١٢ ـ كيف تبنى "روبوت" حقيقى؟

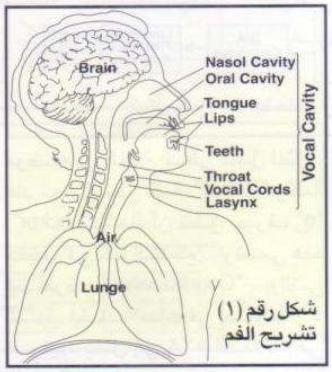
الاتصال الصوتى Voice Communication

د. علاء خميس
كلية هندسة البترول - جامعة قناة السويس

يعتبر الاتصال الصوتى من أكثر وسائل الاتصال فاعلية فى حياة الإنسان.. حيث لاتقتصر لغة الحديث على تبادل معلومات بين البشر عن أحداث الحاضر أو الماضى أو المستقبل.. ولكن اللغة المنطوقة تتضمن الكثير من المعلومات الاخرى.. مثل عمر المتحدث وجنسه بالإضافة الى حالته النفسية.

يحاول كثير من مصممى المنظومات الحاسوبية والروبوتية الآن توفير وسائل الاتصال الصوتى لجذب المستخدمين وبصفة خاصة ذوى الاحتياجات الخاصة والمعوقين مثل المكفوفين وفاقدى القدرة على التحكم فى أدوات الاتصال التقليدية مثل الماوس ولوحة المفاتيح وذلك بسبب اضطرابات حركية كمرضى الباركنسون Parkinson أو الشلل الرعاش حيث يعانى المريض من أعراض بطء الحركة بالإضافة الى التصلب أو التخشب الذى ينتج عنه فقدان الاتزان والسيطرة على الحركة. لذا.. يعتبر الاتصال الصوتى وسيلة فعالة لحل هذه المشاكل بالإضافة إلى ما يوفره من سهولة فى الاتصال وإعطاء الأوامر. يستخدم الاتصال الصوتى أيضًا فى بعض ألعاب الحاسب المصممه لمساعدة الاطفال المصابين بمشكلات لغوية.

وفى مجال الاتصال الصوتى بالحاسب.. يعتبر برنامج ViaVoice الذى أنتجته شركة IBM من أكفأ البرامج التى توفر إمكانية التفاعل مع الحاسب باستخدام الأوامر المنطوقة.. ويوجد منه إصدار باللغة العربية. وقد قام باحثو المعهد الوطنى الأميركي للمعايير والتكنولوجيا مع باحثين من مراكز مماثلة من فرنسا وألمانيا واليابان وبريطانيا.. بتطوير لغة برمجة جديدة أطلق عليها اسم "أيزو ١٨٦٢٩" لها القدرة على الاتصال الصوتى مع المشغل البشرى لاستخدامها فى التطبيقات الصناعية. وفى الحاسبات الشخصية.. يمكن أيضًا إستخدام برنامج مثل Festivalلتحويل النصوص إلى صوت.. فى حين يمكن استخدام C-voice لتحويل الصوت إلى نص.



شكل رقم (١) تشريح الفم

وتحاول شركة «ميكروسوفت» الآن تحسين إمكانيات الاتصال الصوتى فى إصدار النوافذ الجديد والذى يحمل إسمًا تجريبيًا "Longhorn" وذلك من خلال وصلات اتصال طبيعية Natural User Interfaces معتمدة على الصوت. كما تحاول كثير من الشركات التى توفر خدمات البحث على الانترنت مثل «جوجل» و«ياهوو» توفير خدمة البحث بالصوت بدلاً من الاعتماد على كتابة النص.. مما يسمح بالتفاعل مع محتويات شبكة الانترنت بطريقة طبيعية عن طريق إدماج الكلام مع الأنماط الأخرى من المدخلات والمخرجات. ويذكر.. أن Op-era browser هى اول أداة تصفح تسمح للمستخدمين بتصفح الشبكة ووضع نماذج عليها تعمل بالصوت من خلال التحدث مع أجهزة الكمبيوتر الخاصة بهم.. بالإضافة إلى قيام محتويات المواقع على الشبكة بإعادة القراءة عليهم. وتعتمد هذه الأداة على استخدام تقنية الكلام من شركة IBM التي يطلق عليها Embedded Via-Voice. وبالإضافة إلى مساعدة ذوي الاحتياجات الخاصة.. فإن مثل هذه الأدوات تسهل عملية التفاعل مع الحاسب لكافة المستخدمين. فعلى سبيل المثال.. يمكن للمستخدم القيام بمهام على الكمبيوتر أثناء قيادة السيارة من خلال الصوت والحصول على رد فعل صوتي بدلاً من النظر إلى الشاشة. ومن الاتجاهات الجديدة أيضا فى مجال الاتصال الصوتى..استخدام الصوت للتحكم فى الأجهزة المنزلية كالتليفزيون والفيديو والمكيفات.. وظهر فى الأسواق أجهزة «ريموت كنترول» جديدة تعمل بالصوت مثل InVoca و Hands-Free Accenda Voice و Promptu .

