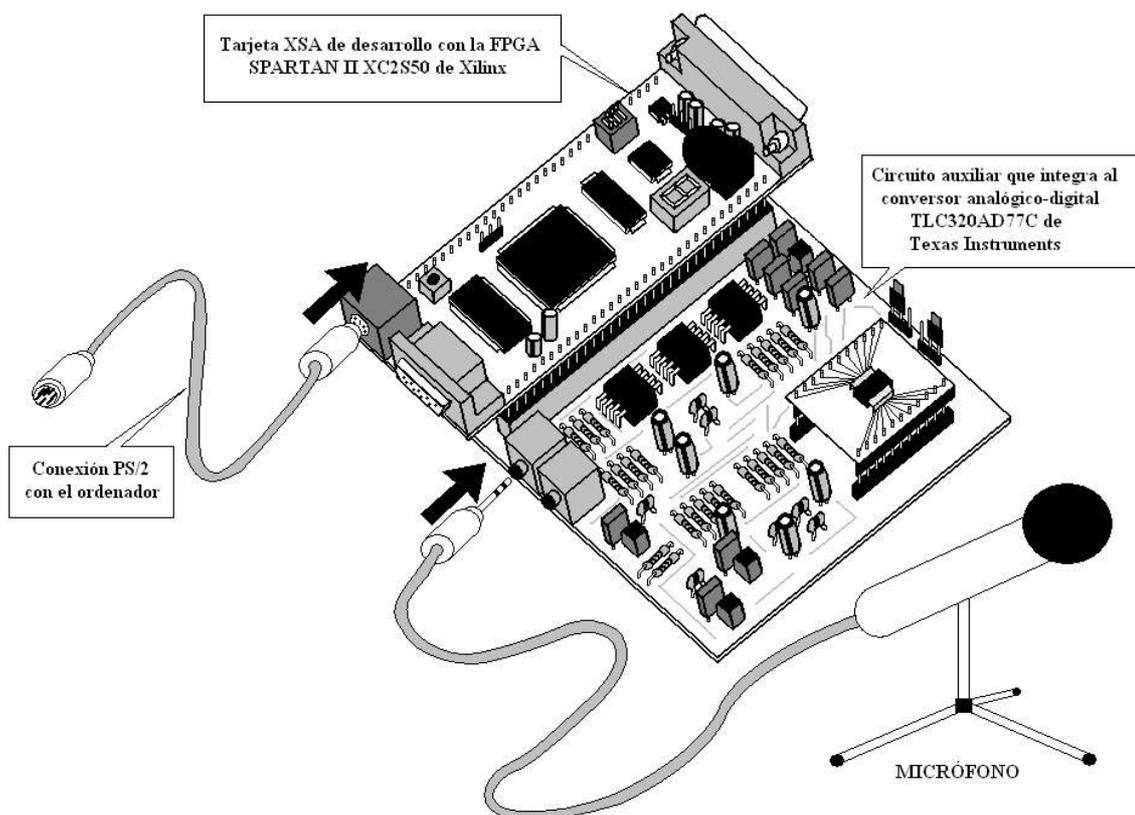
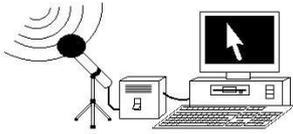


