ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
CENTRO DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA

"Telemando Electrónico Activado por Voz"

J. Llaque, M. Yapur
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, Campus Gustavo Galindo, Km 30.5 vía Perimetral,
Guayaquil, Ecuador
jllaque@espol.edu.ec, myapur@espol.edu.ec

Resumen

Este proyecto nace de la necesidad de compensar los impedimentos de personas con limitaciones físicas o motoras. La finalidad es obtener un dispositivo que reemplace, de cierta forma, el uso de las manos por la voz para hacer funcionar aparatos electro-mecánicos.

Las palabras serán receptadas por medio de un micrófono, luego de lo cual la señal recibida será digitalizada, almacenada y asociada con una acción específica. En funcionamiento normal, las palabras que ingresen serán comparadas con las palabras guardadas. De acuerdo a esto, el sistema tomará la decisión de acción. Para efectos de simulación, este proyecto utilizará un vehículo a control remoto inalámbrico, el cual será dirigido con las palabras: alto, adelante, izquierda y derecha.

Palabras Claves: *Comandos de Voz, Voz Humana, HM2007.*

Abstract

This project stems from the need to compensate for impairments of persons with physical or motor limitations. The aim is to obtain a device that replaces, in some ways, the use of hands to operate the voice electro-mechanical devices.

The words will be receptive through a microphone, after which, the received signal is digitized, stored and associated with a specific action. In normal operation, the words entered will be compared with the stored words. According to this, the system will make the decision for action. For purposes of simulation, this project will use a wireless remote control vehicle, which will be addressed with the words: high, front, left and right.

Key Words: *Voice Commands, Human Voice, HM2007.*

1. Introducción

El presente documento comienza haciendo una revisión teórica básica de los componentes que se tomaron en cuenta para la elaboración del proyecto. Establecidos los lineamientos teóricos, se describe la forma en que están construidos los diferentes módulos que van desde la captura de la voz, el almacenamiento, procesamiento y comparación y definición de la salida del sistema.

Una vez tomada la salida del sistema, se debe lograr adaptar el sistema de reconocimiento de voz al circuito del control remoto del simulador. En esta etapa se tiene que identificar el funcionamiento de los comandos que activan al simulador.

Finalmente, después de construido el sistema se debe realizar pruebas para verificar el funcionamiento del dispositivo y determinar su confiabilidad.

2. Fundamentos Teóricos

2.1 Principios Básicos De La Digitalización De La Voz

DIGITALIZACIÓN se conoce al paso de información de forma analógica a digital. Se debe de considerar que la CODIFICACIÓN es la conversión de un formato de datos a otro distinto, sin afectar el contenido. Con esto se infiere que, la información resultante es equivalente a la información de origen. Se debe considerar a las palabras emitidas a través de la voz, como los paquetes que contienen la información. Esta información se encuentra en formato analógico. Un sistema de reconocimiento de voz podrá operar identificando palabras aisladas y/o fonemas.

La digitalización, también llamada CONVERSIÓN ANALÓGICA-DIGITAL (ADC), se basa en la medición periódica de la amplitud de una señal. Ésta es, por lo general, captada por medio de un micrófono si se desea digitalizar audio o voz. Luego, se aproximan sus valores a un conjunto de niveles preestablecidos de AMPLITUD, también conocidos como niveles de cuantificación, para luego almacenarlos como palabras de bits en una unidad de almacenamiento. El ADC se compone de forma general en las siguientes etapas:

Muestreo.- Consiste en medir, de manera periódica, muestras de la amplitud de la señal. La velocidad con que se toma esta muestra, es decir, el número de muestras por segundo se la llama "frecuencia de muestreo".
Retención.- Se realiza por un circuito de retención. Es el tiempo que se requiere para evaluar el nivel de la amplitud de las muestras.
Cuantificación.- Proceso que mide el nivel de voltaje o amplitud de cada una de las muestras. Consiste en asignar un valor de una señal analizada a un único nivel de salida. Es a partir de este punto en que a la señal se la puede considerar digital
Codificación.- Consiste en convertir los valores cuantificados en un número codificado. El código más utilizado es el binario, pero también existen otros tipos de códigos que también son utilizados.