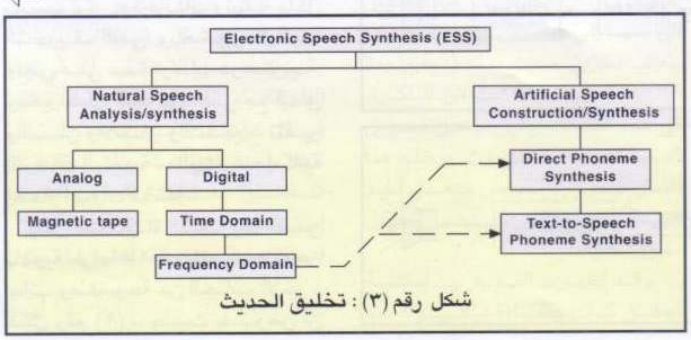
ويستطيع Accenda مثلاً التعرف على أكثر من ٥٠ أمر لتشغيل أو إيقاف التليفزيون أو الفيديو بالإضافة الى وظائف أخرى مثل تغيير محطة التليفزيون أو رفع الصوت أو خفضه أو بدء التسجيل على الفيديو إلخ.. . ويجدر بالذكر.. أن المشكلة الأساسية التى تواجه هذه الأجهزة هى كيفية جعل جهاز الريموت كنترول قادرًا على التفريق بين صوت المتحدث والصوت الصادر من التليفزيون.. وهو ما يتطلب تصميم دوائر خاصة لإزالة الضوضاء.

وفى المنظومات الروبوتية.. يمثل الاتصال الصوتى تحديًا كبيرًا على الرغم من الجهود التى يبذلها الباحثون لحل مشكلة فهم اللغات الحية لتحقيق الاتصال الطبيعي بين الروبوت والإنسان. وحتى إذا ما تم حل مشكلة الفهم الكامل للكلام.. فسوف تستمر مشكلتا التخمين والحدس اللتان يتمكن بهما الإنسان من تخمين بعض الكلمات إذا أساء نطقها المتحدث أوتخمين الغرض من الحديث أو الأمر المطلوب تنفيذه. وحاليًا.. يتم تبسيط المشكلة من الفهم الكامل للغة إلى محاولة فهم بعض الأوامر الصوتية المختزنة مسبقًا فى قائمة Voice Tags (أوامر مثل.. تحرك للأمام _ اذهب إلى غرفة المكتب _ توقف إلخ..) وهذا شبيه بما يحدث فى منظومات الهواتف المحمولة والخدمات الصوتية. ويتم ذلك بتحويل الصوت إلى نص يمكن استخدامه كمدخل لبرنامج التحكم فى حركة الروبوت. وإذا تطلب الأمر توفير استجابة صوتية من الروبوت للإنسان.. يتم استخدام عملية تسمى تخليق الكلامSpeech Synthesis.

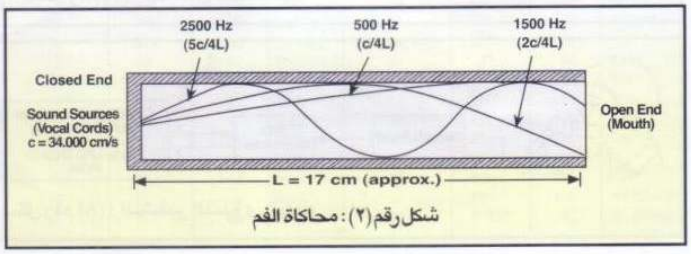
وتتضمن عملية الاتصال الصوتى عمليتين أساسيتين هما.. تخليق الكلام Speech Synthesisوفهم الكلام Speech Recognition. ويعتبر التخليق أسهل بكثير من فهم الكلام.. فكما هو معروف فى معظم اللغات.. يمكن أن يتغير معنى الكلمة حسب موقعها فى الجملة.. فإذا تخيلنا روبوت مبرمجًا على فهم اللغة الانجليزية جرت مخاطبته بالعبارة "Time Flies Like an Arrow".. فسنلاحظ هنا أن الكلمات الاولى "Time" و "Fly" و "Like" فى هذه العبارة يصلح كل منها ليكون فعلاً للجملة. وعلى الرغم من قدرة الإنسان على فهم معنى هذه الجملة بسهولة.. إلا

جدول رقم (١): تقسيمات الصوت

Voiced Sounds		Unvoiced Sounds	
Pure	a, e, i, o, u uh, aa, ee, er, uu, ar, aw,	Fricative	s, sh, f, th
Fricative	z, zh, v, dh	Polsive (stop)	p, t, k, h
Polsive (stop)	b, d, g	Affricate	ch
Nasal	m, n, ng	Aspirate	h
Affricate	j		
Glides	R, w, l, y		



شكل رقم (٣) : تخليق الحديث



شكل رقم (٢) : محاكاة الفم

أن منظومة فهم الكلام المزود fها الروبوت يمكن أن تفهم هذه العبارة بثلاث طرق مختلفة هى "الوقت يطير مثل السهم" و "ذبابات الوقت تحب السهم" و "وقت للذبابات مثل السهم". وهناك مشكلة أخرى.. تتمثل فى قدرة الروبوت على الفهم الدقيق للأوامر التي يصدرها المشغل البشرى بصورة مباشرة. فمثلاً.. ثمة اختلاف بين أمر "اغسل الزجاجات قبل تعبئتها" وأمر "أدر جهاز التكييف قبل تعبئة الزجاجات".. لا تستطيع المنظومة الحاسوبية للروبوت أن تدركه. فكلمة "قبل" في الأمر الأول.. تعني أن غسل الزجاجات يجب أن يُنجز وينتهي قبل البدء فى عملية التعبئة. أما "قبل" في الأمر الثاني.. فتعني أن تشغيل جهاز التكييف يسبق زمنيًا تعبئة الزجاجات.. ولكنه سيظل مستمرًا خلال عملية التعبئة. فى هذه الدراسة.. يتم شرح تقنيات تخليق الكلام على أن يتم شرح عملية فهم الكلام فى مقال تال بمشيئة الله.