CAPÍTULO 1

INTRODUCCIÓN

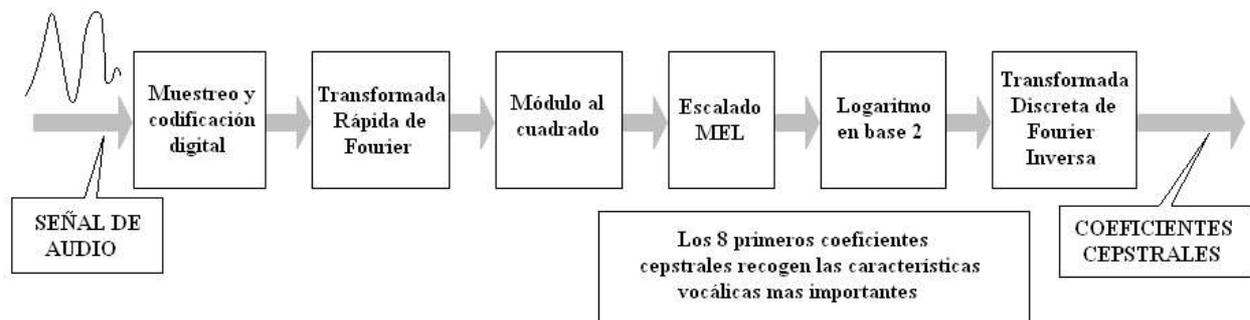
El objetivo del Proyecto Fin de Carrera “Sistema autónomo para accionamiento por voz de un ratón de ordenador” es el diseño e implementación de un ratón de ordenador PS/2 controlado por la voz. Podrá ser utilizado como un ratón normal por personas con incapacidad de movimiento o en situaciones donde el movimiento no sea posible. La electrónica del diseño se implementa sobre la FPGA Spartan II XC2S50 de Xilinx gracias a una tarjeta XSA de desarrollo. El usuario utiliza su voz para controlar los movimientos del dispositivo y el estado de pulsación de los botones del ratón, por lo que es necesario traducir la señal de voz a señales digitales que entienda la electrónica de la FPGA. El convertor analógico-digital TLC320AD77C de Texas Instruments realiza esta operación, para ello, la tarjeta XSA se acopla sobre un circuito que integra el convertor analógico digital.



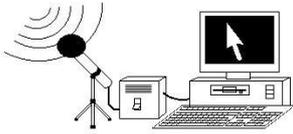


Las señales de voz con las que se manejará el dispositivo serán sólo las 5 vocales ('A', 'E', 'I', 'O' y 'U'). El dispositivo es capaz de reconocer las distintas vocales pronunciadas y de acuerdo con un protocolo que debe ser conocido por el usuario, traducirá la pronunciación de las vocales en eventos de movimiento virtuales equivalentes. Así, el núcleo básico del proyecto es el reconocimiento de las vocales pronunciadas por el usuario. Primero se obtienen unos parámetros que caractericen de forma adecuada a la señal de audio y en este diseño se utilizan los coeficientes cepstrales para este objetivo. La obtención de los coeficientes cepstrales puede resumirse en los siguientes pasos:

