Cristina Villoria -k idatzia
Asteartea, 2009(e)ko martxoa(r)en 31-(e)an 10:11etan

There are no translations available.
La accesibilidad en el mundo de la informática es la tarea prioritaria a la hora de desarrollar nuevos programas y componentes para nuestros ordenadores...

INTRODUCCIÓN

La accesibilidad en el mundo de la informática es la tarea prioritaria a la hora de desarrollar nuevos programas y componentes para nuestros ordenadores. Gracias a esto, se puede expandir el uso de un ordenador a todo el mundo, incluidas las personas que tienen diversas discapacidades físicas, como pueden ser personas con deficiencia visual a las que les es prácticamente imposible utilizar un periférico tan necesario como es la pantalla, o personas con discapacidad motriz a las que utilizar el ratón o el teclado les puede resultar una tarea realmente costosa.

Existen lupas de aumento para la pantalla, teclados especiales, teclados en pantalla, pero esto no es suficiente si una persona tiene una ceguera total. Una de las posibles soluciones que se están desarrollando es el reconocimiento y la síntesis de voz. Esto es, que la persona mediante la voz pueda manejar el ordenador, y a su vez el ordenador pueda comunicarse con la persona emitiendo sonidos inteligibles. A esta interacción se la conoce como comunicación hombre-máquina mediante la voz y son muchas las empresas que se dedican a mejorarla.

Hoy en día se puede ver como Windows Vista ya incorpora en sus sistemas operativos la posibilidad de ejecutar una aplicación de reconocimiento y síntesis de voz con la que se pueden manejar todas las opciones del sistema, desde redactar un documento hasta leer un e-mail y por supuesto ejecutar casi cualquier orden que se ejecute con el teclado.

SINTESIS DE VOZ

La Síntesis de Voz, también conocida como Conversión de Texto a Voz (CTV), consiste en dotar al sistema de la capacidad de convertir un texto dado en voz. Esto se puede hacer mediante grabaciones realizadas anteriormente por personas.



La voz del ordenador puede generarse uniendo las grabaciones que se han hecho, ya sean de palabras enteras, o fonemas, pero siempre intentando que el sonido producido parezca lo mas natural posible e inteligible, encadenando correctamente los sonidos dentro del discurso. El sistema tiene que ser capaz, además de todo esto, de sintetizar cualquier texto aleatorio, no uno establecido por defecto.

Existen dos formas de realizar esta síntesis:

Síntesis Concatenativa

La Síntesis Concatenativa se basa en la unión de segmentos de voz grabados. Este método produce una síntesis más natural, pero se pierde a causa de las variaciones del habla.

Existen tres métodos para realizar la Síntesis Concatenativa. La llamada Síntesis por selección de unidades utiliza una base de datos en la que se encuentran grabaciones de voz tanto de fonemas, silabas, palabras, frases y oraciones. Este método es el que produce un sonido mas natural, pero estas bases de datos pueden alcanzar un tamaño muy grande.

Pero este método no es el único que existe en cuanto a Síntesis Concatenativa se refiere. La síntesis por difonemas utiliza una base de datos minima en la que se ha guardado un unido ejemplo de difonemas (en español existen aproximadamente 800 difonemas distintos), pero este método produce una voz robótica por lo que esta prácticamente en desuso.

Otro método de Síntesis Concatenativa es la síntesis específica para un dominio que une frases y palabras para crear salidas completas. Este método se utiliza para ámbitos muy limitados como por ejemplo, en gasolineras.

Síntesis de formantes

Este método no utiliza muestras de habla humana en tiempo de ejecución como los anteriores, sino que se utiliza un modelo acústico. Se crea una o-nda de habla artificial. Este método produce un sonido robótico y nunca se podría confundir con la voz humana, pero tiene la

ventaja de que producen programas más pequeños ya que no necesitan de una base de datos de muestras grabadas como los métodos de concatenación.

RECONOCIMIENTO DE VOZ

El reconocimiento del Habla permite a un ser humano comunicarse con un ordenador.

A grandes rasgos, consiste en que el ordenador captura la señal de voz que emite una persona a través de un micrófono, convirtiéndola en información digital. El motor de voz debe ser capaz de reconocer las silabas de entre un conjunto de fonemas que ha recibido, y combinarlas para formar las palabras que se habían dicho anteriormente por el usuario.

