

في إطار السمنارات العلمية لطلاب الدفعة 26 هندسة الحاسوب بكلية  
الهندسة والتكنولوجيا بجامعة الجزيرة (السودان) :

بحث بعنوان:

تحدث مع حاسبك الشخصي  
(عن طريق **الفجوال بيسك**)

Speak to your PC  
(via **VB 6.0**)

إعداد :

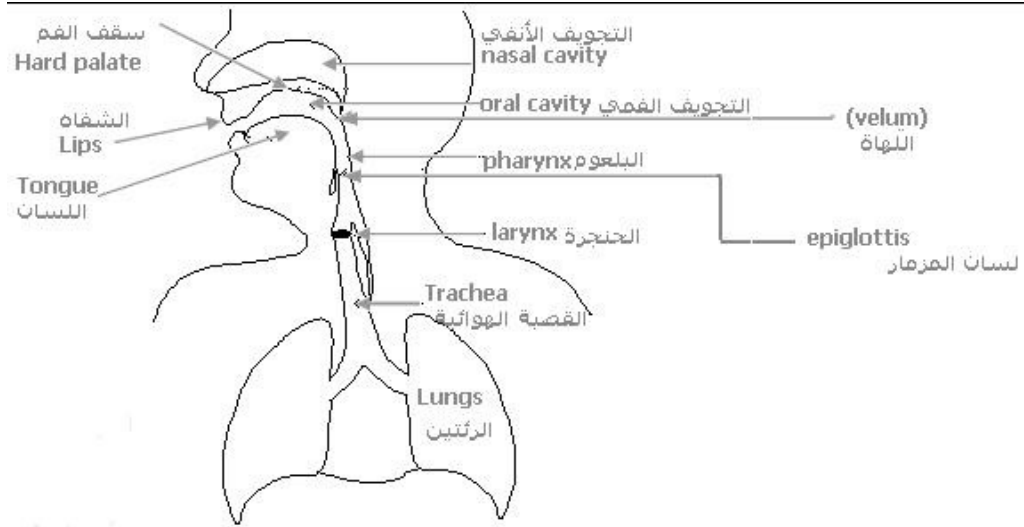
**محمد عفيفي إبراهيم**

## التخاطب لدى الإنسان :

تتلخص عملية التخاطب لدى الإنسان في إنتاج الكلام بواسطة أعضاء النطق وسماعه بواسطة الأذن .

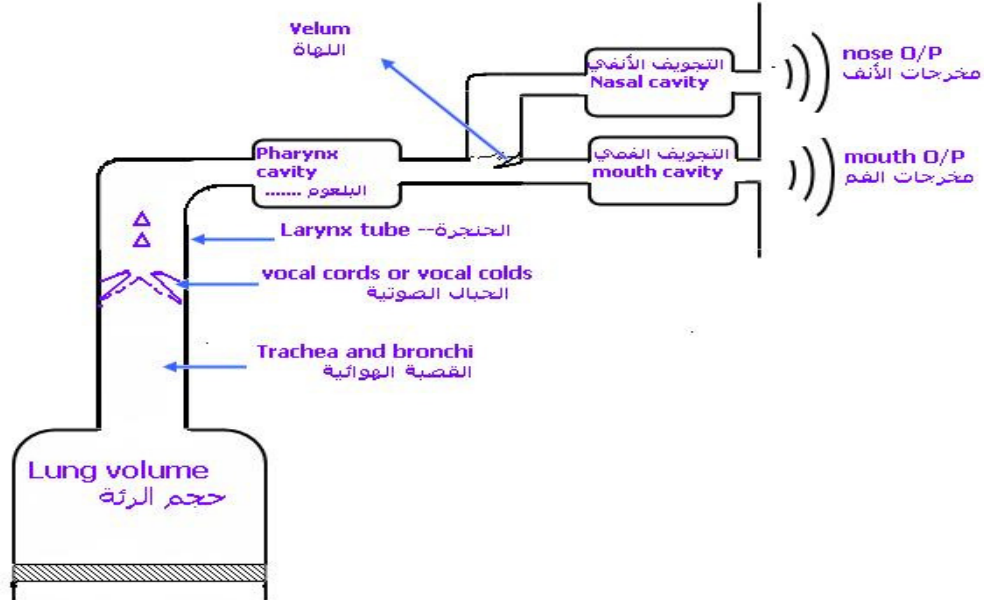
### آلية النطق ( speech mechanism )

يتكون جهاز النطق لدى الإنسان من ثلاثة أجزاء رئيسية هي الحلق و الأنف والفم , وكل منها يحتوي على أعضاء فرعية. وأهم هذه الأجزاء هو الحلق حيث يحتوي على الحبال الصوتية . الشكل يوضح أعضاء النطق:



### مراحل إنتاج الكلام :

عندما يريد الإنسان الكلام يفكر في الرسالة التي يريد النطق بها وبعد ذلك يرسل الدماغ رسائل عصبية إلى أعضاء النطق , حيث تبدأ العملية بملاً الرئتين بالهواء وذلك حسب شدة الصوت وطول الكلام المراد نطقه , بعد ذلك يخرج الهواء مندفعاً من الرئتين ماراً بالحبال الصوتية مما يسبب ارتطامها ببعضها البعض وذلك الارتطام هو الذي ينتج التردد الصوتي حتى هذه المرحلة لم ينتج الكلام , بعد هذه المرحلة يخرج الهواء الذي أصبح صوتاً إلى الفم أو الأنف لتحديد الحرف المراد . ويوضح الشكل التالي هذه العملية :



