Informe técnico sobre los sistemas de reconocimiento de voz

Autor: Jorge Hierro Álvarez
Madrid, 28 de junio de 2004

Índice

1. **Historia del sistema de reconocimiento de voz:**

La tecnología a lo largo de los años ha evolucionado de forma significativa en lograr que el esquema de comunicación de los seres humanos haya conseguido una representación efectiva a través de la conjunción de una aplicación informática y su desarrollo en una interfaz sencilla de manejar, para que el usuario final pueda llevar a cabo las tareas cotidianas desde su casa y simplemente utilizando la voz como plataforma para la consecución de su objetivo final: ***la compra de una entrada o billete de avión, la redacción de una carta, o bien, la consulta de los datos bancarios.***

A continuación aparecen las fechas más significativas, así como las distintas investigaciones desarrolladas en el campo del reconocimiento de voz, como producto de mercado utilizado por las entidades financieras, compañías aéreas, organizaciones sin ánimo de lucro y fundaciones como la **ONCE,** o bien, empresas de servicios que han creado programas para hacer más sencillo el manejo del ordenador en el entorno familiar. La **domótica** y la ayuda en carretera son dos eslabones al alcance de la mano a partir de ahora.

- **1870's Alexander Graham Bell:** quería construir un sistema/dispositivo que hiciera el habla visible a las personas con problemas auditivos. El resultado fue el teléfono.
- **1880's Tihamir Nemes:** solicita permiso para una patente para desarrollar un sistema de transcripción automática que identificara secuencias de sonidos y los imprimiera (texto). Este programa fue rechazado como **"Proyecto no Realista".**
- **30 años después AT&T Bell Laboratories:** construye la primera máquina capaz de reconocer voz (basada en **Templates)** de los 10 dígitos del **Inglés.** Requería extenso reajuste a la voz de una persona, pero una vez logrado tenía un **99%** de aciertos. Por lo tanto, surge la esperanza de que el reconocimiento de voz resulte simple y directo.
- **A mediados de los sesenta,** la mayoría de los investigadores reconoce que era un proceso mucho más intrincado y sutil de lo que habían anticipado. Empiezan a reducir los alcances y se enfocan a sistemas más específicos:

    - Dependientes del Locutor.
    - Flujo discreto de habla (con espacios / pausa entre palabras)
    - Vocabulario pequeño (menor o igual a 50 palabras)

Estos sistemas empiezan a incorporar técnicas de normalización del tiempo (minimizar diferencia en velocidad del habla). Además, ya no buscaban una exactitud perfecta en el reconocimiento. Más tarde, **IBM** y **CMV** trabajan en reconocimiento de voz continuo pero no se ven resultados hasta la década de los 70.

- **1970-1980:** nace el primer producto de reconocimiento de voz, el **VIP100** de **Threshold Technology Inc.** Utilizaba un vocabulario pequeño, dependiente del locutor, y reconocía palabras discretas. Asimismo, gana el **U.S. National Award en 1972.** Nace el interés de **ARPA,** organismo que

pertenece al **Departamento de Defensa de los Estados Unidos de América,** por los que nos precipitamos a la época de la inteligencia artificial. El proyecto financiado por esta institución buscó el reconocimiento del habla continua, de la ampliación del vocabulario. Impulsa a los investigadores para que se centren en el entendimiento del habla. Los sistemas empiezan a incorporar los siguientes módulos:

- Análisis léxico (conocimiento léxico)
- Análisis sintáctico (estructura de palabras)
- Análisis semántico
- Análisis pragmático

Este proyecto termina en **1976** con el resultado esperado y las empresas contratadas **CMU, SRI y MIT,** por medio de sus investigadores crean los siguientes sistemas para **ARPA,** organismo que pertenece al **Ministerio de Defensa de los Estados Unidos.**

A) CMU:

- Harpy
- Dragon

B) HWIM:

- Hearsay II -> Votan/Dragon Systems->PC

En la década de los 80 será IBM la que remonta el vuelo. No olvidemos que es la empresa que inventa el ordenador personal en agosto de 1981.

- **IBM** desarrolla **N-grams,** lo cual forma la base de la mayoría de los sistemas actuales comerciales.

En estos mismos años, surgen los sistemas de vocabulario amplio, que ahora son la norma. (>1000 palabras) . Adicionalmente bajan los precios de estos sistemas.

Empresas importantes actualmente que desarrollan aplicaciones de reconocimiento de voz

A) Philips
B) Lernout & Hauspie
C) Sensory Circuits
D) Dragon Systems
E) Speechworks
F) Vocalis
G) Dialogic
H) Novell
I) Microsoft

J) NEC
K) Siemens
L) **Intel (apoyo / soporte técnico)**

Por otro lado, es necesario recordar que los siguientes productos de reconocimiento de voz son utilizados por la empresa **Microsoft:**

A) "Dragon Dictate and Naturally speaking" de Dragon Systems
B) Los sigtes productos son compatibles con los sistemas Dragon para asegurar la alta fidelidad del sonido.
C) Freespeech by Philips
D) Kolvox Lawtalk 2.0 para Windows by Kolvox
E) Learout and Hauspie V R technology
F) Metroplex Voice computing: desarrollo hands free programs, los cuales utilizan Dragon Systems Speech recognition para dictar matemáticas
G) "Voice Pilot" para Windows by Voice Pilot, Inc.
H) Via Voice para Windows de IBM

Grabadores digitales

A) Dragon Naturally Mobile
B) Olympus D1000 digital recorder
C) Norcom Dictation Systems
D) Sony MZ-R55
E) Voice it Products (este producto esta asociado para su uso con los sistemas Dragon)

***Nota:** el primer software que se diseñó con el sistema de reconocimiento de voz para **PC,** con la función de dictado, la desarrolló la empresa **Dragon Systems (Dragon Dictate for Winows 1.0),** en el año **1994**. En **1996, IBM,** decidió sacar su propio software de reconocimiento llamado **MedSpeak/Radiology**. Con motivo de la competencia surgida, **Dragon** lanza al mercado en junio de **1997 Naturally Speaking** e **IBM** responde con **ViaVoice (Punto 5, Caso de IBM en España y Estados Unidos y listado de productos)**

***Nota II:** el cuadro con la historia está resumido en el **Anexo IV del informe.**

Fuente: Internet
Fecha: 28 de junio de 2004
Link en la red: http://www.dei.uc.edu.py/tai2000/reconocedor/Historia.htm

2. **¿Qué es un sistema de reconocimiento de voz?**

El reconocimiento de voz generalmente es utilizado como una interfaz entre el ser humano y la computadora a través del algún software.

Debe cumplir con las siguientes tareas:

- Preprocesamiento: convierte la entrada de voz a una forma que el reconocedor pueda procesar.
- Reconocimiento: identifica lo que se dijo (traducción de señal a texto).
- Comunicación: envía lo reconocido al sistema **(software/hardware)** que lo requiere.

Componentes en una aplicación

Existe una comunicación bilateral en aplicaciones, en las que la interfaz de voz está íntimamente relacionada al resto de la aplicación. Estas pueden guiar al reconocedor especificando las palabras o estructuras que el sistema puede utilizar. Otros sistemas sólo tienen una comunicación unilateral.

Los procesos de pre-procesamiento, reconocimiento y comunicación deberían ser invisibles al usuario de la interfaz.
El usuario lo nota de manera indirecta como: *certeza* en el reconocimiento y *velocidad*. Estas características las utiliza para evaluar una interfaz de reconocimiento de voz.

2.1 Los Datos del Reconocimiento de Voz

Los sistemas de reconocimiento de voz se enfocan en las palabras y los sonidos que distinguen una palabra de la otra en un idioma. Estas son los fonemas. Por ejemplo, **"tapa", "capa", "mapa", "napa"** son palabras diferentes puesto que su sonido inicial se reconocen como fonemas diferentes en **Español.**
Existen varias maneras para analizar y describir el habla. Los enfoques más comúnmente usados son:

1. **Articulación:** Análisis de cómo el humano ***produce*** los sonidos del habla.
2. **Acústica**: Análisis de la ***señal*** de voz como una secuencia de sonidos.
3. **Percepción Auditiva:** Análisis de cómo el humano ***procesa*** el habla.

Los tres enfoques proveen ideas y herramientas para obtener mejores y mas eficientes resultados en el reconocimiento.

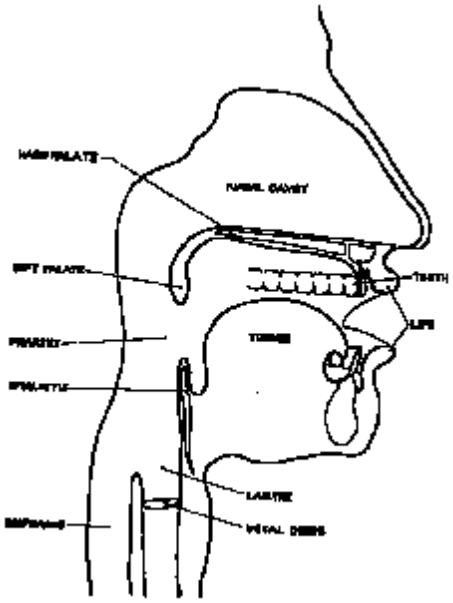
2.1.1 Articulación

La articulación centra su atención en el aparato vocal: Garganta, boca, nariz, en donde se producen los sonidos del habla.

- **Garganta:** Contiene las cuerdas vocales, cuya vibración produce los fonemas "hablados".
- **Boca y Nariz:** "Cavidades de resonancia" porque refuerzan ciertas frecuencias sonoras.
  - Cuando el paladar suave baja y deja el aire pasar por la nariz se generan los fonemas nasales (/m/ /n/)
  - La boca consiste de:
    - Puntos de articulación
      - Dientes
      - Puente alveolar (puente óseo atrás de los dientes superiores)
      - Paladar duro
      - Paladar suave o velum

    y de

  - articuladores
      - Labios
      - Lengua



Clasificación de los fonemas

Para la consistencia en el análisis y facilitar la comunicación entre los investigadores se han definido diferentes sistemas notacionales para clasificar los sonidos del habla de todas las lenguas del mundo.

Sistemas notacionales:

- ARPABET (1970's)
- WORLDBET
- IPA

CONSONANTES	Labios	Labios+Dientes	(Entre los labios)	Puente Alveolar	Paladar Duro	Vellum	Gloths
Stops	P, B			T, D		K, G	
Fricativos		F, V	D, H (that)	S, (Z) tzin	Sh	X	
Africativos					Ch		
Nasales	M			N		Ng	
Semivocales	W			L	J		
Flaps (Vibrantes)				R simple RR múltiple			

Diptongos:

- /ei/
- /oi/
- /ai/

Nasal:

- labial : M
- Velar : N

Alveolar

nj / n

Vibrantes:

- r / rr

Africativo:

- ch

Semivocales:

- /w/ /j/ (siempre con una vocal)

  Cuatro = kc k w a tc t r o
  Cuento = kc k w e n tc t o

- /l/ forma lateral

Fricativos:

- labial /f/
- alveolar /s/
- velar /x/

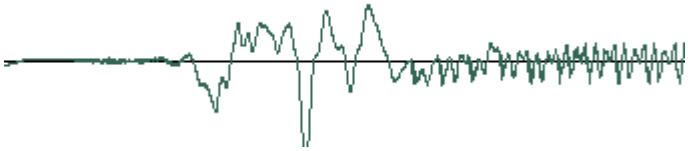
Oclusivos:

- Sonoro: /b/ /d/ /g/
- Sordo: /p/ /t/ /k/

En resumen, la articulación prevé información valiosa sobre la forma de producción de la Voz.

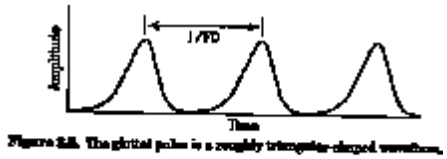
2.1.2 La Señal (Acústica)

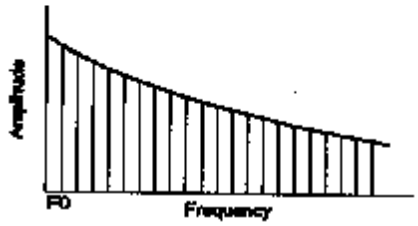
Un reconocedor no puede analizar los movimientos en la boca. En su lugar, la fuente de información es la señal de voz misma. El Habla es una señal **analógica**, es decir, un flujo continuo de ondas sonoras y silencios.



El conocimiento de la ciencia de la acústica se utiliza para identificar y describir los atributos del habla que son necesarios para un reconocimiento de voz efectivo.
Cuatro características importantes del análisis acústico son:

- **Frecuencias**
- **Amplitud**
- **Estructura Armónica (tono versus ruido)**
- **Resonancia**

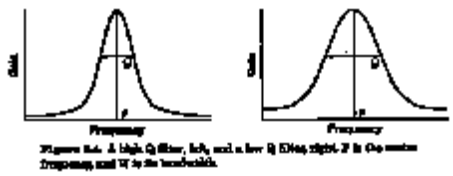


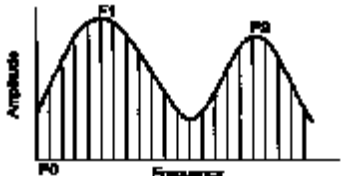


<u>Frecuencia y amplitud</u>

Los sonidos más simples son los sonidos puros (pure tones)
Se pueden representar gráficamente por una onda sinoidal.

Es un patrón simple y cíclico. La amplitud de una onda sonora fisiológicamente representa (corresponde) al movimiento del tímpano de oído.





Frecuencia: número de vibraciones del tono por segundo 100 por segundo = 100 ciclos/segundo = 100 Hz.

Tonos altos = Mayor frecuencia
Tonos bajos = Menor frecuencia
El volumen de un sonido refleja la cantidad de aire que es forzada a moverse.
Se describe y representa como amplitud de la onda y se mide en decibeles **DB**.

Resonancia

La mayoría de los sonidos incluyendo del habla tienen una frecuencia dominante llamada **frecuencia fundamental.** La percibimos como el **pitch (tono)** combinado con frecuencias secundarias. En el habla, la frecuencia fundamental es la velocidad a la que vibran las cuerdas vocales al producir un fonema sonoro.
Sumadas a la frecuencia fundamental hay otras frecuencias que contribuyen al timbre del sonido (Son las que nos permiten distinguir una trompeta de un violín, etc. o las voces de diferentes personas). Algunas bandas de la frecuencia secundarias juegan un rol importante en la distinción de un fonema de otro. Se les llama formantes y son producidas por la resonancia.

Por otro lado, la resonancia se define como la habilidad que tiene una fuente vibrante de sonido de causar que otro objeto vibre (por ejemplo en una fábrica, una máquina hace que vibre el piso). Las cámaras de resonancia en instrumentos de música responden a frecuencias específicas o anchos de banda específicos. Al ser estas cajas o cámaras de resonancia más grandes que la fuente del sonido amplifican las frecuencias a las que responden.

