بسم الله الرحمن الرحيم

التعرف على الصوت

إعداد:

أ.زاهر

م.نزار محمد المعتصم

تم تحميل هذا الكتاب من موقع كتب
www.kutub.info
للزيد من الكتب في جميع مجالات التقنية ، تفضلوا بزيارتنا

المنارة للاستشارات
www.manaraa.com

بسم الله الرحمن الرحيم

شهدت التسعينيات أول تعامل مع الأنظمة المنطوقة بعد ما بات في وقت من الأوقات حلم مبرمجي الحاسوب الوصول إلى الطريقة التي تمكنهم من التحدث مع الحاسوب والتخاطب معه , هذه الطريقة التي أصبحت تستخدم لمساعدة من يواجهون صعوبة في استخدام لوحة المفاتيح أو الماوس ليتعاملوا مع الحاسوب , وهي مع ذلك وسيلة للرفاهية والراحة لدى الإنسان مما علمه الله لبني البشر...

كان هذا العلم هو التعرف على الكلام الذي صار في بعض الدول المتقدمة وسيلة لإنجاز امتحانات الطلاب وحقق كثير من الأمن بعد ما استخدمته الجهات الأمنية.....

وأيضا التعرف على الكلام هو أحد الخصائص التي أتت بها ميكروسوفت أوفيس XP والتي جعلت من الممكن للمستخدمين إدخال النصوص وتنفيذ الأوامر والتحكم في النوافذ عن طريق التحدث في المايكروفون .....

بداية حديثنا سيكون عن فروع علم الحاسب الآلي كالتالي :

- 1-علوم الحاسب الآلي(Computer Science **CS**)
- 2-نظم المعلومات(Information System **IS**)

وعلوم الحاسب الآلي **(CS)**بحد ذاتها يندرج تحتها عدة فروع نذكر منها:

- معالجة الصور (Image Processing)
- معالجة أو تشغيل الكلام(Speech Processing)

والذي نريد التوصل إليه أن **(Speech Processing)** معالجة الكلام هو أحد الأفرع الرئيسية لعلوم الحاسب الآلي

والحاسب الآلي لكي يقوم بعملية معالجة الأصوات (الكلام بالأخص) يجب في البداية أن يقوم بالتعرف علي هذه الأصوات وهو ما يسمى ب **(speech recognition)**التعرف على الكلام وعليه سيكون محور حديثنا في هذه السلسلة

كلمة: **(Speech)** تعني الكلام

لكن : **(voice)** تعني الأصوات عموما ليس فقط الأصوات البشرية بل تتعدى ذلك إلى أصوات العصافير وأصوات الأجهزة وغير ذلك.

مثلا لدينا مصنع به العديد من الأجهزة فيهم جهاز جديد قمنا بتسجيل صوته عند أول تشغيل له وبذلك قد أخذنا بصمة له وهو جديد وبعد عمل هذا الجهاز **1000** ساعة مثلا عندها يحتاج الجهاز إلى صيانة نسجل صوته مرة أخرى ونأخذ بصمة أخرى له.

تبقى هذه البصمات الصوتية محفوظة لدينا.

المنارة للاستشارات
www.manaraa.com

عندما يصادفنا جهاز في المصنع به عطل نأخذ منه بصمة صوتية (أي تسجيل صوتي لهذا الجهاز وهو يعمل) ونقارنها مع البصمات الصوتية المحفوظة لدينا مسبقا وعندها نقرر هل هذا الجهاز جديد أم أنه يحتاج إلى صيانة.

ولذلك ما نعنيه في هذه السلسلة التعرف على الكلام **(speech recognition)** هو تعرف الحاسوب على صوت الإنسان .

نظام التعرف على الكلام: **(speech recognition system)**

هذا النظام ينجز 3 مهام أساسية منذ دخول الصوت إلى الحاسوب ومن ثم معالجته وحتى الاستفادة منه وهي:

1- **Pre-processing**

2- **Recognition**

3-**Communication**

Pre-processing

وهي العمليات التي تسبق المعالجة الأساسية للصوت وهي عبارة تحويل الصوت المدخل إلى الحاسوب إلى شكل يستطيع الـــ (Recognizer)التعامل معه

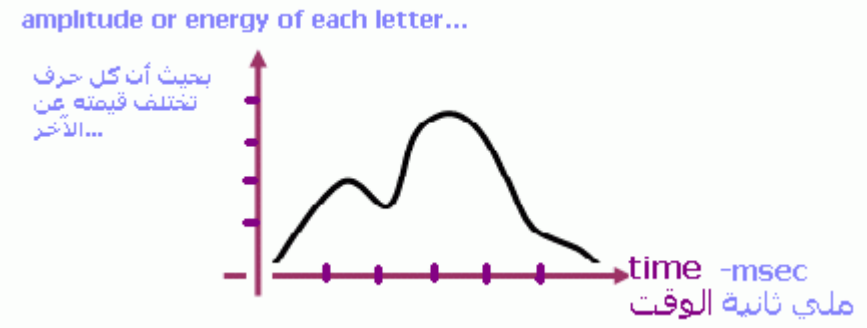
وال (Recognizer) متمثل في جزأين (s/w & h/w)

الـــ s/w وهي البرامج التي تتعامل مع الإشارات الصوتية والخاصة بالتعرف على الصوتيات وال h/w هي الحاسوب نفسه

وبالتالي لكي يقوم الحاسوب بمعالجة الإشارات الصوتية والتعامل معها لا بد أن تتم عملية

الـــ **Pre-processing** التي تتلخص في تحويل الإشارات الصوتية المدخلة إلى شكل ثنائي.