## عمل الحبال الصوتية :

إن عمل الحبال الصوتية يتلخص في إعطاء التردد للصوت ويعتمد هذا التردد على عدة عوامل منها:

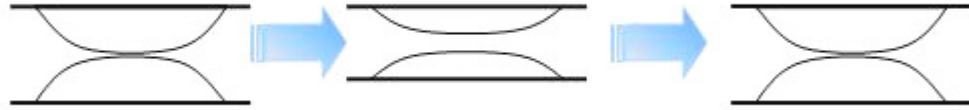
- 1- **طول الحبال الصوتية** : فكلما قل طولها زاد التردد.
- 2- **كتلة الحبال الصوتية**: وكلما زادت الكتلة زاد التردد.
- 3- **درجة شد الحبال الصوتية**: فكلما زاد الشد زاد التردد.

والتردد الناتج من الحبال الصوتية عادة يتراوح من 60 – 400 هرتز . فيبلغ متوسط التردد للنساء البالغين 180 هرتز ويبلغ لدى الرجال البالغين 100 هرتز .

ويمكن تقسيم الأصوات التي تنتج إلى **قسمين**:

### 1\ الأصوات المسموعة (Voiced Sounds) :

ويقصد بها الأصوات التي يمكن سماعها بوضوح عالي و في هذه الحالة تحرك الحبال الصوتية ( والتي تكون في أغلب الأحيان مقفلة وساكنة) بواسطة ضغط الهواء الآتي من الرئتين مما يسبب ارتطامها ببعضها البعض لتقوم بعمل شيء من الاهتزاز ثم تعود مرة أخرى إلى حالتها المقفلة تتكرر هذه الحركة أثناء إنتاج الكلام بشكل دوري مما يسبب إنتاج طنين معين الذي بدوره يسبب إنتاج التردد, (**frequency**) وطبعاً طالما الهواء لم يتعدى الحبال الصوتية إذن لم يتم إنتاج الصوت بعد حتى يتعدها . الشكل يوضح هذا :



### 2\ الأصوات غير المسموعة (Unvoiced Sounds OR Voiceless Sounds) :

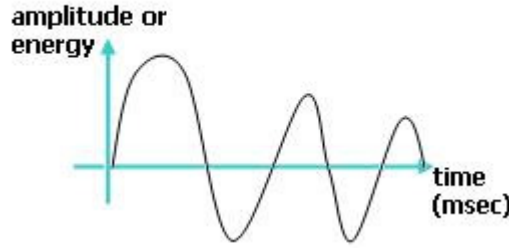
وليس المعنى أنها ليست مسموعة تماماً ولكن هي أصوات طاقتها لها قيمة منخفضة و عند إنتاجها تكون الحبال الصوتية مفتوحة قليلاً عن بعضها البعض.

**وهناك** نوع آخر من الأصوات وهو الذي ينتج من عملية التنفس حيث تكون الحبال الصوتية مفتوحة بشكل كبير أي تكون المسافة بينها واسعة وبسبب انفتاحها الواسع يمر الهواء بينها دون عمل اهتزاز ولا ينتج صوت حينها .

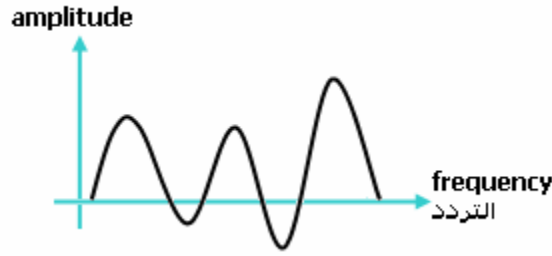
## تمثيل موجة الصوت :

يوجد نوعين رئيسيين لتمثيل الموجة :

**1 / WaveForm :** وفي هذا النوع يتم تمثيل موجة الصوت بواسطة مقدار الطاقة Energy Amplitude والزمن Time ويكون بالمللي ثانية ويعتبر هذا النوع من التمثيل من أسهل الأنواع .

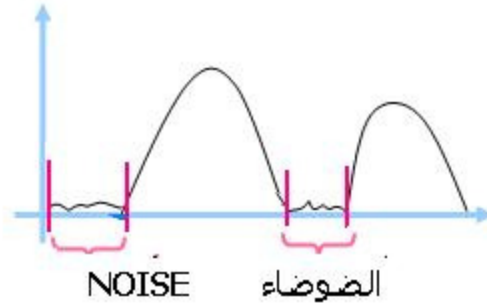


**2 / Spectrum :** وفي هذا النوع يتم تمثيل الموجة بواسطة مقدار الطاقة Energy Amplitude والتردد Frequency وبعد هذا النوع مفيد في عملية معالجة الإشارة الصوتية.