La garganta, boca y nariz son cámaras de resonancia que amplifican las bandas o frecuencias formantes contenidas en el sonido generado por las cuerdas vocales. Estas formantes amplificadas dependen del tamaño y forma de la boca y si el aire pasa o no por la nariz.

Los patrones de las formantes son más fuertes (distinguibles) para vocales que para las consonantes no sonoras.

Estructura Armónica y Ruido

El habla no es un tono puro es continuación de múltiples frecuencias y se representa como una onda compleja. Vocales se componen de 2 o más ondas simples son ricos en frecuencias secundarias y contienen estructuras internas que incluyen ondas cíclicas y aciclicas.

Las ondas acíclicas no tienen patrones repetitivos generalmente llamados ruido forman parte de todos los fonemas sonoros, consonantes y semivocales.
Las frecuencias y características de los patrones acíclicos proveen información importante sobre la identidad de los fonemas. La identidad de las consonantes también se revela por el cambio en las formantes que resultan cuando las articuladores se mueven de un fonema anterior a la consonante y de ella al siguiente fonema llamadas transiciones de formantes. Estas se analizan utilizando técnicas como la transformada rápida de Fourier (FFT) generando espectrogramas.
La complejidad de las formas de onda de los fonemas y las constantes transiciones de un patrón a otro dificultan el análisis de los patrones utilizando las representaciones complejas de las ondas.
Los patrones armónicos y de ruido se muestran con más claridad utilizando los espectrogramas de banda ancha.
La localización (la distancia entre ellas) y cambio en las formantes ayudan a identificar fonemas y palabras.

2.1.3 Coarticulación

Los fonemas aparentemente tienen parámetros acústicos claramente definidos, pero más bien:
Los fonemas tienden a ser abstracciones implícitamente definidas por la pronunciación de palabras en un lenguaje.
La forma acústica de un fonema depende fuertemente del contexto acústico en el que sucede. A éste efecto se le llama coarticulación.
Todo
Tipo Estornudo
Investigadores, utilizan este concepto para distinguir entre la característica conceptual de un sonido del habla (fonema) y una instancia o pronunciación específica de ese fonema (tono).

2.1.4 Percepción Auditiva

La variabilidad del habla producida por coarticulación y otros factores hacen del análisis de la voz extremadamente difícil.
La facilidad del humano en superar estas dificultades sugiere que un sistema basado en la percepción auditiva podría ser un buen enfoque. Desafortunadamente nuestro conocimiento de la percepción humana es incompleto.
Lo que sabemos es que el sistema auditivo está adaptado a la percepción de la de voz.

El oído humano detecta frecuencias de 20Hz a 20,000 Hz pero es más sensible al rango entre 1000 y 6000 Hz. También es más sensible a cambios pequeños en la frecuencia en el ancho de banda crítico para el habla. Además el patrón de sensibilidad a cambios en el tono (pitch) no corresponde a la escala lineal de frecuencias de ciclos por segundo de la acústica.
Para representar mejor al patrón de percepción del oído humano, se desarrollo una escala llamada mel-scale, la cual es una escala logarítmica.
Estudios recientes muestran que el humano no procesa frecuencias individuales independientemente, como lo sugiere el análisis acústico. En su lugar escuchamos grupos de frecuencias y somos capaces de distinguirlas de ruidos alrededor.

Speech coding

Métodos para codificar digitalmente el habla para utilizarlo en diversos ambientes, desde juguetes parlantes, CD's hasta transmisiones vía telefónica.
Para utilizar la voz como dato en aplicaciones tan diversas como el voice mail, anotaciones en un texto o un directorio parlante, es necesario almacenar la voz de manera que una computadora pueda recuperarla.
La presentación digital de la voz nos provee también con las bases para el reconocimiento y síntesis de voz.
El método "convencional" o secuencial de almacenamiento de datos, la cinta magnética requiere de que se le adelante y regrese hasta encontrar la posición buscada. Es propensa a daño mecánico, no se pueden editar (cut/paste) y no duran mucho tiempo en uso.
Alternativa: métodos de codificación digital, manipular el sonido en memoria.
Por lo tanto la meta es capturar fielmente la señal con el menor número de bits. Esto provoca que se tengan que tomar en cuenta diversos factores:

- Memoria y Ancho de banda necesario para flexibilidad de uso
- Costo de transmisión.
- Diversos rangos de calidad
- Depende de la aplicación.
- Codificadores de Voz (algoritmo de…)

Waveform coders: Aprovechan el conocimiento sobre el habla en sí.
Source coders: Modelan la señal en términos de los órganos vocales (vocal tract).

Muestreo y Cuantización

El habla es una señal continua y varia en el tiempo.
Las variaciones en la presión del aire se irradian desde la cabeza y se transmiten por el aire.
Un micrófono convierte esas variaciones en presión del aire a variaciones en voltaje : señal análoga.
Análoga: la señal se puede transmitir a través de un circuito telefónico (voltaje) o almacenados en una cinta magnética (flujo magnético).
Mundo Real: los estímulos sensoriales son análogos.

Si embargo, para las computadoras: es necesario digitalizar la señal (primera fase del procesamiento de la señal)
Series de valores numéricos con una frecuencia regular (frecuencia de muestreo).
El número posible de valores está limitado por el número de bits disponible para representar a cada muestra.
PCM Pulse Code Modulation (simple rep.)
Algoritmos de codificación
Codificación y codificación

Consideraciones del codificador

Inteligibilidad: Todos quieren la mayor calidad posible!
Error e inteligibilidad.
Edición: Simple?
Eliminación del silencio (ahorrar espacio)
Time-scaling: rep. mas rápido que lo que se tardó en reproducirse o más lento.
Adaptación de Velocidad
Robustez.

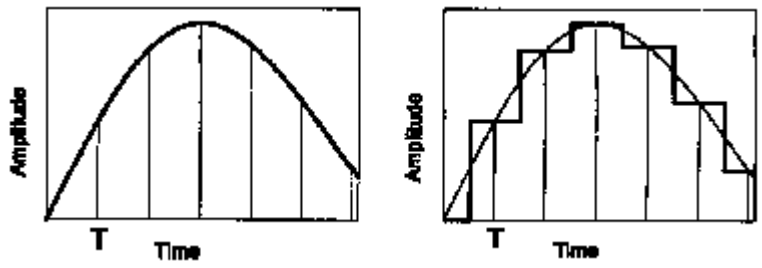
Muestreo (Sampling)

Asigna un valor numérico a la señal en unidades discretas de tiempo (constante)
La frecuencia Nyquist: Específica la frecuencia máxima de una señal que puede reproducirse completamente. Esta estabvlece que

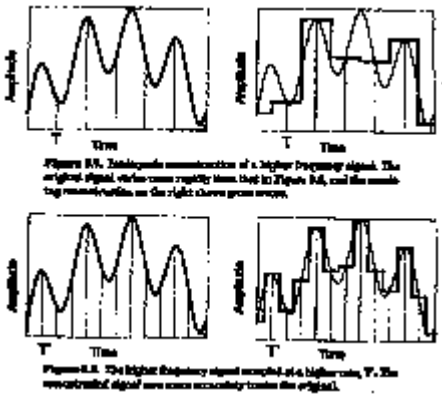
Frecuencia muestreo = 2 veces la frecuencia máxima de la señal

Para poder reproducir la señal análoga debe pasar por un filtro ***pasa-bajas*** a la frecuencia de muestreo (quitar ruidos creados por el muestreo).
Al no cumplirse estas condiciones sucede el fenómeno de ***aliasing***

Señal que varia lentamente: muestreada en una frecuencia T



Si la señal varía más rápido se requiere una T más pequeña por lo tanto un menor ancho de banda de frecuencias.



El humano produce señales de Voz desde los 100(hombre)-400(mujer) Hertz hasta los 15000Hz.

Teléfono : 3100Hz por lo tanto se muestrea a 8000 Hz, inteligible pero baja calidad.
Comparado con un CD, se muestrea a 44.1Hz
20Khz
El ancho de banda es mayor para instrumentos que para voz. Pero la diferencia es audible!!
Por lo tanto se requiere mayor espacio para almacenar y transmitirla.

Resolución: Cuantificación

Cada muestra se representa con un valor digital limitando el rango de valores discretos correspondiente al original.
Ejemplo:
Utilizando 4 bits se pueden representar 16 valores diferentes. Con 8 bits ya son 256 valores.
Esto se puede ver como la resolución:

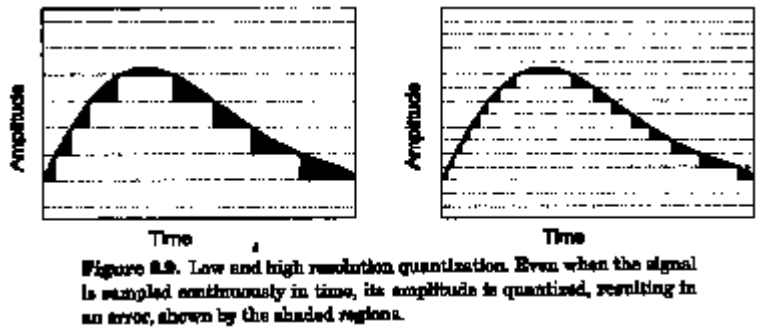


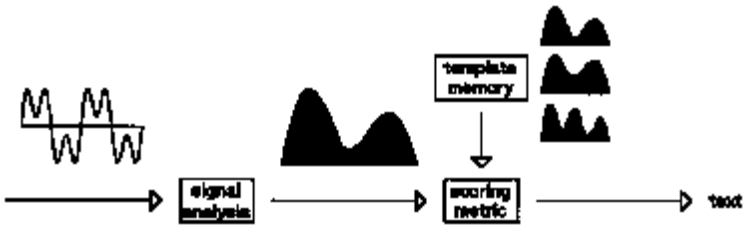
Figure 8.9. Low and high resolution quantization. Even when the signal is sampled continuously in time, its amplitude is quantized, resulting in an error, shown by the shaded regions.

El error o diferencia entre la señal original y la reconstruida se percibe como **ruido**.
Por lo tanto, a mayor resolución mayor cuantificación y menor ruido como consecuencia.
La resolución del "cuantizer" (# de bits por muestra) se describe generalmente en términos de la relación señal-a-ruido (signal to noise ratio o SNR)
A mayor SNR es mayor la fidelidad de la señal digitalizada.

SNR aprox 2^{B} (B=bits/muestra)
Es independiente de la frecuencia de muestreo.
Teléfono: 8bits/muestra , es decir, si muestreamos a 8kHz tenemos 8000 muestras por segundo y así 8000 x 8 = 64000 bits por segundo.
CD: 16bits/muestra, por lo tanto 44100 muestras por segundo x 16bits = 705600 bits por segundo (mono) para 2 canales (estéreo) se duplica.

Template-based Approach

Teoría que ha dado lugar a toda una familia de técnicas de reconocimiento de voz, causando un considerable avance durante las décadas de los 70's y 80's.



El principio es simple: Se almacenan patrones de voz típicos (templates) como modelos de referencia en un diccionario de palabras candidato. El reconocimiento se lleva a cabo comparando la expresión desconocida a los patrones almacenados (templates) seleccionando el que más se le parece.
Usualmente se construye patrones para las palabras completas.
Ventaja: Se evitan errores debidos a la segmentación o clasificación de unidades pequeñas que pueden variar mucho acústicamente, como los fonemas.
Desventaja: Cada palabra requiere de su patrón y el tiempo de preparar y compararlos se vuelve demasiado al incrementarse el tamaño del vocabulario.
Inicialmente se pensaba que se restringía sólo a reconocimiento dependiente del locutor. Sin embargo se logro un reconocimiento independiente del loc. Utilizando técnicas de "Clustering" para generar automáticamente grupos de patrones para cada palabra del vocabulario.
También pasó de ser para el reconocimiento de palabras aisladas a habla continua utilizando técnicas de programación dinámica para encontrar la mejor cadena de patrones.

A) Medición de características:

Se trata básicamente de una técnica de reducción de datos en la cual el gran número de datos en la señal grabada es transformados en un grupo más pequeño de carac's que describen las propiedades importantes de la señal.
El tipo de características que se calculan depende de:

1. Tiempo para calcularlas
2. Espacio para almacenarlas
3. Facilidad de Implementación.

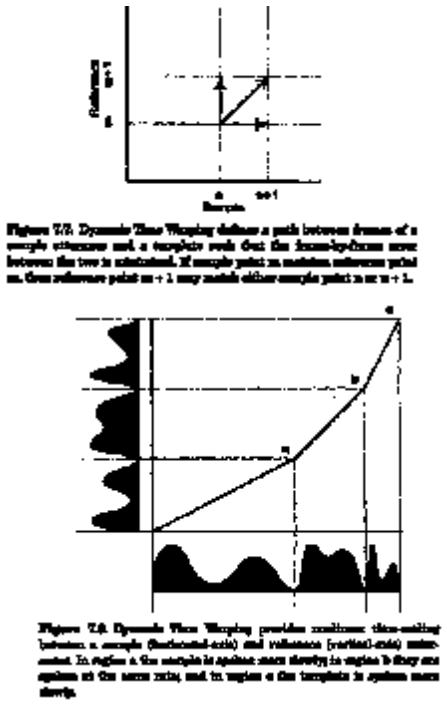
Análisis mediante "banco de filtros" (Filter Bank)

Una muestra de voz se puede aproximar como una combinación lineal de muestras anteriores.

B) Determinación de similitud entere patrón de referencia de voz y serial capturada.

R(t) T(t)
Se busca una función de alienamiento w(t) que mapé R y las partes correspondientes de T.
El criterio de correspondencia es la minimización de una medida de distancia D(T,R)
Por lo tanto se busca una w(t) tal que mística ***d(T(t), R(W(t))) función de ponderación deriva de w(t)***
D(T,R)= min ***INTEGRAL***
dw(t)
conjunto de funciones diferenciables que se incrementan monotónicamente
complicado ***snore*** puede ***resalren*** en general por lo tanto se debe ***discretizar***.
Se buscan una alineación en tiempo "optima" a través de una curva que relacione el eje de tiempo m de R a el eje de tiempo del patrón T.
m=w(n)
Restricciones w(I)=1
W(NT)=NR
Para determinar el tipo de la alineación se propusieron diversas técnicas:

1. alineamiento lineal m=w(n)=(n-1)[(NR-1)/(NT-1)]+1
2. Mapeo de eventos temporales (significativos) en ambos patrones
3. Máxima Correlación se varía la función para maximizar la correl entre RyT
4. Dynamic Time Warping (-se calcula la distancia entre R y T)



Se debe encontrar una medida que indique qué tan similares con los patrones R y T. Para ello es necesario alinearlos de alguna forma.

C**) Reglas de decisión**

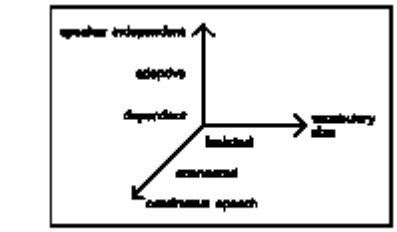
Nearest neighbors el candidato con la menor distancia o una lista ordenada por distancias (de ***menos a mayor***)

Se usa cuando se tienen varios patrones de referencia para cada cantidad R.

