

**SISTEMA DE RECONOCIMIENTO DE VOZ HUMANA POR
HARDWARE**

**AUTOR:
JAIME CAMARGO SERRANO**

**DIRECTOR:
REYNALDO CLAROS LAMUS
INGENIERO ELECTRÓNICO**

**UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍA ELECTRONICA
BUCARAMANGA
2010**

*A mi madre querida, Luz Marina;
por estar siempre a mi lado, por creer en mí, por apoyarme y
por darme fuerzas para seguir adelante.
Gracias madrecita linda*

*A mi padre, Jaime; por la paciencia y el apoyo
que me ha brindado siempre.*

*A mi hija, Isabella; por ser esa fuente inspiradora
que me impulsa a luchar y a ser mejor cada día.*

*A mi tía, Lita; por creer en mí y
apoyarme cuando más lo necesite.*

*A toda mi familia por apoyarme y
aconsejarme durante el transcurso de mi carrera.
Gracias*

AGRADECIMIENTOS

Quien elaboró este proyecto agradece a todas las personas que me sirvieron de apoyo, que me colaboraron y fueron de gran ayuda durante el desarrollo de este proyecto; especialmente a mi director de trabajo de grado Reinaldo Claros, por confiar en mi y por su gran apoyo brindado, al ingeniero Alex Monclud por su gran soporte.

Deseo agradecer también a los ingenieros Héctor Andrés Cruz y Sergio Andrés Zabala, cuyos comentarios y recomendaciones fueron de gran importancia para el desarrollo de este documento.

Finalmente, agradecer a mis familiares y amigos que siempre estuvieron ahí, brindándome apoyo constante durante el desarrollo de esta actividad.

TABLA DE CONTENIDO

RESUMEN GENERAL DE TRABAJO DE GRADO	7
INTRODUCCIÓN	8
1. GENERALIDADES	10
1.1 PLANTEAMIENTO DEL PROBLEMA	10
1.2 OBJETIVOS DEL PROYECTO	12
1.3 BREVE JUSTIFICACIÓN	13
1.4 APROXIMACIÓN METODOLÓGICA	13
1.5 CRONOGRAMA DE ACTIVIDADES	15
2. MARCO REFERENCIAL Y ESTADO DEL ARTE	16
2.1 RECONOCIMIENTO AUTOMÁTICO DE VOZ- GENERALIDADES	16
2.2 MARCO HISTÓRICO DEL RECONOCIMIENTO AUTOMÁTICO DE HABLA O VOZ	18
2.3 CONSIDERACIONES GENERALES DE IMPLEMENTACIÓN DE UN SISTEMA DE RECONOCIMIENTO DE VOZ	21
2.4 ELEMENTOS PRINCIPALES DE LA ETAPA DE RECONOCIMIENTO DE VOZ	23
2.4.1 Módulo de adquisición de datos	24
2.4.2 Módulo de recolección y extracción de características de la voz	24
2.4.3 Módulo de cuantificación de los registros de sonido	25
2.4.4 Módulo o etapa de identificación y reconocimiento	25
2.5 SINTESIS DE LAS METODOLOGÍAS DE RECONOCIMIENTO	26
2.5.1 Técnicas de comparación de patrones.	26
2.5.2 Técnicas de reconocimiento por Modelos Ocultos de Markov	28
2.5.3 Técnicas de reconocimiento de voz por Redes Neuronales Artificiales	29
2.6. SINTESIS DEL ESTADO DEL ARTE	35

2.6.1. Aplicaciones En Colombia	39
3. ELECCIÓN E IMPLEMENTACIÓN DEL SISTEMA DE RECONOCIMIENTO	50
• HM2007 (HMC semiconductors)	51
3.1 SISTEMAS RAH POR HARDWARE COMPARADOS	52
3.1.1 HM2007 [30]	52
3.1.2 MSM6679 (OKI semiconductors) [31]	54
3.1.3 RSC-164/300/464 [33]	56
3.2 TABLA COMPARATIVA DE CIRCUITOS INTEGRADOS REVISADOS	59
3.3 IMPLEMENTACIÓN DEL SISTEMA CON HM2007	60
3.2.1 Brazo robótico (periférico final)	62
3.2.2 Tarjeta de Reconocimiento de voz.	63
3.2.3 Tarjeta de Teclado- Configuración de parámetros	67
3.2.4 Tarjeta de Display o visualización	67
3.2.5 Tarjeta Microcontrolada para la manipulación de motores.	69
4. PRUEBAS DE FUNCIONAMIENTO DEL HARDWARE.....	72
4.1 METODOLOGÍA DE CONFIGURACIÓN DEL SISTEMA	73
4.2 PRUEBAS ELÉCTRICAS Y DE FUNCIONAMIENTO	76
4.2.1 Pruebas de funcionamiento con el mismo hablante	77
4.2.2 Pruebas de funcionamiento con diferente hablante	79
CONCLUSIONES	83
RECOMENDACIONES	85
BIBLIOGRAFIA	86
ANEXOS	90
ANEXO A – Circuitos esquemáticos en herramienta Proteus	91
ANEXO B- SINTESIS DE HOJAS DE DATOS DEL FABRICANTE	93