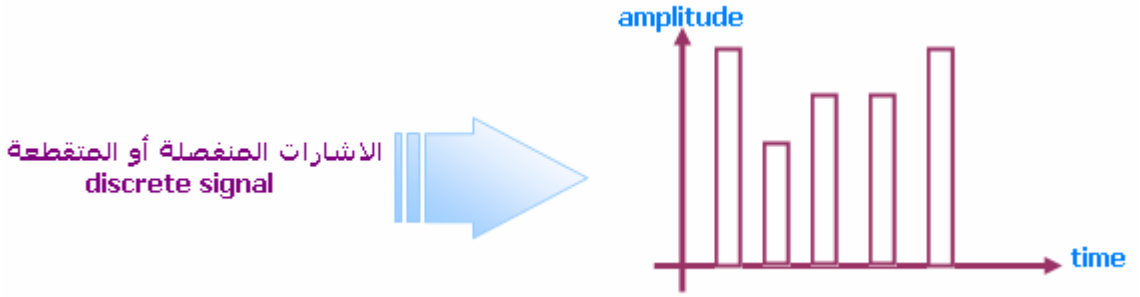
لنبدأ العملية من البداية س من الناس قام بتسجيل صوته عن طريق المايكروفون للحاسب مجموعة الكلمات التي قام بتسجيلها هي عبارة سلسلة من الإشارات المتصلة والتي تعرف ب **Analog signal** وممكن تسميتها ب **Continuous Signal** والتي تمثل بالشكل التالي-:



المنارة للاستشارات
www.manaraa.com

ومعنى أنها إشارات متصلة أي في كل لحظة زمنية توجد قيمة للصوت حتى لو توقف الشخص عن التحدث فإنه توجد إشارات بسيطة.

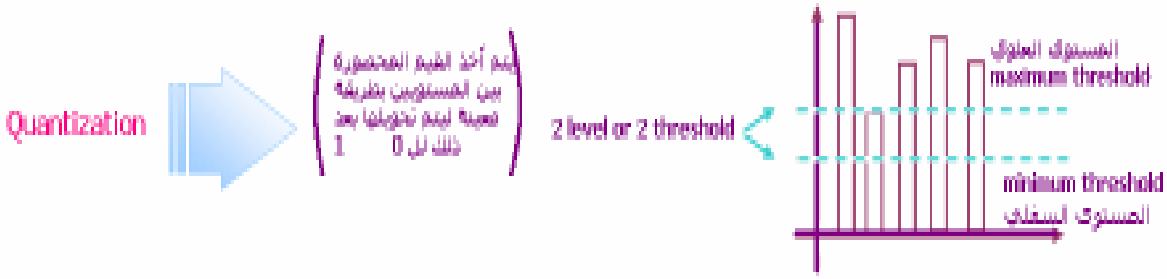
طيب تتحول هذه الإشارات المتصلة إلى إشارات منفصلة أو متقطعة والتي تعرف بـــ **discrete signal** كالتالي:



ولكي تتم عملية التحويل من **discrete signal** إلى ثنائي وما يعرف ب **digital signal** أو **binary form** حتى تكون في الصورة القابلة بأن يتعامل معها الحاسوب فإنه يتم عمل عملية تسمى

بـــ **Quantization**

وهي أخذ القيم المحصورة بين مستويين من الشكل السابق(**2level**) أو ما يسمى علميا ب (**threshold**) وتكون هذه القيم المحصورة قيما عشرية طبعا ويتم تحويلها إلى شكل ثنائي انظر الشكل التالي:



وبذلك تم تحويل الإشارات المتصلة (**Analog Signal**) إلى الإشارات المتقطعة (**Discrete Signal**) وبعدها عملية الـــ (**Quantization**)التي بدورها تنتج الصيغة الثنائية وهي الصورة الوحيدة التي يستطيع الحاسوب التعامل معها وبالتالي تبدأ عملية المعالجة (Processing)

وهذا هو دور كرت الصوت مبدئيا.

والعرض التالي يوضح العملية باختصار....

ملاحظة : في الرسم البياني المصطلح **Amplitude** يعني قيمة الحرف عند لحظة زمنية معينة , حيث كل حرف له **amplitude** مختلف عن الآخر، أو هو مقدار الطاقة الموجودة في الموجة والتي تمثل بارتفاع الموجة .....ويتضح معنى الـــ **amplitude**

ومن الرسم أيضا يتضح أن المسافة بين قمتين هي طول الموجه....

المنارة للاستشارات
www.manaraa.com

* أما بالنسبة للمصطلح **threshold** فهو علميا يشبه **level** في المعنى تقريبا ولكن بالتحديد هي مستوى محدد ومفروض لا يزيد ولا يقل عنه وال threshold ليست متساوية في جميع الإشارات بل الإشارات الصوتية لها **threshold** معينة و **image** لها **threshold** معينة وهكذا....

استكمالا للدرس السابق والذي تحدثنا فيه عن علم التعرف على الأصوات وذكرنا المهام الثلاثة الرئيسة التي ينجزها نظام التعرف على الكلام **(Speech Recognition System)**

وهي pre-processing - recognition - communication :

بعد مرحلة الـــ pre-processing التي تم فيها تحويل الإشارات المتصلة**(analog signal)** إلى عدد من ال**(digital signal)** 1 , 0 لكي تصبح جاهزة للتخاطب مع الحاسوب تأتي مرحلة.....

recognition:

هي مرحلة التعرف على الصوت المدخل وتنقسم هذه المرحلة إلى قسمين:

(identification & verification)

وفي كل من هذين القسمين يجب فيها إجراء نوع من الاختبار وأوضح أنـــه يجـــب فـــي البدايـــة عمـــل الـــ **identification** ثم الـــ**verification**

لنبدأ بإيضاح معنى الـــ **identification** بالمثال التالي:

نفرض أنه يوجد لدينا مجموعة من الطلبة موجودين بمكان واحد وكانت الطالبة X أحد الموجودين , ولدينا بصمة لصوت الطالبة X طبعا متمثلة بالثنائي ونحن نريد معرفة هل الطالبة X هي ضمن هذه المجموعة أم لا ؟؟ماذا سنفعل ؟؟

سنقوم في البداية بأخذ بصمة صوتية البصمة الصوتية هي تسجيل صوتي دقيق للشخص لكل شخص من الموجودين ونقارنها وحدا واحدا مع بصمة الطالبة X ولكن طريقة المقارنة ستتم بطريقة خاصة تسمى بـــ **Distance measurement** وهي أنني أقوم بعمل مقارنة بين بصمة الطالبة X وبين كل بصمة داخل المجموعة فإذا وجدت بصمة في هذه المجموعة أقرب ما يكون إلى بصمة الطالبة X عندها نستطيع القول أن الطالبة X موجودة ضمن هذه المجموعة..