Creación de un reconocedor de voz

Pasos

1) Entrenamiento- adquisición de los conjuntos de características para cada palabra en el vocabulario.
2) Clustering- Creación de los patrones de referencia.
3) Pruebas.



Fuente: Internet
Fecha: 28 de junio de 2004
Link en la red: http://www.udlap.mx/~ingrid/ingrid/RV/Proc_de_Voz.html

4. Definición de reconocimiento de voz

Es la conversión de la emisión vocal de una persona en señales digitales. La mayoría de los programas utilizados tienen que estar **"entrenados"** para reconocer los comandos que el usuario da verbalmente. El reconocimiento de voz se usa en la profesión médica para permitir a los doctores compilar rápidamente reportes. Más de **300 sistemas Kurzweil Voicemed** están instalados actualmente en más de **200 Hospitales en Estados Unidos.** Este novedoso sistema de reconocimiento fónico utiliza tecnología de independencia del hablante. Esto significa que una computadora no tiene que ser entrenada para reconocer el lenguaje o tono de voz de una sola persona. Puede reconocer la misma palabra dicha por varios individuos.

En la actualidad, este tipo de sistemas se utilizan para la compra de entradas de cine, billetes de avión, verificación de datos en cualquier entidad bancaria, así como a diario y a través de los **call centers** instalados, tanto en la **Administración Pública** como en las empresas privadas para la verificación de datos.

Por otro lado, recordemos que la **"síntesis de voz"** es un proceso que tiene mejor acogida en otros idiomas, en especial en el inglés más que en el castellano, bien por la existencia de reglas mejor estructuradas o porque no tienen tantas variaciones en sus fonemas. Uno de los grandes retos para las compañías españolas y grupos de investigación es crear una interfaz en castellano, capaz de cubrir el mayor número de cambios posibles (educación, negocios, telecomunicaciones, construcción).

En **México,** el **Laboratorio Tlatoa** tiene un grupo de investigación de la **Universidad de las Américas,** el cual está enfocado en el desarrollo de **Sistemas de Reconocimiento de Voz** para el español mejicano. Pertenecen al **Centro de Investigación en Tecnologías de la Información y Automatización (CENTIA),** dentro del **Departamento de Ingeniería en Sistemas Computacionales de la Universidad de las Américas – Puebla.**

5. Caso de IBM en España y Estados Unidos:

IBM es la empresa pionera en este tipo de sistemas, aunque no es la única. A continuación se resumen, a través de las notas de prensa de esta compañía, algunas de las innovaciones llevadas a cabo en el sector del reconocimiento de voz:

IBM consolida su liderazgo en sistemas de reconocimiento de voz

La Compañía hace públicos nuevos productos, así como contratos y acuerdos con Nokia, Daimler-Chrysler, Johnson Controls, Compaq y Legend.

Madrid, 29 de octubre de 2001

IBM ha anunciado hoy nuevos productos dentro de su oferta de sistemas de reconocimiento de voz. La Compañía ha hecho públicos además una serie de acuerdos con clientes y empresas del sector, que confirman un claro liderazgo dentro del segmento del e-business móvil.Según datos de la consultora Kelsey Group, el mercado de aplicaciones de voz superará los 40.000 millones de dólares (casi 45.000 millones de euros o tres cuartos de billón de pesetas) en 2004. La consultora considera que el negocio de soluciones telemáticas representará, uniendo los mercados de Europa y Estados Unidos, casi 6.500 millones de dólares en el año 2006.

Entre los productos anunciados hoy destaca WebSphere Voice Server 2.0, una nueva versión de este software que reside en un servidor Web y permite el acceso a Internet y a bases de datos de empresas con sólo pronunciar órdenes verbalmente a través de un teléfono mientras que el sistema responde con una voz computerizada. La nueva versión incluye un sintetizador de voz que asemeja a la de una persona real, y está disponible en un mayor número de lenguajes, entre los que se incluye el español (también chino, italiano y japonés, que se añaden a los ya disponibles: inglés, francés y alemán).

IBM anuncia hoy también WebSphere Voice Response, última versión del sistema de respuesta interactiva (IVR, Interactive Voice Response) de IBM, que acepta indistintamente comandos de voz y de ordenador. Los sistemas IVR permiten automatizar diversos servicios, como gestión de pedidos o banca por teléfono, y permiten manejar un número mucho mayor de llamadas.

IBM ha integrado este sistema con su plataforma e-business WebSphere para crear una infraestructura de software que incluya el teléfono como un elemento más de la red. Además, se trata del único sistema de este tipo en todo el mundo que funciona con estándares como VoiceXML y Java, lo que permite a cada empresa personalizar sus aplicaciones y adaptarlas por completo a sus necesidades.

Contratos y alianzas

El lanzamiento de los nuevos sistemas de reconocimiento de voz para empresas coincide con importantes acuerdos, de un lado, con empresas clientes y, de otro, con importantes firmas del sector para el desarrollo de soluciones conjuntas.

En el apartado de clientes, IBM ha firmado acuerdos con empresas como Johnson Controls, uno de los mayores fabricantes del sector de automoción que va a implantar el sistema ViaVoice en vehículos de Daimler-Chrysler a partir de 2003. También la japonesa Honda está trabajando ya con la plataforma de voz de IBM para la integración de productos de este tipo en sus productos. Nokia, por su parte, usará WebSphere Voice Server en su sede central de Finlandia, con más de 15.000 empleados.

Por lo que respecta a alianzas, Compaq ha anunciado que utilizará el sistema Embedded ViaVoice Mobility Suite para dotar de reconocimiento de voz a sus ordenadores de bolsillo Pocket PC. Idéntica decisión ha tomado Legend, el mayor fabricante de ordenadores de China, cuyos primeros productos con ViaVoice aparecerán en los próximos meses.IBM es el mayor suministrador mundial de tecnología, con 80 años de liderazgo ayudando a las empresas a innovar.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: 29 de octubre de 2001
Link en la red: http://www-5.ibm.com/es/press/notas/2001/octubre/webspherevoiceserver.html

IBM, líder en investigación de tecnologías de reconocimiento de voz

IBM lleva más de 25 años trabajando en el reconocimiento automático del habla por ordenador. En su Centro de Investigación Thomas J. Watson de Nueva York, y en colaboración con diversos equipos de investigadores en varios países europeos, entre ellos España, IBM acumula una amplia experiencia en el ámbito de la tecnología informática de la lengua.

La historia de la industria informática es un vertiginoso viaje hacia lo sorprendente. Desde su origen, las modernas tecnologías de la información viven en un permanente estado de ebullición y despliegan una capacidad innovadora de alcance e intensidad sin precedentes.

El compromiso que esta industria mantiene con el futuro ha impulsado asombrosos avances en los que las ideas y proyectos que en su origen rozaban la ciencia ficción se están convirtiendo en realidad. Hoy, uno de esos grandes mitos, la posibilidad de que un ordenador sea capaz de recoger y procesar el lenguaje hablado es ya, gracias al esfuerzo de IBM, una realidad en el mercado y a un precio asequible.

Lo natural

El desarrollo de la capacidad y prestaciones de los sistemas informáticos es espectacular. Los actuales ordenadores personales superan con creces la potencia que ofrecían los enormes y costosos sistemas de hace tan sólo unas décadas. La información contenida en todos los ordenadores de Europa hace 25 años podría guardarse hoy en solo chip.

La aparición en 1981 del ordenador personal (PC) IBM supuso una revolución y no solamente en términos tecnológicos. Por primera vez de manera general, los sistemas informáticos se acercaban a millones de usuarios potenciales, iniciándose así un espectacular crecimiento y popularización de los ordenadores y la consiguiente transformación del modo de entender las relaciones profesionales y, en cierto sentido, la vida cotidiana.

Desde el mundo de los códigos y comandos empezaron a surgir, de forma tímida, los menús de ayuda y guía y, más recientemente, las ventanas e iconos, que han supuesto un uso cada vez más intuitivo del ordenador.

Esta necesidad de hacer cada vez más fluida la relación entre el ser humano y el ordenador constituye una de las preocupaciones más importantes de la actual industria informática. El nacimiento del término "informática natural" alude a los esfuerzos de investigación y desarrollo que se están realizando con el objetivo de ir logrando paulatinamente que el usuario pueda comunicarse con la máquina de forma tan natural como cuando habla, escribe a mano e, incluso, gesticula.

Al habla

IBM lleva trabajando desde hace 25 años en el reconocimiento automático del habla por el ordenador. En su Centro de Investigación Thomas J. Watson de Nueva York, y en colaboración con diversos equipos de investigadores en varios países europeos, IBM acumula una amplia experiencia en el ámbito de la tecnología informática de la lengua.

Este área se ocupa del tratamiento del lenguaje natural a través del ordenador y constituye una de las líneas de investigación más estratégicas y vanguardistas de la actual industria informática. El fin último es conseguir que el usuario pueda conversar con el ordenador de la manera más aproximada posible a como lo hace con otras personas, y en el camino surgen aplicaciones tan útiles como herramientas lingüísticas de ayuda a la escritura, la traducción o la enseñanza de idiomas.

Esta tecnología representa no sólo un salto cualitativo crucial en el modo en que se utiliza el ordenador, y por tanto en que se extienden sus posibilidades, sino que resultará vital para el valor de un idioma dentro de la creciente globalización.

El trabajo de desarrollo realizado por IBM España en este campo sirve para potenciar que la lengua española y las tecnologías a ella asociadas. Tal esfuerzo, iniciado en 1990 y liderado por un equipo multidisciplinar de profesionales, ha situado el castellano en una posición de vanguardia en lo que a las tecnologías informáticas del habla se refiere.

Dicho y hecho

Los logros conseguidos son ya importantes. En los últimos años, la mejora sustancial en la capacidad de proceso del hardware y en el desarrollo de los algoritmos (conjunto de instrucciones que indica al sistema cómo interpretar lo que "escucha") han permitido avanzar hasta conseguir las primeras aplicaciones de reconocimiento del habla por ordenador.

A mediados de 1992 se anunciaba en España el sistema de dictado automático IBM Speech Server Series diseñado para operar sobre una potente estación de trabajo IBM Risc System/6000.

Pocos meses después IBM consiguió llevar las mismas prestaciones a un ordenador personal, poniendo a disposición de millones de usuarios el sistema de reconocimiento de voz más completo y avanzado del mundo. Hoy, los productos de reconocimiento de voz de IBM -ViaVoice Standard Millennium, ViaVoice Web Millennium y ViaVoice Pro Millennium - han ido superándose a sí mismos, llevando el software de reconocimiento del habla a Internet.

El sistema de reconocimiento de voz está basado en complejos métodos probabilísticos y modelos lingüísticos. La conversión de la palabra hablada en texto se realiza a través de sofisticados algoritmos que aíslan, identifican e interpretan los componentes fonéticos individuales del habla humana. Este proceso resulta altamente eficaz y permite alcanzar una tasa media de aciertos del 96 por ciento. El ordenador es capaz, por ejemplo, de elegir correctamente entre palabras homófonas, como "a" y "ha", y diferenciar la palabra "coma" del signo de puntuación ",".

Las aplicaciones del sistema son múltiples. Además de permitir a cualquier usuario sustanciales mejoras de comodidad y ahorro de tiempo en la habitual tarea de introducir textos en el ordenador, el sistema de reconocimiento de voz de IBM resulta de extraordinaria utilidad para los profesionales que elaboran informes.

Así, el radiólogo puede dictar sus conclusiones mientras examina con total libertad una radiografía y el periodista escribir un reportaje al tiempo que consulta otros documentos. Y cualquier usuario podría preparar al sistema para que al decirle "buenos días" éste abriera automáticamente las aplicaciones.

ViaVoice Pro Millennium Edition 

- Permite crear documentos sin necesidad de teclear, simplemente hablando de forma natural y relajada.
- Puede crear documentos para la gran mayoría de las aplicaciones de Windows a través de Speakpad.
- Se puede dar formato al texto a través de sencillas órdenes habladas como, por ejemplo "activar cursiva".
- El análisis de la voz aumenta la precisión del dictado al actualizar los modelos personalizados.
- La opción ViaVoice Outloud permite escuchar los documentos sin necesidad de leerlos. Esto resulta de gran utilidad, por ejemplo, al recibir correo electrónico.
- El asistente Woodrow y la opción "¿Qué puedo decir?" ofrecen una amplia gama de posibilidades a la hora de resolver problemas, además de hacer sugerencias sobre los comandos más adecuados para las distintas aplicaciones.
- La opción "Analizar mis documentos" busca palabras nuevas en documentos previamente existentes y las añade al vocabulario personal.
- Durante el dictado, y de acuerdo con el contexto, las palabras que suenan de forma igual o parecida son escritas correctamente.
- Permite crea macros para insertar textos de uso frecuente.
- ViaVoice ha sido optimizado para procesadores Intel Pentium II/III y AMD Athlon
- Cuenta con un Vocabulario activo de más de 100.000 palabras, incluyendo terminología jurídica. Además, ofrece la posibilidad de añadir 64.000 términos adicionales.
- Dispone de un diccionario de respaldo con 475.000 palabras más.
- Permite conectarse a Internet mediante el uso de comandos como "Navegar por la Web" o "Saltar a <favoritos>".
- Navega por Internet hablando con Internet Explorer, simplemente leyendo los links.
- Ofrece la posibilidad de "chatear" en la web, así como de enviar y recibir correos electrónicos con la voz.
- La función de "Voice Mouse" permite utilizar el ratón sin usar las manos.
- Aplicaciones como Netscape Communicator 4.5, Outlook 97,98,2000 y Outlook Express 4 y 5 pueden ser utilizadas con la voz.
- Cuenta con diferentes funciones de búsqueda por el escritorio del PC.
- Incluye la posibilidad de dictado directo en Microsoft Word 97 y 2000, así como en la mayoría de las aplicaciones de Windows.
- Existe un módulo adicional con vocabulario médico disponible por separado.
- Utiliza mandatos naturales para decirle al sistema las acciones que ha de realizar.
- Se pueden activar con la voz plantillas de dictado para documentos estándar.
- Permite realizar funciones de Comando y Control del sistema mediante la voz usando comandos sencillos, del tipo "abrir programa", "imprimir" o "guardar".
- Permite corregir los textos directamente con la voz o delegar esta tarea en otra persona.