مفاهيم أساسية

قبل البدء فى شرح كيفية عمل منظومات تخليق الكلام الالكترونية.. لابد من فهم الآلية التى تتم بها هذه العملية فى الإنسان.. حيث تعتبر المنظومات الاصطناعية محاكاة غير كاملة لآلية الكلام التى خلقها الله عز وجل فى الإنسان.

يظهر الشكل رقم (١) تشريح فم الإنسان حيث يتكون من ثلاث وحدات أساسية.. هي اللسان والحنجرة والتجويف الفمى.

تخيل الآن أنك فى عيادة طبيب حيث يطلب منك الطبيب أن تنطق كلمة "aaaah". لنطق هذه الكلمة أو لإحداث هذا الصوت.. تبدأ الرئتان فى توفير القدرة اللازمة للمنظومة بدفع الهواء خلال الحنجرة إلى التجويف الفمى.. ويتولد الصوت بواسطة الأحبال الصوتية الموجودة فى الحنجرة خلف تفاحة آدم والتي تتكون من طبقات جلدية تهتز عند مرور الهواء بها مما يسبب تولد عدة ترددات رنينية داخل التجويف الفمى. ويحتوى كل تردد رنينى على عدة ترددات هرمونية. وبتغير شكل التجويف الفمى مع الحلق واللسان والأسنان والشفتين.. تتغير الترددات الرنينية مما ينتج عنه إمكانية إصدار أصوات مختلفة.

يمكن محاكاة التجويف الفمى بأنبوبة طولها ١٧ سم تقريبًا مغلقة من جانب ومفتوحة من الجانب الآخر - شكل رقم (٢) - حيث يفترض أن الاحبال الصوتية هى الجانب المغلق بينما يمثل الجانب المفتوح الشفتين. فى هذه الحالة.. تحدث حالة الرنين عند الترددات الفردية التالي 1/4(c/L), 3/4 (c/L), 5/4(c/L), 7/4(c/L). حيث c سرعة الصوت (٣٤٠٠٠سم/ث) و L طول الانبوبة (١٧ سم).. وبالتالى تكون ترددات الرنين هى ٥٠٠ و ١٥٠٠ و ٢٥٠٠ و ٣٥٠٠ على التوالي. وبتقريب مثال الأنبوبة بحالة التجويف الفمى.. نجد أن هذه الترددات هى الترددات الرنينية التى تجعلك تستطيع إصدار الصوت "aaaah".

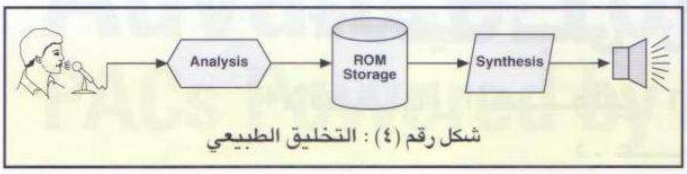
يمكن تقسيم الاصوات التى يصدرها الإنسان إلى قسمين أساسيين هما.. أصوات مصدرها اهتزاز الأحبال الصوتية تسمى Voiced Sounds مثل أصوات الحروف المتحركة فى اللغة الانجليزية (a, e, i, o, u).. وأصوات تنتج من جعل الأحبال الصوتية مفتوحة مع دفع الهواء خلال التجويف الفمى أو الانفى وتسمى Unvoiced Sound مثل أصوات الحروف الساكنة (s, f, p, t). ويتم تقسيم هذين النوعين الى أنواع أخرى كما هو مبين بالجدول رقم (١).

تعتمد عملية تخليق الكلام Speech Synthesis على تجميع عناصر صوتية تخليقية يطلق عليها المصوتات «الفُونيمات» Phonemes.. وهى أبسط وحدات اللغة التى عادة ما تتكون من حرف أو حرفين _ جدول رقم (١)- فى حالة اللغة الانجليزية والتى تحتوى على ٢٦ حرفاً و ٤٠ «فُونيم». وتسمى عملية ربط «الفُونيمات» لتكوين كلمة ما "سلسلة" أو Phoneme Concatenation وهى عملية غير بسيطة بسبب اختلاف نطق كل «فُونيم» حسب موضعه فى الكلمة. فعلى سبيل المثال.. عند محاولة نطق كلمة "Robot" وكلمة "Locker"نلاحظ أن نطق حرف "o" يختلف فى كلتا الكلمتين.. وتسمى هذه الظاهرة "Coarticulation" والتى تسبب اختلافات فى أصوات ٤٠ «فُونيم» يصل إلى ١٢٨ اختلافًا فى اللغة الانجليزية. تسمى هذه الاختلافات "تباينات النطق" أو Allo-phonesوالتى يتم استخدامها فى عملية تخليق الكلام بدلاً من «الفُونيمات» للتغلب على مشكلة اختلاف نطق كل «فُونيم» حسب موضعه فى الكلمة. عندَ تكوين جملة من مجموعة من الكلمات.. تظهر مشكلة أخرى تتمثل فى ضرورة اختلاف نطق الكلمة أو نبرة الصوت حسب نوع الجملة من حيث كونها استفهامية أو خبرية. فمثلاً.. يجب أن يتغير نطق كلمة "Right" فى حالة إستخدامها للإستفهام مثل "Right?" أو للتعبير عن الدهشة مثل "Right!". ويتم حل هذه المشكلة بالتحكم فى نطق الكلمة الموجودة فى نهاية الجملة بتغيير درجة الصوت Pitch والتوقيت Timing أو الإيقاع Rhythm. يمكن تغيير درجة الصوت بتغيير التردد.. مثلما يحدث بزيادة التردد عند نهاية الجملة الاستفهامية. ويسمى القالب Pattern الذى يتم فيه تغير الدرجة بالترنيم In-tonation. و يتم ضبط التوقيت أو إيقاع الجملة بإدراج وقفات Pauses بين تباينات النطق. فعلى سبيل المثال.. يساعد ضبط الإيقاع على تغير نطق كلمة "Approximate" فى الجملتين "The Approximate Value is .." و "To Approximate the Value, you must.." حيث تمثل فى الجملة الأولى صفة و فى الثانية فعلاً. وفى حالة عدم استخدام التوقيت.. نجد أن الصوت المتولد اصطناعياً يكون غير طبيعى.