إذن هذه المرحلة identification توضح لي بطريقة القياس الخاصة **Distance measurement** هل الشخص موجود أم لا ضمن كثافة معينة ولكن نحن في هذه الحالة لسنا متأكدين هل البصمة التي ظهرت نتيجة الاختبار السابق هي X فعلا أم لا ..إذن لكي أتأكد من أنها هي الطالبة X بعينها نقوم بعملية الـــ **verification** وهي مرحلة التحقق كيف؟؟

verification في هذه المرحلة نقوم بأخذ العنصر الذي نتج لدينا من الاختبار السابق أنه أقرب ما يكون لبصمة صوت الطالبة (X منفردا وعندها نتحقق ونتأكد هل هو فعلا X أم لا

وبذلك تنتهي مرحلة الـــ**recognition** ولكن بعد كل عمليات التعرف هذه لابد تطبيقات تستفيد من هذه الصوتيات المدخلة للحاسوب إذن تأتي مرحلة الـــ**Communication**

المنارة للاستشارات
www.manaraa.com

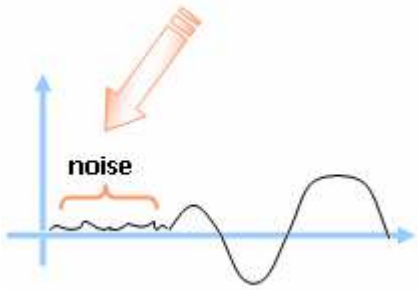
Communication:

هي الاتصال أو إرسال المُدخل بعد أنّ تم التعرف عليه إلى أنظمة الـ S/W أو H/W التي تحتاجه وأمثلة على ذلك:

Security وكما يتضح من الاسم أننا ممكن نستخدم الصوت في تحقيق الأمن , مثلا نفرض أننا قمنا بعمل نظام صوتي للدخول لإحدى القاعات وقمنا بأخذ بصمة صوتية لشخص معين ووضعناه في الحاسوب الموضوع عليه النظام بحيث لو قال" سبحان الله "مثلا يتم فتح باب القاعة لو أتى الشخص وأدخل الجملة "سبحان الله " إلى النظام بواسطة ميكروفون أو غيره سيقوم كرت الصوت مبدئيا كما قلنا بتحويل الإشارات المتصلة إلى إشارات رقمية وبعدها تتم عملية التعرف بمقارنة البصمة المسجلة في النظام مع البصمة الجديدة المدخلة حاليا فإذا تطابقت البصمتين أو مع وجود نسبة خطأ بسيطة جدا مسموح بها فإنه سيتم فتح باب القاعة

طبعا الـ security ممكن يتم عمله بعده تقنيات منها بصمة العين والشفاه!! وغيرها ولكن الصوت من الأشياء الصعب وجود شبيه لها حتى لو أتى نفس الشخص المدخل صوته في النظام وسجل صوته في كاسيت وعند دخوله للقاعة أدخل الصوت للنظام بواسطة هذا الكاسيت فإنه لن يتم فتح الباب بسبب عدم تطابق البصمتين لماذا؟؟ لأنه نتج من تسجيل الصوت ما يسمى بـ white noise وهي الضوضاء الناتجة عن عملية التسجيل حيث شريط الكاسيت عبارة عن مكونات إلكترونية فعندما أقوم بتسجيل الصوت تضاف فوق الإشارات الصوتية إشارات أخرى تسبب الـ noise وبالتالي تتسبب في الاختلاف عن الصوت الأصلي .

تلميح:عندما نلاحظ الإشارات الصوتية لشخص يتحدث نلاحظ فترات السكوت تمثل أيضا بواسطة موجات لكنها بسيطة جدا تسمى noise لاحظ الشكل:



education ممكن نستخدم الـspeech في التعليم أيضا..

control ممكن التحكم في الأجهزة الكهربائية بواسطة الصوت أيضا..

Diagnosis ومعناها تشخيص الأمراض أعني أنه يمكن بواسطة ميكانيزم معين عندما ينطق الشخص عدة كلمات يتم تشخيص حالته..

وبالمناسبة تشخيص الأمراض في علوم الحاسب يشكل فرع هام جدا نسميه ب Bio informatics المعلوماتية العضوية وهي الخاصة بأعضاء الإنسان والتي يتم فيها تشخيص الأمراض بواسطة طريقتين ,,,الأولى وهو ما عرفناه speech processing

أما الثانية فهي image processing والتي تعتمد على الأشعة وغيرها ....

وأخيرا هذه المراحل الثلاثة communication , recognition , pre-processing هي غير مرئية للمستخدم لكن كل ما يراه المستخدم من عملية التعرف التي تنجز له بواسطة برنامج أو غيره هي : دقة

المنارة للاستشارات
www.manaraa.com

تعرف هذا البرنامج على الكلام و سرعة التعرف أي سرعة تنفيذ البرنامج ومن هذه النقطتين يستطيع المستخدم من عمل تقدير للبرنامج..

ننتقل الآن للتقسيمة الرئيسية لمعالجة الكلام..

تحدثنا بالسابق عن مرحلة الـــ **pre-processing** ووظيفتها. سنذكر الخطوات الأساسية التي تندرج تحتها وهي:

1/ تجميع البيانات واكتسابها (Data collection & acquisition)

ومعناها جمع أشخاص معينين ذكور وإناث في أعمار متقاربة ونأخذ لهم بصمات صوتية وطبعا كما ذكرنا أنه يجب الحصول على هذه الإشارات بشكل يسمح للحاسب بأن يتعامل معها..

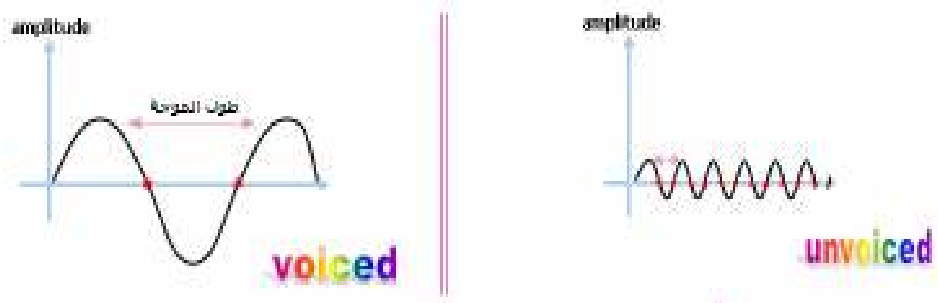
2/ اكتشاف المسموع والغير مسموع (voiced & unvoiced detection)

من الطبيعي أن يكون لدينا في الكلام أصوات مسموعة وغير مسموعة...