2.2 Elementos Básicos Para La Digitalización De La Voz

En la digitalización de una señal, en general se recurre a elementos electrónicos que permiten desarrollar cada una de las etapas del proceso. La clasificación más básica de los componentes es en elementos pasivos y activos.

2.2.1 Elementos Pasivos

Los elementos pasivos son aquellos que no generan ninguna ganancia de voltaje a la señal de entrada. En lugar de esto, consumen voltaje induciendo un decremento a la amplitud de la señal. Dentro de los elementos pasivos están los resistores, capacitores e inductores. Se debe tener en cuenta que el cableado presenta resistencias e inductancias que en sistemas de precisión debe ser cuidadosamente considerada para evitar distorsiones y malfuncionamientos.

2.2.2 Elementos Activos

Los elementos activos son aquellos que pueden generar una ganancia o incrementar la amplitud de una señal. Los primeros trabajos con los tubos al vacío fueron los impulsores de los elementos activos. Entre ellos están los transistores.

Luego, con los conocimientos adquiridos en semiconductores se consigue la miniaturización de los elementos logrando construir los circuitos integrados.

2.3 Almacenamiento Digital Y Procesador De Señales.

2.3.1 Dispositivos De Almacenamiento De Datos

Como su nombre lo indica, los dispositivos de almacenamiento de datos sirven para guardar información. Se puede referenciar los primeros circuitos de almacenamiento digital con los **dispositivos biestables** o Flip-Flop. Los Flip-Flop son dispositivos con la capacidad de permanecer en un estado determinado (Alto o Bajo) durante un tiempo indefinido.
Dependiendo de esta característica se tiene los **dispositivos de almacenamientos volátiles y no volátiles**. Dentro de los dispositivos volátiles también llamados MEMORIAS, existe un amplio espectro de variedades.
SRAM es el acrónimo de Static Random Access Memory. Esta memoria de acceso aleatorio estático tiene la particularidad de mantener la información almacenada sin necesidad de estarla recargando constantemente desde el bus de entrada de datos. Se debe entender como "refrescar" a la acción de estar enviando constantemente la información que debe estar almacenada. Sin embargo, al ser un memoria volátil, si la alimentación eléctrica es interrumpida, la información guardada se pierde. Al mismo tiempo que una SRAM tenga la característica de acceso aleatorio, significa que las ubicaciones en la memoria pueden ser escritas en cualquier orden, sin importar cual fue la última ubicación de la memoria accedida.

La SRAM tiene tres estados de funcionamiento:
- Modo de espera cuando el circuito está inactivo.

- Modo de lectura cuando los datos se solicitan.
- Modo de escritura cuando se actualiza el contenido.

2.3.2 Dispositivos De Comparación De Señales Digitales

La **comparación**, como su nombre lo indica, no es otra cosa que comparar una señal A con otra señal B. De esta operación lógica se puede determinar si dos cadenas de dígitos son iguales o no. También se puede considerar una comparación digital para determinar si una señal es mayor o menor que otra. Principalmente, los comparadores digitales están compuestos de puertas lógicas AND y OR con sus respectivas variantes NAND y NOR.

2.3.3 Microcontroladores

Los microcontroladores son dispositivos que son utilizados para controlar la operación de un sistema. Aunque el aspecto es similar a los clásicos circuitos integrados, se pueden considerar micro-computadoras al disponer todas las etapas de un sistema: entrada, procesamiento y salida en un sistema cerrado. Se lo denomina sistema cerrado debido a que a un microcontrolador solo se requiere adaptar los periféricos de entrada-salida. Por medio de los microcontroladores se puede realizar, de una forma condensada, una gran cantidad de operaciones que solo se lograrían con el uso de varios elementos TTL, como lo son comparar, sumar, etc. al tiempo de definir varios tipos de operaciones “inteligentes” de acuerdo a las variables ingresadas.