## الضوضاء Noise :

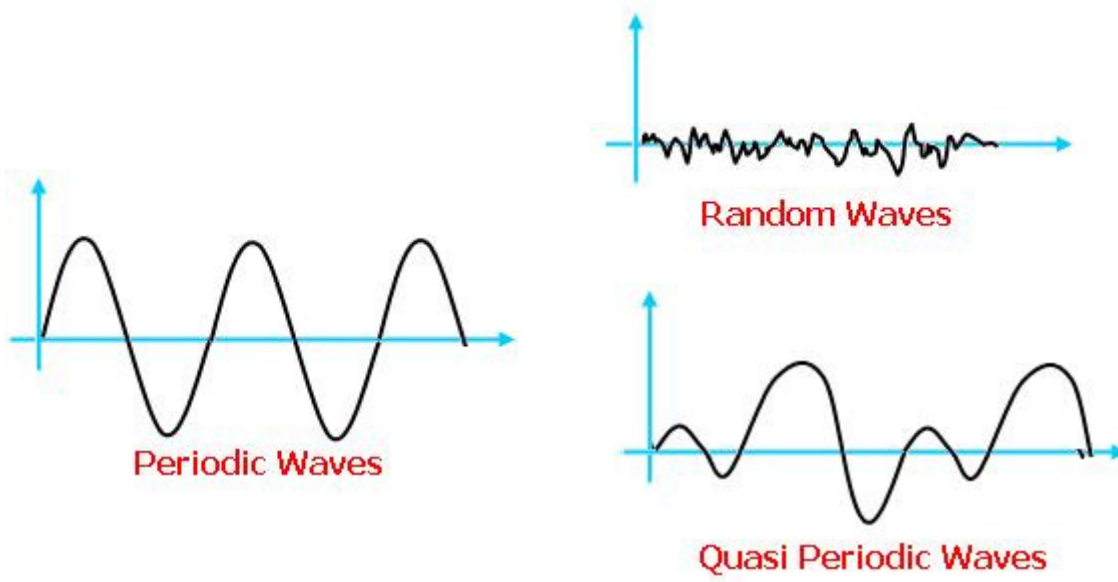
الضوضاء هي الأصوات غير المرغوب فيها و التي تكون متداخلة مع الكلام والناجمة من أصوات المحيط الخارجي مثل الهواء و أصوات الآلات والسيارات وغيرها , وتتميز بأن لها طاقة Energy منخفضة حيث يمكن تمييزها عن الكلام , ويتم التخلص منها في المعالجة بواسطة مرشحات Filters خاصة لهذه العملية.



## تمييز موجة الصوت من غيرها:

يقصد بالصوت جميع الأصوات سواء أكانت أصوات بشرية أو غيرها . ويمكن تمييز موجة الصوت من شكل الموجات حيث تتميز موجة الصوت أنها غير متشابهة (ما يسمى **Quasi Periodic Wave** ) ولا يمكن أن تعبر الموجات المتشابهة ( ما يسمى **Periodic Wave** ) عن الصوت , كما أن موجة الصوت تخلف عن الموجات الصادرة من الضوضاء ( تسمى **Random Wave** ) .

الشكل يوضح الموجات الثلاثة :



## عملية التعرف على الكلام :

يمر التعرف على الكلام في الحاسوب بعدة مراحل هي :

- 1- مرحلة ما قبل المعالجة **Pre-Processing** : وفي هذه المرحلة يتم تجهيز الإشارة الصوتية ليتم معالجتها .
- 2- مرحلة التعرف **Recognition** : وهي أهم المراحل ويتم فيها المعالجة والتعرف على الكلام.
- 3- مرحلة الاتصال **Communication** : ويتم فيها إرسال النتائج إلى جهة معينة.

### **\*\* Pre-Processing :**

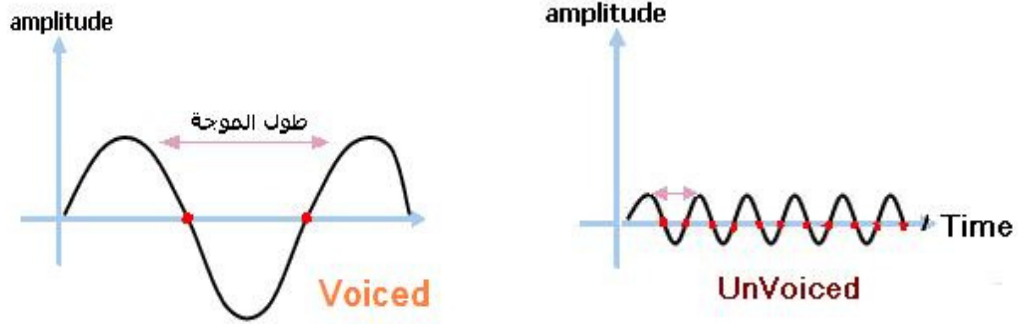
وتقسم إلى مراحل :

#### 1- تجميع البيانات **Data Collection** :

وفي هذه المرحلة يتم تجميع البيانات المدخلة وتجهيزها .