عندنا مثلا حروف المد (ا , و , ي) نسميها حروفا مسموعة حيث الـــ**amplitude** لها اكبر ما يمكن

وبالمقابل لدينا حروف أخرى مثل ال (س , ش) الـــ لها أقل ما يمكن وتسمى غير مسموعة أو بالأصح غير المجهورة ليس بسبب أنها لا تسمع ولكن بسبب أننا حينما نلاحظ شكل الموجات عند النطق بهذه الحروف قريبة جدا من حالة الـــ **noise**

ولمعرفة الفرق بين هذه الأصوات المسموعة **Voiced** والغير مسموعة **Unvoiced** سنتطرق إلى مفهوم الـــ **Zero Crossing** لنشاهد الفرق بيت الموجات المسموعة وغير المسموعة



المنارة للاستشارات
www.manaraa.com

من الرسم نلاحظ الأصوات المسموعة **voiced**

مقدار الــ amplitude لها كبير جدا والموجات أطول

مما يسبب أن عدد مرات إلتقاء أو تقاطع موجاتها مع الخط الأفقي horizontal access قليل جدا "كما نلاحظ النقاط الحمراء في الرسم هي نقاط تقاطع الموجات مع الخط الأفقي " وهذا ما نسميه بمعدل zero crossing الذي هو عبارة عن معدل تقاطع الموجات مع الخط الأفقي ,وجاءت كلمة zero بسبب أن قيمة الموجة أو الإشارة عند التقاطع تكون مساوية للصفر

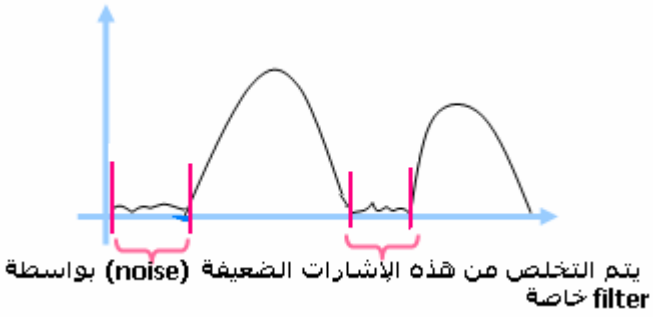
حيث يمكننا أن نقول بأن معدل ال Zero Crossing منخفض في الــvoiced speech

أما الــ**unvoiced**

قيمة ال amplitude لها قليل جدا, والموجات أقصر مما يسبب أن معدل التقائها مع الخط الأفقي كبير وسريع وبالتالي فإن معدل ال zero crossing كبير هنا.....

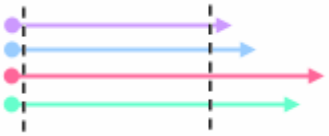
3/ تحديد بداية ونهاية الكلام المفيد end -point-detection

طبعا عند النطق هناك فترات توقف عن الكلام لكن فترات التوقف هذه كما قلنا تحتوي على إشارات signal ضعيفة noise ومن ضمن الأشياء المستهدفة في المعالجة Processing هي تقليل كمية الحسابات والتي تتم عن طريق التخلص من الإشارات الغير ضرورية مثل الــ noise وهي تتم بواسطة filter متخصصة للتخلص من الــ noise بطريقة معينة...



4/ Time Wrapping يتضح معنى هذا المصطلح من خلال المثال التالي_

مثلا لو كان لدينا مجموعة من الطالبات نطقوا كلمة واحدة مثلا) الحمد لله (عندما نأخذ تسجيل صوتي لكل واحدة منهن فإننا نلاحظ أن كل طالبة تنطق الكلمة في فترة مختلفة عن الأخرى كالتالي:



إذن الواجب علينا وضع جميع البصمات في طول واحد لكي نستطيع التعامل معها ككلمة واحدة دون الإخلال بالكلمة وطبعا هذه لها ميكانيزم معين وطريقة algorithm خاصة لهذه الأغراض.....

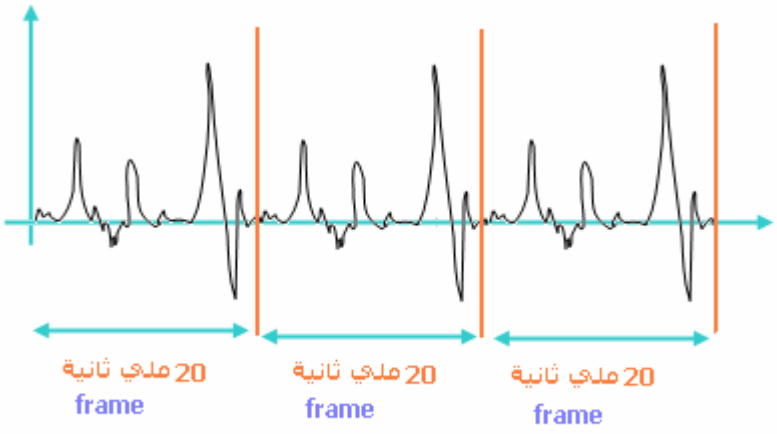
المنارة للاستشارات
www.manaraa.com

Framming /5

مثلا أثناء النطق بحرف معين فإنه خلال فترة زمنية بسيطة جدا ستكون أعضاء النطق في ثبات غير ملحوظ وهذه الفترة هي 20 ملي ثانية.

إذن الـ **speech** يحدث فيها ثبات كل (20 ملي ثانية) وهي فترة صغيرة جدا تكاد لا تذكر , وبالتالي فإننا سوف نقسم الـ speech إلى مجموعة من الـ frames بحيث أن كل frame عبارة عن (20 ملي ثانية) ثم نأخذ من كل frame عينة (sample) تعبر عن خصائص هذا الـ frame وذلك بدل أن نأخذ الـ frame بكامله وهذا مما يقل الحسابات أثناء المعالجة ومما يقل أيضا من زمن المعالجة processing.