RESUMEN GENERAL DE TRABAJO DE GRADO

TITULO: SISTEMA DE RECONOCIMIENTO DE VOZ HUMANA POR HARDWARE

AUTOR(ES): Jaime Camargo Serrano

FACULTAD: Facultad de Ingeniería Electrónica

DIRECTOR(A): Reynaldo Claros Lamus

RESUMEN

El crecimiento o avance de la tecnología ha generado la utilización de nuevas maneras de interactuar con sistemas robóticos en diferentes campos como la industria automotriz, la medicina, la industria militar, las comunicaciones etc., de ahí que se puedan operar brazos robóticos en la industria automotriz, hasta operar un brazo en la luna; también en medicina para realizar procedimientos tele operados. En el presente trabajo se realiza la descripción de un sistema de Reconocimiento Autónomo de Habla (RAH), el cual permite registrar, validar, identificar y generar acciones a partir de palabras o frases básicas. Específicamente en el caso del proyecto desarrollado, se ha elegido un sistema basado en hardware (circuitos integrados especializados), los cuales permiten desarrollar aplicaciones stand-alone (independientes). En este orden de ideas, se ha realizado una fuerte revisión del estado del arte sobre el tema de Reconocimiento de Voz; destacando tres topologías: Comparación de Patrones, Modelos Ocultos de Markov y Redes Neuronales Artificiales. Los trabajos más destacados a nivel Colombia se centran en la detección de voz para controles robóticos, parametrización del habla para detección de patologías, aplicaciones para el apoyo a personas con discapacidad auditiva, entre otros. Basado en la revisión descrita anteriormente, el proyecto permitió verificar cerca de 5 opciones de dispositivos hardware para cumplir con los requerimientos; donde los criterios de elección fueron bajo consumo de potencia, bajo costo y capacidad para gestionar hasta 20 palabras independientes. En ese sentido, se ha escogido el integrado HM2007, el cual cumple con un funcionamiento bastante acertado para el control del brazo robótico (eje central del apoyo a los discapacitados). Como resultados, observar que es recomendable utilizar pruebas con el mismo hablante, que garantiza una efectividad de funcionamiento superior al 95%. Además, las palabras utilizadas deben tener diferencias fonéticas considerables, para evitar problemas de funcionamiento.

PALABRAS CLAVES: Reconocimiento de voz, HM2007, discapacitados, robótica

V°B°DIRECTOR DE TRABAJO DE GRADO

RESUMEN GENERAL DE TRABAJO DE GRADO

TITULO: HUMAN VOICE RECOGNITION SYSTEM USING HARDWARE

AUTOR(ES): Jaime Camargo Serrano

FACULTAD: Facultad de Ingeniería Electrónica

DIRECTOR(A): Reynaldo Claros Lamus

RESUMEN

The growth or progress of technology has led to the use of new ways to interact with robotic systems in different fields such as automotive, medical, military, communications and so on., So they are able to operate robotic arms industry automobile, to operate an arm on the moon, also in medicine for tele-operated procedures. In the present study is the description of a system of Self Speech Recognition (ASR), which allows you to record, validate, identify and generate action from basic words or phrases. Specifically in the case of the project developed, we have chosen a system based on hardware (specialized integrated circuits), which allows develop applications stand-alone (independent). In this vein, it has made a major revision of the state of the art on the subject of voice recognition, highlighting three topologies: Comparison of Patterns, Hidden Markov Models and Artificial Neural Networks. Work Colombia's foremost focus on the detection of voice for robotic controls, customizing speech pathology detection, applications for support for people with hearing disabilities, among others. Based on the review described above, the project allowed us to verify about 5 choices of hardware devices to meet the requirements, where the selection criteria were low power consumption, low cost and ability to manage up to 20 independent words. In this regard, we have chosen the integrated HM2007, which fulfills a very successful operation to control the robotic arm (central axis of support for the disabled). As a result, note that you should use tests with the same speaker, ensuring operational effectiveness above 95%. Furthermore, the words used must have considerable phonetic differences, to avoid operational problems.

PALABRAS CLAVES:

Speech Recognition, HM2007, robotics, Patterns

V°B°DIRECTOR DE TRABAJO DE GRADO

INTRODUCCIÓN

El crecimiento o avance de la tecnología ha generado la utilización de nuevas maneras de interactuar con sistemas robóticos en diferentes campos como la industria automotriz, la medicina, la industria militar, las comunicaciones etc., de ahí que se puedan operar brazos robóticos en la industria automotriz, hasta operar un brazo en la luna; también en medicina para realizar procedimientos tele operados.

En el presente trabajo se realiza la descripción de un sistema de Reconocimiento Autónomo de Habla (RAH), el cual permite registrar, validar, identificar y generar acciones a partir de palabras o frases básicas. Específicamente en el caso del proyecto descrito en este documento, se ha elegido un sistema basado en hardware (circuitos integrados especializados), los cuales permiten desarrollar aplicaciones stand-alone (independientes).

En este orden de ideas, en el capítulo 1 se describe con detalle las condiciones iniciales del trabajo, donde se destacan generalidades, objetivos, justificación, referencia conceptual, cronograma y metodología para desarrollar el proyecto descrito. En el capítulo 2 se establecen las bases del conocimiento, marco referencial, histórico y estado del arte del tema de Reconocimiento de Habla; sobre todo destacando los proyectos más cercanos a las ideas planteadas en el presente documento.

Como complemento a lo anterior, la comparación entre opciones de circuitos integrados para el reconocimiento, así como las reglas de diseño utilizadas en el trabajo; son presentados en el capítulo 3. En dicho capítulo se ha desglosado cada una de las tarjetas utilizadas, describiendo ventajas, desventajas, entre otros ejemplos.

En el capítulo 4 se presenta pruebas básicas de la eficiencia del comportamiento del sistema implementado para reconocimiento, donde se destaca la verificación

realizada con diferentes conjuntos de palabras, diferentes hablantes y concordancias fonéticas de mayor o menor dificultad.

Finalmente, se presentan recomendaciones y conclusiones sobre el trabajo realizado; donde se realiza énfasis en el cumplimiento de los objetivos del proyecto y en las posibilidades de mejoramiento del mismo.

1. GENERALIDADES

En el presente capítulo se describe, de manera sucinta, las bases conceptuales y metodológicas del proyecto; destacando el planteamiento del problema, la justificación de su ejecución, los objetivos puntuales planteados, la metodología para la obtención de los logros, entre otros.

1.1 PLANTEAMIENTO DEL PROBLEMA

Es evidente que a partir de los avances del siglo XX y los desarrollos de vanguardia del siglo XXI, se ha gestado el crecimiento constante de la tecnología, generalmente en beneficio de la humanidad. Uno de los referentes y predicciones mas destacas en el área del avance tecnológico es la famosa Ley de Moore¹ [1], la cual se sigue revalidando, posicionando a la microelectrónica y la nanoelectrónica como las ciencias de vanguardia.