Existen dos grandes campos dentro del reconocimiento del Habla:

- Reconocimiento Automático del Locutor (RAL)
- Reconocimiento Automático del Habla (RAH)

Reconocimiento Automático del Locutor

Un sistema de reconocimiento automático del locutor permite al sistema comprobar si la persona que ha emitido la señal de voz, es en verdad quien dice ser. Para ello hay realizar un entrenamiento al sistema. El locutor debe introducir muestras de voz para que el sistema pueda crear una serie de patrones. Una vez que se ha hecho esto, se dice que el sistema esta entrenado y esta listo para reconocer al locutor.

Este método se esta utilizando mucho en criminalística para identificar a criminales mediante grabaciones de su voz. Otra aplicación de este campo del reconociendo de voz es la seguridad. Cada persona genera unos parámetros diferentes a la hora de hablar y es complicado que dos usuarios generen los mismos patrones aunque al oído humano parezcan

la misma persona.



Reconocimiento Automático del Habla

El reconocimiento automático del habla presenta varias modalidades producto de una serie de restricciones que se le imponen a la tarea de reconocimiento con el fin de simplificarla. Los principales parámetros que se utilizan para realizar este tipo de reconocimiento son los siguientes: modalidad de habla, estilo de habla, entrenamiento, tamaño del vocabulario, modelo del lenguaje, etc.

Un sistema de Reconocimiento Automático del Habla ideal, es aquel que funciona en entornos con ruido de fondo muy alto, y es capaz de reconocer el habla de cualquier locutor, corrige los errores producidos por la mala pronunciación de éste y además es insensible a las variaciones inducidas por los canales de comunicación, pero aún no se ha conseguido este sistema.

Algunas de las modalidades que presenta el RAH son:

- Reconocimiento de palabras aisladas (RPA): la entrada vocal se realiza palabra a palabra o bien mediante la detección de bordes para determinar el inicio y fin de las mismas y compararlas con una base de datos, creada previo entrenamiento, para elegir la que más se aproxima.
- Detección de palabras Clave (DPC): se detecta la entrada de una palabra en concreto que se encuentra inmersa dentro de un discurso hablado.
- Reconocimiento de palabras conectadas (RPC): se admiten como entradas vocales, secuencias de un conjunto finito de palabras que forman un vocabulario de tamaño moderado o pequeño.
- Reconocimiento automático del habla continua (RAHC): debe ser capaz, evitando ligeros matices, de reconocer habla continua y espontánea como la que se utiliza de manera natural en la vida cotidiana. Esto aún no es posible.



APLICACIONES DENTRO DE LA EDUCACIÓN:

Como ya se ha mencionado antes, son muchas las aplicaciones que se le pueden dar tanto al Reconocimiento como a la Síntesis de Voz, pero nos vamos a centrar en una en especial: la Educación Especial.

El "**Proyecte Fressa 2009. Software para la Educación Especial**", se centra en crear aplicaciones informáticas que facilitan la comunicación de personas con discapacidad motora, visual y auditiva, con el entorno mediante el uso del ordenador. El software que nos encontramos en la página del "
Proyecte Fressa. Software para la Educación Especial
"
http://www.xtec.cat/~jlagares/indexcastella.htm
, es libre, por lo que se permite su copia y su uso para todos los usuarios.

Uno de los programas que se ha desarrollo que realiza un Reconocimiento de Palabras Aisladas es **"Reconocimiento de Fonemas**".

Esta aplicación facilita la discriminación de fonemas en personas con dificultades auditivas o de habla.

Su funcionamiento es básicamente el siguiente:

Se realiza un entrenamiento de la aplicación, con lo que se crean una serie de patrones de los fonemas.

Una vez que el sistema está entrenado, se carga este patrón que se ha creado.

Existen 4 juegos que ayudan a la personas con estos tipos de dificultades a realizar la discriminación de fonemas. Cada uno de los juegos funciona con distintos números de fonemas.

- El juego del "Futbol" funciona con 2 fonemas. El usuario escoge los fonemas que quiere discriminar y la acción que ejecuta. En el juego del futbol solo se permite subir y bajar.
- El juego de los "Platillos Volantes" funciona con 3 fonemas: Izquierda, Derecha y Dispara.
- El juego del "Coche" funciona con 4 fonemas: Izquierda, Derecha, Sube y Baja.
- El juego del "Bonk. Juego del Matamoscas" funciona con 5 fonemas: Izquierda, Derecha, Sube, Baja y Golpe de Martillo.

Toda la información sobre el **"Proyecto Fressa. Software para la Educación Especial**" se puede encontrar en la siguiente pagina:

http://www.xtec.cat/~jlagares/indexcastella.htm