١٢ ـ كيف تبنى "روبوت" حقيقى؟

فهـــم الكـــلام Speech Understanding

د. عـــلاء خميس

كلية هندسة البترول - جامعة قناة السويس

كما أوضحنا فى المقال السابق.. فإن عـمليـة الاتصـال الصـوتى تتـضـمن عمليتين أساسـيتين هما.. تخليق الكلام Speech Synthesis .. وإدراك وفـهم الكـلام Speech Recognition and Understanding. وفى هـذا المقـــال.. نستكمل عرض منظومات تخليق الكلام بشرح كيفيـة إضافة وحدة تخليق كلام إلى الروبـوت وبعـد ذلك نسـتـكمل مـوضوع الاتـصال الصـوتى بشـرح تقنيات إدراك وفهم الكلام.

إضافة وحدة تخليق إلى الروبوت

يمكن إكـسـاب الروبوت القـدرة على تخليق الـكلام باسـتـخـدام الـرقـاقـة الالكتـــرونيـــة SP0256-AL2 التى أنتـجـتـها شـركـة «General Instruments» وهى عبارة عن دائرة متكاملة 28- pin تعمل بجهد ٥ ف يتم توصيله على نقـطتى التـوصـيـل ١ و ٧ . يمكن توصيل هذه الرقـاقة مبـاشرة بخطوط البـيانات والـتحكم فى مـيناء توصـيل توازى كالمستخـدم فى توصيل الطابعة بالحاسب Parallel Printer Port .

وكـمـا هو مـبين بالشكل رقم (١).. يتم ربط خطوط البـيـانات من A1 إلى A6 فى الرقاقة بخطوط خـرج البيانات (٠) إلى (٥) . ولإظهار «فونيم».. يضع الحـاسـب المزود به الروبوت شـفـرة عـشـريـة مـا بـين (٠) إلى (٦٤) على خطوط البيانات فى ميناء التوصيل.. ثم يعطى الخط STROBE قـيـمـة LOW. يعـمل هذا الخط على بدء تشـغـيل نقطة التـوصـيـل على الرقاقـة للإشـارة إلى وصول بـيـانات.. وعندئذ تقوم الرقـاقة بنطق «الفـونيم». بعد إرسال «الـفونيم» يقوم الحاسـب بسرعة باختبار وإعادة اختبار حـالة خط التوصيل BUSY فى مـيناء التوصـيـل المتـوازى. يلاحظ.. أن خط التـوصيل BUSY متـصل بنقطة التوصـيـل فى الرقاقة وهى التى تتغـير حـالتـهـا إلى HIGH خـــلال نطق «الفُونيم».. وبالتالى لا يستطيع الحاسب إرسال «فونيم» جديد الا بعد الانتهاء من نطق «الفـونيم» الـحـالى. تسـتـمـر هذه العـملية حـتى يتم الانتهـاء من نطق كل «الفُونيمـات». يوضح الجدول المبين على الشكل رقم (١) المكونات الـضـــروريـة لوحدة تخليق الكلام فى الروبوت.

بالاسـتـعانة بجـدول بيـانات Data Sheet الرقـاقة المحتـوى على شـفرات «الفُـونيمـات» لعدد من الكلمـات.. يمكن كتابة بَرنامج BASIC بسيط كالموضح بالشكـل رقم (٢) لجـعل الدائـرة تنطق بكلمـات مثل _ one _ two _ testing _ three - four_ وذلك بغرض التـجريب. فى حـــالة حـدوث عطل.. يجـب إعـادة اختـبار الدائـرة وبصفـة خاصة نـقاط التوصـيـل بين الرقاقـة وميناء تـوصيل التوازى. وفى حالة صدور ضوضاء من الدائرة.. يجب فـصـل الـتـغـذية وإعـادة الـتـشـغـيل. بإضـافـة الدائرة المـتكاملة CTS256-AL2 التى قـامت «General Instruments» بإنتـاجـهـا أيضًـا إلى الرقاقة SP0256.. يمكن بناء دائرة لها القدرة علـى تحويل النصوص المـكتوبة إلى حديث مسموع كـتلك المستخدمه فى برامج الترجمة وتعلم اللغات.