Ahora bien, con el crecimiento de la tecnología es evidente observar la utilización de nuevas maneras de interactuar con sistemas robóticos en diferentes campos como la industria automotriz, la medicina, la industria militar, las comunicaciones etc., de ahí que se puedan operar brazos robóticos en la industria automotriz, hasta operar un brazo en la luna; también en medicina para realizar procedimientos tele operados.

En este sentido, la robótica se ha convertido en uno de los elementos más destacados como herramienta de apoyo en la relación entre el ser humano y el entorno. Existen innumerables ejemplos de esto: Robots aplicados a la medicina [2], sistemas aplicados a la detección de minas de manera autónoma [3], prototipos para el monitoreo de fallos estructurales, robotica para aplicaciones publicitarias [4], entre otros.

¹ Planteamiento realizado por G. Moore, cofundador de Intel, el cual proyecta que el número de transistores se duplica en un chip alrededor de cada dos años.

En este orden de ideas, y basados en las necesidades propias de Colombia, se ha evidenciado la importancia de conocer e identificar los principales aportes que puede realizar la robótica a la población discapacitada. Según las cifras del DANE² cerca del 8% de los colombianos sufren alguna discapacidad (más de 3'000.000), además los expertos en el tema indican que Colombia no está diseñada ni pensada para los discapacitados [5]. Siendo consecuentes con lo anterior, surge la pregunta de investigación general: *¿Puede la robótica y mecatrónica generar conocimientos y productos que faciliten la vida a los discapacitados?*. La respuesta es SI, y con gran cantidad de ejemplos que se encuentran en el estado del arte.

Aterrizando la investigación para el ejercicio académico del presente trabajo de pregrado, se ha determinado *la revisión de alguna estrategia para el control de brazos robóticos de apoyo a los discapacitados, partiendo de premisas de usabilidad, bajo coste, sencillez de implementación y consumo de potencia aceptable en los equipos hardware*. Las operaciones que tienen que realizar las personas con discapacidades físicas cotidianamente son muy diversas, algunas de ellas son tan complejas que es casi imposible programarlas en los brazos robóticos industriales de la actualidad.

Hoy día es muy difícil llegar a imaginarse poder disponer un brazo domestico capaz de ayudar a una persona severamente discapacitada a efectuar todas las funciones que necesita realizar. Sin embargo, si es posible crear y programar un brazo mecánico que pueda realizar un limitado número de funciones básicas de uso diario, como lo son el apartar y acercar objetos, ayudar a comer o beber, ayudar a la persona con discapacidad en su higiene personal, a pasar hojas de un libro, etc. Existen otras aplicaciones para el control del entorno que ayudan a estas personas a realizar tareas del diario mediante la utilización de la voz, como lo son, levantar y bajar persianas, encender-apagar la radio, encender-apagar la

² Departamento Administrativo Nacional de Estadística.

TV, encender-apagar las luces, etc. Estas operaciones son efectuadas más eficientemente pero no todos tienen acceso por sus altos costos.

En síntesis, este proyecto plantea una solución eficiente empezando por el control de un brazo mecánico ya construido y controlado por voz humana, sin la necesidad de utilizar un PC o software para su manejo. Igualmente se propone diseñar y crear un dispositivo por hardware que permita el eficaz funcionamiento del brazo. Con este proyecto se busca innovar con un producto de interés en la región y cuya revisión ha mostrado que el desarrollo en la región ha sido incipiente. Lo segundo es reducir notablemente los costos que implican el control por voz humana de un brazo mecánico y así proponer una aplicación muy útil, como lo es el que personas con discapacidades físicas (amputadas) y de bajos recursos puedan beneficiarse con esta nueva tecnología.

1.2 OBJETIVOS DEL PROYECTO

Objetivo General

- Diseñar un sistema de reconocimiento y análisis de voz humana por hardware mediante un microprocesador para implementarlo en un brazo mecánico.

Objetivos Específicos

- Realizar un análisis comparativo de los diferentes sistemas integrados de reconocimiento de voz, la fisiología, la acústica, el procesamiento de señal y la inteligencia artificial.
- Analizar el procesamiento y respuesta del dispositivo seleccionado para el reconocimiento en el género masculino y femenino.
- Implementar acciones de movimiento a un brazo electromecánico por medio de la voz.

1.3 BREVE JUSTIFICACIÓN

Basado en la premisa descrita en la sección 1.1, de la definición de estrategias o metodologías que apoyen a la población discapacitada en Colombia y soportado en las cifras oficiales del DANE; se ha decidido generar un proyecto que brinde apoyo directo a la comunidad, desde el punto de vista funcional, de bajo coste; pero también de carácter motivacional. Cabe resaltar que el presente proyecto constituye una pequeña parte de la iniciativa de impacto real, ya que su alcance llega hasta el sistema de reconocimiento de voz y el procesamiento del mismo, realizando pruebas con brazos existentes de varios grados de libertad.

Este proyecto brinda un apoyo muy grande a la comunidad, tanto en la parte económica como en la parte moral, ya que se muestra el desarrollo tecnológico aplicado a la ayuda de personas con limitaciones. Este proyecto brinda además la posibilidad que los estudiantes y profesionales de la Universidad Pontificia Bolivariana, puedan mostrar a la comunidad sus competencia y capacidades para la implementación de una nueva tecnología, como lo es el reconocimiento de voz humana mediante la utilización de hardware; algo muy novedoso en nuestra ciudad. El trabajo cuenta con el acople de diversos conceptos teórico-prácticos vistos en la universidad, sobre áreas complementarias como programación, control, comunicación entre otros.