Una variable de microcontrolador es el Chip **HM2007**. Este Chip realiza un procedimiento de digitalización de señales analógicas las cuales pueden ser almacenadas en una memoria SRAM. Las palabras almacenadas se relacionan con un número, el cual es usado como apuntador de dirección en la memoria. Cada vez que el chip HM2007 sensa una palabra y la digitaliza, puede comparar el patrón digital con los patrones almacenados en la memoria. En caso de que los patrones coincidan, el integrado coloca en el bus de salida el código BCD respectivo asociado al patrón.

2.4 Dispositivos De Salida

Los dispositivos de salida permiten mostrar el resultado de un proceso realizado. Las formas de presentar la información procesada pueden ser visual, sonora, auditiva e inclusive por medio de una excitación de los sentidos del gusto y olfato. Aunque las dos últimas requieren integración con la química, las tres primeras requieren en la mayoría de casos solo circuitos electrónicos.
Otra forma de mostrar la salida de un proceso, es el envío de la señal resultado a otro sistema. Esto formaría una especie de sistema conectado en cascada, que a su vez lo tomaría como señal de entrada.

2.4.1 Despliegue Visual

Uno de los más tradicionales y utilizados dispositivos de salida son los despliegues visuales. Estos usan leds para mostrar el resultado de una operación que a su vez sirve para que el operador o usuario interprete la información proporcionada. Entre los más antiguos encontramos el despliegue visual de 7 segmentos. Este elemento consta de 7 leds arreglados de forma que se puedan formar los 10 caracteres del sistema decimal, haciendo muy amigable y fácil de interpretar la información a la persona que lo utiliza.

Todos los despliegues utilizan circuitos que traducen la información digital a una forma tal que se pueda mostrar. Estos circuitos son los llamados “drivers”.

2.4.2 Señales Inalámbricas

Si bien una señal inalámbrica no puede ser considerada un dispositivo de salida por sí misma, se puede utilizar como “puente” entre la salida de un sistema y su interconexión a una siguiente etapa.

Las señales inalámbricas dependen de dos circuitos: emisor y receptor. Existen varias variables a considerar en la transmisión inalámbrica: la frecuencia a la que se establece la comunicación, la potencia irradiada, la frecuencia de modulación, entre las más elementales. Es por esto que de acuerdo al escenario y uso la señal inalámbrica, puede alcanzar diferentes distancias y transmitir varios tipos de información.

2.4.3 Tipos De Modulación

Cualquier tipo de información a transmitir por medio de una señal inalámbrica debe ser modulada. En telecomunicaciones, el término modulación encierra las técnicas para transportar información sobre una onda portadora, generalmente una onda sinusoidal. Gracias a esto, se permite aprovechar el enlace y optimizar la cantidad y calidad de información, protegiéndola de posibles interferencias y ruidos.

La modulación se basa en modificar un parámetro de la onda portadora en función de las variaciones de la señal moduladora, que es la información que queremos transmitir. Existen varios tipos de modulación que dependen del parámetro que la señal modifique; por ejemplo, para nombrar unas cuantas variantes tenemos: modulación en doble banda lateral (DSB), modulación de amplitud (AM), modulación de fase (PM), modulación de frecuencia (FM), modulación de amplitud en cuadratura (QAM) y modulación por división ortogonal de frecuencia (OFDM), entre las más utilizadas.

3 Descripción De Los Componentes Del Sistema

3.1 Componentes Del Circuito De Entrada De La Señal

El circuito de entrada se compone de un micrófono y un filtro pasa-altos.

3.1.1. Filtro Pasa Altos

La señal es tomada por medio de un micrófono excitado con VCC. A más de energizar el micrófono, el voltaje genera un nivel offset que permite que la señal ingrese de una forma íntegra al circuito integrado que está alimentado entre VCC y GND. El esquema del filtro es mostrado en la figura 3.1.