ViaVoice Web Millennium Edition

- Permite crear documentos sin necesidad de escribir, simplemente hablando de forma natural y relajada al ordenador.
- Puede crear documentos para la gran mayoría de las aplicaciones de Windows a través de Speakpad.
- Se puede dar formato al texto a través de sencillas órdenes habladas como, por ejemplo "activar cursiva".
- El análisis de la voz aumenta la precisión del dictado al actualizar los modelos personalizados.
- La opción ViaVoice Outloud permite escuchar los documentos sin necesidad de leerlos. Esto resulta de gran utilidad, por ejemplo, al recibir correo electrónico.
- El asistente Woodrow y la opción "¿Qué puedo decir?" ofrecen una amplia gama de posibilidades a la hora de resolver problemas, además de hacer sugerencias sobre los comandos más adecuados para las distintas aplicaciones.
- La opción "Analizar mis documentos" busca palabras nuevas en documentos previamente existentes y las añade al vocabulario personal.
- Durante el dictado, y de acuerdo con el contexto, las palabras que suenan de forma igual o parecida son escritas correctamente.
- Permite crea macros para insertar textos de uso frecuente.
- ViaVoice ha sido optimizado para procesadores Intel Pentium II/III y AMD Athlon
- Cuenta con un Vocabulario activo de más de 100.000 palabras, incluyendo terminología jurídica. Además, ofrece la posibilidad de añadir 64.000 términos adicionales.
- Dispone de un diccionario de respaldo con 475.000 palabras más.
- Permite conectarse a Internet mediante el uso de comandos como "Navegar por la Web" o "Saltar a <favoritos>".
- Navega por Internet hablando con Internet Explorer, simplemente leyendo los links.
- Ofrece la posibilidad de "chatear" en la web, así como de enviar y recibir correos electrónicos con la voz.
- La función de "Voice Mouse" permite utilizar el ratón sin usar las manos.
- Aplicaciones como Netscape Communicator 4.5, Outlook 97,98,2000 y Outlook Express 4 y 5 pueden ser utilizadas con la voz.

ViaVoice Standard Millennium Edition

- Permite crear documentos sin necesidad de escribir, simplemente hablando de forma natural y relajada al ordenador.
- Puede crear documentos para la gran mayoría de las aplicaciones de Windows a través de Speakpad.
- Se puede dar formato al texto a través de sencillas órdenes habladas como, por ejemplo "activar cursiva".
- El análisis de la voz aumenta la precisión del dictado al actualizar los modelos personalizados.
- La opción ViaVoice Outloud permite escuchar los documentos sin necesidad de leerlos. Esto resulta de gran utilidad, por ejemplo, al recibir correo electrónico.
- El asistente Woodrow y la opción "¿Qué puedo decir?" ofrecen una amplia gama de posibilidades a la hora de resolver problemas, además de hacer sugerencias sobre los comandos más adecuados para las distintas aplicaciones.
- La opción "Analizar mis documentos" busca palabras nuevas en documentos previamente existentes y las añade al vocabulario personal.
- Durante el dictado, y de acuerdo con el contexto, las palabras que suenan de forma igual o parecida son escritas correctamente.
- Permite crea macros para insertar textos de uso frecuente.
- ViaVoice ha sido optimizado para procesadores Intel Pentium II/III y AMD Athlon
- Cuenta con un Vocabulario activo de más de 100.000 palabras, incluyendo terminología jurídica. Además, ofrece la posibilidad de añadir 64.000 términos adicionales.
- Dispone de un diccionario de respaldo con 475.000 palabras más.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: año 2004
Link en la red: http://www-5.ibm.com/es/press/informes/voz.html

6. Informe de Intel Corporation:

Otra de las grandes empresas que se ha metido en el análisis y estudio del sistema de reconocimiento de voz ha sido **Intel.**

Las computadoras están aprendiendo a leer los labios
Mejoramiento del software de reconocimiento de voz

"El nuevo software de fuente abierta de Intel permite que las computadoras lean los labios de los usuarios, un gran avance en las aplicaciones de reconocimiento de voz".
*–"**Han nacido las computadoras que leen los labios**,"*
Internet Magazine, 29 de abril de 2003

Imagine la hora de mayor tráfico en la estación del metro del Zoológico de Berlín, o cualquier otro agitado metro o estación de metro en el mundo. Usted se dirige a una máquina para comprar un boleto. La máquina le pregunta su destino. La estación está llena de ruido de trenes, conversaciones entre los pasajeros y hasta los eternos músicos. Usted dice el nombre de la estación. Casi no puede oírse a sí mismo, pero la máquina reconoce el nombre e imprime el boleto. Aunque no es posible en la actualidad (el software de reconocimiento de voz no funciona muy bien cuando hay ruido), la situación descrita podría ser realidad muy pronto.

Nuevo software de fuente abierta

Los expertos del Centro de investigación de Intel en China (ICRC), que es parte de los Laboratorios de investigación de microprocesadores de Intel, han dado un gran paso para acercarse a que este escenario se haga realidad. En el Foro de Desarrolladores de Intel (IDF) que se llevó a cabo en Berlín en abril de 2003, Intel anunció el lanzamiento al mercado de software de reconocimiento de voz y audiovisuales (AVSR) de fuente abierta como parte de la prestigiosa Biblioteca de visión informática de fuente abierta (OpenCV).

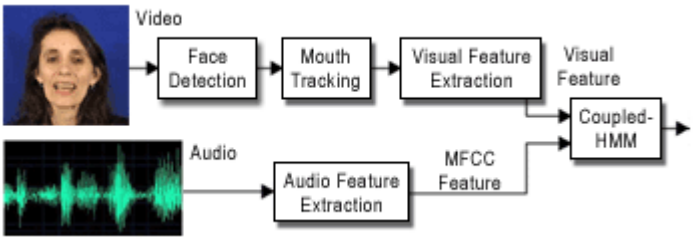


Figura 1. La tecnología de reconocimiento de voz y audiovisuales (AVSR) está basada en la recopilación de información tanto de voz como visual, la cual es la entrada de un motor de reconocimiento. La combinación de información de una variedad de entradas mejora considerablemente la tasa de reconocimiento de las computadoras.

El Software AVSR permite que los desarrolladores creen PC con capacidad no sólo de reconocer la voz, sino también de 'ver' y 'leer los labios' del mismo modo que las personas. Las tecnologías de Interfaz de usuario nuevas (NUI), como

AVSR, representan un área de investigación importante en el Centro de investigación de Intel en China.

Justin Rattner, experto de Intel y Director de los laboratorios de investigación de microprocesadores, expresó "Intel tiene el propósito de lograr que las computadoras tengan la capacidad de interactuar con el mundo del modo que nosotros lo hacemos. Tomamos decisiones al combinar información de una variedad de fuentes; muy raras veces el reconocimiento se basa en un solo tipo de información. La adición del código de reconocimiento de voz, sonido y vídeo a la biblioteca OpenCV seguramente aumentará la investigación y el desarrollo de reconocimiento de voz asistida por la visión".

La realidad del ruido

Aunque los poderosos algoritmos de reconocimiento de voz de la actualidad tienden a funcionar bien cuando se elimina el ruido de fondo o cuando se utiliza un auricular bien sintonizado, la exactitud se degrada con rapidez cuando las aplicaciones se enfrentan a entornos ruidosos.

En combinación con los algoritmos de reconocimiento de voz de la biblioteca OpenCV de Intel, el software AVSR permite que las computadoras no solo reconozcan la voz, sino que también detecten el rostro del usuario y sigan los movimientos de la boca, para permitir un reconocimiento de voz más preciso. Vea las ilustraciones dos y tres a continuación.

Figura 2. El sistema preparado para AVSR realiza un proceso para detectar al hablante y clasificar las regiones de la imagen del hablante en regiones del rostro y regiones que no son del rostro.

Figura 3. Una vez que se identificado una región del rostro, el sistema preparado para AVSR busca la detección de la voz en la región de la boca del rostro. El software AVSR puede detectar y leer los labios con o sin vellos faciales.

Con AVSR, los investigadores de Intel han logrado un 55 por ciento de aumento en la exactitud de las situaciones en las que el nivel del ruido llega a un 50 por ciento, lo cual prepara el terreno para una amplia variedad de aplicaciones de voz en entornos ruidosos, tales como los aeropuertos o centros comerciales.

La visión informática y las aplicaciones del mundo real

En los próximos diez años, se espera que la visión informática desempeñe un papel importante en la interacción entre las computadoras y los usuarios. Al equipar las computadoras con las rudimentarias funciones de la vista, éstas tendrán más conciencia de sus alrededores. Los desarrolladores han utilizado nuestro código en diversas aplicaciones que van desde la seguridad hasta la exploración espacial y desde los juguetes hasta la fabricación industrial.

Por ejemplo, veamos la seguridad. Una computadora puede observar el estacionamiento de un aeropuerto. Con software de reconocimiento de patrones, se crea una base de datos con los patrones de actividad normales. El sistema en sí no observa a las personas, sino que observa los patrones de las personas que van y vienen. Cuando alguien entra en el estacionamiento y quebranta el patrón, por ejemplo si va de un auto a otro, el sistema puede enviar una alerta de que algo poco habitual está sucediendo. En la industria, esta tecnología se utiliza para el control de calidad, ya que detecta los productos defectuosos antes de que se empaquen.
A medida que las aplicaciones de visión informática se vuelven más sofisticadas, podrían utilizar el reconocimiento de gestos para alertar a los grupos de seguridad sobre comportamientos agresivos o detectar los rostros de terroristas conocidos en las cámaras de seguridad. O bien, una cámara inteligente montada en lo alto de una piscina pública, podría servir como un segundo par de ojos del salvavidas para reconocer conductas peligrosas.

¿Qué es OpenCV?

OpenCV, es decir Visión informática de fuente abierta, es una plataforma abierta sobre la cual pueden crearse aplicaciones comerciales. La biblioteca de software OpenCV es una kit de herramientas con más de 500 funciones de imagen en formato de código fuente que ayuda a los investigadores a desarrollar aplicaciones de visión informática, tales como el reconocimiento del rostro, de gestos y de audiovisuales.

La biblioteca permite que los investigadores en academias y en el sector avancen las últimas tecnologías mediante el uso de una base de código estable y optimizado para el desempeño. Desde su lanzamiento en 2000, OpenCV ha experimentado más de 500.000 descargas de código y ha atraído más de 5.000 miembros inscritos en el grupo de usuarios. Las cifras son considerables, si se toma en cuenta la especialización de este campo.

¿Por qué la fuente abierta?

La biblioteca, la cual puede utilizarse sin regalías en un producto cerrado y lucrativo, permitirá el amplio desarrollo y utilización de funciones de visión informática más sofisticadas en aplicaciones existentes y nuevas.
Debido a que la investigación aumenta al compartir las ideas, esperamos que al compartir nuestra investigación con la comunidad lograremos:
- Incentivar que otras entidades compartan sus descubrimientos.
- Aumentar la tasa de innovación en el reconocimiento de voz.
- Hacer que sea viable de forma comercial como interfaz informática humana.

Más información

Marque el sitio de investigación y desarrollo en Intel y manténgase al día con estas y otras noticias sobre los avances tecnológicos de Intel.

Obtenga un panorama más técnico de la investigación AVSR aquí. Busque el encabezado "Software" a fin de obtener instrucciones para la descarga del software de fuente abierta. Obtenga más información sobre la biblioteca OpenCV aquí.

Elija estos vínculos selectos para explorar las perspectivas del sector e independientes:

CNET News.com, Análisis de Intel sobre el software de lectura de los labios

Electronic Business Online, Lea mis labios: Intel da ojos al software de reconocimiento de voz

Internet Magazine, Han nacido las computadoras que leen los labios

LinuxDevices.com, Intel lanza al mercado software de lectura de los labios de fuente abierta

The Register, Intel pone a la disposición software de reconocimiento de voz que lee los labios

VNUNet.com, Intel enseña a las computadoras a leer los labios

Lea el anuncio corporativo de Intel en la conferencia IDF de abril de 2003 en Berlín, Alemania.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha del informe: 29 de abril de 2003
Link en la red: http://www.intel.com/espanol/labs/features/sw04034.htm

7. **Noticias de Dialogic:**

Otra de las empresas que estuvieron desde el principio trabajando en el sistema, gracias a la creación de las tarjetas que hicieron posible la emisión de la señal, llegó a un acuerdo con **Intel** para la distribución de las mismas.

Migración a las tarjetas Intel® Dialogic JCT

Callware Voice Technologies S.A., Intel® Authorised Distributor y mayorista de componentes para la industria de telefonía por ordenador o Computer Telephony, anuncia la migración a tarjetas Intel® Dialogic JTC de todos los modelos anteriores.

Las tarjetas Intel® Dialogic JCT ofrecen interfaces de red digitales y analógicos, con formato PCI universal, y equipados con Bus H.100. Las tarjetas poseen magníficas características de media, tales como procesado de voz, software de reconocimiento de voz, fax, señalización y procesamiento de llamadas, detección y generación de tonos, haciéndolas ideales para proveedores de servicios y grandes empresas.

Las tarjetas Intel® Dialogic D/41JCT-LS, D/120JCT-LS, D/240JCT-T1, D/300JCT-E1, D/480JCT-1T1 y D/600JCT-2E1 son los productos de próxima generación basados en el firmware SpringWare. Las tarjetas son ideales para los desarrolladores que buscan ofrecer aplicaciones de comunicaciones económicas, muy escalables y de elevada densidad, que requieren fuentes multimedia, tales como voz, reconocimiento de voz basado en software, fax e interfaz telefónico en un solo slot. Estas tarjetas ofrecen un enorme número de características: soporte para tecnología de procesamiento de señal digital (DSP), tecnología de bus PCI para comunicación con el PC y CTBus para comunicación entre tarjetas.

El soporte de la innovadora tecnología de proceso de voz continuo (Continuous Speech Processing) permite la integración de software de reconocimiento de voz basado en la tecnología desarrollada por los principales fabricantes del mercado. El fax basado en DSP y el soporte para el reconocimiento de voz sobre CPU permiten a los desarrolladores maximizar el número de tarjetas en el sistema para aplicaciones de comunicaciones multimedia, tales como call centers, portales de voz, mensajería unificada o sistemas para respuesta de voz interactiva (IVR).

La opción de usar nuevos codificadores de voz, tales como GSM y G.729, ofrece la capacidad de realizar soluciones de mensajería unificada al mismo tiempo que se trabaja con los actuales sistemas de mensajería tradicionales. Además, el soporte de GlobalCall como interfaz universal de programación para Telefonía facilita el despliegue global y añade la flexibilidad de escalar sistemas para cumplir con las crecientes necesidades de cada proyecto.

Las tarjetas Intel® Dialogic JTC son soportadas por el SDK para Windows NT, Windows 2000 y Linux.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: año 2004

Link en la red:
http://www.truedataonline.com/xq/asp/rvdymtx.trfpwrdcvp/file.home/catID.114/qx/index.htm

8. Caso de Intermec:

En busca de la compatibilidad de los sistemas, en **España** el reconocimiento de voz se impuso en la década de los **90.** La mayoría de las sociedades han apostado por crear herramientas, tanto de software como de hardware, para garantizar una plataforma común de trabajo.

Intermec añade reconocimiento de voz a sus sistemas portátiles Pocket PC 700

Los sistemas de reconocimiento de voz y su aplicación en los diferentes procesos de la cadena logística cada vez están más extendidos, especialmente en operaciones de preparación de pedidos y en la logística del frío.