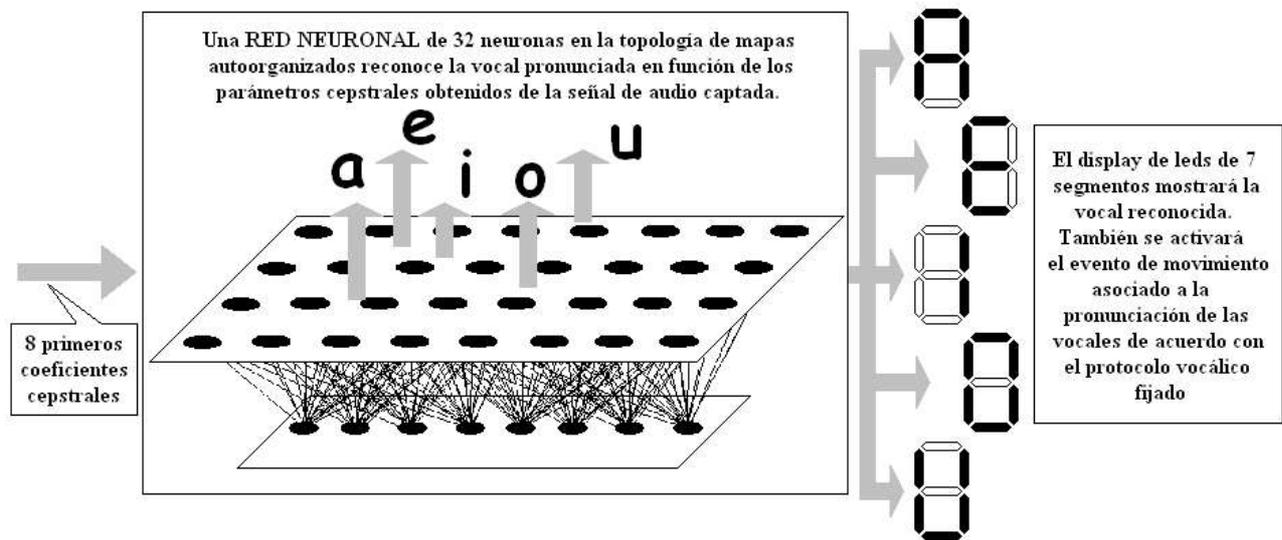
- Cálculo de la Transformada Rápida de Fourier de 512 muestras de la señal de audio procedentes del conversor analógico-digital externo.
- Se resuelve el módulo al cuadrado de los coeficientes espectrales ya que la información de la fase no es relevante en el reconocimiento de las vocales.
- Se realiza un escalado Mel para obtener unos parámetros más eficientes y más cercanos a las señales que crea el caracol del oído humano para informar al cerebro de las frecuencias de los sonidos captados.
- Cálculo del logaritmo en base 2 de los coeficientes mel anteriores para desacoplar por un lado la aportación del tono creado por las cuerdas vocálicas y por otro lado la aportación que el tracto vocálico proporciona al tono de las cuerdas vocálicas (en este miembro reside la información de la vocal pronunciada).
- Se realiza la Transformada Discreta de Fourier Inversa para obtener los coeficientes cepstrales. Los coeficientes cepstrales de las primeras frecuencias resumen la información que el tracto vocálico aporta al tono, por eso se escogen los 8 primeros coeficientes cepstrales como los parámetros más característicos de la señal vocálica. Los coeficientes cepstrales de las frecuencias más altas engloban la información relativa al tono y se desechan como fuente de información vocálica.



El sistema reconocedor detecta cuál ha sido la vocal pronunciada a través de una red neuronal de 32 neuronas en una topología de mapas autoorganizados rectangular de 4 filas y 8 columnas. Al excitar las neuronas con los parámetros cepstrales, se activará la neurona que tenga los pesos sinápticos que más sintonicen con los parámetros cepstrales. Cada neurona tiene asignada una vocal. Cuando una neurona se activa, se propondrá la vocal asignada a esa neurona como la vocal reconocida siempre que la energía de la señal de audio supere el umbral de energía a partir del

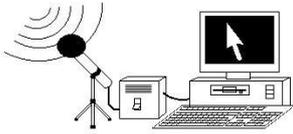


cual se considera que se ha pronunciado una vocal. En esta fase, el display de leds mostrará la vocal reconocida para que el usuario compruebe el correcto funcionamiento del dispositivo.



Los pesos sinápticos de las neuronas se obtienen de forma externa en un ordenador mediante un algoritmo de aprendizaje desarrollado en lenguaje Matlab®. En este algoritmo de aprendizaje la red neuronal es excitada de forma sucesiva con los parámetros cepstrales de las distintas vocales mostradas de forma aleatoria. La red neuronal modifica los pesos sinápticos hasta que en las neuronas se produce una ordenación topológica capaz de discriminar los patrones vocálicos. Llegado este momento se asignará a cada neurona la vocal que más activaciones provoca en la misma. Los información relativa a los pesos sinápticos y la correspondencia entre las neuronas y las vocales se almacena en la memoria FLASH externa AT49F002 de ATMEL presente en la tarjeta XSA. El diseño implementado en la FPGA leerá y almacenará estos parámetros de la red neuronal en un proceso de inicialización.