1.4 APROXIMACIÓN METODOLÓGICA

El ejercicio académico para desarrollar el presente proyecto de grado, centra su desarrollo en cinco pasos metodológicos que serán descritos a continuación, y que permiten el logro de los objetivos propuestas. Dichas etapas son:

- **Realización de marco conceptual y estado del arte:** En este punto se verifica el marco conceptual, referencial y teórico que enmarca los sistemas

de Reconocimiento de Voz y procesamiento de la misma. A su vez, se describe de manera detallada, pero concreta, un estado del arte de la tecnología en esta área; enfocándose principalmente en la situación colombiana sobre este tema.

- **Comparación y selección del dispositivo para reconocimiento de voz:** Luego de la revisión del estado del arte, se construye un listado de las opciones existentes en el mercado, eligiendo una de ellas para su análisis e implementación. Para dicho fin, se determinan criterios de fácil consecución en el mercado local, bajo coste, bajo consumo de potencia y modularidad para su implementación.
- **Implementación del prototipo hardware de reconocimiento de voz:** En esta etapa se implementa el sistema de reconocimiento de voz, realizando el diseño de la tarjeta y la simulación de los resultados en aplicativos software con esta finalidad (Proteus DS™, Eagle Layout Editor, entre otros).
- **Implementación y pruebas del hardware con brazo robótico:** Posterior al diseño e implementación de la tarjeta, se pretende utilizar un brazo robótico ya construido; con el fin de verificar el correcto funcionamiento del sistema de reconocimiento de voz. En esta etapa del desarrollo, el monitoreo del movimiento del brazo, el número de éxitos en el experimento, entre otros; será debidamente registrado, tabulado y analizado.
- **Documentación y conclusión del proceso:** Finalmente, se procede a documentar lo realizado, realizando pruebas de funcionamiento, documentándolas, y generando un artículo de investigación que se pueda proyectar a publicación en evento o revista de carácter nacional.

1.5 CRONOGRAMA DE ACTIVIDADES

ACTIVIDADES GENERALES	SEMANAS																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Realización de marco conceptual y estado del arte	■	■	■	■	■	■																			
2. Comparación y selección del dispositivo para reconocimiento de voz							■	■	■	■	■	■													
3. Implementación del prototipo hardware de reconocimiento de voz													■	■	■	■	■	■	■	■	■	■			
4. Implementación y pruebas del hardware con brazo robótico																							■	■	■
5. Documentación y conclusión del proceso					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

* El desarrollo del proyecto se establece con una duración de 25 semanas, donde cabe resaltar que la etapa de documentación se ejecuta a partir de la finalización de la semana 4.

2. MARCO REFERENCIAL Y ESTADO DEL ARTE

En el presente capítulo, se busca que el lector se familiarice, de la forma más sencilla y concisa, sobre el tema del Reconocimiento Automático de Voz; las principales metodologías utilizadas y el estado del arte a nivel mundial. Cabe resaltar que no se pretende realizar una revisión tan compleja al respecto, sino fundamentalmente reconocer, basado en documentación en línea, textos guías y bases de datos (IEEE, Elsevier, entre otros) los principales trabajos realizados en este tópico.

2.1 RECONOCIMIENTO AUTOMÁTICO DE VOZ- GENERALIDADES

El concepto de Reconocimiento Automático de Voz, o como es mejor conocido en la literatura asociada como Reconocimiento Automático del Habla (RAH), no es más que una parte importante de los sistemas de Inteligencia Artificial, cuyo objetivo central es permitir la comunicación entre dos entes: el ser humano y los sistemas electrónicos [6]. Básicamente se cuenta con un problema que involucra, no solo el registro de la información, sino la búsqueda de concordancia entre parámetros fonéticos, léxicos, sintácticos, acústicos, entre otros; donde las condiciones del ambiente, generalmente son difíciles de controlar generando amplias incertidumbres al proceso.

Entre las definiciones o conceptos más relevantes de los sistemas RAH, se tienen los siguientes:

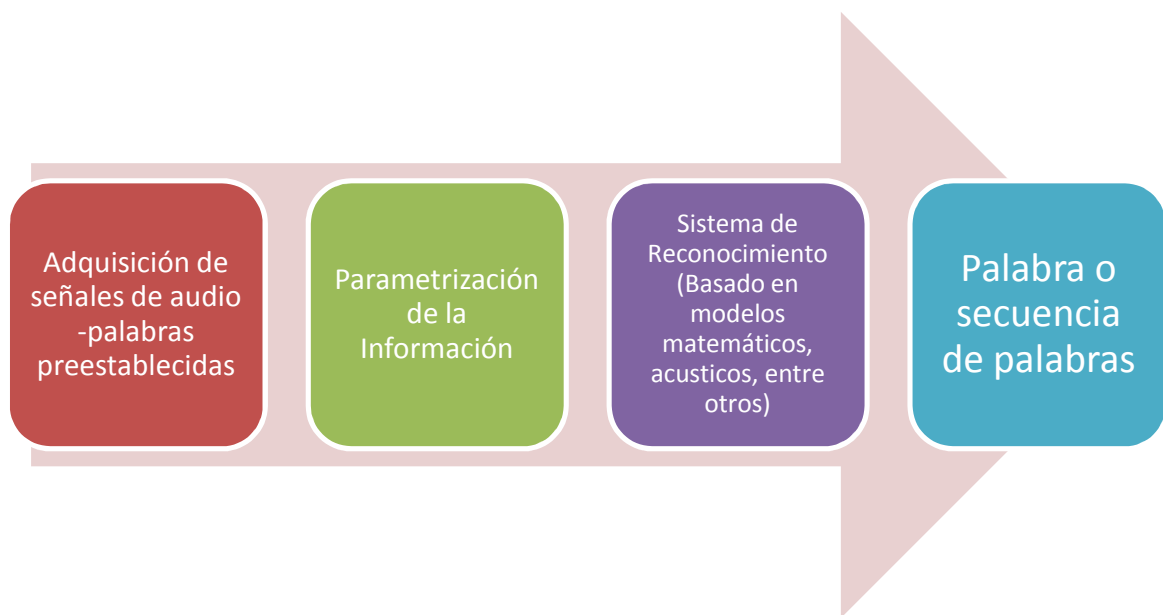
- **Reconocimiento de voz:** Metodología de conversión de palabras habladas a escritas; es decir una conversión de sonido a texto digitalizado.