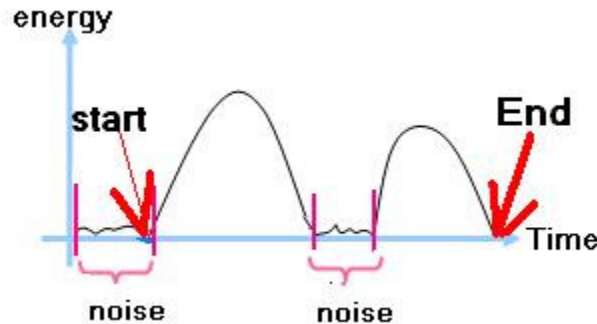
#### 2- تمييز المسموع من الغير مسموع ( Voiced & Unvoiced Detection )

ويتم في هذه المرحلة تمييز الكلام من الضوضاء وهناك حروف لها طاقة مقارنة لطاقة الضوضاء فمثلا الحروف الثابتة لها طاقة منخفضة بينما حروف المد مثل aa,ae,ou لها طاقة عالية .  
ويتم ذلك بطريقة **Zero Crossing** حيث أن الأصوات (الحروف) المسموعة يكون تقاطعها مع المحور السيني Time قليل بعكس الغير مسموع وبالتالي قيمة Zero Crossing لها منخفض، بعكس الغير مسموع حيث أن قيمة Zero Crossing لها عالي، وسميت هذه الطريقة بهذا الاسم لأن قيمة الطاقة عند التقاطع مع المحور السيني تكون مساوية Zero .



#### 3- تحديد بداية ونهاية الكلام المفيد ( End-Point Detection ) :

ويتم فيها تحديد بداية ونهاية الكلام بواسطة الـ Noise حيث يسبق النطق Noise وأيضاً بعد الانتهاء من الكلام تكون هناك Noise .



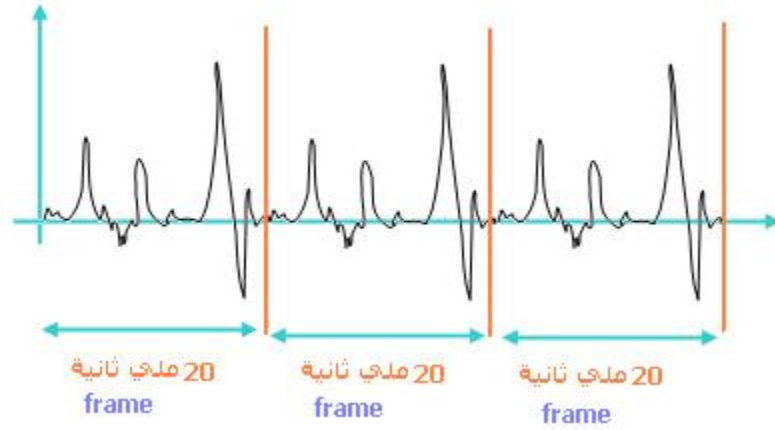
#### -4 : Time Wrapping

و يتضح ذلك عندما ينطق عدة أشخاص منفردين بنفس الكلمة فنلاحظ اختلاف في زمن النطق فيتم في هذه المرحلة مساواة الزمن بدون تغيير خصائص الكلمة ولها طريقة و خوارزمية محددة.



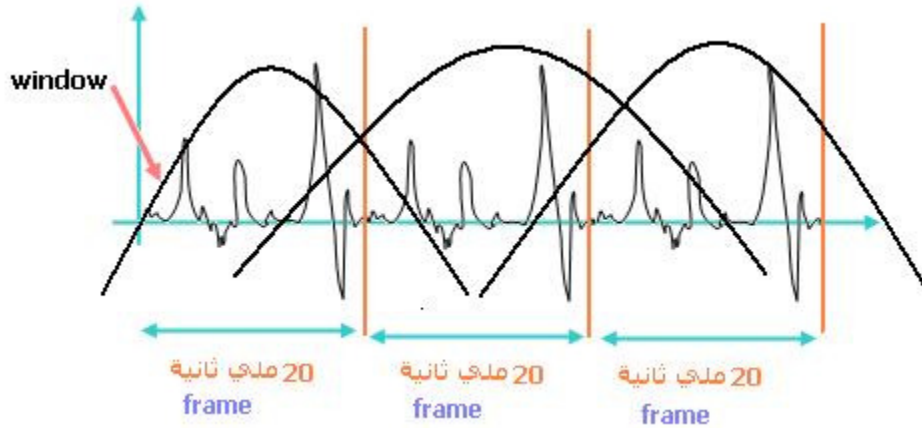
#### -5 : Framing

لقد لوحظ أن موجة الصوت تكون في ثبات كل 20 ملي ثانية لذا تقوم هذه العملية بتقسيم الموجة إلى إطارات Frames مدة كل منها 20 ملي ثانية ثم أخذ عينة من كل إطار وإدخالها في الحسابات وذلك بنسبة خطأ بسيط نسبة لأن العينة لا تكون كافية للتعبير عن الإطار Frame بأكمله.



#### -6 : Windowing

ويتم بواسطة هذه العملية تقليل نسبة الخطأ من العملية السابقة (Framing). وتتخلص بعمل نافذة في كل إطار تحتوي على 50% من الإطار ويتم أخذ محتويات هذه النوافذ في الحسابات وبهذا يكون معظم قيم الإطارات قد أدخلت في الحسابات.