Madrid, 21 de abril de 2004 - Intermec Technologies ha anunciado una nueva versión de SyVox, su software logístico de voz, al que ha integrado la tecnología ViaVoice de IBM. Ambas aplicaciones han sido incorporadas a sus terminales de la serie 700, por lo que los usuarios de estos equipos pueden implantar tecnología de reconocimiento de voz en sus operaciones.

Esta última versión de SyVox incluye el motor de voz de IBM que presenta soporte para comandos de lenguaje natural y una extraordinaria capacidad para modificar y adaptarse dinámicamente a aspectos gramaticales o dialectos locales. Intermec ha anunciado la disponibilidad inmediata de estas aplicaciones en inglés y español, mientras que las versiones para otros idiomas se lanzarán al mercado a finales del presente año.

El reconocimiento de voz permite a los operarios manejar el terminal y las operaciones que con él desarrolla mediante instrucciones verbales, así como comunicarse con los sistemas ERP ó SGA (Gestión de almacenes), de tal forma que sus manos quedan libres para efectuar las labores en las que se requiere el uso de ambas manos. Estas aplicaciones software pueden integrarse fácilmente con los actuales sistemas de gestión de almacenes o ERPs, utilizando las redes inalámbricas por radiofrecuencia instaladas en los centros de trabajo, por lo que no es necesario modificar o implantar nuevas infraestructuras.

La utilización de sistemas de reconocimiento de voz es ideal para aplicaciones típicas de preparación de pedidos, control de inventarios, reposiciones o ubicaciones, mejorando notablemente los tiempos empleados en estas operaciones y manteniendo niveles óptimos de fiabilidad y seguridad. Son asimismo, particularmente interesantes en aquellas tareas donde la utilización de la identificación o lectura por códigos de barras puede ser compleja o resulta poco eficiente.

Algunos ejemplos de aplicaciones donde los sistemas de reconocimiento de voz pueden ofrecer grandes ventajas son aquellas en lo que es necesario realizar manipulaciones sobre un gran número de ítems, típicas del sector retail o alimentación; el trabajo en cámaras frigoríficas, donde los operarios deben utilizar guantes y resulta engorroso e incómodo la entrada de datos mediante terminales

portátiles; y en otras operaciones donde no sea posible utilizar la codificación por barras.

La tecnología de reconocimiento de voz ha alcanzado una gran madurez técnica y llega en unos momentos donde se analiza profundamente todas las operaciones de la cadena de suministro para ahorrar minutos e incluso segundos en las tareas más comunes. Dar órdenes habladas a un terminal es mucho más rápido que introducir información mediante teclado o en la propia pantalla. Esta innovadora tecnología viene a aportar un nuevo valor añadido a los terminales 700 de Intermec. La gama 700 de Intermec integra los sistemas operativos de Microsoft® con soporte para el estándar de audio AC 97 por lo que los usuarios obtendrán una alta calidad de sonido.

Intermec Technologies Corporation, compañía subsidiaria de UNOVA Inc., es el líder mundial en desarrollo, fabricación e integración de sistemas de captura automatizada de datos e informática móvil. Los productos y servicios de la compañía permiten a clientes de múltiples sectores mejorar su productividad, calidad y capacidad de respuesta de sus operaciones empresariales, desde la gestión de suministros y la planificación de recursos, hasta las ventas en campo y servicios. Visite nuestra web: www.intermec.es

Para más información: **spain.press@intermec.com**

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: 21 de abril de 2004
Link en la red: http://www.intermec.com/eprise/main/Intermec/Content/About/NewsPages/pressRelease?pressID=559

9. Caso de Nortel Networks S.A.:

Nortel apostó por la integración de los sistemas junto a **Philips.**

Nortel Networks Integra reconocimiento de voz de Philips Speech Processing

Agrega asistencia de directorio con voz, aplicaciones para atención al cliente y ofertas eBusiness

SAN JOSE, Calif. - Nortel Networks* [NYSE/TSE: NT] integró el sistema de reconocimiento de voz SpeechPearl* de Philips Speech Processing, unidad empresarial de Royal Philips* Electronics, en su plataforma de interacción de voz OSCAR (Cómputo de Señal Abierta y Recursos de Análisis) para el eBusiness, que permite el uso de asistencia de directorio con voz y aplicaciones para atención al cliente.

OSCAR es la plataforma de procesamiento de voz dedicada en la cual se crean las soluciones de respuesta de voz interactiva (IVR) y de integración de telefonía de cómputo (CTI) de Nortel Networks. La máquina SpeechPearl permite a los usuarios emplear diálogos naturales con sistemas IVR en lugar de respuestas touch-tone a los menúes. Está diseñado para soportar 20 idiomas y un diccionario con 200,000 palabras.

"Nortel Networks ahora puede ofrecer uno de los sistemas de reconocimiento de voz líderes de la industria en una plataforma que ayuda a que el eBusiness pueda crecer con rapidez", dijo Richard Rosinski, director ejecutivo de procesamiento avanzado de voz, Aplicaciones eBusiness Clarify*, Nortel Networks.

Forrester Research describe a las aplicaciones de voz como "la gran oportunidad para obtener utilidades y mejorar las relaciones con el cliente" para las empresas y los proveedores de servicios. Forrester predice que estas compañías invertirán $1.3 trillones de dólares para introducir aplicaciones de voz en el Internet en 2003.

"Philips y Nortel Networks cuentan con experiencia global, experiencia en telecomunicaciones e instalaciones extensas, que se complementan con el liderazgo de Philips en tecnología de idioma natural y número de idiomas soportado", dijo Peter Foster, vicepresidente ejecutivo de Philips Speech Processing.

Philips Speech Processing, unidad empresarial de Royal Philips Electronics, es pionera y líder mundial en reconocimiento de voz, diálogo natural y tecnologías para comprensión de idioma, con más de 40 años de experiencia en el desarrollo y la mercadotecnia de productos de voz. Philips ofrece un amplio portafolio de soluciones de tecnología de voz para las industrias de telecomunicaciones, TI, automóviles y electrónicos para el consumidor. Como desarrollador y proveedor de tecnologías de voz en idiomas múltiples, Philips tiene la base instalada de sistemas de reconocimiento de voz y comprensión natural de idiomas más grande de Europa, y es un proveedor importante de tecnología de voz en América y otras regiones del mundo. Sus instalaciones se componen de automatización de servicios en centros telefónicos, aplicaciones telefónicas, soluciones para dictado profesional, dispositivos controlados por voz, y aplicaciones Internet. Para mayor información, visite www.speech.philips.com**.

Nortel Networks es un líder mundial en Internet y comunicaciones, con capacidades que abarcan sistemas ópticos, inalámbricos, de Internet local y de negocios electrónicos. La compañía tuvo ingresos en 1999 de US$21.300 millones (contabilidad estadounidense) y sirve a clientes en todo el mundo que incluyen compañías telefónicas, proveedores de servicio, y empresas. Hoy, Nortel Networks está creando una Internet de alto rendimiento que es más confiable y más rápida que nunca. Está redefiniendo los aspectos económicos y la calidad de la operación en red y en Internet, prometiendo una nueva era de colaboración, comunicaciones y comercio. Visítenos en www.nortelnetworks.com.

* Nortel Networks, el logotipo de Nortel Networks y la marca del Globo son marcas registradas de Nortel Networks. Philips, el logotipo de Philips y SearchPearl son marcas comerciales de Royal Philips Electronics.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la redacción de la nota de prensa: año 2000
Link en la red: http://www.nortelnetworks.com/index.html

10. Caso de Telefónica:

La operadora de telecomunicaciones **Telefónica** tampoco se ha quedado atrás. Según la última noticia que ha aparecido en las **Agencias de Noticias,** ha instalado en **Iberdrola** un sistema de reconocimiento de voz en los centros de llamadas.

Telefónica Empresas, a través de Telefónica Soluciones, ha implantado sistemas automáticos de reconocimiento de voz en los centros de atención de llamadas de Iberdrola (Madrid: IBE.MC - noticias - foros) , informó hoy la compañía.

Esta tecnología permitirá agilizar algunas demandas de los usuarios como duplicados de factura, modificación o alta de contrato, cambio de cuenta, avería, lectura de contador u otras operaciones. Además, es posible identificar al cliente y el motivo de la llamada por medio de diálogos sencillos.

De este modo, el flujo de información entre el operador y el usuario no sólo es más rápido, sino también más eficaz, al automatizar las operaciones más sencillas y permitir liberar a los agentes que atienden las llamadas para que dediquen su tiempo a operaciones que requieren en mayor medida su atención.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: 31 de mayo de 2004
Link en la red: http://es.biz.yahoo.com/040531/4/3g2n6.html

11. Reconocimiento de la voz en la Unión Europea:

Con la ampliación de la **UE,** uno de los problemas que tienen en la **Comisión** es la implementación de un sistema seguro que haga las traducciones a todos los idiomas de las reuniones y conclusiones acontecidas en las instituciones.

Bruselas, 23 oct (EFE).- La empresa española, radicada en Bruselas, Speech Recognition Ware, está revolucionando la forma de trabajar de las instituciones europeas, gracias a su novedoso sistema de reconocimiento vocal, calificado por expertos como "la respuesta a las decepciones de los programas de dictado vocal".

Los traductores de instituciones como la ONU, la Comisión Europea, el Tribunal de Justicia de la UE, el Parlamento Europeo, el Comité de las Regiones o el Consejo de Europa ya trabajan con este sistema de reconocimiento vocal, que podría convertirse en el mejor aliado para la Torre de Babel en que se va a transformar la UE tras su ampliación.

El director general de Speech Recognition Ware (SRW), Jesús María Boccio, señaló hoy a EFE que "algo está pasando en el mercado de reconocimiento de voz para que organismos tan sólidos y serios como la Comisión Europea o la ONU compren esta tecnología y digan públicamente que supone un cambio absolutamente espectacular en su forma de trabajar."

El responsable del proyecto de reconocimiento vocal en la Comisión Europea, el español Antonio Ballesteros, confirmó su éxito asegurando que 250 de los 1.200 traductores con los que cuentan ya utilizan el sistema.

Dicha tecnología permite generar texto con facilidad y velocidad (180 palabras por minuto), en diversas lenguas (español, francés, italiano, alemán, neerlandés, chino y cuatro variantes de inglés), y sin necesidad de teclear, simplemente dictando al ordenador.

"La lengua que mejor funciona de todos nuestros proyectos es el español, porque es un idioma que escribimos igual que hablamos, cosa que no ocurre con el inglés o francés", señaló Boccio.
Otro de sus logros es su "gran precisión" en el reconocimiento de la voz al dictar texto libre (entre el 97 y el 98 por ciento).

Según Boccio, es la primera tecnología del mundo con la que "puedes crear un perfil en cuatro minutos, llegando ya al 95 o 96 por ciento de precisión garantizada. Al cabo de un par de semanas con un cierto uso pasas a niveles estratosféricos".
También ofrece la posibilidad de dictar en una grabadora digital, una agenda

personal e incluso un teléfono móvil, fuera de la oficina, para transcribir el texto con su ordenador más tarde.

Así como controlar íntegramente con la voz todos los menús y funciones en la mayoría de las aplicaciones más utilizadas como Microsoft Office y Windows, además de gestionar íntegramente el correo electrónico y navegar por internet. Los sistemas actuales se basan en que cada usuario tiene que corregir los errores para educar al programa. Sin embargo, SRW, mediante el programa Transcription Aid, permite al mismo concentrarse sólo en el dictado y que otra persona pueda corregir el texto escuchando al mismo tiempo la voz.

El interés que este sistema está suscitando coincide con la inminente entrada en vigor de una directiva (ley-marco) europea, a finales de 2003, por la que todas las entidades que operen en la UE deberán garantizar la estricta igualdad de oportunidades entre sus trabajadores.

Esto significa proporcionar a los discapacitados los medios tecnológicos necesarios, como por ejemplo el reconocimiento vocal, para superar su minusvalía.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: 23 de octubre de 2003
Link en la red: http://www.efe.es

12. Reconocimiento de la voz en los bancos:

Las instituciones bancarias cuyos servicios son cada vez más solicitados, se han visto obligados a satisfacer las necesidades de sus Usuarios. Las demandas de nosotros, los clientes, tienen que ver, en buena medida, con el poco tiempo de que disponemos ahora para realizar nuestras operaciones bancarias de manera convencional.

En efecto, el ritmo de la vida moderna nos deja cada vez menos oportunidades para ir al banco a hacer largas filas para cambiar un cheque, depositar, pagar servicios, etc.

Esto ha hecho que nuestras demandas vayan cambiando y que seamos más exigentes y más selectivos con quienes nos prestan los diversos servicios que utilizamos cotidianamente.

Por ello, los bancos han tenido que ajustarse a las necesidades de sus cuentahabientes buscando nuevas formas de facilitarles el uso de los servicios financieros y ofreciéndoles mayores beneficios que los mantengan satisfechos y complacidos.

La banca por teléfono es el servicio perfecto para que desde la comodidad de su casa u oficina, y con una llamada telefónica, pueda realizar, en un abrir y cerrar de ojos, las operaciones bancarias que anteriormente lo obligaban a perder valiosos minutos del día.

¿Cómo opera?

La banca por teléfono es un sistema que los bancos han puesto en marcha para ayudar a sus clientes a optimizar su tiempo, ya que sólo con marcar el teléfono pueden tener acceso a todos los servicios que les ofrece una sucursal. El objetivo de la banca por teléfono es acercar los servicios del banco a sus Usuarios, evitando que éstos tengan que visitarlo físicamente.

Está disponible las 24 horas del día, los 365 días del año. Desde cualquier teléfono, y con su clave de acceso, usted puede utilizar este sistema computarizado con sólo marcar el número telefónico que para este efecto le indicó su banco.

Cada institución opera sus servicios de banca por teléfono de manera particular, pero, para todos los casos, es muy sencillo utilizarla. Desde cualquier aparato telefónico (fijo, móvil, de tonos, o de pulsos) usted se comunica al número que el ejecutivo del banco le indicó para la utilización de este sistema, y dando la clave que le dio la propia institución bancaria puede utilizar los diferentes servicios.

Algunos bancos cuentan con la modalidad del sistema automático para los servicios de identificación y consultas, que le atiende mediante procedimientos

de reconocimiento de voz, (donde el cliente opera el sistema de forma verbal) y de reconocimiento de tonos (donde el cliente opera con el sistema usando el teclado de su teléfono). Otra modalidad para la prestación de este servicio es el estilo mixto, donde el cliente puede utilizar combinaciones de los dos sistemas para realizar sus consultas.

Además, existe otra modalidad de la banca por teléfono que requiere de una operadora para que el trato sea más personal. Ello dependerá del tipo de operaciones que se pretendan realizar.

Los servicios que ofrece el sistema de banca por teléfono varían de acuerdo con cada banco, al tipo de cuenta (o cuentas) que usted maneja dentro de su banco y a las operaciones que necesite realizar.

Estos servicios son, por lo general, los siguientes:

· Consulta de saldos y movimientos de sus cuentas (fechas y montos).

· Traspasos entre cuentas.

· Pago de servicios.

