00	198	162	usvs00	112	196	ucvs00	152	169	0000vs	258	
ذ	acvc00	178	145	usvc00	81	175	ucvc00	135	164	0000vc	200	149
ج	acjc00	200	154	usjc00	78	175	ucjc00	175	175	0000jc	285	135
ش	acjs00	204	151	usjs00	130	196	ucjs00	165	169	0000js	261	
ح	achc00	211	157	ushc00	104	192	uchc00	206	169	0000hc	262	
ع	accs00	175	149	uscs00	80	189	uccs00	186	170	0000cs	137	154
خ	acxs00	207	150	usxs00	104	188	ucxs00	189	154	0000xs	244	
غ	acxc00	166	156	usxc00	104	192	ucxc00	166	203	0000xc	194	161
ر	acrs00	169	150	usrs00	74	185	ucrs00	140	165	0000rs	178	189
س	acss00	217	154	usss00	106	169	ucss00	176	158	0000ss	283	
ز	aczs00	206	162	uszs00	82	189	uczs00	185	166	0000zs	208	135
ص	acsc00	202	152	ussc00	118	201	ucsc00	161	173	0000sc	312	
ض	acdc00	196	158	usdc00	100	200	ucdc00	157	173	0000dc	190	169
ظ	aczc00	209	156	uszc00	100	205	uczc00	167	196	0000zc	200	141
ف	acfs00	192	164	usfs00	119	212	ucfs00	175	208	0000fs	250	
ق	acqs00	186	161	usqs00	98	185	ucqs00	177	189	0000qs	228	
ك	acks00	190	160	usks00	103	179	ucks00	166	176	0000ks	266	
ل	acls00	156	157	usls00	79	176	ucls00	132	175	0000ls	240	167
م	acms00	164	160	usms00	88	196	ucms00	138	172	0000ms	256	167
ن	acns00	145	163	usns00	81	194	ucns00	142	189	0000ns	220	176
و	acws00	172	161	usws00	97	213	ucws00	193	182			
ي	acys00	168	151	usys00	93	190	ucys00	156	200			
هـ	achs00	172	157	ushs00	88	182	uchs00	174	176	0000hs	176	
Mean		185	156		96	191		166	176		228	157
SD		21	5		14	11		20	14		46	18
Total Duration		5171			2691			4638			5919	

Total Duration in MS is

59741

That is 59.74 seconds