#### 7- التحويل Modeling :

ويتم فيها توصيف إشارة الصوت بواسطة الملامح الظاهرة فيه.

#### 8- استخلاص الخصائص (Features Extraction) :

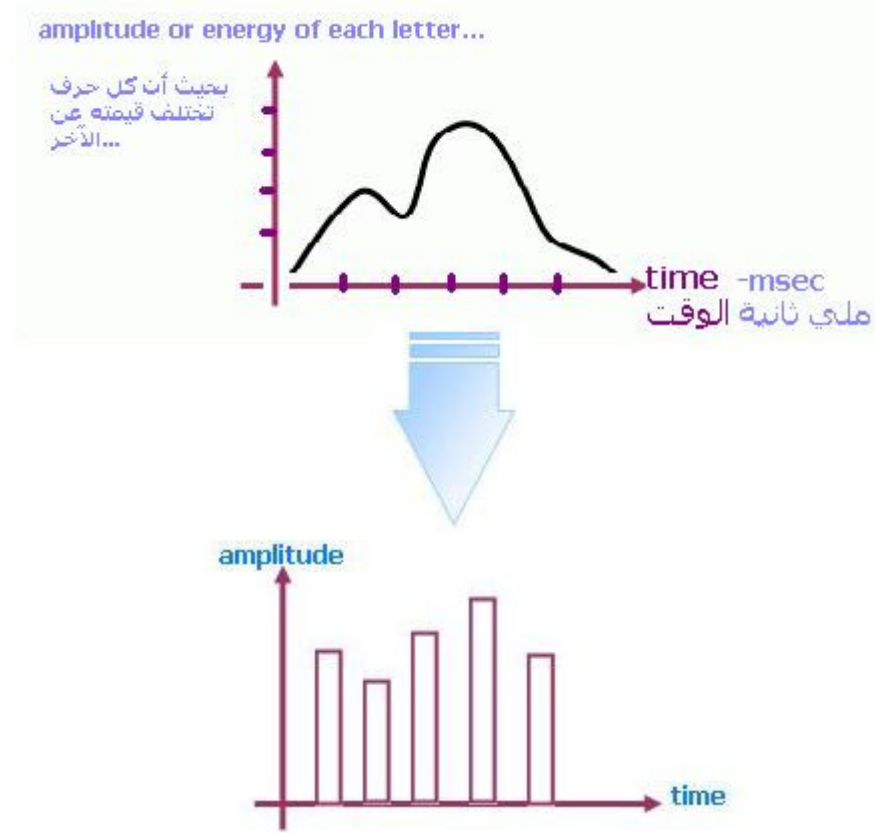
وتتم في هذه المرحلة توصيف الإشارة بواسطة أوصاف ومعالج أكثر حيث تمثل إكمال للعملية السابقة.

#### 9- تحويل الإشارة إلى رقمي (Analog to Digital) :

إن إشارة الصوت هي إشارة مستمرة Continues وبما أن الحاسوب لا يفهم هذا النوع من البيانات فإن الصوت يجب أن يحول إلى إشارة رقمية 0's & 1's . وتتم هذه العملية بمرحلتين :

#### أ\ تقطيع الإشارة :

تتلخص هذه العملية في تقطيع الإشارة المتصلة إلى إشارة متقطعة تمهيداً لتحويلها إلى الإشارة الرقمية



#### ب\ تحويل الإشارة المقطعة إلى رقمية (THRESHOLD LEVEL-DIGITAL CONVERT) :

وهنا يتم أخذ مستويين Two Levels على حدود معينة من الإشارة المقطعة يسمى كل منها (Threshold) ويتم التحويل على هذا الأساس وتسمى هذه العملية بـ Quantization . ولكل نوع من الإشارات Threshold معين فالصوت له threshold يتم التحويل على أساسه و الصور أيضاً لها threshold خاص ومعروف.



والشكل التالي يوضح المستويين :



وأغلب عمليات هذه المرحلة يقوم بها كرت الصوت.

## **\*\* عملية التعرف ( Recognition ) :**

بعد أن تم تحويل الصوت إلى بيانات رقمية هنا يأتي دور محركات التعرف على الكلام **Speech Recognition Engines**.

### **\*\* محركات التعرف على الكلام (SRE) Speech Recognition Engines :**

وهي المسئولة من تحويل الإشارة الصوتية إلى كلمات وحروف و تتكون من جزأين

#### **1/ البرمجيات (S/W) Softwares :**

وهي عبارة عن مجموعة من المكتبات البرمجية (مكتبات الربط الديناميكي DLL Dynamic Link Libraries) والتي تحتوي على مجموعة من الدوال (Functions) والخوارزميات (Algorithms) والتي تتبع طرق خاصة وعمليات حسابية بواسطتها تستطيع التعرف على الكلام .

#### **2/ العتاد (H/W) Hardware :**

وهو عبارة عن المعالجات (Processors) والذاكر (Memories) التي تستخدم في عمليات الحسابات والتخزين المؤقت و غالباً ما يكون العتاد هو جهاز الحاسب الشخصي (أي تستخدم معالج وذاكرة الحاسب).