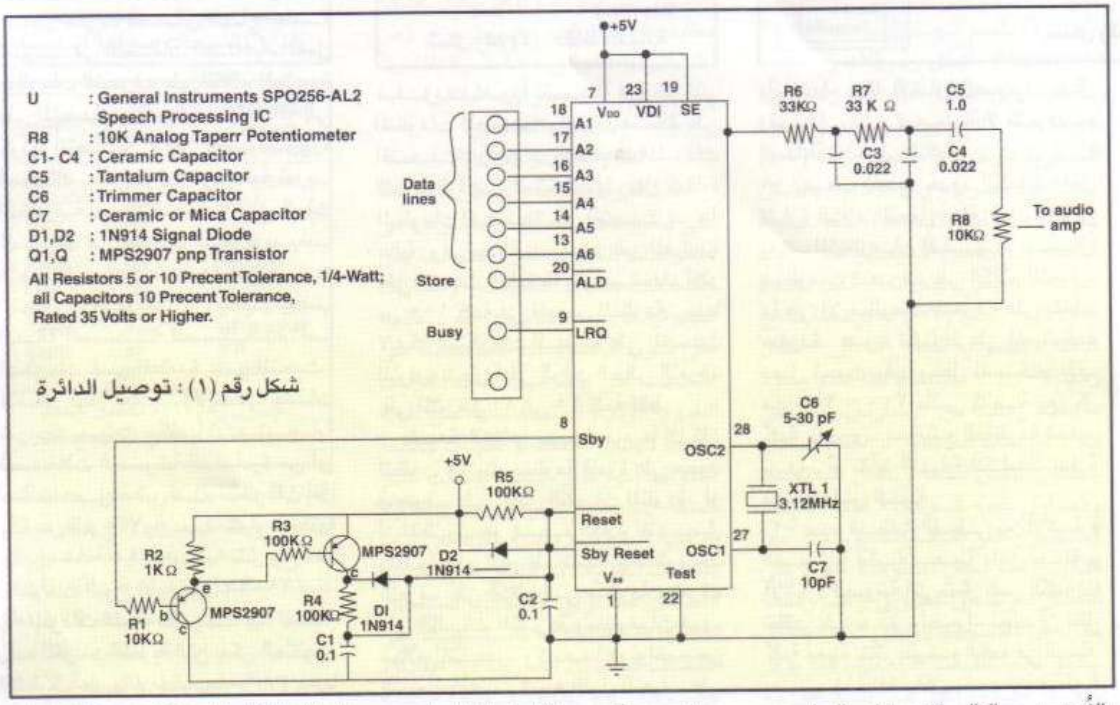
إدراك وفهـم الكـلام

تعتبر عمليـة إدراك وفهم الكلام.. عملية عكسـية لعملية تخلـيـق الكلام (أنظر العدد الـسـابق) تمـكن المنظومـة الروبـوتيـة من تفسـير وفهم مـا يقال.. مما يسـاعد على استخدام التخاطب الصوتى كوسيلة تفاعل مع الروبوت أو وسـيلة إدخـال للحـاسب تفوق فى قدرتها وسائل الإدخال التقليدية مثل لوحـة المفاتيح والماوس.. وتمكن ذوى الاحـتـيـاجات الخـاصـة من التفـاعل مع المنظومات الحاسوبية بسهولة ويسر. قبل البدء فى شرح كـيفية فـهم الكلام بطريقة آلية.. يجب الإشارة إلـى أن هناك ثمة فرق بين عـمليـتـى الـتـعـرف أو إدراك الكلام Speech Recognition وعـمليـة فـهم الـكـلام Speech Understanding.. إذ تعـتـبـر العمليـة الأولى أبسط بكثـيـر من الثانية.. حيث يتطلـب من المنظومة التعرف على بعض المفـردات اللغـوية أو الكلمـات المنفـــردة Isolated-Words.. أمـــا فى العمليـة الثانيـة فيجب أن تكـون المنظومة قادرة على فهم الكلام المتصل Connect-ed-Speech والغـــرض من الجـــمل المنطوقـة. ويتوفـر حاليًـا شرائح تسـمى شرائح الكلام Speech Chips لها القدرة على إجراء عـمليتى تخليق والتـعرف على الكلام. وباعـتـبـار أن كـلـتـا الـعـمليـتين تسـتـخـدمان طريقـة الـتـشـفـير التنبـؤى الخطى LPC (أنظر العدد السابق) لضغط البيانات لتقليل السعـة التخزينية المطلوبة.. فقد تم إدماج وحدة تخليق LPC مع محلل سمعى LPC لإنتاج هذه الشرائح. يوضح الشكل رقم (٣) الوحـدات الأسـاسـيـة فى شريحة الكلام.

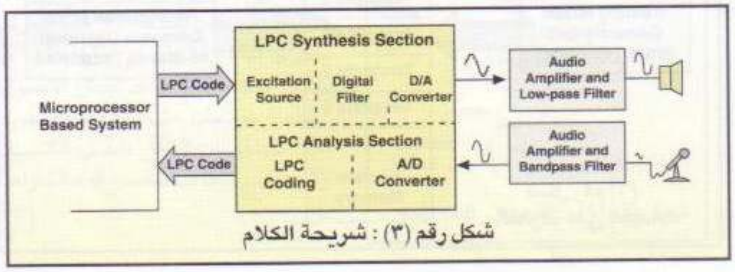
منظومات التعرف على الكلام

كـمـا هو الحـال فى كـل المجـالات الهندسـية.. يجب البـدء بتحـديد طبيـعة التطبيـق قبل الشروع فى بناء مـنظومة التعرف على الكلام لضمان بناء منظومة تحقق الغرض المطلوب وفى نفس الوقت تكون اقتـصادية. تحدد المـعايير التـالية المدى الذى يمكن اسـتـخـدام منظومـة التعرف على الكلام فيه:

– الاعـتـمـاد على المتـحـدث: يحدد هذا المعـيار مـدى اعتـمـاد المنظومة أو عـدم اعتمادها على طبيـعة صوت المتحدث أو قابليـتها للتعـامل مع أكثر من متـحدث. لذا.. نجـد أن هذا المعيار يحدد طبـيعـة المنظومـة من حيث اعتـمادها أو عـدم اعـتـمـادها على المتـحـدث Speaker-Dependent or Independent.. ومن



شكل رقم (١) : توصيل الدائرة



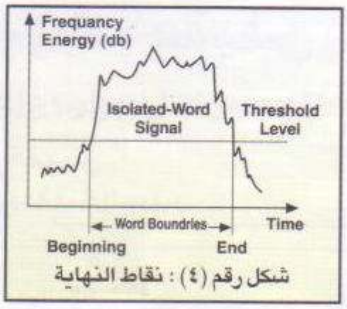
شكل رقم (٣) : شريحة الكلام

```
 5  CLS
10  Y=0
20  READ X
30  IF X=100 THEN GOTO 100
40  OUT 888, X
50  Y=INP ( 889 )
60  IF Y=143 THEN GOTO 70 ELSE 50
70  OUT 890,127
80  OUT 890,128
90  GOTO 20
100 OUT 888,0
110 OUT 890,127
120 OUT 890,128
130 REM Testing
140 DATA 13,7,55,2,13,12,44,4,4
150 REM One
160 DATA 46,15,15,11,2
170 REM Two
180 DATA 13,21,2
190 REM Three
200 DATA 29,14,19,2
210 REM Four
220 DATA 40,40,58,1,1,1,1,100
```