· Información sobre tasas de interés, tipos de cambio y divisas.

· Solicitudes y aclaraciones.

· Cambio de domicilio.

· Solicitud de estados de cuenta.

· Solicitud de suspensión de pago de cheque(s).

· Cambio de número confidencial.

· Compra y venta de valores.

· Fondos de inversión.

· Depósitos.

· Consultas sobre su Tarjeta de Crédito (saldo disponible, fecha de pago, monto pago mínimo, monto pago de contado, pago de tarjeta, límite de crédito, disposiciones, cambio de clave, etc.).

· Transferencias entre cuentas o bien, a su fondo de pensión.

· Denuncia de robo o extravío de tarjeta.

· Solicitud de préstamos (personales o hipotecarios).

· Apertura de cuentas corrientes, de ahorro, etc.

· Solicitud de billetes y cheques de viajero.

· Planes de pensiones.

· Pago de impuestos.

Todos los clientes de los bancos tienen acceso a estos servicios mediante la banca por teléfono. Si usted aún no está registrado, puede llamar a su banco y solicitar que lo den de alta y le faciliten su clave de acceso.

Los beneficios de la banca por teléfono

· Comodidad para efectuar sus operaciones bancarias sin tener que acudir a la sucursal en los casos en los que no es necesario.

· El servicio es gratuito.

· Seguridad al contar con medidas de acceso que evitarán el mal uso de sus recursos e información.

· Flexibilidad al poder realizar sus operaciones bancarias y financieras aun fuera del horario de sucursal.

· Facilidad de empleo del sistema automático ya que le guía paso a paso para realizar la operación que usted quiera.

· Se puede utilizar desde cualquier punto geográfico, desde su casa, el extranjero, su automóvil, su trabajo.

· Dependiendo de las políticas de cada banco, los servicios automáticos están disponibles todos los días del año las 24 horas.

· Los servicios con operadora: lunes a viernes laborables: de 8 a 22 horas.

· Sábados hábiles para la banca: de 8 a 15 horas.

· Máxima seguridad en sus operaciones.

· Un sistema automático que guía al cliente en las acciones que debe realizar para sus operaciones y consultas.

Seguridad en la línea

Para su seguridad, el sistema automático realiza un proceso de identificación personal de cada Usuario. Esta identificación se compone de un **IDENTIFICATIVO** (clave de identificación) y de una **CLAVE PERSONAL** adicional.

· Identificación por voz en horario de operadora: Nombre y apellidos + CLAVE.

· Identificación por voz sin presencia de operadora: IDENTIFICATIVO + CLAVE.

· Identificación por tonos en cualquier horario: IDENTIFICATIVO + CLAVE.

¿Cómo puede hacerse Usuario de la banca por teléfono?

Sólo pregunte en su banco y solicite el servicio. La Institución Financiera le dará una guía para la utilización de este servicio y una clave de seguridad para que pueda realizar con toda tranquilidad sus consultas y operaciones.

Sin duda, cada vez son más los avances de la tecnología en diversos ámbitos del conocimiento humano, y las instituciones financieras definitivamente están aprovechando las ventajas de ésta para optimizar sus servicios , mantenerse a la vanguardia y complacer a sus clientes.

Fuente: Internet
Fecha: 28 de junio de 2004

Implementación de sistemas de reconocimiento de voz en los bancos: Natural Vox

Natural Vox, con más de diez años trabajando en el sector de las tecnologías de reconocimiento de voz, le sigue ofreciendo su tecnología, experiencia y últimas innovaciones, para mejorar de forma definitiva el servicio a sus clientes.

Hasta la fecha se han desarrollado sistemas de telefonía interactiva para clientes de reconocido prestigio en el sector de las finanzas, administración, tecnología, pymes, etc. Natural Vox es líder indiscutible a nivel nacional en implantación de sistemas de telefonía interactiva para la Banca. Muchas de las principales corporaciones financieras han optado por la implantación de nuestro producto estrella, BpT, para mejorar la calidad en el servicio telefónico a sus clientes.

El excelente funcionamiento de nuestros servicios de soporte técnico, y la extrema calidad en la atención a nuestros clientes, han permitido que todos los que han elegido a Natural Vox como proveedor de sistemas de reconocimiento de voz continúen confiándonos la implantación de sus sistemas de telefonía automática y mejora de sus centros de llamadas. Es la calidad de los sistemas desarrollados y los excelentes servicios de soporte técnico, lo que hace destacar a Natural Vox sobre los demás proveedores existentes en el mercado.

Nuestra mejor tarjeta de presentación es, sin duda, el conjunto de instalaciones ya consolidadas que diariamente dan servicio a centenas de transacciones telefónicas. Natural Vox estará encantado de proporcionarle referencias de primera mano por parte de nuestros clientes, para que usted pueda comprobar, más en profundidad, el excelente funcionamiento de los sistemas implantados.

A continuación, se presenta una breve descripción de algunas de las más exitosas instalaciones implantadas en clientes de reconocido prestigio:

- **Caja Madrid**
- **Banco Bilbao Vizcaya Argentaria**
- **Banco Atlántico**
- **Patagon - OpenBank**
- **IVESUR**
- **Iteuve Euskadi**
- **Dirección General de Tráfico**
- **Diputación Foral de Guipúzkoa**
- **Telefónica Móviles**
- **Banesto**
- **Meteológica**

1. **Caja Madrid.**

   Caja Madrid, una de las primeras cajas de ahorros a nivel nacional, tiene instalado un Sistema de Telefonía Interactiva (STI) suministrado por Natural Vox que está diseñado para atender aproximadamente 25.000 comunicaciones diarias procedentes de cualquier lugar del territorio nacional.

   La conversación entre el usuario y el sistema está basada en la comunicación oral, es decir, predomina el reconocimiento de voz. No obstante, aquellos clientes que dispongan de un terminal telefónico capaz de emitir tonos multifrecuencia, pueden acceder tecleando aquellos datos numéricos que se soliciten.

   El sistema implantado en Caja Madrid permite que sus clientes accedan de forma ágil y eficaz, a información relacionada con sus cuentas. También les permite disfrutar de diferentes servicios proporcionados por la entidad, evitando la necesidad de desplazarse a una de las oficinas.

2. **Banco Bilbao Vizcaya Argentaria (BBVA).**

   El Sistema de Telefonía Interactiva instalado en BBVA, una de las principales corporaciones bancarias a nivel nacional e internacional, está diseñado para realizar por teléfono un amplio número de operaciones bancarias.

   Al igual que otros sistemas BpT diseñados por Natural Vox, está basado en reconocimiento de voz. La conversación se inicia con la pregunta: ¿Qué desea?. Esta innovadora interfaz de usuario, da un toque realmente humano al sistema, y deja patente la filosofía de funcionamiento de Natural Vox: interactuar con el usuario de la forma más natural posible.

   Mediante este sistema la entidad bancaria ofrece a sus clientes un servicio de gran calidad y fiabilidad, que les permite realizar vía telefónica, las 24 horas

del día y 365 días al año, una gran parte de las operaciones relacionadas con la gestión de sus cuentas.

3. **Banco Atlántico.**

   Banco Atlántico, otro de los bancos más importantes a nivel nacional, ha optado por ofrecer a sus usuarios un sistema de información bancaria hot-line basado en el reconocimiento vocal que desarrolla Natural Vox. Esta instalación, plenamente consolidada, permite a los clientes del banco acceder desde cualquier teléfono a la información de sus cuentas, y realizar un amplio número de operaciones, sin necesidad de desplazarse hasta una sucursal.

4. **Patagon - OpenBank.**

   Patagon - OpenBank, entidad financiera perteneciente al Grupo BSCH que sigue una innovadora filosofía de atención al usuario al no utilizar sucursales, ha decidido utilizar como complemento en sus líneas de atención a los clientes el sistema BpT proporcionado por Natural Vox.

   El sistema implantado, basado en las más modernas tecnologías de reconocimiento fonético, permite a los clientes de la entidad acceder a numerosos servicios de forma cómoda, ágil y eficaz.

5. **IVESUR (Inspecciones de vehículos del Sur, S.A.)**

   IVESUR, primera empresa española dedicada a ITV que recibió el certificado de registro de empresa AENOR, conforme a la norma ISO 9002, dispone de un sistema telefónico de cita previa desarrollado por Natural Vox, que permite ofrecer un servicio de mayor calidad a sus clientes.

   Este sistema está desarrollado utilizando tecnología de reconocimiento de voz, y permite a los clientes de IVESUR pedir cita para realizar la inspección técnica de su vehículo las 24 horas del día.

6. **Iteuve Euskadi.**

   El sistema de telefonía automática desarrollado para Iteuve Euskadi, permite al igual que el implantado en IVESUR, coger cita telefónica las 24 horas del día, para realizar la ITV en diferentes estaciones. Mediante este aplicativo, Iteuve Euskadi ha podido reducir en gran medida los costes derivados de la atención para recogida de citas, permitiendo a sus clientes disfrutar de un sistema de cita previa eficaz y de sencilla utilización.

7. **Dirección General de Tráfico.**

   Uno de los servicios más solicitados y útiles proporcionados por la Dirección General de Tráfico es el de información sobre el estado de las carreteras. Este servicio, disponible las 24 horas del día y 365 días al año, es utilizado de forma masiva en operaciones de salida y retorno de vacaciones, y en días con situaciones meteorológicas de riesgo para la conducción.

Con el objetivo de mejorar la calidad de este servicio, y permitir una gestión más eficaz de los picos de llamadas que se producen en determinados días del año, la DGT ha instalado un sistema de reconocimiento vocal desarrollado por Natural Vox. Este sistema permite que un mayor número de personas pueda conocer telefónicamente la información sobre el estado de las carreteras de forma simultanea, aumentando así la accesibilidad al servicio.

8. **Diputación Foral de Guipúzcoa.**

Cada año, durante la campaña de renta, un gran número de personas se ponen en contacto con la Diputación Foral de Guipúzcoa para realizar su declaración. El sistema de telefonía interactiva suministrado por Natural Vox para la Diputación Foral de Guipúzcoa, permite atender aproximadamente 4000 comunicaciones diarias.

De esta forma, se reducen notablemente las inversiones en personal, tiempo, y dinero, de la Diputación a la hora de hacer frente a este periodo del año. El sistema de telefonía automática implantado permite, a los contribuyentes, concertar cita las 24 horas del día, y posibilita a la Diputación atender a un mayor número de comunicaciones simultáneas.

9. **Telefónica Móviles.**

Telefónica Móviles, es líder en el desarrollo y gestión de actividades relacionadas con la telefonía móvil y los servicios de radiobúsqueda y radiotelefonía. Esta empresa ha seleccionado a Natural Vox para desarrollar proyectos relacionados con las tecnologías del habla, que le permitan ofrecer un mejor servicio a sus usuarios, y mejorar sus actividades corporativas.

10. **Banesto**

En Enero de 2002, Banesto decidió instalar un STI de Natural Vox diseñado para realizar operaciones bancarias por teléfono. Esta aplicación es la primera que Natural Vox ha desarrollado íntegramente utilizando su Generador de Aplicaciones GAP. Queda patente la flexibilidad de GAP para desarrollar aplicaciones, desde las más sencillas a las más complejas, como es el caso de las aplicaciones bancarias.

En este Sistema predomina el reconocimiento en la interacción entre el usuario y el sistema. El sistema de Banesto cuenta con la última tecnología tanto en reconocimiento, con Reconocimiento Avanzado, como en Síntesis en tiempo real, ya que cuenta con la última versión de síntesis desarrollada por Natural Vox.

Cabe destacar que Banesto es el primer cliente que cuenta con nuestro Generador gráfico de Aplicaciones, GAP. La potencialidad que ofrece esta herramienta es enorme, pudiendo Banesto diseñar y modificar sus propias aplicaciones, de una forma muy sencilla, de tal forma que incluso personal no experto en programación IVR puede hacer sus propias aplicaciones. Una vez

que la aplicación está diseñada y es enviada a compilar, pasan menos de 5 minutos hasta que se devuelven los ejecutables a instalar en el STI.

11. **Meteológica**

Este Sistema de Telefonía Interactiva está diseñado para dar información meteorológica. La interacción entre el usuario y el sistema se inicia con una pregunta abierta que permite al usuario expresarse con absoluta libertad y obtener de manera fluida la información meteorológica que requiera, acerca de cualquier municipio, estación de esquí o zona costera.

El Sistema dispone de información para 10 días, si se trata de municipios y estaciones de esquí, y de 5 días para las zonas de costa.

El usuario puede solicitar información sobre el tiempo en general o sobre las siguientes especificaciones: niebla, nubosidad, precipitaciones, tormenta, nieve, temperatura, viento, estado de las pistas (en caso de solicitar información de estaciones de esquí) y estado de la mar (en caso de solicitar información sobre zonas de costa).

También puede pedir la información relativa a un día en concreto o escuchar la información para todo el intervalo de que dispone el sistema, siendo ésta una característica que no está presente en otros sistemas de información meteorológica.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la nota de prensa: junio de 2004
Link en la red: http://www.natvox.es/clientes.html

13. **Opiniones y conclusiones:**

La mayoría de los expertos han valorado de forma muy positiva la incorporación de este sistema en los procesos de producción. A continuación destacamos algunas de las consecuencias del reconocimiento de voz aplicado al trabajo.

“La revolución de los sistemas de reconocimiento de voz”
Por Heriberto Covarrubias, Gerente General de Switch

No cabe ninguna duda que la voz es la interface natural del ser humano, es decir, el medio más espontáneo y simple que tenemos para comunicarnos. Ejemplo de ello es que para hacernos entender no requerimos de ningún dispositivo especial más que nuestra voz. Además, mientras hablamos podemos estar realizando diferentes tareas al mismo tiempo.

Pese a que la evolución en el ámbito de las comunicaciones ha sido revolucionaria en el último tiempo, aún estamos acostumbrados a la necesidad de utilizar diferentes intermediarios para lograr una comunicación a distancia. Por ejemplo, estamos habituados a discar un aparato para hablar por teléfono y a usar el teclado o el mouse para comunicarnos con nuestro computador. Sin embargo, esta realidad ya está cambiando, gracias al vertiginoso avance de las tecnologías.

Entendiendo que el desarrollo tecnológico tiene como principal objetivo hacernos la vida cada día más fácil, es que los avances de los últimos años se han concentrado en rescatar a la voz como la interface idónea para lograr que nuestros sistemas de comunicación resulten más eficientes y cómodos. Se trata de un salto cualitativo que se viene desarrollando con ímpetu, más aún si consideramos que el tamaño físico de nuestros dedos es la principal limitante para no poder reducir aún más el tamaño de teléfonos, teclados y otras interfaces existentes. Por lo mismo, es que la implementación de la voz se convierte en el medio idóneo para acceder a dispositivos de comunicación remotos cada vez más pequeños.