## +أنواع محركات التعرف على الكلام:

- 1/ **محركات التعرف على الكلمات (ISWR) Isolated Word Recognition Engines :**  
وتستطيع التعرف على كلمات منفصلة فقط ولا تستطيع التعرف على جمل كاملة.
- 2/ **محركات التعرف على الكلمات المتصلة (CWRE) Connected Word Recognition Engines :**  
وتستطيع التعرف على كلمات متصلة لكن بوضع فواصل بين الكلمات (Stops) .
- 3/ **محركات التعرف على الكلام (CPRE) Continues Speech Recognition Engines :**  
وتستطيع التعرف على الكلام المتصل ولكنها تحتاج إلى تدريب Training من قبل المتكلم وإلى الهدوء أثناء التكلم وما زال هذا النوع تحت التطوير .
- 4/ **محركات فهم الكلام (SUE) Speech Understanding Engines :**  
وتستطيع فهم وتمييز أي نوع من الكلام ويمكنها بعد ذلك تحويلها إلى نص (STT) Speech to Text .
- 5/ **محركات التعرف على المتكلم (SrRE) Speaker Recognition Engines :**  
وتستطيع التعرف على المتكلم وتمييز صوت الناطق بالكلمة من غيره وذلك بالإضافة إلى التعرف على الكلمة مثلاً إذا نطق أحمد بكلمة welcome فإن نتيجة التعرف تكون (:: قال احمد welcome ::) وتكون على مرحلتين :  
أ/ **تحديد الهوية Identification :** ويتم فيها مقارنة صوت المتكلم بمجموعة من الأصوات المخزنة وتحديد أقرب صوت من الأصوات لصوت المتكلم وذلك بطريقة تسمى Distance Measurement .  
ب/ **التحقق Verification :** ويتم فيها مقارنة الصوت المحدد في المرحلة السابقة مقارنة دقيقة مع صوت المتكلم .
- 6/ **محركات التنقيب عن كلمة (WSE) Word Spotting Engines :**  
وتستخدم للتعرف على كلمة منطوقة من مجموعة من الكلمات المدخلة ويعد من أسهل الأنواع.

**وهناك** نوع آخر من المحركات لعمل العملية العكسية

**\*\* محركات نطق الكلام (TTS) Text to Speech Engines :**  
وهي تستطيع نطق الكلام المكتوب أي تحويل النص إلى كلام وإخراجه عن طريق كرت الصوت .

## **\*\* عملية الاتصال Communication :**

وهي العملية الأخيرة من عملية التعرف وتتلخص في إرسال النتائج أي الكلام إلى الجهة الطالبة (غالباً برنامج مبرمج بإحدى لغات البرمجة) وتكون النتائج عبارة عن نص Text .

## التعرف على الكلام بواسطة لغة Visual Basic

تعتبر لغة Visual Basic من أوسع اللغات انتشاراً وذلك لسهولة تعلمها وقوتها في برمجة التطبيقات وتدعم لغة Visual Basic التعرف على الكلام وذلك بواسطة محرك Microsoft للتعرف على الكلام .

### **\*\* محرك Microsoft للتعرف على الكلام :**

وهو محرك طورته شركة Microsoft على مدى سنوات وذلك للتعرف على الكلام و هو مجاني ويمكن تحميله من موقع الشركة على الشبكة العالمية , ويكون المحرك على شكل ملفات OCX , ActiveX بحيث يتيح للمبرمجين استخدامه في برامجهم , ويدعم المحرك عدة ميزات :

- Speech Recognition \*
- Speech to Text \*
- Text to Speech\*
- Voice Commands \*

ولقد استخدمت Microsoft كل ميزات هذا المحرك في حزمة برامج Office XP . والمحرك يدعم 11 لغة هي الإنجليزية الأمريكية والإنجليزية البريطانية و الألمانية والهولندية و الفرنسية والإيطالية واليابانية والكورية والبرتغالية والروسية والأسبانية وذلك في ميزة Text to Speech ويدعم حتى الآن الإنجليزية الأمريكية فقط في ميزة Speech Recognition وجاري تطوير المحرك بواسطة عدة فرق من المبرمجين من كل الجنسيات.

وصلة التحميل [هنا](http://www.microsoft.com/downloads/details.aspx?familyid=7D13964C-06FD-4BF9-B49C-814FAA6A86EA&displaylang=en)

<http://www.microsoft.com/downloads/details.aspx?familyid=7D13964C-06FD-4BF9-B49C-814FAA6A86EA&displaylang=en>

### **عمل مشروع للتخاطب مع الحاسوب:**

فكرة المشروع بسيطة وهي أن تتكلم مع الحاسوب بواسطة الميكروفون ويقوم الحاسوب بالرد عبر السماعات.