شكل رقم (٢) : برنامج Basic

سوف يستخدم المنظومة . فى المنظومات المعتمدة على المتحدث.. يجب أن يقوم المستخدم بتدريب المنظومة على التعرف على صوته من خلال جلسات تدريب يتم فيها نطق كل كلمة يجب على المنظومة التعرف عليها لاحقاً على الأقل خمس مرات لضمان نجاح جلسة التدريب. فى التطبيقات الصناعية.. تعمل منظومات التعرف على الكلام المعتمدة على المتحدث بدقة عالية بالإضافة إلى كونها اقتصادية.. ولكن يعيبها ضرورة استهلاك بعض الوقت فى عملية تدريب المنظومة. وبالإضافة إلى الوقت المستهلك فى جلسات التدريب.. يعيب هذه المنظومات اعتمادها على صوت المتحدث والذي يمكن أن يتغير حسب الحالة الصحية للمستخدم. أما المنظومات غير المعتمدة على المتحدث.. فتعتبر أكثر تعقيداً ولاتتطلب جلسات تدريب.. حيث يمكنها الاستجابة للمدخلات الصوتية الصادرة من أى مستخدم. وتعتبر هذه المنظومات أقل دقة من النوع الأول حيث تتراوح دقتها ما بين ٨٥– ٩٥٪.. بينما تتراوح دقة منظومات النوع الأول بين ٩١–٩٩٪.

– **المفردات اللغوية:** يحدد هذا المعيار عدد المفردات اللغوية التى يمكن للمنظومة التعرف عليها أو فهمها. بزيادة عدد المفردات اللغوية.. تزداد درجة تعقد المنظومة والذاكرة المطلوبة وزمن الاستجابة والتكلفة بصورة مطردة.. لذا ينصح دائماً باستخدام اقل عدد ممكن من المفردات اللغوية التى تمكن المنظومة الروبوتية من أداء عملها. ولعله من غير المنطقى تصميم منظومة للتعرف على مئات الكلمات.. بينما يكفى التعرف على بضع كلمات فقط لإتمام عملية الاتصال الصوتى بين الإنسان والروبوت فى الحدود المطلوبة. فعلى سبيل المثال.. يمكن لعدة تطبيقات صناعية الاعتماد على التعرف على كلمات بسيطة مثل «نعم» و«لا» و «إبدأ» و «قف» والأرقام من صفر إلى ٩ لأداء المهمة. يمكن استخدام مفردات لغوية بسيطة فى حالتى المنظومات المعتمدة وغير المعتمدة على المتحدث.. ولكن بزيادة عدد المفردات اللغوية الواجب استخدامها.. تقل كفاءة المنظومات المعتمدة على المتحدث بدرجة عالية. وفى حالة المنظومات غير المعتمدة على المتحدث.. عادة ما يتم استخدام أكثر من ١٠٠ كلمة فى قاموس المنظومة.. بينما لايمكن للمنظومات المعتمدة على المتحدث المتوفره تجارياً فى الوقت الحالى التعرف على اكثر من ١٠ الى ٢٥ كلمة فقط.



شكل رقم (٤) : نقاط النهاية

– **طبيعة الكلام:** يجب تحديد ما إذا كان المطلوب هو بناء منظومة قادرة على مجرد التعرف على بعض الكلمات المنفردة.. أو أن المطلوب هو فهم الكلام المتصل والمتواصل. لذا.. يتم تقسيم منظومات التعرف على الكلام إلى منظومات تعرف على الكلمات المنفردة ومنظومات فهم الكلام المتصل. وكما هو واضح من التسمية.. يصمم النوع الأول للتعرف على بعض الكلمات.. ويمكن لهذه المنظومات أن تكون معتمدة أو غير معتمدة على المتحدث. ويتم استخدام مثل هذا النوع فى كثير من التطبيقات مثل الهواتف المحمولة. وفى النوع الثانى.. تسمى الإشارة الصوتية المحتوية على جمل كاملة وعبارات بالكلام المستمر أو المتصل أو الطبيعى. وباعتبار أن هذا الكلام المتصل يتكون من عدة كلمات منفردة.. يمكن تفسير وفهم هذا الكلام بتقسيم الإشارة الصوتية الممثلة له إلى عدة إشارات تمثل كلمات منفردة.. وعندئذ يمكن استخدام تقنيات التعرف على الكلمات المنفردة لفهم الجملة. فى معظم منظومات التعرف على الكلام.. يتم تحديد نقاط النهاية للكلمة باستخدام طاقة الإشارة الصوتية –شكل رقم (٤). ولكن العملية لا تتم بهذه البساطة للأسباب التالية:

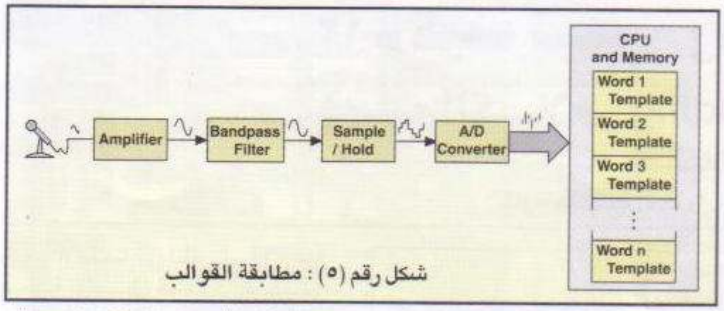
١– يصعب تحديد حدود الكلمات داخل إشارة الكلام المتصل بدقة. لذا.. نجد أن بعض المنظومات المسماة مجازاً بمنظومات التعرف على الكلام المتصل.. ما هى إلا منظومات تعرف على كلمات منفردة.. حيث تشترط على المستخدم معدل تحدث مقيد يجعل المستخدم يقف من ٢٠٠ – ٣٠٠ مللى ثانية بين كل كلمة وأخرى.. مما يتيح للمنظومة تحديد حدود أو نقاط النهاية للكلمات بدقة مقبولة داخل الجملة.