La configuración RC conformada establece un filtro pasa-altos que busca eliminar el voltaje offset utilizado. Como se puede observar, durante la frecuencia de corte el micrófono y el resistor R se encuentran en paralelo y a su vez, este resultado Req, está en serie con el capacitor C.
El mismo establece una frecuencia de corte en fc:

$$fc = 1 / (2 \pi ReqC)$$

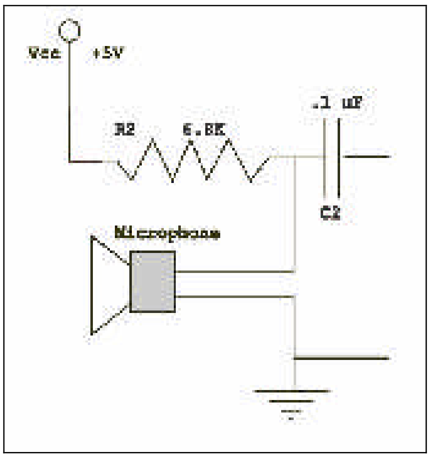

Figura 3.1.- Circuito de entrada

3.1.2 Selección De Comandos Por Teclado Matricial

Las palabras ingresadas a través del micrófono deben relacionarse con diferentes acciones ordenadas al simulador. Para esto se debe identificar las palabras almacenadas por medio de números. La selección de los números se lo establece por medio de un teclado matricial como se muestra en la figura 3.3.
El teclado matricial no es más que un arreglo de interruptores que permiten una combinación de datos binarios. Esta combinación de datos es la que se ingresa al circuito integrado HM2007, el cual interpreta el número y lo relaciona con una palabra digitalizada almacenada en la SRAM.

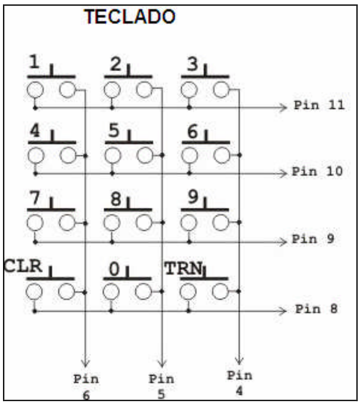

Figura 3.2.- Teclado Matricial

3.2 Etapa De Procesamiento De La Señal

3.2.1 Conversión Analógica-Digital

La conversión analógica-digital la realiza el circuito integrado HM2007. Este elemento toma la señal analógica provista desde el micrófono, realiza un muestreo y la digitaliza a una palabra de 13 bits de largo. Esto proporciona una cantidad de 8192 posibles sonidos reconocidos. Se debe considerar que la frecuencia de reloj a la que trabaja el integrado HM2007 es de 3.67 MHz. Lamentablemente no existe mayor información del fabricante sobre el funcionamiento de este integrado, por el cual no es posible abundar en detalles.

3.2.2 Módulo Digital De Almacenamiento Y Procesamiento

Para realizar el proceso de almacenamiento se debe de ingresar por medio del teclado la secuencia de números a los cuales se va a relacionar la palabra. Los números son mostrados en el despliegue visual de 7 segmentos. Para ordenar al integrado HM2007 que reconozca la voz y almacene en la SRAM se debe, una vez ingresados los números, digitar la tecla TRN. Una vez que la palabra captada mediante el micrófono es digitalizada, esta señal es colocada en un bus de datos que está conectado a una memoria SRAM. Una vez transcurridos 1.96 s. el HM2007 genera la señal que es la que sirve para almacenar la palabra en el SRAM.

A continuación, el pin Ready entra en señal alta por 240 ms, lo que indica que la palabra ha sido almacenada en la SRAM. Las palabras son:

01: Adelante
02: Retrocede
03: Izquierda
04: Derecha
05: Para

Para borrar las palabras asociadas a uno de los números, se debe por medio del teclado ingresar el número y a continuación digitar la tecla CLR.

Desde el integrado HM2007 se dispone de dos buses de datos S y D que son utilizados para la transmisión del patrón de la palabra digitalizada y el número BCD asociado a esta palabra, respectivamente.

Para el proceso de comparación, el ingreso de la palabra de voz es a través del micrófono de una forma similar al proceso de almacenamiento. Una vez que las palabras digitales son colocadas en el bus S, el patrón es comparado con las palabras almacenadas en la SRAM. El resultado de la comparación es mostrado en el bus de datos D, de números decimales en formato BCD. Los resultados posibles pueden ser:
AB: Muestra el número AB relacionado con la palabra AB.

55: Se ha detectado una palabra muy corta
66: Se ha detectado una palabra muy larga
77: La palabra detectada no coincide con las almacenadas.