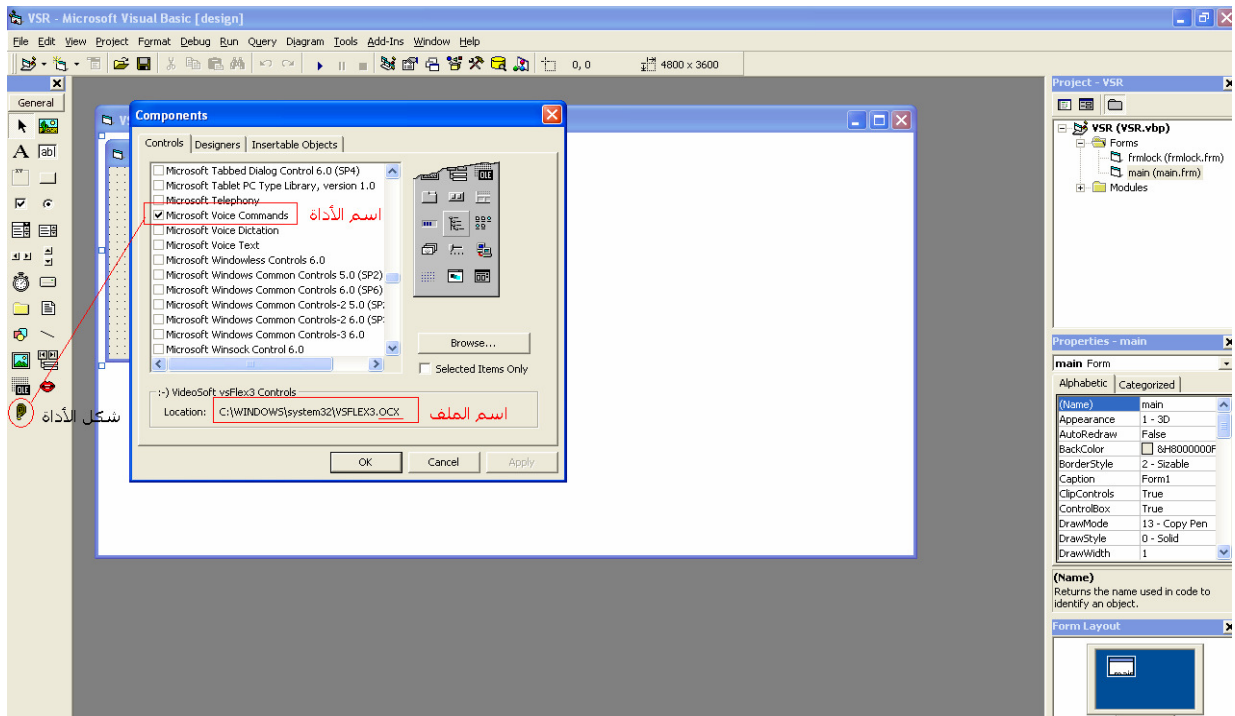
#### **\*\* كيفية العمل:**

فتح مشروع جديد في Visual Basic وإضافة محرك Microsoft للتعرف على الكلام .

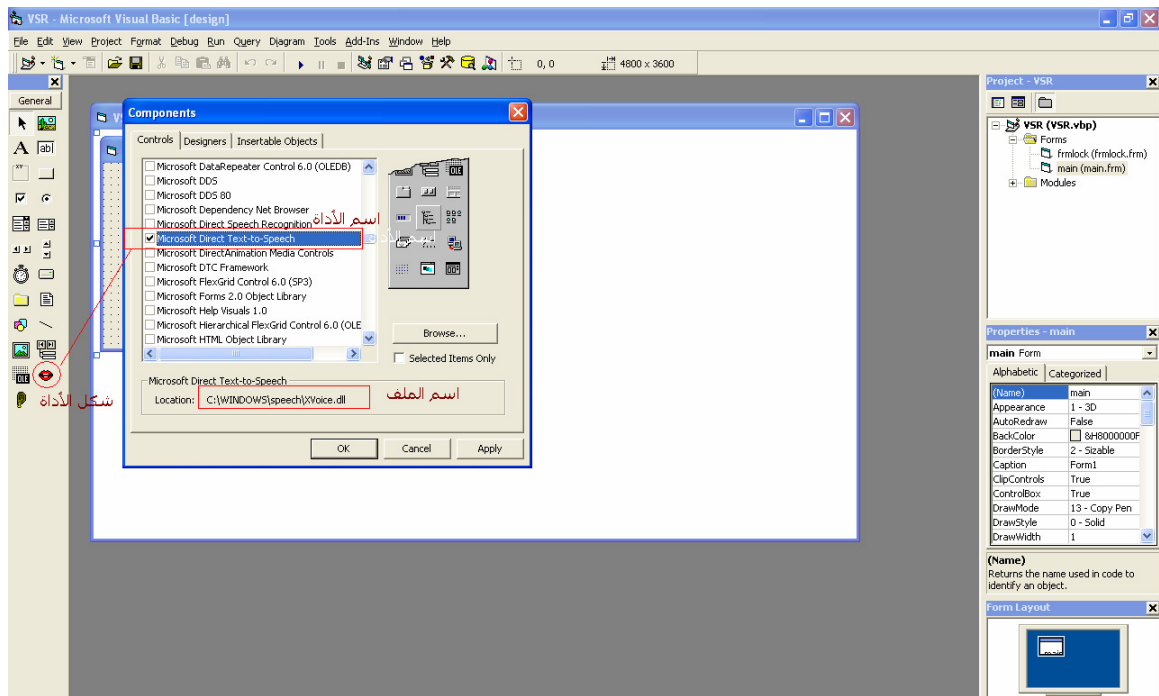
كما سبق فإن المحرك عبارة عن ملفات OCX , ActiveX وسنتعامل مع ميزتان فقط هما :

1/ Voice Commands : وهي تمكن من فهم الأوامر الصوتية و يمكن إضافتها من قائمة Component للأدوات البرمجية للغة Visual Basic وسيكون أسمها Microsoft Voice Commands ويظهر شكلها كشكل **أذن**

2/ Text to speech : وهي تمكن الحاسوب من الكلام أي تحويل النص إلى كلام ويتم إضافتها بنفس الطريقة ويكون أسمها Microsoft Direct Text to Speech ويكون شكلها يشبه **الشفاه** .