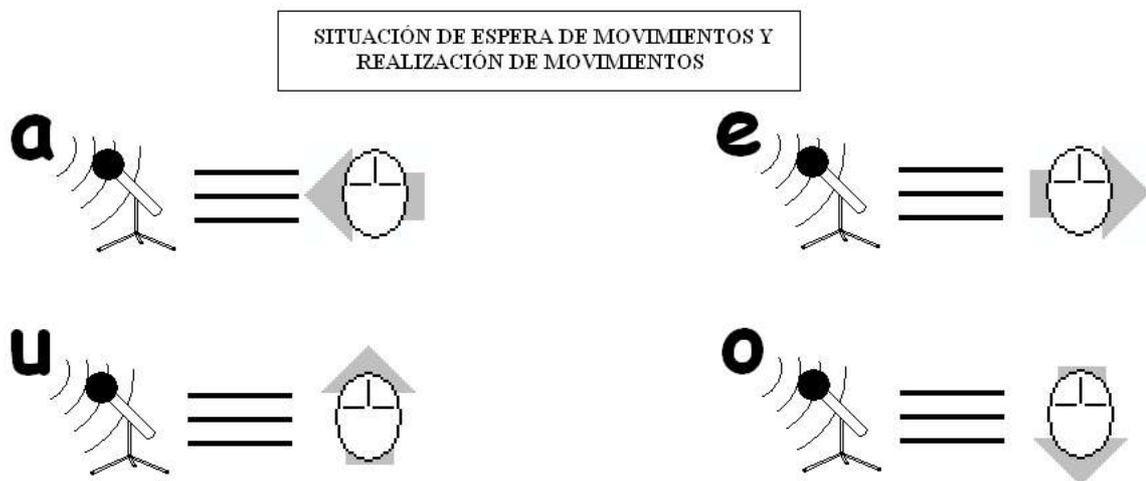
Un ratón de ordenador PS/2 utiliza el protocolo serie PS/2 para transmitir los eventos de movimiento (movimiento hacia arriba, movimiento hacia abajo, movimiento hacia la derecha, movimiento hacia la izquierda, pulsación del botón derecho y pulsación del botón izquierdo). La tarjeta XSA dispone de un puerto PS/2 cuyas líneas de datos y de reloj serán utilizadas para la comunicación bidireccional con el ordenador a través del protocolo serie PS/2. A través de la línea de datos se envían bytes de datos en la comunicación bidireccional con el ordenador. Sucesivamente se envían por la línea de datos y sincronizados con un reloj que el ratón transmite a través de la línea de reloj: un bit de inicio ('0' lógico), los 8 bits del byte enviado empezando por el menos significativo, un bit de paridad impar asociado al byte enviado y un bit de parada ('1' lógico). Mediante el protocolo serie el ordenador puede enviar comandos para provocar cambios en el funcionamiento del dispositivo o para pedir al dispositivo que le informe de determinadas variables internas. El dispositivo utiliza el protocolo serie para responder a los comandos recibidos desde el ordenador o para enviar datos que informen al ordenador de los eventos de movimiento producidos.