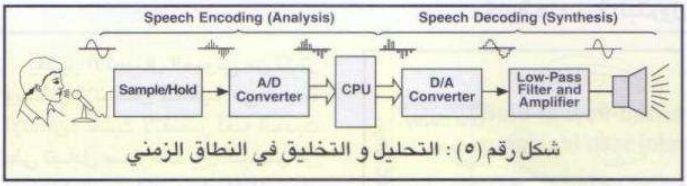
تخليق الكلام الطبيعى

كما هو موضح بالشكل رقم (٣).. هناك طريقتان لتخليق الكلام هما.. طريقة تحليل وتخليق الكلام الطبيعى Speech/Natural Speech Analysis.. وطريقة بناء أو تخليق الكلام الاصطناعى Artificial Construc-tion/ Synthesis.

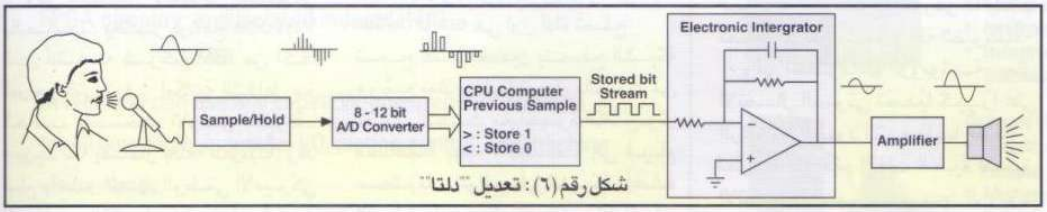
تسمى الطريقة الأولى بتقنية "التخليق الطبيعى".. لأنها تتضمن



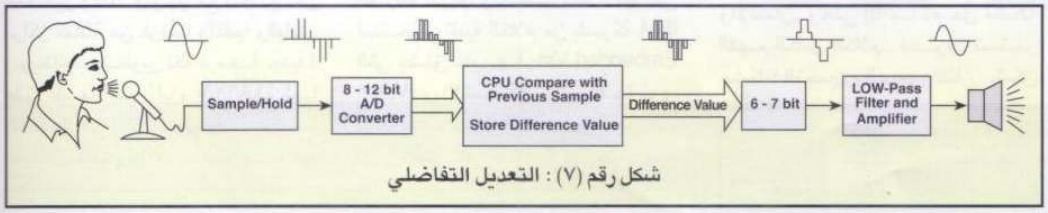
شكل رقم (٤) : التخليق الطبيعي



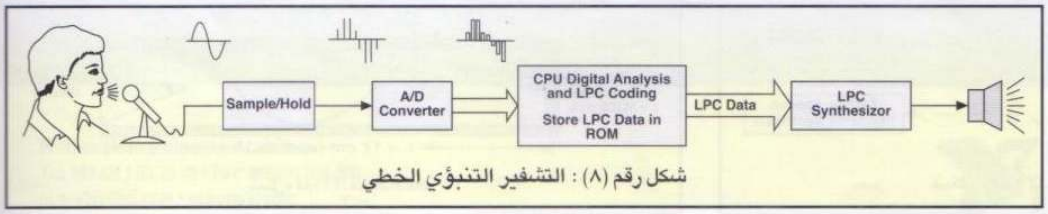
شكل رقم (٥) : التحليل و التخليق في النطاق الزمني



شكل رقم (٦) : تعديل "دلتا"