٢– عدم التطابق الفعلى بين الإشارة الصوتية لكلمة معينة داخل جملة و الإشارة الصوتية التى تمثل نفس الكلمة إذا نطقت على حدة. وأخيراً.. نجد أن نطق كلمة معينة يتأثر بموضع الكلمة فى الجملة. بسبب هذه المشاكل.. نجد أن عملية التعرف أو فهم الكلام المتصل.. عملية صعبة بالمقارنة بعملية التعرف على الكلمات المنفردة.

التعرف على الكلمات المنفردة

تعتمد منظومات التعرف على الكلمات المنفردة والمعتمدة على المتحدث.. على مطابقة الاشارات الصوتية للكلمات مع قوالب مختزنة Template. لذا.. تسمى هذه المنظومات فى بعض الأحيان منظومات مطابقة الاشارة Signal-Matching أو مطابقة القوالب Tem-plate-Matching. وتعانى منظومات مطابقة القوالب غير المعتمدة على المتحدث من مشكلة اختلاف الإشارات الصوتية من مستخدم إلى آخر. ولحل هذه المشكلة.. يتم استخدام تقنيات التحليل الصوتى Phonetic Analysis والذى يتم من خلالها تقسيم الصوت إلى عدة تقسيمات صوتية تستخدم فى تحديد مجموعة صغيرة من الكلمات قريبة الشبه.. وتتم العملية بشكل تكرارى حتى يتم مطابقة الكلمة بالكامل.

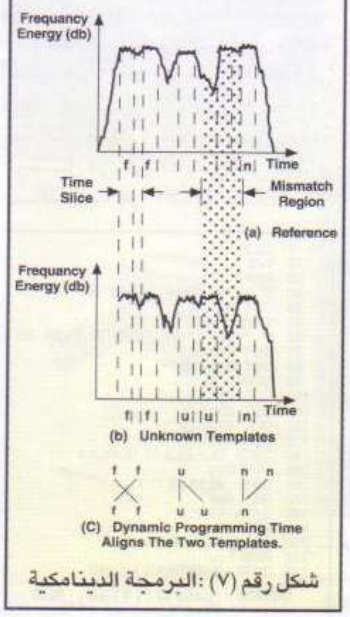
– **مطابقة القوالب Template Matching :** تتميز هذه الطريقة – شكل رقم (٥) – بالكفاءة العالية والدقة بالإضافة إلى انخفاض تكلفة بناء المنظومة نظراً لبساطتها.. حيث تتكون الوحدة من ميكروفون ومكبر ومرشح ودائرة أخذ عينات وإيقاف ومحول تناظرى/رقمى. ويعتبر الميكروفون أهم جزء فى الوحدة.. مما يتطلب فيه أن يتسم بنسبة إشارة إلى ضوضاء عالية (على الأقل ٢٥ ديسبل) لتقليل احتمال الخطأ فى التعرف على الكلمة نتيجة الضوضاء المحيطة.



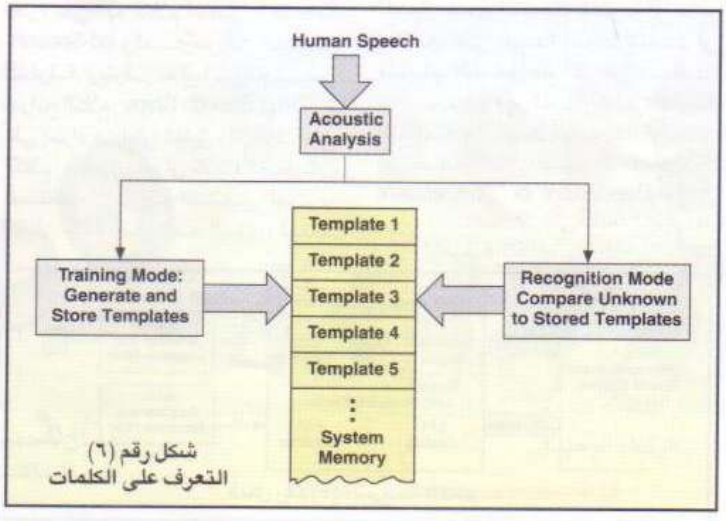
شكل رقم (٥) : مطابقة القوالب

وتعتمد فكرة العمل.. على تكبير الإشارة الصوتية الصادرة من الميكروفون وترشيحها للحصول على إشارة فى مدى تردد بين ٢٠٠ – ٣٢٠٠ هرتز.. وذلك لزيادة مستوى الإشارة ولإزالة الضوضاء. وباستخدام دائرة أخذ العينات والإيقاف والمحول التناظرى/الرقمى (٨ بت).. يتم تحويل هذه الإشارة التناظرية إلى رقمية. وهنا.. نلاحظ أن معدل أخذ العينات يجب أن يكون على الأقل ضعف أعلى تردد لمركبات الإشارة.. أى ٢ x ٣٢٠٠ أى ٦٤٠٠ عينة / ث.. وعادة ما يتم استخدام معدل أخذ عينات يتراوح ما بين ٨٠٠٠ – ١٢٠٠٠ عينة / ث. وبمجرد تحويل الاشارة الصوتية إلى صورة رقمية.. يمكن استخدام الحاسب والاستعانة ببعض البرامج الخاصة لإتمام عملية التعرف على الكلمة.. حيث يتم إجراء عمليتين أساسيتين – شكل رقم (٦)– هما.. التدريب Training والتعرف Recognition.