El número binario es transportado por el bus de datos hacia los circuitos de salida, los cuales se detallan a continuación.

3.3 Dispositivos De Salida

Los números binarios que se obtienen como salida del Integrado HM2007 son tomados por el circuito integrado 74LS373, el cual los convierte en una palabra de dos dígitos decimales en formato binario. Este número pasa por medio del bus D, a los dispositivos de salida del sistema.

Los dispositivos de salida se componen de un despliegue visual de siete segmentos y un convertidor de números BCD a decimal, el cual brindará la señal para manejar el simulador.

3.3.1 Manejo Del Despliegue Visual

La señal es recogida por el integrado 74LS373 y enviada a dos sistemas. El primero de éstos contiene dos manejadores de despliegues de 7 segmentos, 74LS48. Su función es decodificar la señal BCD enviada desde la etapa anterior, a un formato que sea entendible por los despliegues.

La señal decodificada es llevada por medio de resistores de 200 ohmios a los despliegues de cátodo común, como se muestra en la figura 3.3.
De esta manera, el despliegue visual construye el número, brindando al usuario una información fácil de interpretar.

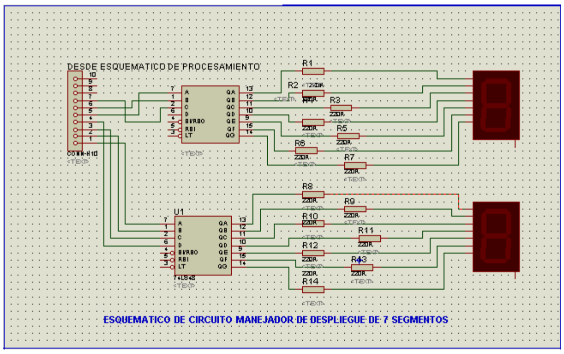

Figura 3.3.- Circuito Controlador de despliegues visuales.

3.3.2 Controlador De Salida A Dispositivo Inalámbrico

La señal BCD que sale del 74LS373, es separada en sus dígitos más y menos significativo. El dígito más significativo ingresa a un arreglo de puertas lógicas OR y NAND. La salida de esta combinación brinda una señal baja al ingresar un número diferente de cero. Por el contrario, si la señal que ingresa es un cero, manda como salida una señal alta.

Este arreglo busca habilitar un segundo integrado 74LS373 al momento que el número más significativo es diferente a 0. El efecto de habilitar o deshabilitar se consigue conectando la salida del arreglo de puertas NAND al pin LE del integrado 74LS373 que toma el número decimal menos significativo. Como salida, este integrado brinda un número BCD que pasa a un decodificador BCD-Decimal. Por medio de este decodificador, se tienen 10 diferentes salidas correspondientes a los 10 primeros números almacenados en la memoria, los cuales fueron ingresados por el teclado.

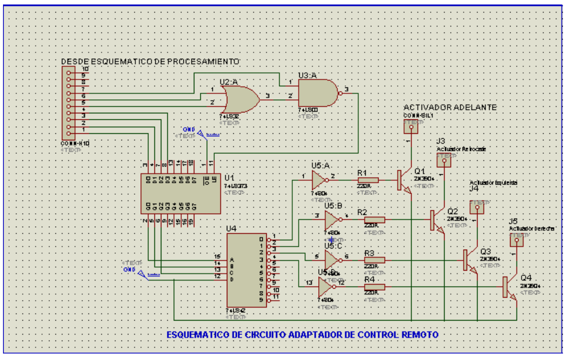

Figura 3.4.- Circuito controlador de salida a dispositivo inalámbrico

Al tener el 74LS42 sus salidas con lógica negativa, se opta por colocar un inversor 74LS04 a cada una de las señales. Las salidas correspondientes a los números 1, 2, 3 y 4 del 74LS04 alimentan a transistores NPN en configuración base común. Estos transistores actúan a manera de interruptores electrónicos para el manejo de los comandos del simulador. Para el diseño de los interruptores, se utilizó transistores 2N3904.

El control remoto del simulador actúa como carga conectada al colector del transistor. Al momento que la base del transistor recibe un voltaje alto del inversor, el transistor entra en estado de saturación. Esto permite que el circuito de activación del mando en el control remoto sea cerrado y se genere la orden deseada.