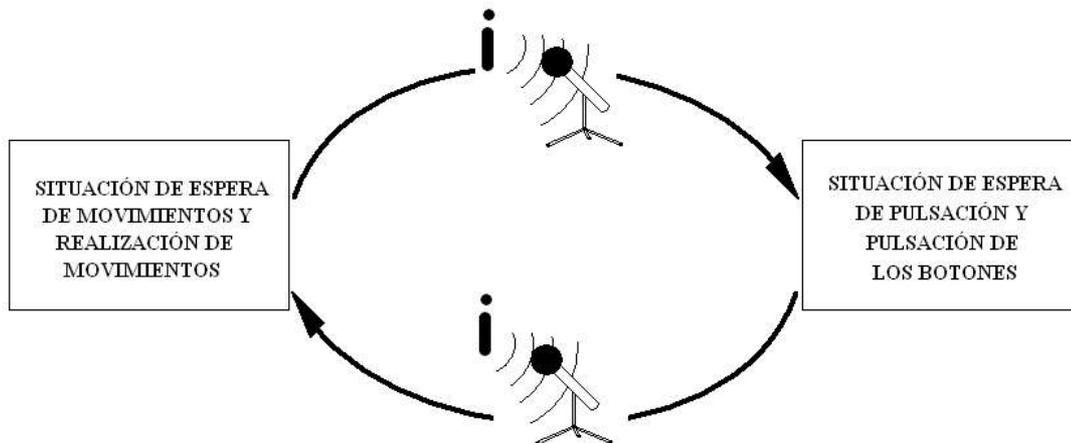
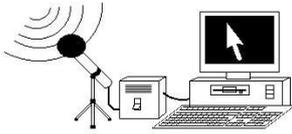
al ordenador. Un micrófono transforma la señal vocálica en una señal eléctrica que el conversor analógico-digital muestrearán en datos digitales a la frecuencia de muestreo propuesta por el emulador de ratón. A partir de las muestras digitales y tras un proceso de reconocimiento, se detecta la vocal pronunciada, donde reside la información de la intención del usuario del dispositivo de acuerdo a un protocolo vocálico concertado para el uso del dispositivo. De acuerdo con el protocolo vocálico, el dispositivo puede encontrarse en dos situaciones diferentes que definen su funcionamiento:

- Situación de espera de movimientos y realización de movimientos: arriba, abajo, izquierda y derecha.
- Situación de espera de pulsación y pulsación de los botones: botón derecho y botón izquierdo.

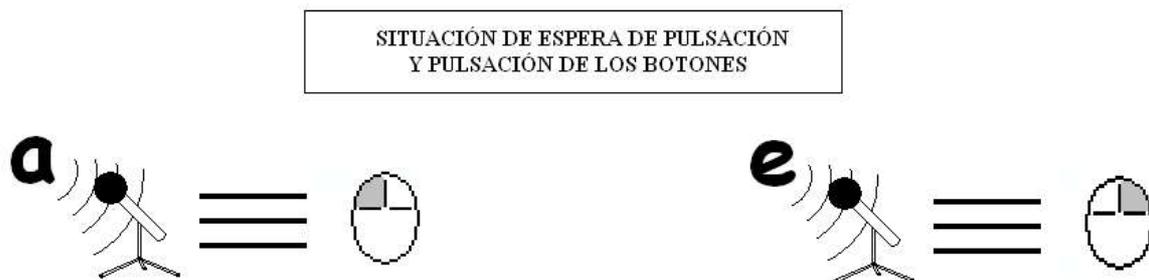
Tras el inicio del dispositivo se encontrará en la primera de las funcionalidades (Situación de espera de movimientos y realización de movimientos). En este momento, la pronunciación de la vocal “U” se traducirá en un movimiento efectivo del ratón hacia “ARRIBA”. La pronunciación de la vocal “O” se traducirá en un movimiento efectivo del ratón hacia “ABAJO”. La pronunciación de la vocal “E” se traducirá en un movimiento efectivo del ratón hacia la “DERECHA” y la pronunciación de la vocal “A” se traducirá en un movimiento efectivo del ratón hacia la “IZQUIERDA”.



Si se pronuncia la vocal “I” se realizará un cambio de funcionalidad. Es decir, si se encuentra en situación de espera de movimientos y realización de movimientos, pasará a la situación de espera de pulsación y pulsación de los botones. Y a la inversa, si se encuentra en situación de espera de pulsación y pulsación de los botones, pasará a la situación de espera de movimientos y realización de movimientos.

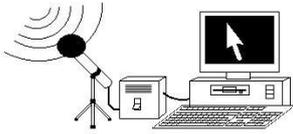


Cuando el dispositivo se encuentra en situación de espera de pulsación y pulsación de los botones, la pronunciación de la vocal “A” se traducirá en la pulsación efectiva del botón de la izquierda (CLICK). La pronunciación de la vocal “E” se traducirá en la pulsación efectiva del botón de la derecha (CLACK).