فى مرحلة التدريب.. يتم القيام بعملية تحليل سمعى Acoustic Analysis وتوليد قوالب. وهنا.. يجب على المستخدم نطق المفرد اللغوى أمام المنظومة عدة مرات لضمان نجاح عملية التدريب.. ثم تقوم المنظومة بتحليل



شكل رقم (٧) : البرمجة الدينامكية



شكل رقم (٦)
التعرف على الكلمات

إشارة الصوت وتوليد كلمات يتم تخزينها فى الذاكرة ليتم استخدامها لاحقاً فى مطابقة النطق غير المعلوم. وخلال جلسة التدريب.. يقوم المحول التناظرى/الرقمى بتحويل الإشارة بمعدل ثابت. ولإتمام عملية التشفير.. عادة ما تستخدم طريقة التشفير التنبؤى الخطى Linear Predictive Coding (LPC) التى لاتتطلب سعة تخزينية كبيرة. وبمجرد انتهاء عملية التشفير.. يتم تخزين نماذج الكلمات فى ذاكرة RAM . وطبقاً لنوع التشفير المستخدم.. يمكن أن يتطلب قالب كلمة واحدة ما بين ١٠ - ١٠٠٠ بايت من الذاكرة. وفى حالة استخدام طريقة LPC.. تكون السعة المطلوبة ١٠٠ بايت لكل كلمة.. ويتم تخزين البيانات فى مصفوفة يتم استخدامها لمطابقة الكلمات المنطوقة مع القوالب التى تم تخزينها. فى مرحلة التعرف.. يتم تحويل وتحليل الاشارة الصوتية كما فى مرحلة التدريب. ثم يتم تخزين البيانات المشفرة باستخدام LPC بصورة مؤقتة فى مصفوفة لاستخدامها فى المقارنة مع قوالب تم توليدها خلال مرحلة التدريب وتؤخذ كمرجع. وتتم عملية المقارنة بشكل مباشر عن طريق حساب معامل الاختلاف بين المرجع والمصفوفة غير المعلومة.. وهو مجموع مربعات الفروق بين العناصر المتناظرة فى كلتا المصفوفتين.. ويحدث التطابق عند أصغر معامل اختلاف. وفى الحالة المثالية.. يجب مقارنة المصفوفة غير المعلومة مع كل المصفوفات المرجعية المتوفرة للحصول على أصغر معامل اختلاف ممكن.. والذي يكون على حساب زمن الاستجابة. لذا.. عادة ما يتم استخدام لوغاريتمات بحث خاصة لتقليل زمن الاستجابة مع ضمان نسبة خطأ مقبولة فى التعرف على الكلمة. ومن اللوغاريتمات الشائع استخدامها فى المقارنة.. لوغاريتم البرمجة الديناميكية Dynamic Programming بالإضافة إلى تقنية تسمى القاعدة الزمنية أو Time-Warping لتقليل أخطاء التعرف على الكلمات وحل المشكلة المعروفة باسم المحاذاة الزمنية Time Alignment - شكل رقم (٧).

بفرض أن الكلمة المراد التعرف عليها هى كلمة _fun_ وبفرض أن الإشارة التناظرية الموضحة بالشكل رقم (٧-أ) تستخدم كمرجع وأن الاشارة بالشكل رقم (٧-ب) تمثل الاشارة الصوتية للكلمة فى بعض الأحيان عند توليد القالب المرجعى.. يتم إنتاج قالب لكلمة _ffunn_ -شكل (٧-أ) - بينما تنتج الإشارة الصوتية كلمة _ffuun_ - شكل رقم (٧-ب) - مما ينتج عنه عدم وجود محاذاة زمنية بين الإشارتين وبالتالى نجد أن المنظومة يمكن أن تنتج قالبين مختلفين لنفس الكلمة ولنفس المستخدم. وتحاول تقنية البرمجة الديناميكية.. الحصول على افضل تطابق ممكن بين القالبين مما يجعل المنظومة قادرة على التعامل مع المدخل الصوتى وعدم رفضه لمجرد أن المستخدم لاينطق الكلمة بشكل مطابق تماماً للقالب المرجعى.. مما يوفر مرونة وقدرة أعلى فى التعامل مع المدخلات. يمكن تمثيل البرمجة الديناميكية باستخدام خط المحاذاة الزمنية الموضح بالشكل رقم (٧-ج).. حيث يلاحظ قدرة هذه التقنية على التعامل مع الاختلافات بين القالبين.. ويتم ذلك باستخدام لوغاريتم مطابقة القوالب.. وهو عبارة عن تحليل مصفوفى يعتمد على حساب كل التوافقيات الممكنة للمحاذاة الزمنية بين القوالب المرجعية والقوالب المجهولة للحصول على أفضل تطابق ممكن. من أكبر مزايا البرمجة الديناميكية.. عدم الحاجة إلى معرفة نقاط النهاية للكلمات كما هو الحال فى مقارنة المصفوفات المباشرة. أما أبرز عيوبها.. فيتمثل في بطء زمن الاستجابة نتيجة الحسابات المعقدة التى يتم إجراؤها خلال عملية مطابقة القوالب.. حيث يزيد زمن الاستجابة بحوالى ١٠٠ مرة بالمقارنة بطريقة مقارنة المصفوفات المباشرة.