3.4 Selección De Señales De Salida

Con el sistema utilizado de reconocimiento de voz, se puede almacenar un total de 20 palabras de 1.96 s de duración en la SRAM 8k x 8. De los 20 números que se pueden asociar con los patrones de voz, se seleccionan 5 palabras para el control del simulador. La selección de señales de salida se establece tomando en cuenta que se van a utilizar los 4 primeros dígitos enteros para los comandos. Esto debido a que la activación se produce al generarse un voltaje a la salida del decodificador, que a su vez está conectado a la base del transistor.

Cualquier palabra que esté almacenada, pero que no se encuentre relacionada con los primeros cuatro números decimales, no activará los controles del simulador. En el número 5 se relacionará la palabra de mando para detener el simulador. El criterio usado, de rechazo de palabras ubicadas en otro número, evita la posibilidad que el simulador se active con la activación involuntaria de los números 11, 12, 13 y 14 cuyos números menos significativos son los mismos que activan los transistores,

3.5 Transmisión Inalámbrica

La señal de mando al aparato simulador se la va a realizar por medio de un carro dirigido por control remoto. El control remoto inalámbrico genera una señal a 27 MHz en FM y dispone de un sistema que modula los comandos enviados al vehículo, para que sean trasmitidos por medio de una antena colocada en el mando y en el simulador respectivamente.

4. Controlador De Dispositivo Simulador

4.1 Elemento Electrónico A Controlar

Para este proyecto se presenta un dispositivo simulador que comprende un carro a control remoto inalámbrico. Se debe considerar que la operatividad inicialmente diseñada del control remoto es manual por medio de palancas de mando.
El circuito del control remoto presenta la las señales en estado de "hold" presentes en circuito abierto. La activación de cada una de las señales se produce cuando al manipular las palancas manuales, se realiza contacto entre el Terminal que contiene la señal y tierra. De esta manera se cierra el circuito, tal y como se muestra en la figura 3.1. El circuito del control remoto se encarga de enviar la señal al vehículo.

4.2 Tipo De Señal Inalámbrica A Utilizar

La señal inalámbrica que se utilizará es una señal en frecuencia modulada, con una frecuencia de portadora de 27 MHz. La misma tiene un alcance de 4 metros aproximadamente.

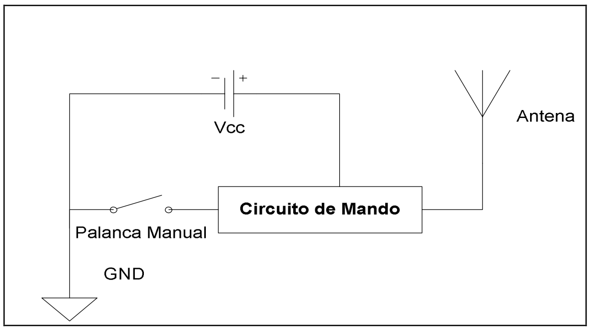

Figura 4.1.- Diagrama de circuito de Control Remoto Manual

4.3 Adaptación De Comandos Manuales A Comandos De Controlador

Una vez que se ha determinado que la activación de las órdenes que se ejecutan por medio de la palanca de mando es a través de un interruptor a tierra, se considera la opción de diseñar un interruptor por medio de transistores. Para esto se ha elaborado la siguiente configuración:

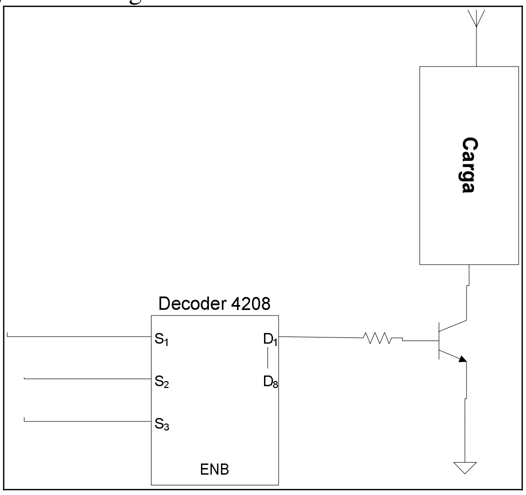

Figura 4.2.- Adaptación de Control Remoto de interruptor manual a electrónico

4.4 Esquemático General Del Proyecto
Una vez que se ha realizado el análisis de cada una de las etapas del proyecto, en la figura 4.3 se muestra el gráfico del esquemático general del proyecto.