- **Sectores de aplicación:** Entre los sectores más destacados, como ya se profundizará, se tiene la domótica, la robótica, la implementación de sistemas de inteligencia artificial, entre otros.
- **Posibilidad de reconocimiento:** El sistema de reconocimiento puede identificar desde un set limitado de palabras o vocablos, comparándolos con unos patrones predefinidos; hasta un conjunto de palabras completo basado en metodologías de inteligencia artificial (la intención es lograr identificar frases completas y en casos más complejos generar sentido y significado a la misma).
- **Posibilidad de operación:** Los modelos de operación básicos del sistema RAH se fundamentan en la medición de **palabras aisladas** o, en un caso más complejo un sistema de reconocimiento de **fonemas** (ejemplo a desarrollar en el presente proyecto).

Ahora bien, cabe resaltar que el proceso de reconocimiento y de interfaz entre el hombre y la máquina se puede realizar con la integración de sistemas hardware de adquisición con herramientas computacionales de diversas complejidades. En este punto se hace evidente que el problema no es solo de registro (transducción de señal), sino que se complementa con un proceso complejo de parametrización de los datos para luego permitir la identificación o autenticación del comando enviado [7]. Es fácil observar un símil al revisar este proceso con el tratamiento clásico de imágenes para reconocimiento de patrones, el cual ha sido ampliamente utilizado en sistemas de identificación por huella, características faciales, biometría de la mano, entre otros.

Adicionalmente a lo descrito, se invita a revisar la imagen presentada a continuación (Figura 1.), la cual es una síntesis en diagrama de bloques de un proceso de Reconocimiento Automático del Habla (RAH).

Figura 1. Esquema ilustrativo del RAH



Fuente. Autor del Proyecto.

De la anterior imagen se pueden observar cuatro pasos o etapas claves de los sistemas RAH, las cuales serán descritos en la siguiente sección con el detalle necesario.

2.2 MARCO HISTÓRICO DEL RECONOCIMIENTO AUTOMÁTICO DE HABLA O VOZ

Para el desarrollo de esta sección, se ha realizado una revisión detallada de los principales hitos de la historia del reconocimiento y transmisión de voz; sin embargo la revisión se ha concentrado en los elementos más destacados **[10]**:

- 1870 → Construcción de un dispositivo que permitiera facilitar la comunicación entre personas (**A. Graham Bell**). El resultado de estos trabajos fue el teléfono.
- 1880 → Se bosqueja un sistema de transcripción automática de información, que permita identificar secuencias de sonidos y genera, de alguna forma, texto. No fue posible realizarlo (**T. Nemes**).
- 1910 → Se construye la primera máquina, que utiliza como base un conjunto de plantillas, las cuales logran ser capaces de reconocer varios dígitos.
- 1960-1970 → Se implementan sistemas que incorporan criterios selectivos, donde se destaca la dependencia de flujos discretos de palabras o de vocabularios reducidos. Se mantienen procesos demasiados lentos para ser aplicados en dictados y conversaciones reales.
- 1970-1980 → Entre los elementos a destacar, se produce el primer sistema de reconocimiento de voz lanzado de manera comercial, denominada VIP100 de TT Inc. Primeros acercamientos destacados en el tema de la inteligencia artificial y su aplicación a los sistemas de reconocimiento. En esta década aparecen también los sistemas de análisis léxico, semántico, sintáctico y análisis.
- 1980-2000 → Surgen los sistemas con capacidad de procesar más de 1000 palabras, además se observa una fuerte reducción de los precios de los sistemas. Entre las empresas más destacables que han desarrollado

sobre esa aplicación son Philips, Sensory Circuits, Speechworks, Dialogic, Novell, Siemens, NEC e Intel.

Adicional a esto, es importante destacar los avances constantes en el nivel de integración de los circuitos, generando sistemas de menor tamaño, menor consumo de potencia y mayor eficiencia. Otro punto clave en la cronología de los sistemas de reconocimiento son los avances de las redes neuronales artificiales, los cuales se destacan de manera bastante puntual a continuación [11]:

- 1949 → Hebb generaliza los conceptos fundamentales de las Redes Neuronales Artificiales, realizando un símil entre las características fisiológicas y las computacionales. Describe de manera sucinta como el proceso de sinapsis³ permite el aprendizaje, además se planteó un conjunto de reglas que sigue siendo la base de los modelos que se utilizan en la actualidad.
- 1960 → Se plantean las teorías sobre la adaptación neuronal. El autor más destacable en estos temas fue C. Widrow. En este año se realizó la primera aplicación de interés para el presente proyecto, un filtro adaptativo para eliminar ecos en los procesos de comunicación telefónica (problema real, el concepto fue más allá de la teoría).
- 1962→ Se desarrolla un elemento de identificación de patrones ópticos con aplicaciones binarias. El término utilizado para definir este sistema fue el Perceptrón, desarrollado por Roseblatt. Este sistema tiene una naturaleza meramente lineal, lo que generó muchas limitaciones y críticas para su aplicación en problemas pertinente.

³ Proceso de comunicación entre neuronas- fundamento del aprendizaje de las redes establecidas

- 1982 → Se plantea un modelo de red neuronal por parte de Hopfield, el cual se encuentra constituido de unidades de proceso interconectadas que alcanzaron mínimos energéticos. Este punto tiene una fuerte vinculación con la estabilidad del proceso.
- A finales de la década de los ochentas del siglo pasado e inicio de los noventa, se plantearon sistemas adaptativos (Hinton & Sejnowski), Grossberg desarrolló modelos para justificar o sustentar la estabilidad de redes neuronales y se implementaron nuevas metodología provenientes de conceptos biológicos (M. Kohonen).
- En los últimos años se ha desarrollado sistemas de reconocimiento de patrones, reconocimiento de voz y reconocimiento de video. También se han creado sistemas de predicción económicos (Bolsa de Valores), apoyo a la medicina, gestión de grandes datos, entre otros.