Una década de avances

La evolución en los sistemas de reconocimiento de voz ha sido realmente impresionante. En sólo 10 años pasaron de ser sistemas discretos -que reconocían palabra por palabra y número por número- a sistemas continuos y naturales. Esto significa que si el computador le pregunta al usuario cuál es el origen y destino de su viaje y éste le responde de Antofagasta a Arica, lo que realmente le interesa al computador son los conceptos "Antofagasta y Arica", independientemente de las demás palabras que contenga la respuesta.

Otro avance es el que se ha dado con los números, puesto que en la actualidad el PC es capaz de reconocerlos en forma continua, es decir, desde el simple 1, 2 y 3 hasta la interpretación correcta de 1.234.

Existen, además, otros sistemas que han evolucionado notablemente en el último tiempo, como son el *Text to Speach* y la *Verificación de Voz*. El primero es una tecnología que permite convertir cualquier texto en voz. Un salto cualitativo se ha

dado en este ámbito, ya que antiguamente la voz del equipo era muy robótica -ya que traducía palabra por palabra- mientras que hoy es capaz de dar la entonación correspondiente a cada frase, siendo cada vez más parecida a la voz humana. En tanto, la Verificación de Voz es la tecnología que permite reconocer la identidad de la persona que está hablando. Se aplica con frecuencia en sistemas de seguridad y, si bien aún no es 100% segura, no cabe duda de que pronto alcanzará niveles de verificación más eficientes.

El mercado de los servicios

El reconocimiento de voz es, en definitiva, la tecnología que se está imponiendo en el mundo moderno. La madurez alcanzada por este tipo de herramientas se traduce en que éstas se están aplicando ampliamente en el mercado de los servicios, como, por ejemplo, en líneas aéreas, bancos, empresas de seguros, agencias de viajes y bolsas de valores. Gracias a ello es que muy pronto tendremos acceso a la información que necesitemos, sin importar dónde estemos, qué estemos haciendo o qué tan lejos nos encontremos de los centros de información a los que queramos acceder. Todo esto con sólo hablarle a un dispositivo más pequeño que un lápiz.

Fuente: Internet
Fecha: 28 de junio de 2004
Fecha de la redacción del artículo: 28 de junio de 2004
Link en la red: http://www.gerencia.cl/articulo.mv?sec=3&num=103

Consecuencias del tratamiento de la voz

La voz va a ser, según todas las investigaciones, una **herramienta** clave, que llegará a permitirnos **"hablar con las máquinas"**, y facilitar así a muchas personas la accesibilidad a los principales avances de la vida cotidiana; ya se aplica, por ejemplo, en sistemas de información telefónica o programas de dictado. La cuestión ahora es crear un corpus donde se registren la mayor cantidad de **variantes contextuales** de **sonidos**. Con los **sintetizadores de voz** las personas ciegas o con graves problemas de visión pueden reconocer los mensajes de la pantalla de un **ordenador**, ajustando la velocidad, el timbre de voz o el tono.

En la **Universidad Politécnica** de Madrid investigan un sistema para dotar de **"emociones"** a la **voz sintética** y, en general, los sistemas de reconocimiento de voz van mejorando sensiblemente, aunque siguen derivando problemas las voces distorsionadas. *"El principal **avance** será que, un día, el reconocimiento pueda recoger no sólo una voz maravillosa y clara, sino **todo tipo de voces**. Sobre todo para aquellas personas con **problemas en el habla**, que en definitiva son las que más lo necesitan"*, matiza Cristina Rodríguez.

Para ella, que lleva muchos años trabajando en el campo de las **ayudas técnicas**, el futuro exigirá ir aumentando **opciones**; es decir, que se pueda dar la orden de encender o apagar una luz, pero manteniendo la opción de pulsar. *"Siempre habrá personas -explica- que, por costumbre, por cultura, etc. quiera seguir funcionando de un **modo tradicional**, y nadie tiene por qué imponer estos avances"*

Las tecnologías de reconocimiento de voz avanzan rápidamente, los especialistas destacan su potencial y aseguran que, con la reducción de los precios de las nuevas tecnologías, los sistemas de reconocimiento de voz se convertirán en una industria multimillonaria a principios del próximo siglo y, seguramente, en una forma de acceso de los discapacitados a las nuevas tecnologías. La expansión de este mercado, que ya se ha iniciado, es posible gracias al descenso en los costes de implementación de los sistemas de reconocimiento de voz en cada chip. Esta reducción puede ser la principal aliada para el avance de la tecnología y su popularización.

Sin embargo, a pesar del optimismo de muchos especialistas y de las conclusiones de algunos estudios, Nicholas Negroponte asegura que la tecnología de reconocimiento de voz ha avanzado más lentamente de lo que se esperaba. Varios estudios aseguran que los sistemas de reconocimiento de voz empezarán a popularizarse en los ordenadores domésticos y aparatos electrónicos en los próximos años. Un ejemplo de esto es la Corporate America, una empresa que utiliza un software que permite a los clientes obtener informaciones a través del teléfono. Un sistema similar está siendo desarrollado por American Airlines, por el cual el usuario puede realizar reservas de billetes aéreos por teléfono a través de una máquina.

Fuente: Internet
Fecha: 28 de junio de 2004
Link en la red: http://greysis.tripod.com/reconocimiento_de_voz.htm

14. **Links en la red de empresas que se dedican al tratamiento de la voz:**

A) **Tuvox:** http://www.tuvox.com/
B) **Philips:** http://www.philips.es/InformationCenter/NO/FArticleDetail.asp?lArticleId=2877
C) **Scansoft:** http://spain.scansoft.com/embedded/
D) **Sodels:** http://www.sodels.com
E) **Albisa:** http://www.albisa-solutions.com/inicio.asp
F) **VocalTec:** http://www.vocaltec.com/

***Notas:**

A) La empresa **ART** de **Israel** sacó en **1990** su programa de reconocimiento de voz, como los que se usan en algunos teléfonos celulares con los cuales sólo se necesita decir el nombre de la persona a la que se quiere llamar. Esta tecnología fue explotada por **Samsung** y **Lucent.**
B) Algunas empresas han utilizado la tecnología **CTI (Computer Telephony Integration)** para la compatibilidad en los sistemas de reconocimiento de voz.
C) En la **Edición de los Premios Internacionales de I+D de 2002,** se conoció el fallo con respecto a la **Edición de los Premios Internacionales I+D de 2001,** que concede la **Organización Nacional de Ciegos Española (ONCE),** por lo que la **Fundación Bosch Gimpera de Barcelona (España)** y la **Universidad Oxford Brookes** recibieron el galardón y tercer premio por **"I Speak: un juego de acceso a Internet basado en el reconocimiento de la voz".**
D) Por su parte, la **ONCE,** el **4 de noviembre de 1999,** ya anunciaba y precisaba a las empresas de software la eliminación de los obstáculos pertinentes para que los ciegos pudiesen aprovecharse de las ventajas del ordenador. En su día, **Enrique Fernández, Director General de la Organización,** exponía que el sistema de reconocimiento de voz "debería de eliminar las barreras que representan para un ciego el teclado y la pantalla".

15. **Anexo:** sistemas de reconocimiento de voz en ambientes virtuales (ámbito ciencia).

Sistemas de reconocimiento de voz en ambientes virtuales

Resumen

El objetivo de esta ponencia es describir una introducción acerca de la posible aplicación de sistemas de reconocimiento de voz en ambientes de realidad virtual, y explicar la implementación de un sistema de realidad virtual, donde se utilizaron comandos de voz a través una red neuronal basada en hardware, para controlar moléculas virtuales, el cual fue implementado en el Laboratorio de Realidad Virtual de la Universidad de Colima. El trabajo aquí mostrado sienta las bases para futuros desarrollos de reconocimiento de voz en dicho laboratorio.

Palabras clave: redes neuronales, realidad virtual, reconocimiento de voz, biología molecular.

Introducción

Uno de los objetivos de la realidad virtual es el de utilizar más de un sentido sensorial humano para interactuar con la información presentada en un ambiente tridimensional gráfico (Kalawsky, 1993), ya sea con fines educativos, comerciales, o de investigación, entre otros (García Ruiz, 1998). Una característica particular de la realidad virtual es el uso de dispositivos de entrada especiales para manipular directamente información contenida en un ambiente gráfico para facilitar su análisis y comprensión, por ejemplo, guantes de datos y ratones para 3D (Shneiderman, 1998). De acuerdo a Preece et al. (1994), la realidad virtual interactiva se basa en teorías y lineamientos determinados en la Interacción Humano-Computadora. Se pueden utilizar técnicas de reconocimiento de voz para controlar la información, como por ejemplo, activar estados y obtener resultados, entre otras actividades, dentro del ambiente virtual. Los sistemas de reconocimiento de voz pueden estar basados en software y hardware. Los sistemas basados en software, si bien el repertorio de palabras manejadas es amplio y son fácilmente configurables, normalmente consumen mucho procesamiento del CPU de la computadora, memoria y espacio de almacenamiento. Entre los programas para reconocimiento de voz comerciales más utilizados se encuentran el Viavoice de IBM (http://www-3.ibm.com/software/speech/dev/index.shtml), y el Dragon Naturally Speaking (http://www.scansoft.com/naturallyspeaking/developers/).
¿Por qué utilizar reconocimiento de voz en un ambiente de realidad virtual? Es posible que el usuario de un ambiente virtual tenga las manos ocupadas en alguna tarea, por ejemplo, “sosteniendo” algún objeto virtual con guantes de datos, y necesite realizar otra tarea a la vez. En este caso se podría utilizar un comando de voz para realizar esa tarea extra.

Los sistemas de reconocimiento de voz basados en hardware emplean un circuito integrado con una red neuronal interna, integrando además filtros pasivos y activos para el micrófono, puertos de entrada y salida y memoria EEPROM, entre otros componentes. Estos tipos de sistemas son rápidos, con un grado alto de exactitud, y de

costo más económico que los programas de reconocimiento de voz. Hay que considerar que la cantidad de palabras que se pueden reconocer en este tipo de circuitos es mucho menor a las utilizadas en el reconocimiento de voz por software.

Implementación

Uno de los objetivos del Laboratorio de Realidad Virtual de la Universidad de Colima (perteneciente al CEUPROMED) ha sido el de ofrecer un ambiente de trabajo e investigación para el desarrollo de proyectos de tesis de cualquier nivel de estudios. Tal es el caso de la tesis de ingeniería desarrollada por Ceja Castillo y Mendoza Chávez (2003), donde utilizaron un circuito de reconocimiento de voz Voicedirect 364 para controlar con comandos de voz tres representaciones de una molécula (aminoácido), todo esto montado en un visualizador de mundos virtuales llamado DIVE, desarrollado por el Instituto Sueco de Computación (Carlsson y Hagsan, 1993). En la figura 1 se muestra la interfaz gráfica de DIVE. Se decidió abordar el tema de análisis de estructuras moleculares, ya que éste presenta ciertas dificultades para manipular información molecular con interfaces gráficas utilizando dispositivos de entrada convencionales, como el ratón, además de que es un tema de difícil comprensión y visualización (García Ruiz, 2002). Es por esto que se vio la alternativa del uso de reconocimiento de voz.

El circuito fue conectado a través del puerto paralelo de una PC, y fue accedido con un programa escrito en Visual Basic. Las tareas que se idearon para la prueba de manipulación con la voz consistieron en cambiar la representación gráfica de una molécula en particular, siendo el aminoácido alanina el que se utilizó para el desarrollo. Se utilizó la representación gráfica de esta molécula por su simple estructura. Para esto, se utilizaron los comandos verbales “stick”, “CPK”, y “ball and stick” para cambiar su representación. Además, la molécula pudo ser rotada en cuatro direcciones utilizando comandos de voz utilizando los comandos verbales “arriba”, “abajo”, “izquierda”, y “derecha”. Las figuras 1 y 2 muestra el mundo virtual y el circuito utilizados en el proyecto de tesis.

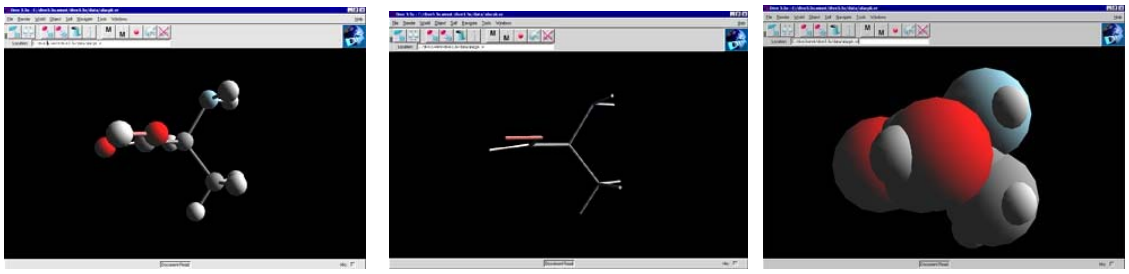


Figura 1. Las tres representaciones de la molécula alanina utilizadas con el circuito reconocedor de voz.

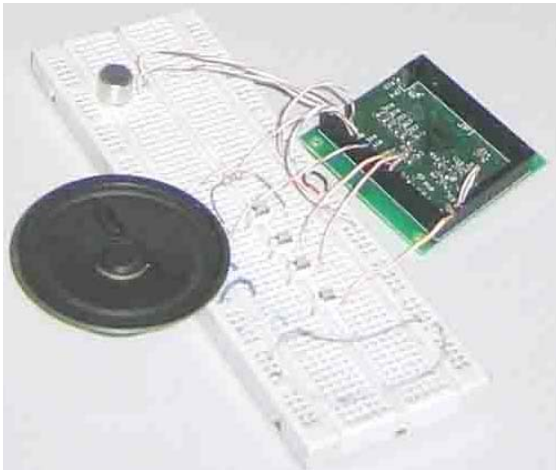


Figura 2. Circuito Voicedirect 364.

Dado que el circuito Voicedirect funciona en modo “dependiente del usuario”, se entrenó al circuito previo a su uso, por medio de la grabación de los comandos de voz directamente en el circuito, utilizando el micrófono provisto.

Previo a la implementación del sistema, las alumnas realizaron una comparación de técnicas de reconocimiento de voz por software y hardware, siendo favorable el uso del circuito de reconocimiento de voz. Más información sobre el circuito Voicedirect 364 puede verse en: www.sensoryinc.com.

Trabajo a Futuro

El paso siguiente será aplicar evaluaciones de usabilidad (Bekker y Vogten, 1999; Nielsen y Mack, 1994) del ambiente virtual y su interacción con el circuito reconocedor de voz, y además llevar a cabo pruebas de desempeño técnico para determinar la eficiencia y tiempo de respuesta exacto del circuito Voicedirect, y con esto mejorar la interconexión e interacción entre el circuito y la computadora.

Conclusiones

Esta ponencia describió una introducción a los sistemas de reconocimiento de voz y su posible uso en la realidad virtual. Además, se explicó el desarrollo e implementación de reconocimiento de voz por hardware en un mundo virtual, y se realizó una comparación de sistemas de reconocimiento de voz por software y hardware, donde el último ofrece ventajas en cuanto a velocidad de reconocimiento y costo. La implementación descrita en esta ponencia servirá de base para futuras aplicaciones, en otros campos y áreas del conocimiento que requieran manipulación directa de información por medio de la voz, por ejemplo, simulación de procesos industriales o tableros de control virtuales.