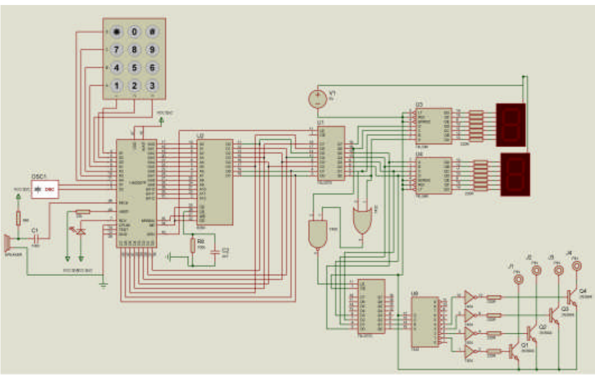

Figura 4.3.- Adaptación de Control Remoto de interruptor manual a electrónico

5.1 Conclusiones

1) El reconocimiento de voz por parte de sistemas electrónicos es una técnica que puede facilitar mucho las tareas de las personas que posean algún tipo de impedimento físico.

2) La mayor cantidad de problemas de funcionamiento que se presentan en los circuitos electrónicos y digitales se debe a conexiones o soldaduras defectuosas.

3) Mientras mayor sea la frecuencia de muestreo, se puede lograr elaborar palabras digitales de mayor precisión.

4) El reconocimiento de la voz se ve limitada y alterada por el ruido ambiental. Debido a la frecuencia de la voz, la elaboración de filtros es compleja, tomando en consideración que para digitalizar una palabra se requiere la mayor cantidad de señal dentro del ancho de banda de la voz.

5.2 Recomendaciones

1) La manipulación de los circuitos integrados debe realizarse tomando en cuenta que éstos son muy susceptibles a cargas electrostáticas. El manejo inadecuado puede provocar daños de los circuitos integrados.

2) Este circuito puede ser mejorado colocando filtros pasa-banda y notch a la entrada de la señal del micrófono para brindar mayor estabilidad a las palabras reconocidas. Al mismo tiempo, con circuitos integrados avanzados, se podría tender al reconocimiento de fonemas en lugar de palabras.

3) Este proyecto debe ser sólo y únicamente utilizado como modelo demostrativo, mas no, como circuito de producción. Esto debido a que el sistema no es completamente confiable. En personas con limitaciones físicas, algún malfuncionamiento del sistema podría generar problemas.

6. Agradecimientos

A Dios por regalarme la vida y salud, a mis padres y hermana por incentivarme y apoyarme desde niño, a mi esposa por su comprensión y a mi hijo por brindarme el empuje final. Al Ing. Miguel Yapur y a Daniel Guzmán por ayudarme desinteresadamente en este proyecto.

Jorge

7. Referencias

[1] Texas Instruments, The TTL Data Book for Design Engineers, Second Edition, 1981
[2] Millman J, Electrónica Integrada, Editorial McGraw Hill, 1991
[3] Boylestad R, Electrónica. Teoria de circuitos, Editorial Pretice Hall, 2000.
[4] Wikipedia, Line Code; http://en.wikipedia.org/wiki/Line_code, 13 de Septiembre del 2009
[5] Zañartu, M, Aplicaciones del análisis acústico en los estudios de la voz humana, http://web.ics.purdue.edu/~mzanartu/Documents/Paper%20-%20Analisis%20voz%20humana%202003.pdf
[6] Gonzáles M, Circuitos digitales; http://www.monografias.com, 15 de Octubre de 2009
[7] © 2007 Images SI, Inc, Build a Speech Recognition Circuit, http://www.imagesco.com/articles/hm2007/SpeechRecognitionTutorial02.html, 10 de Octubre del 2009