2.3 CONSIDERACIONES GENERALES DE IMPLEMENTACIÓN DE UN SISTEMA DE RECONOCIMIENTO DE VOZ

Al momento de hablar de un sistema de identificación de personas a través del reconocimiento de la voz, existen varias condiciones que se deben establecer para el diseño de la misma. La primera de estas es la determinación, prudente y concienzuda, del tipo de sistema a utilizar: dependiente del hablante o independiente del hablante. Por ejemplo, en el caso de las personas discapacitadas [8] o en el control de un sistema de tele-operación para un astronauta, el desarrollo dependiente del hablante es útil y significativo (caso con pocos usuarios).

En contraposición a lo anterior, un sistema automático de control de voz que no sea dependiente del hablante es más utilizado para monitorear y controlar sistemas de respuesta automática, por ejemplo el que utilizan las audio-lineas de los bancos

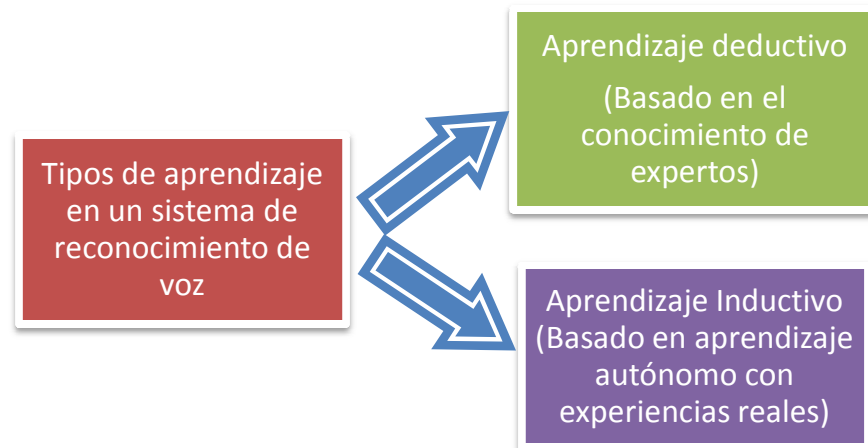
Otro punto importante para destacar, en el momento de implementar un RAH es la determinación de la metodología de aprendizaje del sistema. Como se presenta en la figura 1, el principal logro del sistema es el reconocimiento de palabras, pero a su vez se asocia al reconocimiento de personas; por lo cual se debe crear una base de conocimiento⁴. Para este objetivo, se cuenta con dos tipos de estrategias claramente diferenciadas: Metodología de aprendizaje deductivo y la Metodología de aprendizaje inductivo. La primera fundamenta su operación e importancia en la transferencia de conocimiento de un humano, considerado experto en el tema, al sistema de cómputo (Típico caso de los famosos Sistemas Expertos⁵); mientras que la segunda describe los sistemas que pueden modelar un proceso a partir de experiencias y ejemplos reales sobre la actividad o proceso a realizar, como se puede evidenciar en la figura 2 (Ejemplo clásico de las Redes Neuronales Artificiales y los modelos ocultos de Markov).

Finalmente, los modelos utilizados en el presente trabajo, como en casi todos los procesos reales de implementación de un RAH, combinan los dos tipos de aprendizaje; ya que existe un compromiso o relación entre el proceso deductivo y el proceso inductivo, es decir, prácticamente no es posible contar con un sistema puramente inductivo.

⁴Entiéndase base de conocimiento como un tipo especial de base de datos que permite gestionar, recolectar, organizar y extraer información de forma automatizada y computarizada.

⁵ “Los sistemas expertos son programas que reproducen el proceso intelectual de un experto humano en un campo particular, pudiendo mejorar su productividad, ahorrar tiempo y dinero, conservar sus valiosos conocimientos y difundirlos más fácilmente” [9]

Figura 2: Síntesis de las metodologías de aprendizaje de un RAH



Fuente. Autor del proyecto

En la siguiente sección, se detallan los elementos principales del sistema de reconocimiento; los cuales tienen unas bases teóricas definidas pero una gran variedad de metodologías y formas de implementación.

2.4 ELEMENTOS PRINCIPALES DE LA ETAPA DE RECONOCIMIENTO DE VOZ

Luego de la descripción de las metodologías de aprendizaje, es necesario que el lector del presente documento identifique de manera clara las metodologías de procesamiento de los datos y la obtención de la secuencia de palabras asociadas. En este sentido, la decisión de analizar de primera cuenta la etapa final del proceso se encuentra justificada porque en dicho punto se encuentra el procesamiento e identificación de señales más interesante; posteriormente se revisaran, ya en la descripción del hardware, las metodologías de adquisición de información y registro de la misma.

2.4.1 Módulo de adquisición de datos

Este módulo juega un papel fundamental para el registro de la señal. Básicamente se encuentra asociado a los transductores para la medición de audio, así como las etapas de pre-procesamiento de señal.

En los sistemas de digitalización se implementa un conversor de datos Análogo/Digital, el cual debe contar con parámetros adecuados para la correcta conversión, cumpliendo con criterios de Resolución, Precisión, Rango, entre otros. Aunque es posible utilizar un conversor embebido al sistema microcontrolado, podría llegar a ser necesario el uso de un A/D externo, que cuente con mayor resolución para la edición de la información.

Vinculado al tema del presente proyecto, es importante pensar que los elementos elegidos deben contar con una capacidad de procesamiento adecuada, ya que el sistema no se encuentra asociado a un equipo de computo; entonces el desarrollo y procesamiento se debe realizar en la tarjeta desarrollada (reconocimiento por hardware) [11].