Agradecimientos

Los autores desean agradecer al personal del Centro Universitario de Producción de Medios Didácticos (CEUPROMED) por su apoyo técnico.

Referencias bibliográficas

Bekker, M.M. y Vogten, L.L.M. (1999). Usability of Voice-controlled Product Interfaces. IPO Annual Progress Report, 34. Pp. 111-124.

Ceja Castillo, A.E. y Mendoza Chávez, C.I. (2003). Análisis Grafico de Aminoácidos Utilizando Técnicas de Reconocimiento de Voz. Manipulación de Reconocimiento de Voz por Software y Hardware. Tesis de Ingeniería en Telemática (tesis no publicada). Universidad de Colima, México. Facultad de Telemática.

Carlsson C., y Hagsan, O. (1993). DIVE – A Platform for Multi-User Virtual Environments. Computers and Graphics, 17(6) .

García Ruiz, M.A.(1998). Aplicaciones de la Realidad Virtual en la Educación: Breve Panorama General. Educación 2001, No. 43, pp.37-40.

Garcia-Ruiz, M.A. (2002). Binding Virtual Molecules Sounds Good!: Exploring a Novel Way to Convey Molecular Bonding to Students. Proceedings of E-learn 2002, Association for the Advancement of Computing in Education, 16-19 October, Montreal, Canada.

Kalawsky, R. (1993). The Science of Virtual Reality and Virtual Environments. Addison-Wesley: Wokingham, UK.

Nielsen, J., y Mack, R. L. (Eds.) (1994). Usability Inspection Methods. John Wiley & Sons: New York.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). Human-Computer Interaction. Wokingham: Addison-Wesley.

Shneiderman, B. (1998). Designing the User Interface. Addison-Wesley: Reading, MA.

Estudio realizado:

Miguel Ángel García Ruiz, Adriana Elizabeth Ceja Castillo, Candelaria Isabel Mendoza Chávez
Laboratorio de Realidad Virtual, CEUPROMED. Universidad de Colima. Ave. Universidad 333.
Tel./fax: (312)3161093
Correos electrónicos: mgarcia@ucol.mx, Adriana_ceja@hotmail.com y mcandyc@hotmail.com

16. Anexo II: soluciones y ayudas técnicas: productos y aplicaciones para personas con discapacidad

Los discapacitados son personas que precisan de las **Nueva Tecnologías** para poder llevar a cabo su vida, y sobre todo, de la manera más cómoda posible, por lo que algunas entidades privadas y fundaciones han puesto a su disposición una serie de productos basados en el reconocimiento de voz como medio para cerrar una puerta, una ventana o encender la televisión.

SICARE LIGHT (MÁS LIBERTAD, MÁS INDEPENDENCIA)

Las personas que padecen una discapacidad que limita su movilidad deben mantener la mayor independencia y autonomía personal. Ciertas acciones cotidianas y habituales, como el funcionamiento de una cama hospitalaria, el teléfono, la luz o el sistema de aviso de enfermería, deberían ser realizados sin la necesidad de ayuda externa. Al final de este anexo aparecen los teléfonos de contacto de las compañías involucradas.

- Sicare light es un pequeño y completo mando de control con reconocimiento de voz para personas con discapacidad física.
- Funciona fácilmente con la voz y permite controlar los aparatos que estén provistos de receptor por infrarrojos como, por ejemplo, ventanas, puertas, televisiones, equipos de música y muchos más, usando simplemente su voz.
- Después de una corta y simple fase de entrenamiento, el Sicare light podrá entender cualquier idioma.
- Control total. Por ejemplo, de ventanas, puertas, sistemas de aviso de enfermería, camas eléctricas, teléfonos, televisiones, videos, equipos de música, persianas, luces, calefacciones, aire acondicionado, ventiladores, electrodomésticos (siempre y cuando tengan incorporados receptores de infrarrojos).

CARACTERÍSTICAS TÉCNICAS

- Reconocimiento. Reconoce la voz en 7 idiomas: Alemán, Inglés, Holandés, Francés, Italiano, Español, Portugués.
- Incluye, los códigos infrarrojos más frecuentes.
- Sistema de infrarrojos, con aproximadamente 10 m de distancia.
- Corta y simple fase de entrenamiento.
- Portátil, tamaño de la palma de la mano.
    - Dimensiones: (125 x 88 x 55 mm.)
    - Transmisor de infrarrojos.
    - Incluye:
        - Altavoz.
        - Clip.
        - Pantalla (display).
        - Programador de códigos.
        - Micrófono.

SENIOR PILOT (Para la independencia y autonomía personal)
El único mando a distancia por infrarrojo y universal diseñado específicamente para personas mayores y para ayuda personal.

- Una solución necesaria: Senior Pilot es un equipo de ayuda personal y control del entorno que fue diseñado para potenciar o aumentar la autonomía personal e independencia de personas con discapacidad de bajo nivel de Símbolos limitación,
- El mejor ayudante: Senior Pilot puede controlar la televisión, la minicadena de música o el video, pero además, también podrá abrir y cerrar una puerta o una ventana motorizada, o encender y apagarla luz, o variar la temperatura del aire acondicionado o subir las persianas motorizadas, o avisar pidiendo ayuda, o controlar cualquier equipo que funcione con mando a distancia por infrarrojos.
- Tecnología amigable: Senior Pilot tiene 14 teclas muy grandes, para que sean fáciles de ver y de manejar, con símbolos de colores y las teclas se iluminan si es necesario.
- En caso necesario, tiene la posibilidad de emplear un pulsador automático y las teclas se iluminarían, de una en una.

CARACTERÍSTICAS TÉCNICAS DE SENIOR PILOT

- Funciona con pilas (4 x 1,5 Volt).
- Medidas: 21 cm x 7 cm x 3 cm.

BRS (SISTEMAS DE REHABILITACIÓN) *Sistema electrónico interactivo para el estímulo del reflejo propioceptivo.*
BRS es una plataforma giratoria electrónica que puede rotar 360º (180º en posición simple). El movimiento oscilatorio realizado en la plataforma es captado por un sofisticado sistema de sensores térmicos y se transfiere por radio-frecuencia a una computadora persona.
Cada movimiento del paciente puede controlarse, grabarse y en consecuencia puede analizarse.

Es posible realizar ejercicios de rehabilitación marcando objetivos al paciente:

- **EVALUACIÓN.** Es un instrumento indispensable para obtener una evaluación objetiva de la situación y evolución del apciente. Haciendo uso de la plataforma usted puede controlar el progreso del tratamiento rehabilitador en el paciente.
- **REHABILITACIÓN.** Es posible monitorizar la evolución o progreso y acortar el tiempo de rehabilitación. La plataforma puede emplearse en posición bi-podalica, en mono-podalica y en posición sentada. Usted puede aumentar la seguridad en la deambulación.
- E**NTRENAMIENTO.** Mediante los ejercicios de la plataforma se puede incrementar la capacidad de coordinación motora. El sistema puede emplearse para preparar los centros nerviosos antes de iniciar el entrenamiento para deportistas de alto nivel.
- **PREVENCIÓN.** Es un sistema fácil de utilizar que ayuda a incrementar la capacidad de equilibrio y en consecuencia disminuye la probabilidad de padecer lesiones.

Más información:

Recursos y adaptaciones tecnológicas ATR
Tel.: (800) 770-84 74
Fax: (262) 375-67 77
http://www.adaptivetr.com/

PROINSSA
Promoción de IniciativasSocio-Sanitarias, S.L.L.
Sta. Teresa de Jesús, 10 - 1.º B
28400 COLLADO VILLALBA (Madrid)
Tel.: 91 849 90 69 - Fax 91 849 90 86
E-mail: info@prainssa.com
http://www.proinssa.com/index1024.htm

Buratto Advanced Technology S., I. v
Vía Erizzo, 54
31030 Covolo di Pederobba (TV)
ITALIA
Tel.: (+39)-0423-68 23
Fax: (+39)0423 - 68 80 39
Email: bat@buratto.it
http://bat.buratto.it/

PROMOCIONES DE INICIATIVAS SOCIO-SANITARIAS
DISTRIBUIDOR OFICIAL EN ESPAÑA Y PORTUGAL
C/ Batalla de Bailén, 24
Edificio Europa, oficina 33.
COLLADO VILLALBA - MADRID
Tel.: (91) 949 90 69
Fax: (91) 849 90 86
Email: brs@proinssa.com
http://www.proinssa.com/

Fuente: Internet
Fecha: 28 de junio de 2004
Link en la red: http://www.discapnet.es/

17. **Anexo III:** sector sanitario e innovaciones de **Philips**

Una de las empresas que más han apostado por el reconocimiento de la voz en **España,** ha lanzado a **14 de junio de 2004,** el siguiente producto para todos los hospitales en nuestro país, aunque los primeros beneficiados han sido los de la **Comunidad Autónoma de Madrid.**

Philips introduce en España su solución de reconocimiento de voz para todos los servicios hospitalarios

> Barcelona, España y Viena, Austria - Philips Speech Processing anuncia el lanzamiento de su ConTexto Multimed en español (diccionario profesional de medicina general) para ser utilizado en su software de reconocimiento de voz, SpeechMagic. Este lanzamiento, coincide con la decisión de tres hospitales españoles en ser los primeros en extender el uso del sistema de reconocimiento desde el departamento de Radiología a otras áreas del hospital. En el Universitario Clínico San Carlos de Madrid, 600 doctores tendrán acceso al reconocimiento de voz.

El ConTexto MultiMed cubre la terminología especializada para más del 90% de todas las disciplinas médicas y se basa en más de mil millones palabras obtenidas de informes médicos. Se estructura en varios módulos, incluyendo cardiología, medicina interna, medicina de urgencia, cirugía general etc. La decisión de desarrollar el ConTexto MultiMed se puso en marcha gracias a la aceptación masiva de SpeechMagic en España, donde a lo largo del primer año desde la introducción del ConTexto Radiología se ha alcanzado una penetración de mercado del 20%.

Los primeros que ya han tomado la decisión de implantar SpeechMagic con el ConTexto Multimed en la totalidad de sus servicios son el Universitario Clínico San Carlos de Madrid, el Hospital Universitario Virgen de las Nieves y el Hospital Universitario San Cecilio de Granada.

Como explica José Soto, Gerente del Universitario Clínico San Carlos: "Hace un año implantamos reconocimiento de voz en Radiología, donde hemos obtenido un aumento significativo de la eficiencia del departamento y una mejora en el servicio ofrecido a nuestros pacientes, ya que los informes médicos pueden ser finalizados rápidamente después de haberse realizado el examen médico. Esta positiva experiencia obtenida en Radiología, así como los resultados alcanzados durante la fase de prueba con el ConTexto MultiMed, nos han llevado a tomar la decisión de dar acceso a SpeechMagic a 600 médicos más en otros departamentos del Hospital"

Estas instalaciones serán llevadas a cabo por la empresa NewDoors, socio tecnológico de Philips y líder en el mercado español en la implantación de soluciones profesionales de reconocimiento de voz.

Como comenta Jesús Alvárez Director General de NewDoors: "Estas primeras instalaciones, así como las conversaciones que estamos teniendo con otras instituciones líderes en el sector sanitario nos hacen ser muy optimistas y pensar que SpeechMagic reforzará su posición de liderazgo en el mercado sanitario español".

Fuente: Philips
Fecha: 28 de junio de 2004
Link en la red: http://www.speechrecognition.philips.com/index.php?id=5&full=1034

18. Anexo IV: historia del reconocimiento de voz (cuadro y en inglés)

Cronología

Año	Resumen
1936	AT&T's Bell Labs produced the first electronic speech synthesizer called the Voder (Dudley, Riesz and Watkins). This machine was demonstrated in the 1939 World Fairs by experts that used a keyboard and foot pedals to play the machine and emit speech.
1968	The world-popular science fiction movie *2001: A Space Odyssey* introduced the idea of speech recognition with the space ship computer, HAL.
1969	John Pierce of Bell Labs said automatic speech recognition will not be a reality for several decades because it requires artificial intelligence.
1970	The Hidden Markov Modeling (HMM) approach to speech recognition was invented by Lenny Baum of Princeton University and shared with several ARPA (Advanced Research Projects Agency) contractors including IBM. HMM is a complex mathematical pattern-matching strategy that eventually was adopted by all the leading speech recognition companies including Dragon Systems, IBM, Philips, AT&T and others.
1971	DARPA (Defense Advanced Research Projects Agency) established the Speech Understanding Research (SUR) program to develop a computer system that could understand continuous speech. Lawrence Roberts, who initiated the program, spent $3 million per year of government funds for 5 years. Major SUR project groups were established at CMU, SRI, MIT's Lincoln Laboratory, Systems Development Corporation (SDC), and Bolt, Beranek, and Newman (BBN). It was the largest speech recognition project ever.
1978	The popular toy "Speak and Spell" by Texas Instruments was introduced. Speak and Spell used a speech chip which led to huge strides in development of more human-like digital synthesis sound.
1982	Dragon systems was founded in 1982 by speech industry pioneers Drs. Jim and Janet Baker. Dragon Systems is well known for its long history of speech and language technology innovations and its large patent portfolio.
1984	SpeechWorks, the leading provider of over-the-telephone automated speech recognition (ASR) solutions, was founded.
1995	Dragon released discrete word dictation-level speech recognition software. It was the first time dictation speech recognition technology was available to consumers. IBM and Kurzweil followed a few months later.
1996	Charles Schwab is the first company to devote resources towards developing up a speech recognition IVR system with Nuance. The program, Voice Broker, allows for up to 360 simultaneous customers to call in and get quotes on stock and options... it handles up to 50,000 requests each day. The system was found to be 95% accurate and set the stage for other companies such as Sears, Roebuck and Co., and United Parcel Service of America Inc., and E*Trade Securities to follow in their footsteps.
1996	BellSouth launches the world's first voice portal, called Val and later Info By Voice.
1997	Dragon introduced "Naturally Speaking", the first "continuous speech" dictation software available (meaning you no longer need to pause between words for the computer to understand what you're saying).
1998	Lernout & Hauspie bought Kurzweil. Microsoft invested $45 million in Lernout & Hauspie to form a partnership that will eventually allow Microsoft to use their speech recognition technology in their systems.
1999	Microsoft acquired Entropic, giving Microsoft access to what was known as the "most accurate speech recognition system" in the world.

2000	Lernout & Hauspie acquired Dragon Systems for approximately $460 million.
2000	TellMe introduces first world-wide voice portal.
2000	NetBytel launched the world's first voice enabler, which includes an on-line ordering application with real-time Internet integration for Office Depot.

Fuente: Internet y **Netbytel**
Fecha: 28 de junio de 2004
Link en la red: http://www.netbytel.com/literature/e-gram/technical3.htm

Informe elaborado para ASSIT, por Jorge Hierro Álvarez
Madrid, 28 de junio de 2004