مثلا عندي كلمة فإني سوف أقسمها إلى مجموعة frames حسب طولها وأبدأ بعد ذلك بأخذ العينات التي تعبر لي عن خصائص تلك الكلمة التي تم نطقها.... ليس تماما ولكن بشكل تقريبي ولكن ستكون هناك نسبة خطأ بسيطة.....



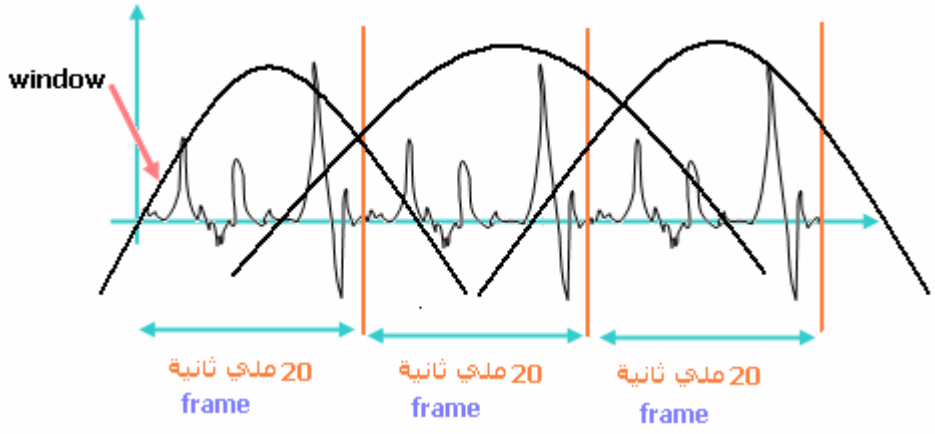
فترات ثبات قمنا بتقسيمها إلى مجموعة إطارات frames ثم سنأخذ من كل frame عينة تعبر عنه بدل ما نأخذ الإشارة بكاملها

Windwing/6 وبواسطة هذه المرحلة يتم التقليل من نسبة الخطأ التي قد تحدث نتيجة تقسيم موجات الكلام إلى frames وأخذ العينات.....

وطريقتها هي أنه يتم عمل window على الـ frame الأول ثم آخر على الـ frame الثاني وهكذا والهدف منها كما قلنا تقليل نسبة الخطأ

وتتقاطع هذه النوافذ الـ windows بنسبة 50% يعني نأخذ 50% من الـ frame الأول و 50% من الثاني وهكذا انظر الشكل التالي--:

المنارة للاستشارات
www.manaraa.com



-

Modeling/7

المشكلة في الــ Speech Signal أنني لا أستطيع أن أجعل الحاسب يتعامل مع الــ Analog Signal ولذلك أقوم بتوصيف هذه الإشارات عن طريق جميع الملامح الثابتة features لها وبذلك أكون عملت modeling لهذه الإشارات.....

8/استخلاص الخصائص Feature extraction

وهي ذات معنى مقارب للــ modeling فهي التعبير عن الصوت بواسطة مجموعة من الصفات وطبعا كل ما زادت الصفات التي أقوم باستخلاصها من الصوت كل ما يكون التعرف على الصوت أسهل... وكان نسبة الخطأ في عملية التعرف أقل.....

يمكننا تصنيف التعرف على الكلام كالتالي:

isolated word recognition I W R ويستخدم للتعرف على كلمات منفصلة ومعزولة عن بعضها وهو أسهل أنواع التعرف وذلك لأننا لا نواجه مشكلة الــ co-articulation وهي التقاء الحرف في نهاية الكلمة الأولى مع الحرف في بداية الكلمة الثانية مما يسبب صعوبة في التعرف..

connected word recognition C W R يستخدم للتعرف على مجموعة من الكلمات بفواصل وذلك بوضع Stops بين الكلمات وهو يشبه النوع السابق لكنه أصعب في التعرف..

continuous speech recognition C S R وهي للتعرف على الكلام المتواصل وإلى الآن مازال به العديد من المشاكل وهذا الصنف يحتاج إلى الكثير من التدريب وهو من أصعب الأنواع في التعرف..

Speech understanding S U وهي عمليات فهم الكلام بواسطة مترجمات خاصة وممكن تحويله إلى نصوص بعد التعرف عليه..

speaker verification S I, S V **speaker identification** والتي تحدثنا عنها.

word spotting w s ويستخدم للتنقيب عن كلمات معينة أو ما يسمى بالــ key word في موضوع معين

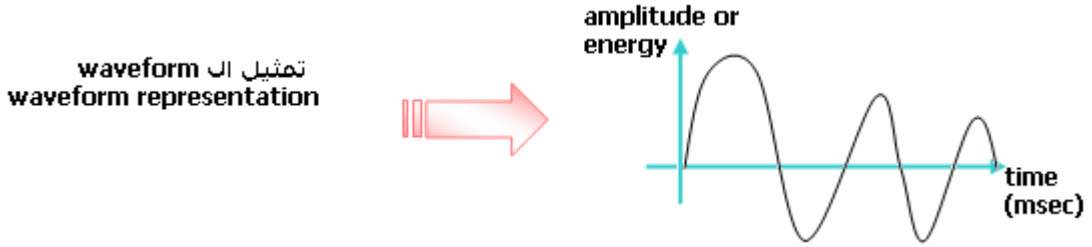
المنارة للاستشارات
www.manaraa.com

وبما أننا في هذه السلسلة نتحدث عن علم التعرف على الكلام فإننا سنتطرق إلى عدة جوانب خاصة بكيفية إنتاج الكلام وأعضاء النطق في الإنسان وبعض المعلومات الخاصة بهذا المجال.....