التحليل الصوتى
Phonetic Analysis

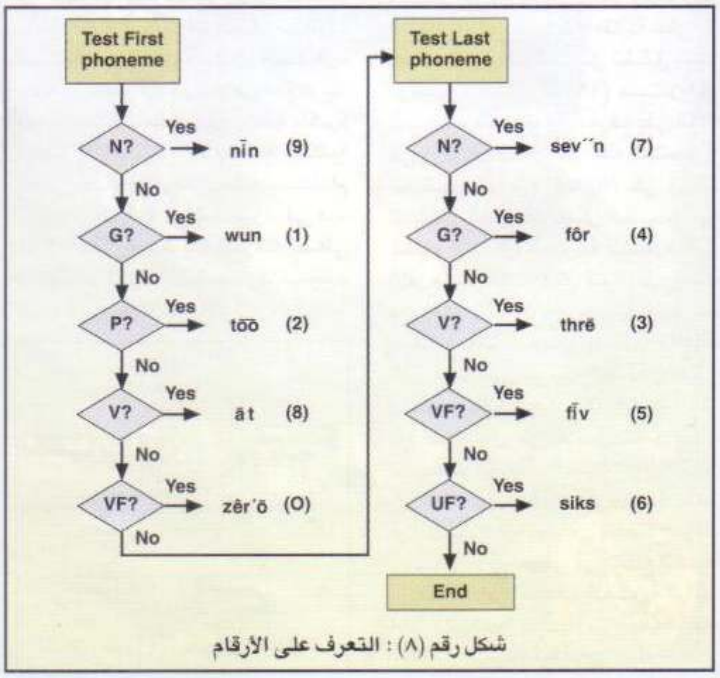
كما ذكرنا سالفاً.. لايمكن الاعتماد على تقنيات مطابقة القوالب فى منظومات التعرف غير المعتمدة على المتحدث بسبب اختلاف الإشارات الصوتية من مستخدم إلى آخر.. علاوة على كون المنظومات المعتمدة على المتحدث التى يزيد فيها عدد المفردات اللغوية اللازم استخدامها على عدة مئات.. تجعل استخدام تقنيات مطابقة القوالب أمراً غير عملى بسبب كبر السعة التخزينية المطلوبة والوقت اللازم لعملية التدريب. وللتعامل مع كثرة المفردات اللغوية.. تم ابتكار تقنية تسمى التحليل الصوتى أو تحليل السمات Feature Analysis.. تعتمد على اكتشاف السمات الصوتية فى الإشارة الصوتية.. يتم بعدها ترجمة الكلمة المنطوقة إلى مجموعة من التقسيمات الصوتية تساعد على تحديد مجموعة صغيرة من الكلمات الممكنة يتم البحث فيها حتى الوصول إلى المطابقة الكاملة. لتوضيح ذلك.. يمكن افتراض أنه تم تقسيم الاصوات إلى ستة تقسيمات صوتية أساسية - الجدول رقم (١). وبفرض تمثيل الأرقام من صفر إلى ٩ باستخدام هذه التقسيمات.. يصبح من الممكن الحصول على تمثيل جديد كما هومبين بالجدول رقم (٢). وبفرض أن منظومة التعرف غير المعتمدة على المتحدث يمكنها التعرف على الأرقام من صفر إلى ٩ وأن المستخدم قام بنطق رقم "واحد" أمام المنظومة.. فإن لوغاريتم التحليل يقوم بترجمة الإشارة الصوتية إلى تتابع صوتى - الجدول رقم (٢) - ثم يقوم لوغاريتم اتخاذ القرار بتحديد أى من العشرة أرقام مطابق للإشارة الصوتية كما هو مبين بخريطة السريان - شكل رقم (٨).

Pure Voiced Vowel (V)	:	a,e,i,o,u,uh,aa, ee,er,uu,ar,aw
Nasal (N)	:	m,n,ng
Voiced Fricative (Vf)	:	z,zh,v,dh
Unvoiced Fricative (Uf)	:	s,sh,f,th
Plosive (P)	:	b,d,g,p,t,k,h
Glide (G)	:	r,w,l,y

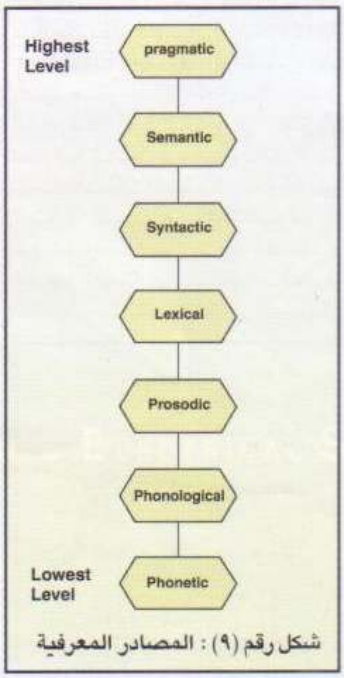
جدول رقم (١) : تقسيم الكلام إلى ٦ تقسيمات صوتية أساسية

Digit	Dictionary Pronunciation	Phonetic Sound Sequence
0 Zero	zêr´ō	VF-V-G-V
1 One	wun	G-V-N
2 Two	tōō	P-V
3 Three	thrē	UF-G-V
4 Four	fôr	UF-V-G
5 Five	fīv	UF-V-VF
6 Six	siks	UF-V-P-UF
7 Seven	sev´´n	UF-V-VF-N
8 Eight	āt	V-P
9 Nine	nīn	N-V-N

جدول رقم (٢) : تمثيل جديد باستخدام نفس التقسيمات



شكل رقم (٨) : التعرف على الأرقام



شكل رقم (٩) : المصادر المعرفية

فهم الكلام المتصل

تعتبر عملية تفسير وفهم الكلام المتصل أو المتواصل عملية صعبة لازالت فى طور البحث والتطوير اعتمادا على تقنيات الذكاء الاصطناعى. لإكساب منظومة روبوتية القدرة على فهم الكلام.. لابد من التعامل مع قاعدة معرفة تتضمن معارف شتى عن الكلام. ويمكن تقسيم مصادر هذه المعارف إلى سبعة مصادر - شكل رقم (٩).

١- المعرفة الصوتية Phonetic Knowledge : تتضمن هذه المعرفة الأصوات الأساسية للكلام التى تحتوى على «الفونيمات» وتباينات النطق Allo-phones (أنظر العدد السابق) والتى تشتق خلال مرحلة التحليل السمعى للإشارات الصوتية كما هو الحال فى منظومات التعرف على الكلمات المنفردة.

٢- المعرفة النطقية Phonological Knowledge : تستخدم هذه المعرفة لتحويل تباينات النطق إلى «فونيمات» من خلال استخدام قواعد تمثل الاصوات الكلامية بعلامات متميزة لتحديد طريقة نطق تباينات النطق داخل الكلمات والجمل.. كما تتضمن هذه القواعد

اختلافات النطق بسبب اختلاف اللهجة.