Voice Command



Text to speech

\*\* التعرف على الأمر :  
 لنجعل الحاسوب يفهم الأمر  
 1/ نضيف الأوامر التي سنتعامل بها مع الحاسوب إلى قائمة Menu .  
 2/ نعرف قائمة الأوامر إلى أداة Voice Command لتستطيع تمييز أي أمر تم النطق به من القائمة.

برمجياً يتم ذلك كالآتي :  
 بعد إضافة الأداة يكون الاسم البرمجي Name الافتراضي لها Vcommand1 نغيره من الخصائص إلى VO  
 فيكون الكود في حدث تحميل النموذج Form\_Load هو :

```
Private Sub Form_Load()  

    VO.Initialized = 1 'تهيئة الأداة  

    My_Menu = VO.MenuCreate(App.EXEName, "Commands", 4) 'إنشاء القائمة  

    VO.Enabled = 1 'تشغيل الأداة  

    ' كل الأسطر القادمة تضيف الأوامر إلى القائمة  

    VO.AddCommand My_Menu, 1, "hi COMPUTER", "hi", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "how are you", "how", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "open my computer", "my computer", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "note pad", "notepad", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "calculator", "calc", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "lock", "lock", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "what is the time", "time", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "say thanks", "thanks", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "end it", "end", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "dada", "call", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "do you listening", "listen", "commands", 0, ""  

    VO.AddCommand My_Menu, 1, "end lock", "endhg", "commands", 0, ""  

    'إضافة قائمة الأوامر إلى الأداة  

    VO.Activate My_Menu  

End Sub
```

وكهذا نكون قد أضفنا الأوامر الصوتية إلى القائمة والتي اسمها My\_Menu . ونكون قد جعلنا الأداة VO في حالة Active أي مستعدة لتلقي الأوامر.

الآن عند النطق بأحد الأوامر من القائمة عن طريق الميكروفون تطلق الأداة VO حدث CommandRecognize وتجعل قيمة المتغير command تساوي الأمر المتعرف عليه وصيغة هذا الحدث هي :

```
VO.CommandRecognize(ByVal ID As Long, ByVal CmdName As String, ByVal Flags As Long,  

    ByVal Action As String, ByVal NumLists As Long, ByVal ListValues As String, ByVal command As  

    String)
```

الآن نختبر قيمة المتغير command بواسطة عبارة Select Case ونضع في كل أمر ما نريد أن يفعله البرنامج

وفيما يلي كود هذه العملية

**Private Sub VO\_CommandRecognize(ByVal ID As Long, ByVal CmdName As String, ByVal Flags As Long, ByVal Action As String, ByVal NumLists As Long, ByVal ListValues As String, ByVal command As String)**

**Select Case UCase(command)**

الدالة Ucase لجعل كل الحروف كابيتال (حروف كبيرة)

**Case "HI COMPUTER"**

هنا نضع كود ما نريد عمله عندما يكون الأمر المنطوق hi computer

**Case "HOW ARE YOU"**

هنا نضع كود ما نريد عمله

**Case "OPEN MY COMPUTER"**

هنا نضع كود ما نريد عمله

**Case "NOTE PAD"**

هنا نضع كود ما نريد عمله

**Case "CALCULATOR"**

هنا نضع كود ما نريد عمله

**Case "LOCK"**

هنا نضع كود ما نريد عمله

**Case "WHAT IS THE TIME"**

هنا نضع كود ما نريد عمله

**Case "SAY THANKS"**

هنا نضع كود ما نريد عمله

**Case "END IT"**

هنا نضع كود ما نريد عمله

**Case "DADA"**

هنا نضع كود ما نريد عمله

**Case "DO YOU LISTENING"**

هنا نضع كود ما نريد عمله

**Case "END LOCK"**

**End Select**

**End Sub**

وهذا هو مثال بسيط فقط .

## **\*\*محركات التعرف على الكلام باللغة العربية:**

للأسف مازالت المحركات العربية قيد الدراسة والتطوير وشركات قليلة هي التي نجحت في ذلك ومثال لهذه الشركات شركة صخر التي أنتجت أول محرك عربي وهو قيد الدراسة وتسعى الشركة للكسب المالي البحت من وراء هذا المحرك ولقد أسمته ASR وهناك أيضا شركة Acapela والتي أنتجت محرك لنطق الكلام العربي . كما أن شركة IBM لها بحوث أيضا في هذا المجال. يجدر الإشارة إلى أن عدة دراسات وأوراق قدمت في هذا المجال.