شكل رقم (٧) : التعديل التفاضلي

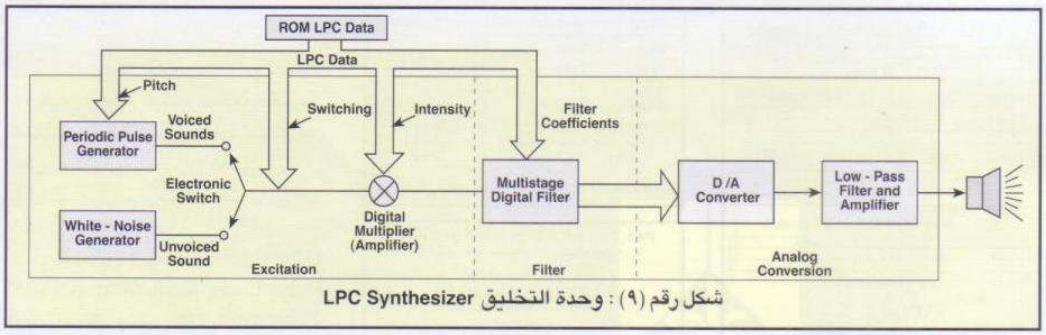


شكل رقم (٨) : التشفير التنبؤي الخطي

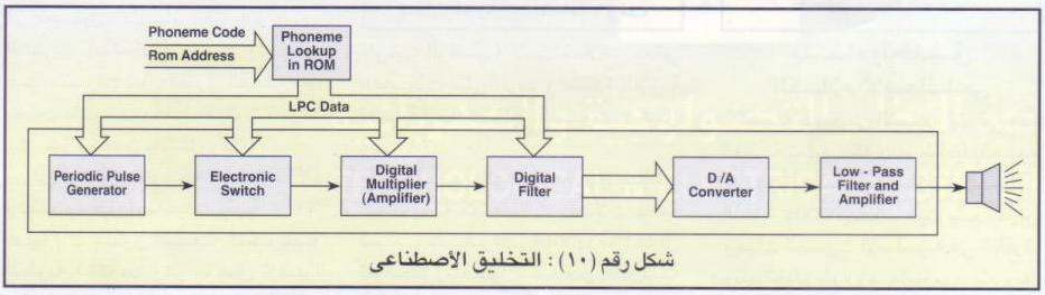
تخزين مقاطع صوتية لأصوات بشرية مسجلة يمكن إعادة استدعائها بواسطة برنامج حاسوبى لتخليق الكلام. يمكن أن تتم عملية التخزين فى صورة تناظرية على شرائط مغناطيسية أو فى صورة رقمية فى ذاكرة الحاسب. وتنتج عملية التخزين وإعادة العرض التناظرى صوتاً شبه طبيعى ولكنها تعتبر غير عملية بسبب اقتصارها على استخدام رسائل مسجلة متسلسلة.. بالإضافة إلى استحالة تسجيل كل الكلمات والجمل التى تتطلبها عملية تخليق الكلام فى المنظومات الروبوتية.. ولكنها تعتبر أفضل اختيار عندما يكون المطلوب إنتاج لغة تخاطب محدودة كما هو الحال فى بعض المعدات او الخدمات الهاتفية أو السيارات أو فى المطارات.

تتضمن طريقة التخزين وإعادة العرض الرقمى مرحلتين أساسيتين هما.. التحليل Analysis والتخليق Synthesis – شكل رقم (٤).

فى مرحلة التحليل.. يتم تحليل الكلام البشرى وتشفيره فى صورة رقمية وتخزينه. وفى مرحلة التخليق يتم استرجاع الكلام الرقمى من الذاكرة وتحويله إلى صورة رقمية لإعادة تخليق الكلام الأصلى. توفر هذه التقنية آلية أكثر مرونة فى تخليق الكلام.. حيث لا تتقيد باستخدام كلمات أو رسائل متسلسلة أو تتابعية كما هو الحال فى الطريقة التناظرية.. بما يتيح استرجاع أى كلمة مختزنة فى الذاكرة. وتتوقف عدد الكلمات على سعة الذاكرة المستخدمه.



شكل رقم (٩) : وحدة التخليق LPC Synthesizer



شكل رقم (١٠) : التخليق الأصطناعى

هناك طرق مختلفة لتشفير الكلام.. مثل طريقة التحليل والتخليق فى النطاق الزمنى Time-Domain Analysis / Synthesis.. التحليل والتخليق فى النطاق الترددى Frequency-Domain Analysis / Synthesis.

– التحليل والتخليق فى النطاق الزمنى:

فى هذه الطريقة.. يتم تحويل التغير فى مقدار موجة الصوت التناظرية Amplitude إلى شفرة رقمية باستخدام محولات تناظرية/رقمية (D/A) فى مرحلة التحليل.. ويتم استخدام محولات رقمية/تناظرية (A/D) لإتمام عملية إعادة التخليق كما هو الحال فى منظومات الهواتف – شكل رقم (٥).

تتطلب هذه التقنية قدرة تخزينية عالية لتخزين المفردات اللغوية اللازمة لتشكيل الكلمات. لذا.. ابتكرت طرق عديدة لتقليل الذاكرة المطلوبة لتخزين الصوت المشفر مثل.. طريقة تعديل الشفرة النبضية البسيطة Simple Pulse-code Modulation (PCM) وتعديل دلتا Delta Modulation.. والتعديل التفاضلى Differentail Pulse-code Modulation (DPCM)..

فى الطريقة الأولى.. يتم التحكم فى كفاءة الصوت المشفر باستخدام معدل أخذ العينات Sampling Rate ودقة المحول التناظرى/الرقمى Resolution. وتزداد كفاءة الصوت المخلق بزيادة معدل أخذ العينات.. ولكن على حساب زيادة سعة الذاكرة المطلوبة لإتمام هذه العملية. وكما هو معروف.. يقع الصوت البشرى فى مدى ترددى كبير يتراوح بين ١٠ – ١٥٠٠٠ هرتز. ولضمان تخزين الصوت بصورة كاملة.. يجب استخدام محول تناظرى/رقمى بمعدل أخذ عينات حوالى ٣٠٠٠٠ عينة/ث. يتم تحويل كل منها إلى شفرة ثنائية ٨ بت. لذا يتطلب تخزين حديث مدته ثانية واحدة ٨×٣٠٠٠ أى ٢٤٠٠٠بت من الذاكرة. فى هذه الحالة.. يكون معدل تحويل البيانات ٢٤٠٠٠بت /ث. يوضح هذا المثال ضرورة إنقاص معدل أخذ العينات للحصول على معدل تحويل مقبول عمليًا.