في عمليات التعرف على الكلام من المهم أن نركز على كيفية إنتاج الكلام وعملية إنتاج الكلام تبدأ بالتفكير في الرسالة التي يريد الشخص النطق بها وهذه الرسالة ستتحول بعد ذلك إلى إشارات عصبية

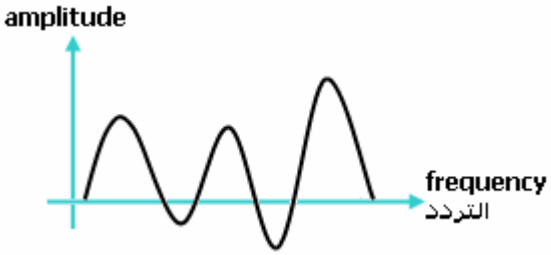
neural signal التي تصبح غالبا إشارات كهربية تحرك أعضاء النطق و بعد أن تتحرك أعضاء النطق سوف تنفذ سلسلة متتابعة من الإشاراتgestures التي ينتج في نهايتها الـ **Waveform** الذي يحتوي على معلومات الرسالة الأصلية وعندها تُسمع الرسالة للطرف الآخر.. أي سوف تُنتج الكلمة أو المنطوق **utterance**.

الـ **Waveform** هي عبارة عن شكل من أشكال الموجات وإحدى الطرق التي نستطيع تمثيل الكلام (**speech**)بواسطتها,..وكل الموجات الصوتية التي رسمناها في الدروس السابقة كانت**Waveform** وهي أسهل الموجات



وإحداثيات الـ **Waveform** السيني هو الوقت **time** والصادي يمثل الطاقة للحرف **Energy** أو الـ**Amplitude**.

وهناك أنواع أخرى للموجات نستطيع بواسطتها تمثيل الكلام ومنها ما يسمى بـ **spectrum** الذي تكون إحداثياته كالتالي: السيني يمثل الترددات **frequencies** والصادي يمثل الـ **Amplitude** وهي أسهل في عملية المعالجة**processing** .



أما الـ **Spectrogram** فإننا نستطيع تمثيل الكلام بها وهي ذات 3 أبعاد : الوقت **time** التردد **frequency** والـ **amplitude** ولكنها صعبة في التعامل مع الـ wave ومعالجتها.

عندما أريد أن أنطق أي كلمة **utterance** في البداية سيتم التنفس من الرئتين ثم يخرج الهواء إلى الحنجرة والذي يسبب عند مروره بها اهتزاز الحبال الصوتية

vocal folds or vocal cords ويستمر في مسيره حتى يصل إلى القناة الصوتية **vocal tract**ومنه إلى التجويف الفمي **oral cavity** وممكن في بعض الحروف مثل م- ن يضطر الهواء

المنارة للاستشارات
www.manaraa.com

للمرور بالتجويف الأنفي **nasal cavity** وبعدها تخرج الكلمة وتصبح مسموعة وتتحول إلى الـــ **wave form** عندها سنطلق عليها مسمى **Acoustic.**

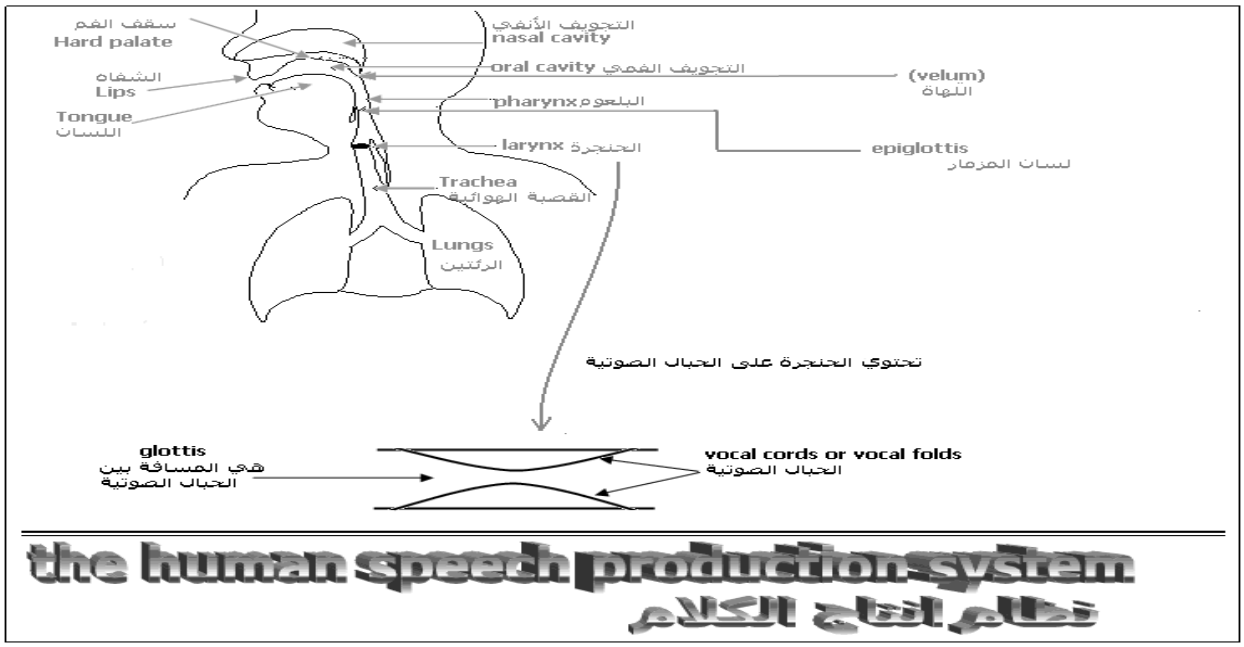
إذن نطلق على الكلمة مصطلح **acoustic** عندما تخرج من الفم وتتحول إلى **Waveform** و تصبح مسموعة.

والعملية العكسية هي الأذن التي تستقبل هذا الكلام حيث تدخل الإشارة إلى طبلة الأذن وتبدأ المطرقة تضرب وتصل الرسالة إلى الشخص الآخر ويسمع الكلمة.

الجهاز الصوتي أو جهاز النطق **vocal apparatus** يتمثل في **3** أعضاء رئيسية هي الحلق **Throat** الفم **mouth** الأنف **nose**

الحلق **throat** يحتوي على الحبال الصوتية **vocal cords** التي تهتز لتنتج لنا الأصوات.