De esta forma, el usuario comunica al dispositivo sus intenciones de movimiento (movimiento del puntero arriba, movimiento del puntero abajo, movimiento del puntero a la derecha, movimiento del puntero a la izquierda, pulsación del botón derecho y pulsación del botón izquierdo) por medio de la pronunciación de las distintas vocales. Cuando se generen los eventos de movimiento se transmitirán al ordenador a través del protocolo serie PS/2. El diseño en VHDL realizado implementa las funciones anteriores y permite realizar un prototipo operativo del dispositivo. En general, las operaciones del dispositivo pueden distribuirse en tres módulos y de la interacción entre las funciones de cada uno de los módulos se consigue el funcionamiento global del dispositivo.

Un primer módulo es el encargado del proceso de reconocimiento, para lo cual pone en funcionamiento el convertor analógico-digital externo a la FPGA para así recibir como datos las muestras digitales de la señal de voz del usuario del prototipo definido en el proyecto. A partir de las muestras digitales recibidas, el componente realiza la Transformada Rápida de Fourier para extraer posteriormente los coeficientes cepstrales. Los coeficientes cepstrales resumen de forma acertada las características vocálicas de las muestras digitales y en función de estos parámetros se

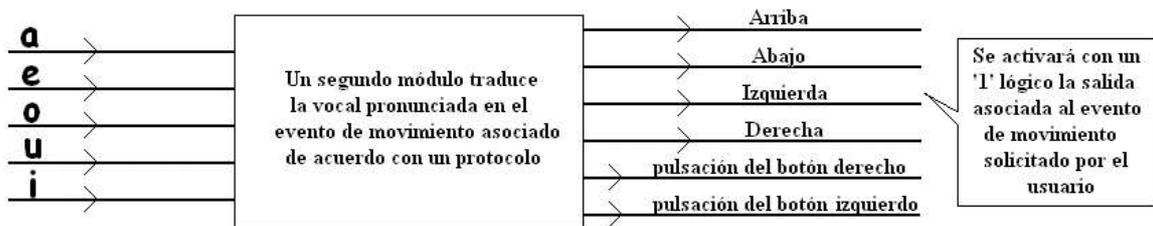


PROYECTO FIN DE CARRERA
Ingeniería de Telecomunicaciones
**SISTEMA AUTÓNOMO PARA ACCIONAMIENTO
POR VOZ DE UN RATÓN DE ORDENADOR**

inicia un proceso de reconocimiento que activa las señales que indican de forma inmediata si se ha pronunciado una vocal y cuál ha sido la vocal pronunciada.



Un segundo módulo traduce la vocal pronunciada a las señales de movimiento que recogen la intención del usuario de movimiento del puntero de pantalla o de pulsación de los botones del ratón. Esta traducción se realiza atendiendo al protocolo vocálico descrito anteriormente. Así se obtienen las señales significativas de los eventos de movimiento que el usuario quiere transmitir al ordenador: movimiento del puntero hacia arriba, movimiento del puntero hacia abajo, movimiento del puntero hacia la derecha, movimiento del puntero hacia la izquierda, pulsación del botón derecho y pulsación del botón izquierdo.



Un último módulo gestiona la comunicación con el ordenador a través del protocolo serie PS/2. El ordenador capta al dispositivo como un ratón normal de ordenador y espera que éste ejecute todo el juego de comandos que un ratón PS/2 es capaz de ejecutar. Por eso, tras la inicialización debe enviar al ordenador una serie de datos que identifiquen al dispositivo, debe ser capaz de responder a los comandos enviados por el ordenador y por supuesto, debe transmitir los eventos de movimiento producidos si el ordenador lo ha solicitado o si ha habilitado al dispositivo para que envíe los datos de movimiento cada vez que se produzcan. Con este módulo se cierra el círculo y se consigue implementar un ratón de ordenador autónomo accionado por la voz.