وقد أظهرت التجارب.. إمكانية تخليق كلام بكفاءة مقبولة عند استخدام معدل أخذ عينات يساوى على الأقل ضعف أعلى تردد للمركبات التى تشكل الموجة الصوتية. وقد وجد.. أن مركبات الكلام البشرى عادة ما تقع بين ٣٠ – ٣٠٠٠ هرتز. لذا.. يمكن تخليق كلام بكفاءة مقبولة عند معدل أخذ عينات ٦٠٠٠ عينة/ث.

وكما ذكرنا.. فإن المعامل الآخر الذى يحدد كفاءة الكلام المخلق هو دقة المحول التناظرى/الرقمى.. حيث يمكن الحصول على صوت مقبول فى حالة استخدام محول ٨ بت. لذا.. يكون معدل التحويل ٤٨٠٠٠ بت/ث فى حالة ضبط معدل أخذ العينات ليكون ٦٠٠٠ عينة/ث. نتيجة لذلك.. يتطلب تخزين ١٠ ثوان من الحديث.. سعة تخزينية مقدارها ٦٠٠٠٠ بايت (تقريبًا ٦٤ ك بايت). ويمكن زيادة كفاءة الصوت باستخدام محول ١٢ بت.. ولكن فى هذه الحالة يجب توفير ٩٦ ك بت من الذاكرة لتخزين ١٠ ثوان.

فى طريقة تعديل «دلتا».. يتم تخزين شفرة واحدة لكل عينة بدلاً من ٨ أو ١٢ كما هو الحال فى الطريقة الأولى.. ويتم مقارنة قيمة كل عينة جديدة بالقيمة القديمة. فإذا كانت القيمة الجديدة أكبر من آخر قيمة.. يتم تخزين "١".. وفى حالة العكس يتم تخزين القيمة "٠" – شكل رقم (٦).

وبالأخذ فى الاعتبار.. أنه فى هذه الطريقة يتم تخزين قيمة واحدة فقط فى كل عينة.. نجد أنه يجب زيادة معدل أخذ العينات لضمان التقاط كل تفاصيل الحديث. ولأنه من الشائع استخدام معدل ٣٢٠٠٠ عينة/ ث.. فإن معدل التحويل يكون

جدول رقم (٢) : خريطة "الفونيمات" وتباينات النطق

Hexadecimal Phoneme Code	Phoneme Symbol	Duration (ms)	Example Word	Hexadecimal Phoneme Code	Phoneme Symbol	Duration (ms)	Example Word
00	EH3	59	Jacket	20	A	185	Day
01	EH2	71	Enlist	21	AY	65	Day
02	EH1	121	Heavy	22	Y1	80	Yard
03	PA0	47	No Sound	23	UH3	47	Mission
04	DT	47	Butter	24	AH	250	Mop
05	A2	71	Made	25	P	103	Past
06	A1	103	Made	26	O	185	Cold
07	ZH	90	Azure	27	I	185	Pin
08	AH2	71	Honest	28	U	185	Move
09	13	55	Inhibit	29	Y	103	Any
0A	12	80	Inhibit	2A	T	71	Tap
0B	I1	121	Inhibit	2B	R	90	Red
0C	M	103	Mat	2C	E	185	Meet
0D	N	80	Sun	2D	W	80	Win
0E	B	71	Bag	2E	AE	185	Dad
0F	V	71	Van	2F	AE1	103	After
10	CH*	71	Chip	30	AW2	90	Salty
11	SH	121	Shop	31	UH2	71	About
12	Z	71	Zoo	32	UH1	103	Uncle
13	AW1	146	Lawful	33	UH	185	Cup
14	NG	121	Thing	34	O2	80	For
15	AH1	146	Father	35	O1	121	Aboard
16	001	103	Looking	36	IU	59	You
17	00	185	Book	37	U1	90	You
18	L	103	Land	38	THV	80	The
19	K	80	Trick	39	TH	71	Thin
1A	J*	47	Judge	3A	ER	146	Bird
1B	H	71	Hello	3B	EH	185	Get
1C	G	71	Get	3C	E1	121	Be
1D	F	103	Fast	3D	AW	250	Call
1E	D	55	Paid	3E	PA1	185	No Sound
1F	S	90	Pass	3F	STOP	47	No Sound

٣٢٠٠٠بت/ث. ولتخزين حديث مدته ١٠ ثوان.. يجب توفير سعة تخزين مقدارها ٣٩ك بت. بالمقارنة مع الطريقة الأولى.. نجد أن تعديل "دلتا" يقلل السعة التخزينية المطلوبة بواقع ٤٠٪.

فى التعديل التفاضلى Differentail Pulse-Code Modulation (DPCM) .. يتم استخدام نفس الفكرة المستخدمة فى تعديل "دلتا" مع زيادة عدد الشفرات المستخدمة من بت واحد إلى عدة شفرات تمثل الفرق الفعلى بين عينتين متتابعتين وغالبًا ما يكون ٦ أو ٧ بت. كما هو مبين بالشكل رقم (٧).