2.4.2 Módulo de recolección y extracción de características de la voz

De manera similar al tratamiento de imágenes, es necesario establecer metodologías para la extracción de propiedades o características de la señal. Entre los parámetros más destacados para la descripción de señal se cuenta con los siguientes:

- Propiedades espectrales – Energía espectral.
- Tono y timbre.
- Componentes vectoriales formantes.
- Valores temporales- Inicio y Finalización de la señal de audio; amplitudes, entre otros.

Cada uno de los anteriores parámetros deben ser extraídos por cada segmentos pronunciado. Ahora bien, entre las técnicas más destacables para análisis espectral se tienen desde algoritmos clásicos como la FFT (Transformada Rápida de Fourier), hasta modelos regresivos y autoregresivos, Modelos de Predicción Línea, entre otros.

2.4.3 Módulo de cuantificación de los registros de sonido

El proceso se fundamenta en la vectorización del audio, donde dicha señal se convierte en una sucesión de valores, donde cada uno representa un vector de propiedades. Es importante que el lector del presente documento comprenda que el vector de propiedades describe, de la forma más detallada posible, el sonido o fonema registrado; sin embargo es un proceso de indexación⁶ y no la señal en el dominio temporal.

2.4.4 Módulo o etapa de identificación y reconocimiento

Esta fase del proceso de extracción de información genera como resultado la comparación contra una base de información existente. En este sentido, se puede identificar tres resultados posibles:

Conocida: Pronunciación o vocablo plenamente identificado con una incertidumbre bastante reducida.

Similar a la conocida: Cuando el vocablo cuenta con una incertidumbre moderada. El sistema tratará de generar un resultado lo más cercano a la base conocida.

Desconocido: Es la tercera opción, donde el sistema de reconocimiento no es capaz de encontrar concordancia con la base de datos preexistente.

⁶ *Indexar* hace referencia al proceso de registrar de manera ordenada y eficiente información para elaborar un índice.

Se proyecta en la siguiente sección una breve descripción de las principales metodologías para el reconocimiento de patrones.

2.5 SINTESIS DE LAS METODOLOGÍAS DE RECONOCIMIENTO

En las siguientes líneas se presentan las principales estrategias en la fase de reconocimiento, entre las que se encuentran:

- Técnicas de comparación de patrones.
- Técnicas de reconocimiento por Modelos Ocultos de Markov⁷.
- Técnicas de identificación por Redes Neuronales Artificiales.

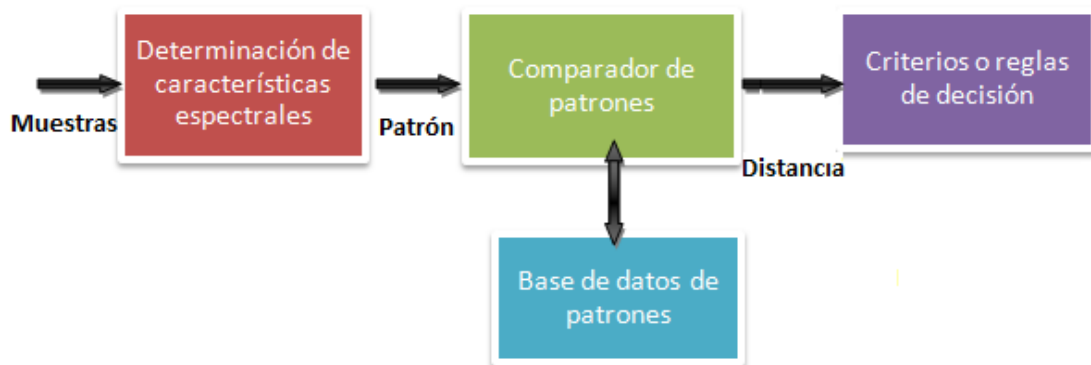
2.5.1 Técnicas de comparación de patrones.

Esta técnica consiste en la comparación directa entre las características espectrales de la señal y las almacenadas en la base de datos. En la figura 3 es posible evidenciar una etapa de muestro (entrada de señal), una etapa de cálculo de propiedades espectrales, para posteriormente comparar con una base de datos de patrones preestablecido. Finalmente, se utilizan criterios o reglas de decisión para llegar a un resultado consecuente y eficiente.

Entre las ventajas más evidentes de esta estrategia se encuentra en que no es estrictamente necesario determinar todas las características espectrales de la voz a un nivel fonético, lo que definitivamente reduce la implementación de secciones complejas, etc. En otras palabras, se compara el vector extraído de la señal adquirida con datos preexistentes en una base de datos.

⁷ Nombre planteado en honor al matemático ruso Andrei Andreevitch Markov (1856-1922).

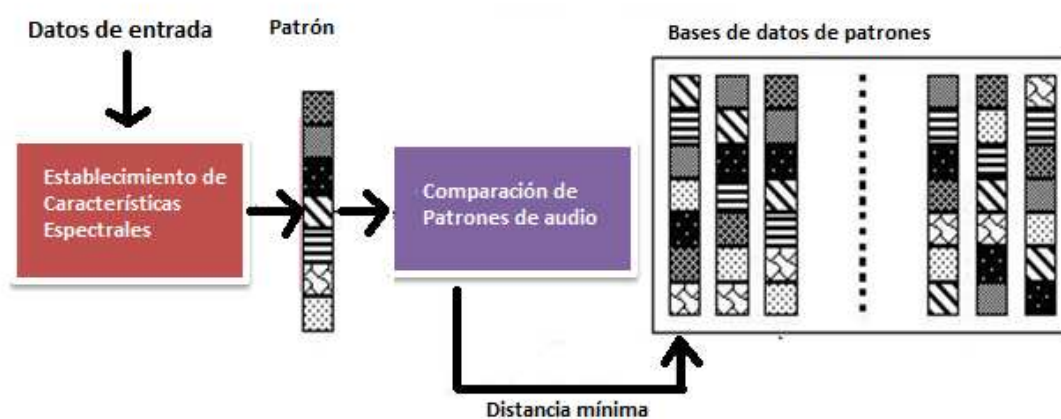
Figura 3: Comparación de patrones



Fuente. Autor del proyecto

En la figura 4 se muestra como se realiza la comparación de patrones, basado en características espectrales. Importante notar que si el patrón no se encuentra en la base de datos, el sistema toma la información cuya diferencia sea mínima al dato de entrada.