وفي الرسم التالي تفصيل لجهاز النطق:

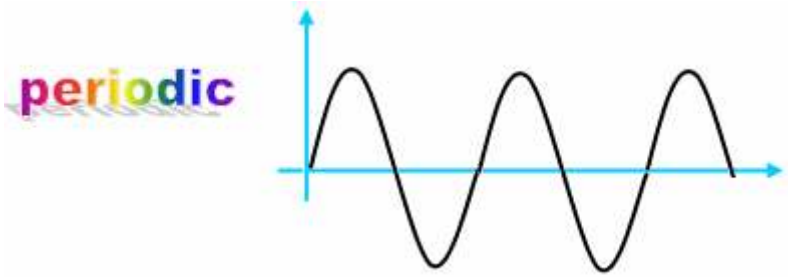


نقاط النطق تتمثل في : الأسنان **teeth** اللسان **Tongue** اللهاة **velum** الجزء العظمي خلف الأسنان العليا **alveolar** الحنك الأعلى **Hard palate: roof of the mouth** أو ما يعرف ب سقف الفم و الشفاه **lips**

نود أن أوضح الآن عدة مفاهيم تختص بالموجات الصوتية.

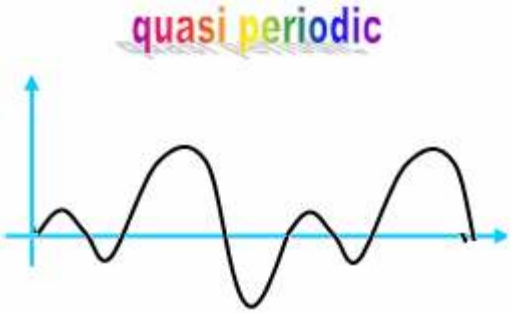
عندنا في الموجات ما يكون كالشكل التالي حيث الموجات متساوية جدا وهذا ما نسميه بـــ **periodic** وهذه الموجات مستحيل أن تعبر عن الموجات الصوتية حيث أن الموجات الصوتية المسموعة **voiced** موجات غير متساوية.

المنارة للاستشارات
www.manaraa.com



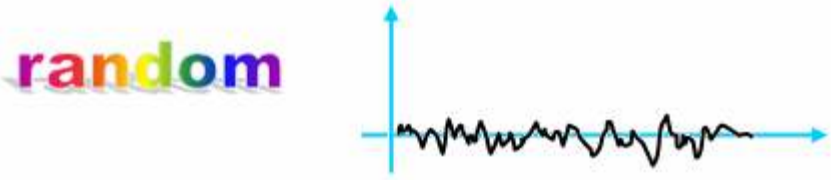
ولذلك فإن موجات الصوت المسموعة **voiced** نمثلها بما يسمى بـــ **quasi periodic**

أو**like periodic** وهي تعني أنها ليست دورة متساوية مثل الشكل السابق وذلك بسبب أن موجات الكلام مستحيل أن تتساوى ويتخللها فترات من الـــ **noise** وشكلها تقريبا كالتالي:



المهم أن تكون الموجات غير متساوية

وهناك مسمى للموجات وهو ما نمثل به الـــ**noise** وهي **random** وذلك بسبب عشوائيتها انظر الشكل :



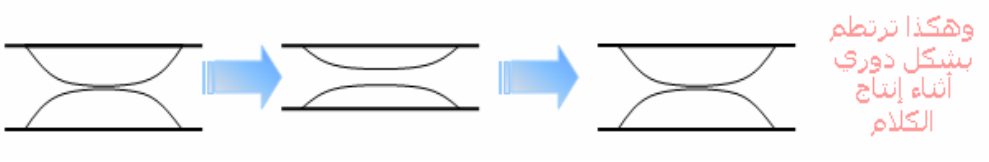
وهي التي تمثل فترات بين الكلام مثل التنفس.

وظيفة الحبال الصوتية عند إخراج الكلام:

نستطيع تقسيم حالات الحبال الصوتية إلى 3 حالات مختلفة بعد أن نتفق على أن الحبال الصوتية في الأحوال العادية تكون مقفلة نوعا ما.

1 بداية في الـــ**voiced sound** الأصوات المسموعة عندها سوف تُجبَر الحبال الصوتية بواسطة ضغط الهواء الآتي لها من الرئتين مما يسبب ارتطامها ببعض ليقوموا بعمل شيء من الاهتزاز ثم يعودوا مرة أخرى إلى حالتهما المقفلة تتكرر هذه الحركة أثناء إنتاج الكلام بشكل دوري مما يسبب إنتاج طنين معين الذي بدوره يسبب إنتاج التردد**frequency** وطبعا طالما الهواء لم يتعدى الحبال الصوتية إذن لم يتم إنتاج الصوت بعد حتى يتعداها انظر الرسم:

المنارة للاستشارات
www.manaraa.com



وقيمة التردد**frequency**التي نتجت تعتمد على عدة عوامل من أهمها:

طول الحبال الصوتية: حيث أنه كلما قصرت الحبال الصوتية كان التردد فيها أعلى

وكتلتها: حيث أنه إذا كانت الكتلة لها كبيرة التردد يكون أقل والعكس صحيح..

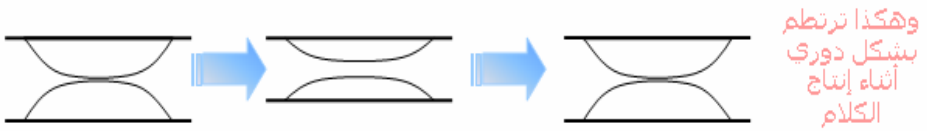
ومقدار الشد فيها: حيث يزداد التردد كلما زاد الشد في الحبال الصوتية..

والتردد الناتج عن اهتزاز الحبال الصوتية عادة يتراوح بين **60 - 400** هرتز حيث يبلغ متوسط التردد في الحبال الصوتية عند النساء البالغين **180** هرتز وعند الرجال البالغين **100** هرتز.

2 الحالة الثانية للحبال الصوتية هي عند ما يسمى بـــــــــــــــ **unvoiced sound**

أو **voiceless sound** وليس المعنى أنها ليست مسموعة تماما ولكن أنها حروف تنطق ولكن عند تمثيلها نلاحظ أنها تشبه حالة الـــ **noise** أي ذات قيمة منخفضة عندها ستكون الحبال الصوتية مفتوحة قليلا عن بعضها البعض.

3 الحالة الثالثة هي كون الحبال الصوتية مفتوحة بشكل كبير أي تكون المسافة بينها واسعة وهذه تحدث أثناء التنفس حيث يسبب انفتاحها الواسع مرور الهواء بينها دون عمل اهتزاز على الحبال الصوتية ولا ينتج صوت حينها....