فى حالة منظومة DPCM-7 bit يستخدم معدل أخذ عينات بقيمة ٦٠٠٠ عينة/ ث وتكون السعة التخزينية المطلوبة ٥١ك بت وهو ما يمثل توفيرًا بنسبة ١٢٪ بالمقارنة بالطريقة الأولى (PCM-8 bit) و ٢٥٪ بالمقارنة بالطريقة الثانية (DPCM-6 bit) .

- التحليل والتخليق فى النطاق الترددى:

فى هذه الطريقة.. يتم تشفير الطيف الترددى للموجة التناظرية بدلاً من تغير المقدار كما هو الحال فى الطريقة الأولى.. ولا يتم تشفير الموجة التناظرية بشكل مباشر باستخدام محول تناظرى/رقمى.. ولكن يتم استخدام نموذج رياضى للطيف الترددى للتحكم فى النموذج الالكترونى للمنطقة الفمية للانسان Vocal Tract. فى مرحلة التحليل.. يتم تحليل الخواص الترددية للصوت البشرى للحصول على سلسلة من المعاملات الرياضية يتم تخزينها لاستخدامها فى عملية التحكم فى وحدة التخليق الالكترونية. تحاكى هذه الوحدة المنطقة الفمية للإنسان باستخدام مولدات تردد ومرشحات يتم ضبطها بواسطة المعاملات المستخلصة من مرحلة التحليل. لإتمام عملية تحليل وتخليق الكلام فى النطاق الترددى.. يمكن استخدام طريقة التشفير التنبؤى الخطى Linear Predictive Coding (LPC) والتى تتميز به من القدرة على إنتاج صوت شبه طبيعى. وقد تم إقتراح هذه الطريقة بواسطة شركة "Texas Instruments" فى لعبة تعليمية تسمى "Speak & Spell" . تعتمد طريقة LPC فى عملها على تشفير موجة الصوت بواسطة محول تناظرى/رقمى باستخدام طريقة تعديل الشفرة النبضية البسيطة PCM المشروحه سالفًا. يتم بعد ذلك تحليل الموجة الرقمية لاستخلاص بعض المعاملات مثل التردد والشدة اللازمة لعملية إعادة تخليق الصوت - شكل رقم (٨).

يتم بعد ذلك.. وضع البيانات المستخلصة كمعاملات لمعادلات خطية تسمى شفرات LPC Codes تمثل نموذجًا رياضيًا للخواص الترددية لموجة الصوت المنطوق. ويتم استخدام تلك المعاملات المستخلصة من مرحلة التحليل.. فى التحكم فى وحدة التخليق الالكترونية والتى تمثل محاكاة أو نمذجة للمنطقة الفمية فى الانسان - شكل رقم (٩). وتتكون هذه الوحدة من مصدر إثارة Excitation Source .. ومرشح رقمى متعدد المراحل.. بالإضافة إلى محول رقمى/تناظرى. يحتوى مصدر الإثارة على مولد نبضات دورية يقوم بمحاكاة عمل الأحبال الصوتية فى الانسان عن طريق توليد نبضات دورية بترددات أصوات حروف متحركة Voiced Sound. وتحدد قيمة التردد درجة أو حدة الصوت Pitch المتولد كما هو الحال فى الانسان.. حيث تتحدد درجة الصوت حسب درجة اهتزاز الأحبال الصوتية. ولتوليد أصوات حروف ساكنة Unvoiced Sounds يتم استخدام مولد الضوضاء الموضح فى شكل رقم (٩). وباستخدام المفتاح الالكترونى.. يمكن اختيار توليد أصوات حروف متحركة أو ساكنة.. ويتم تكبير الكلام المتولد بواسطة مكبر ثم يتم تمريره على وحدة الترشيح لضبط الصوت بتغيير قيم المعاملات المستخلصة من عملية التحليل لإنتاج صوت شبه طبيعى. بعد عملية الترشيح.. يتم تحويل الموجة الرقمية إلى صورة تناظرية بواسطة المحول الرقمى/التناظرى.

Word	Pronunciation	Phoneme Symbols
This	This	TH,I2,S
Text	Tekst	T, EH1, K, S, T
is	iz	I, Z
Great	Grāt	G, R, A1, T

TH,I2,S, T,EH1,K,S,T, I,Z, G,R,AI,T
This Text is Great

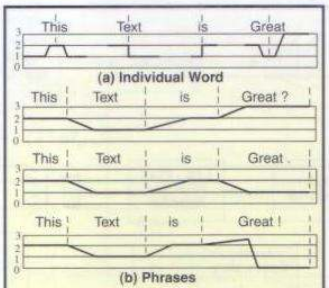
شكل رقم (١١) : الشفرة الرمزية

PA1,TH,12,S,PA1,T,EH1,K,S,T,PA1,
This Text
I,Z,PA1,G,R,A,T,PA1
is Great

شكل رقم (١٢) : إضافة الوقف

PA1, 1/TH, 2/12, 1/S, PA1, 2/T, 1/EH1, 1/K, 1/S, 1/T,
This Text
PA1, 1/I, 2/Z, PA1, 2/G, 1/R, 3/A, 3/T, PA1
is Great?