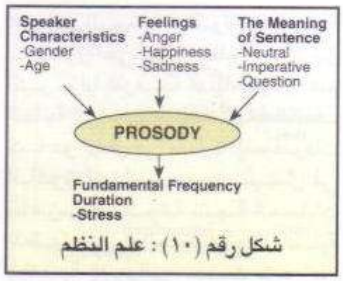
٣- معارف النظم Prosodic Knowledge: تتضمن هذه المعارف معلومات عن شدة الصوت والترنيم والتوقيت التى تساعد على فهم الكلام أو الصوت المتصل واكتشاف حدود الكلمات داخل الجمل. وكما هو مبين بالشكل رقم (١٠).. يوفر علم النظم Prosody كثيراً من المعلومات المشتقة من تحليل الإشارة الصوتية والتى تساعد على معرفة عمر المتحدث وجنسه وحالته النفسية.

٤- المعرفة المعجمية Lexical Knowledge : هى المعرفة المتعلقة بالمعنى المشترك بين الكلمات المشتقة من أصل واحد برغم اختلافها فى الصورة وتتضمن طرق النطق المختلفة للكلمات والتى تختلف باختلاف الموضوع وتركيب الجملة وعلامات الترقيم أو الوقف. تتضمن هذه المعرفة أيضاً قواعد تحدد كون الكلمة جمعاً أو مفرداً وزمن الفعل وضمائر الملكية إلخ... .

٥- المعرفة التركيبية Syntactic Knowledge : يستخدم علم التركيب Syntax لوصف بنيان الجملة أو العبارة وترتيب كلمات الجملة فى أشكالها وعلاقاتها الصحيحة. لذا.. تستخدم القواعد المعرفية المتعلقة بتراكيب الكلام فى منظومات فهم الكلام لوصف التوافقات المسموحه للكلمات التى يمكن ان تستخدم لتكوين جملة أو عبارة.. مما يجعل هذه المعرفة يطلق عليها أحياناً معرفة التقييد Constraining Knowledge. على سبيل المثال.. عبارة مثل _attack the operator! لايمكن أن تكون بأى حال توافقة كلمات مسموحة لمنظومة روبوتية صناعية ولكن يمكن قبولها مثلاً فى تطبيقات عسكرية. بمساعدة هذا النوع من المعارف.. يمكن للمنظومة توقع أو تخمين كلمة لتكملة جملة غير منطوقة بالكامل مثل توقع الأمر _pick and place عند إعطاء أمر غير كامل مثل __ pick and_.

٦- معارف المعان Semantic Knowledge : تتضمن هذه المعارف المعانى الحقيقية للكلمات والجمل ودلالات الألفاظ وتطورها. ويمكن استخدامها كمعرفة تقييد أيضاً بإهمال الجمل والعبارات التى لاتحمل أى معنى على الرغم من كونها صحيحة من الناحية النحوية مثل الأمر _pick the cart and place it on the part_ والترتيب الصحيح للكلمات هو _pick the part and place it on the cart_.

٧- المعرفة الموضوعية Pragmatic Knowledge : يعتبر هذا النوع من المعارف أعلى مستويات المعرفة المستخدمة فى منظومات فهم الكلام ويحتوى على قواعد عن المحادثات العادية ومعلومات عن الحوار. على



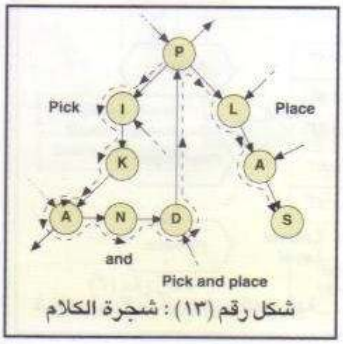
شكل رقم (١٠) : علم النظم

سبيل المثال.. يمكن لهذه القواعد أن تحدد لمنظومة فهم الكلام أن الإجابة على السؤال "ما عمرك؟".. يجب أن تكون إجابة عددية بحيث لا يمكن قبول إجابة مثل "نعم" أو "لا" لمواصلة الحوار.

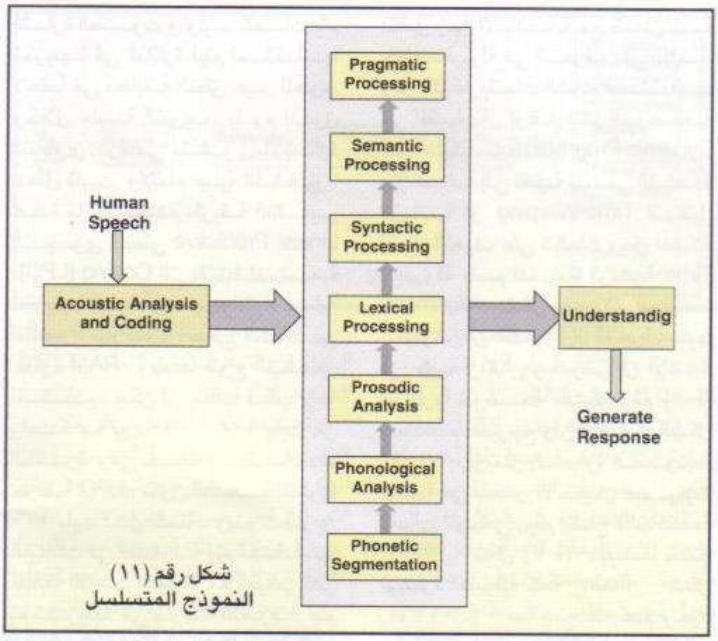
وحتى يتم الاستفادة من هذه المصادر المعرفية.. لابد من تنظيمها فى قاعدة معرفية واختيار طرق مناسبة للبحث والتحكم فى هذه القاعدة بهدف الاستخدام الامثل لها. توجد نماذج عديدة لتنظيم المعارف.. منها النموذج المتسلسل Hierarchical Model ونموذج السبورة Blackboard Model ونموذج الشبكة المصنفة Compiled-Network Model.