تلميح: عندما يكون الشخص لديه التهاب في الحبال الصوتية فإن ذلك يؤدي إلى زيادة سمكها وبالتالي تقل المسافة بين الحبال الصوتية Glottis مما يضيّق مكان مرور الهواء فيها وبالتالي عند تنفس الشخص ينتج صوت ملحوظ للنفس وذلك بعد أن ضاقت المسافة بين الحبال الصوتية فإن الهواء أثناء مروره يسبب اهتزاز للحبال الصوتية مما أدى إلى ظهور الصوت المزعج أثناء التنفس....

بعض التعاريف الهامة:

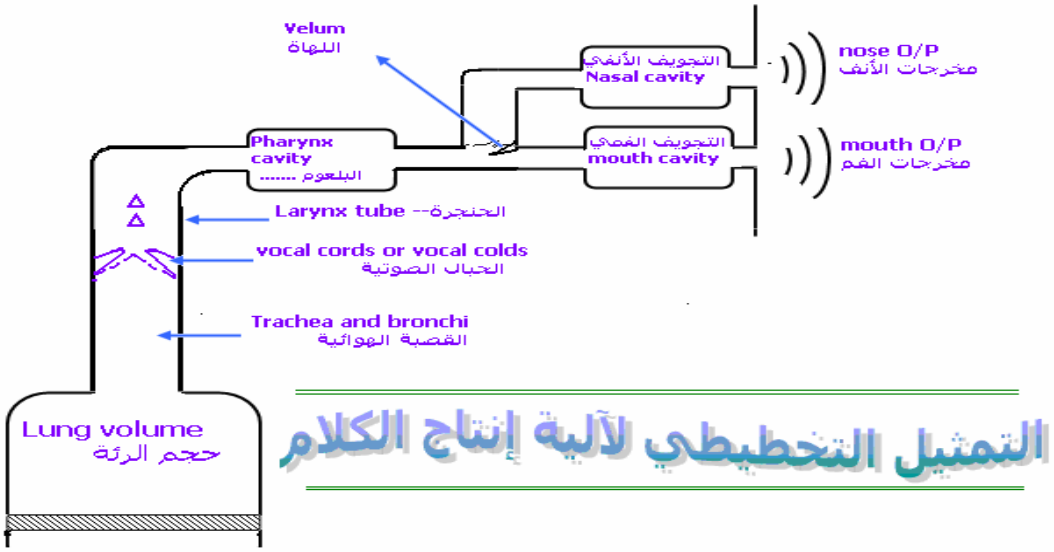
phoneme هي أقل وحدة صوتية وهي عبارة عن مجموعة الرموز التي يتكون منها الصوت وممكن تكون حرف وقد تكون أقل منه.

المنارة للاستشارات
www.manaraa.com

linguistics اللغويات و هي القواعد التي تحكم وترتب الأصوات لكي تنتج لغة... مثلا تأخذ الـــphoneme وتضعها مع بعضها البعض بترتيب وقواعد معينة لينتج لي كلمات التي تجتمع وتكون لي الجمل التي تكون الكلام أو اللغة.....

phonetics علم الأصوات و هي التي تدرس الصوتيات وتصنفها... مثلا تصنف الـــ voiced من الـــ unvoiced وغيره.

لننتقل الآن إلى آلية إنتاج الكلام



نلاحظ أن حجم الرئة يتناسب مع حجم الهواء الذي يكون بداخلها.

نستطع تقسيم الأصوات التي تستطيع القناة الصوتية إخراجها إلى **vowel** و **consonant**

الـــ **vowel** وهي الحروف التي يتم إنتاجها عندما يكون الفم مفتوح قليلا وهي الحروف المدية وهنا تهتز الحبال الصوتية كليا لإنتاجها مثل **a, e ,u , o , aa ,ee , au ,uh**

أما الـــ **consonant** ما نسميها بالسواكن وهنا الحبال الصوتية إما أنها لا تهتز كليا أو أنها تهتز بشكل جزئي وطبعا هذه السواكن تعتمد على الحرف الذي يسبقها والحرف الذي يليها.

وهذه الـــ **consonant** يتم توصيفها اعتمادا على نقطتين هامة

1 مخرج الحرف **place of articulation**

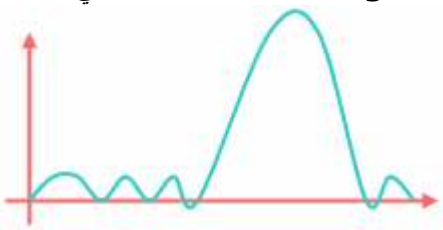
التي قد تكون من الشفاه أو اللسان أو طرف اللسان أو باشتراك بعض نقاط النطق لإخراج هذه الحروف..

2 وطريقة نطق الحرف **manner of articulation**

والتي يندرج تحتها النقاط التالية :

المنارة للاستشارات
www.manaraa.com

- **plosive** وهي الحروف الانفجارية حيث يتم قفل نقطة من نقاط النطق نهائيا مثل الشفاه أو غيرها ثم بعدها يخرج الحرف فجأة مثل حرف **B** في كلمة **bid** وحرف الـــ **G** في كلمة **gate**وتمثل الموجة بالشكل التالي:



- **Fricatives** وهي الحروف الاحتكاكية الناتجة عن احتكاك نقطتين من نقاط النطق أثناء النطق بحرف معين مثل الحروف التالية:

v , th , z, f, s, sh

- **Nasals** وهي الحروف الأنفية مثل **m, n, ng** والتي يشارك الـــ **nasal cavity** بجزء كبير من إخراجها.
- **Affricates** وهي الحروف التي تخرج مع إقفال أحد نقاط النطق تماما ثم يتدرج الهواء في الخروج بعد ذلك مثل حرف الـــ **J** في كلمة Judge
- **Approximant** التي تعني اقتراب أعضاء النطق من بعضها عند إنتاج الحرف ولها نوعين :

semivowels وهي تشبه الحروف المدية **W, Y**

Liquids وهي الحروف المنسابة مثل **R, L**

المنارة للاستشارات
www.manaraa.com