شكل رقم (١٤) : إضافة الترنيم



شكل رقم (١٣)
تنسيق درجة الصوت

بناء وتخليق الكلام الاصطناعى

كما يفهم من الاسم.. يتم فى هذه العملية تخليق الكلام اصطناعياً بتجميع «فونيمات» أو "تباينات النطق" Allophones.. حيث يتم تخزين وحدات الصوت الأساسية فى الذاكرة. وباستخدام خوارزم حاسوبى.. يتم ربط هذه الوحدات لتكوين كلمات يتم ربطها لتكوين جمل.. مع مراعاة خواص الصوت من ترنيم وإيقاع كما سبق شرحه.

لإتمام هذه العملية.. يتم تشفير «الفونيمات» وتباينات النطق باستخدام طريقة تعديل الشفرة النبضية البسيط LPC المشروحة سابقاً.. ثم يتم تخزين هذه الشفرات فى جدول بحث فى وحدة ذاكرة ROM. يوضح الشكل رقم (١٠).. المكونات الأساسية لوحدة تخليق كلام اصطناعى تتكون من ذاكرة بحث و مصدر إثارة ومرشح متعدد المراحل. عند إدخال شفرة «فونيم» Phoneme Code على وحدة التخليق.. تقوم وحدة الذاكرة بترجمة هذه الشفرة إلى مجموعة من معاملات التعديل LPC يتم استخدامها للتحكم فى مصدر الإثارة والمرشح كما هو الحال فى وحدة تخليق الحديث الطبيعى المشروحة سابقًا.

يمكن استخدام هذه الطريقة فى تخليق الكلام.. أو لتحويل النص المكتوب إلى صوت مسموع Text-to-Speech كما هو الحال فى بعض برامج الترجمة وتعلم اللغات. عند تخليق الكلام.. يتم إعداد برنامج حاسوبى لتوليد عدد من شفرات «الفونيمات» الواجب إرسالها إلى وحدة التخليق لتكوين كلمة ما أو جملة ما. تسمى هذه الشفرات سلسلة «الفونيمات» أو Phoneme String.. يتم تخزينها فى وحدة ذاكرة RAM أو ROM كجزء من البرنامج الفرعى المسئول عن عملية التخليق. يوضح الجدول رقم (٢).. استخدام منظومة التشفير الحاسوبية السداسية عشر فى بناء هذه النماذج الصوتية المنطوقة فى برنامج VOTRAX لتخليق الكلام الاصطناعى.

فى هذه القائمة.. يوجد ٦٤ من «الفونيمات» وتباينات النطق يتم استخدامها بواسطة برنامج فرعى لتشكيل سلسلة «الفونيمات» Phoneme String كالتالى:

١- يتم تحديد السلسلة الرمزية «للفونيمات» Phoneme Symbol String اللازمة لنطق كلمة معينة داخل الجملة. على سبيل المثال فى الجملة "This Text is Great!?".. يتم البحث فى القاموس عن النطق الصحيح لكل من كلمات هذه الجملة لاستخدام هذا النطق كدليل بحث فى الجدول رقم (٢).. وبالتالى يمكن تكوين الشفرة الرمزية للجملة كما هو مبين بالشكل رقم (١٤) باستخدام قاموس "Webster's New World" للغة الانجليزية.

٢- إضافة الوقف Pauses بين المقاطع والكلمات حسب الإيقاع المطلوب. وطبقًا للجدول رقم (٢).. هناك نوعان من الوقف حسب المدة الزمنية PA0 (47-ms) و PA1 (185-ms).. يتم إضافة الوقف القصير بين مقاطع بعض الكلمات بينما يتم إدراج الوقف الطويل بين الكلمات.. يوضح الشكل رقم (١٢) الشفرة الرمزية «للفونيمات» بعد إضافة الوقف.

٣- تحديد ترنيم Intonation كل كلمة على حدة.. وترنيم الجملة بالكامل. يعتمد الترنيم على تغير درجة الصوت Pitch أو نبرة الكلام حسب موقع الكلمة ونوع الجملة. لتحديد الترنيم.. يتم تكوين مخطط بيانى لتغير درجة الصوت - شكل رقم (١٣). يتم إضافة الترنيم المستنتج من الرسم البيانى إلى الشفرة الرمزية «للفونيمات» كما هو مبين بالشكل رقم (١٤).

٤- تحويل السلسلة الرمزية للمصوتات إلى سلسلة شفرات «للفونيمات» باستخدام منظومة التشفير الحاسوبية السداسية عشر الموضحة بالجدول رقم (٢). يوضح الشكل رقم (١٥) سلسلة الشفرات الواجب إستخدامها لنطق الجملة.

٥- استخدام سلسلة الشفرات كمدخل لوحدة التخليق الاصطناعى واستماع الجملة المنطوقة و تعديل الشفرة إذا لزم الامر.

This	PA1	= 00 111110= 3E
	1/TH	= 01 111001= 79
	2/12	= 10 001010= 8A
	1/S	= 01 011111= 5F
	PA1	= 00111110= 3E
Text	2/T	= 10 101010= AA
	1/EH1	= 01 000010= 42
	1/K	= 01 011001= 59
	1/S	= 01 011111= 5F
	1/T	= 01 101010= 6A
	PA1	= 00 111110= 3E
is	1/I	= 01 100111= 67
	2/Z	= 10 010010= 92
	PA1	= 00 111110= 3E
Great ?	2/G	= 10 011100= 9C
	1/R	= 01 011011= 5B
	3/A	= 11 100000= EO
	3/T	= 11 101010= EA
	PA1	= 00 111110= 3E

شكل رقم (١٥) : التشفير