يعتبر النموذج الأول – شكل رقم (١١) – أبسط نماذج تمثيل المعرفة الكلامية.. ويتم فيه استخدام المعرفة الصوتية فى المستوى الأسفل إلى أن نصل إلى المعرفة العملية فى المستوى الأعلى باستخدام استراتيجيات استدلال مثل الاستدلال الأمامى Forward Reasoning أو الاستدلال الارتجاعى Backward Reasoning. فى الطريقة الأولى.. يتم البدء بالتحليل السمعى للاشارة الصوتية لتحديد الخواص الصوتية اللازمة لقاعدة المعرفة الصوتية.. ثم يتم استخدام المعرفة النطقية لمزيد من التحليل والتفسير. وتتواصل هذه العملية بشكل متسلسل كما هو مبين بالشكل رقم (١١) .

ورغم بساطة هذه الطريقة.. تشوبها عدة عيوب منها.. إمكانية حدوث ما يطلق علية بالانفجار التوافيقى Combinatorial Explosion عند عدم القدرة على تفسير مدخل صوتى معين مما يستدعى إجراء عملية بحث متكرر فى مستويات المعرفة السفلى.. وعدم القدرة على استخدام المعرفة التركيبية واللفظية كمعرفة تقييد مما يجعل المنظومة أقل كفاءة وسرعة. يلاحظ أيضاً.. أن الأخطاء الحادثة فى مستويات المعرفة السفلى تنتقل بشكل تراكمى إلى



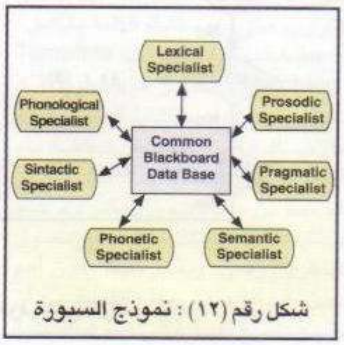
شكل رقم (١٣) : شجرة الكلام



شكل رقم (١١)
النموذج المتسلسل

مستويات المعرفة العليا مما ينتج عنه تفسير غير دقيق للمدخلات الصوتية. فى طريقة الاستدلال الارتجاعى.. يتم عكس اتجاه السريان فى الشكل رقم (١١).. حيث يتم البدء بمستويات المعرفة العليا لتوليد تخمينات أو فروض عن ماهية المدخل الصوتى باستخدام قواعد حدس Heuristic Rules ثم يتم تحليل هذه التخمينات باستخدام مستويات المعرفة السفلى حتى يتم الحصول على تفسير دقيق للمدخل الصوتى. فى أحيان كثيرة.. يتم استخدام طريقتى الاستدلال الأمامى والارتجاعى فى منظومات فهم الكلام التى تستخدم النموذج المتسلسل لتمثيل المعارف للحصول على دقة أعلى.

فى نموذج السبورة – شكل رقم (١٢) – الذى تم استخدامه لأول مرة فى منظومة تعرف على الكلام تسمى HEARSAY .. يتم استخدام المعارف المختلفة عند الحاجة بشكل مستقل.. وعند استخدام مصدر معرفى معين.. يتم وضع نتائج المعالجة فى وحدة ذاكرة مشتركة (سبورة) مما يسمح بامكانية تفاعل مصادر المعرفة المختلفة باستخدام المعلومات المتاحة فى السبورة. فى هذه الحالة.. يمكن لمستويات المعرفة السفلى استخدام معرفة التقييد والتخمينات المتولدة عن طريق مصادر المعرفة العملية



شكل رقم (١٢) : نموذج السبورة

والتركيبية واللفظية لتقييد عملية الاستدلال. وبالمثل.. يمكن لمستويات المعرفة العليا استخدام المعلومات الصوتية للتحقق من الفروض أو لتوليد تخمينات جديدة. نلاحظ أيضاً أنه يمكن لكل مصدر معرفى التعامل مع البيانات بشكل مستقل وغير متزامن.. و بمعنى آخر تكون المعالجة بشكل متواز وليس تتابعياً كما هو الحال فى الطريقة الأولى.. مما يحسن أداء المنظومة. وتتيح استقلالية المصادر المعرفية أيضاً امكانية تحديث أو تغيير القاعدة المعرفية بسهولة.

فى نموذج الشبكة المصنفة.. يتم تصنيف مصادر المعرفة على شكل شجرة تسمى شجرة الكلام Speech Tree – شكل رقم (١٣). لتطبيق معين.. يتم إدخال كل احتمالات النطق الممكنة والعبارات المسموح بها فى الشبكة. ولتفسير مدخل صوتى.. يتم البحث داخل الشبكة باستخدام تقنية مقارنة حتى يتم الحصول على افضل تطابق ممكن. يوضح الشكل رقم (١٣) جزءًا من شجرة كلام لمنظومة تعرف على الكلام فى روبوت صناعى. فى هذه الشجرة.. تحتوى العقد Nodes على قوالب لتباينات النطق التى يمكن استخدامها لمطابقة الإشارة الصوتية المشفرة. تمثل الأفرع Branches كل التوافقات الممكنة بين عقدة وأخرى. عند حدوث تطابق عند عقدة معينة.. تحاول المنظومة مطابقة الفرع المتصل بها مع النقاط الأخرى وتستمر عملية البحث خلال الشجرة حتى يتم الحصول على أفضل تطابق ممكن. يعيب هذه الطريقة.. الحاجة إلى إعادة بناء الشبكة عند إضافة مصادر معرفية جديدة أو عند إجراء أى تغير فى قاعدة المعرفة.. ويتمثل العيب الثانى فى اتساع الشبكة بزيادة عدد المفردات اللغوية الواجب استخدامها مما يتطلب زمن معالجة أطول وبالتالى استجابة أبطأ.