Figura 4: Comparación de patrones



Fuente. Autor del proyecto⁸

⁸ Basado en el documento referencial [11]

Para finalizar los comentarios sobre este tipo de tratamiento, la aplicación del mismo debe ocurrir cuando:

- El número de palabras a trabajar es reducido.
- Los patrones se deben construir basados en un conjunto de sonidos o fonemas cortos, los cuales actuaran como referencia de clasificación.
- Se recomienda registrar o grabar los sonidos a la base de datos, calculando sus características espectrales. La literatura asociada y revisada por el autor del presente documento recomienda el uso de parámetros LPC⁹.
- Es evidente que este tipo de trabajo tiene problemas de normalización de las señales, generando un error considerable si no se restringen las condiciones al registrar las señales.

Este tipo de sistemas es interesante, siempre y cuando el set de palabras sea reducido; si no es así, se recomienda revisar las opciones que aparecen a continuación.

2.5.2 Técnicas de reconocimiento por Modelos Ocultos de Markov

Esta estrategia se basa en el uso de elementos estocásticos, a diferencia del caso anterior [12]. Los siguientes puntos son los más relevantes en la presente metodología:

⁹ LPC- Estimación Espectral basado en Predicción Lineal. Excelente para la parametrización de información en un set de patrones reducido y facilita la reconstrucción adecuada de los fonemas y las palabras utilizadas.

- Se utilizan sistemas autónomos de estados finitos. Estos “modelos” son la herramienta fundamental para construir la base de las pronunciaciones, dada que las señales cuentan con una variabilidad bastante alta. El registro de estas señales contienen alta cantidad de ruido ambiental que debe ser modelado.
- Los modelos de Markov se han venido trabajando desde la década de los cincuentas (siglo XX), y consta de un proceso estocástico con un modelo (visible) y un modulo (oculto); a lo cual debe su nombre. En este punto cabe la pena indicar que aparece el algoritmo de identificación conocido como de Máxima Estimación (ME).
- En la actualidad, estos modelos han evolucionado a representaciones de Baum-Welch y Viterbi. Estos algoritmos están mejorados y adecuados al procesamiento de señales de audio.

2.5.3 Técnicas de reconocimiento de voz por Redes Neuronales Artificiales

A lo largo de los años las Redes Neuronales Artificiales (RNA) han recibido un interés particular en las aplicaciones tecnológicas, ya que ofrece los medios para modelar efectivamente problemas complejos. Estas redes son capaces de encontrar patrones de manera efectiva por medio de una serie de algoritmos y basados en datos existentes. La unidad principal de la RNA es un procesador llamado neurona, que calcula sus entradas y se activa, para enviar una señal a la próxima neurona y así sucesivamente hasta llegar a una salida, se basan en el comportamiento y funcionamiento del cerebro humano, especialmente del sistema nervioso.

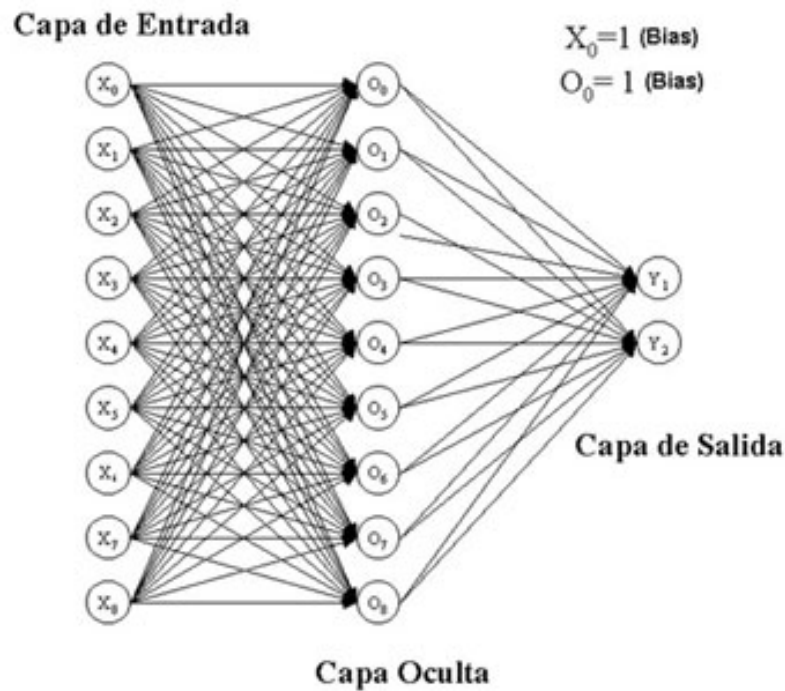
Es importante resaltar que la propiedad más importante de las RNA es su capacidad de aprender a partir de unos patrones generados en un proceso de aprendizaje, que bien puede ser un entrenamiento supervisado o no supervisado, de acuerdo a las necesidades.

- **Entrenamiento Supervisado:** Se necesita un conjunto de datos de entrada clasificado previamente o cuya respuesta u objetivo ya es conocido. Un ejemplo de este tipo de redes es la memoria asociativa bidireccional.
- **Entrenamiento No Supervisado:** No necesitan de conocer la respuesta previamente, son autos organizados. Un ejemplo de este tipo de redes son las redes de aprendizaje competitivo.

Las Redes Neuronales son modelos que intentan simular el comportamiento neuronal biológico a partir de modelos matemáticos recreados artificialmente, para conseguir como objetivo que las maquinas den respuestas similares a las obtenidas por el cerebro humano.

Está compuesta por unidades llamadas Neuronas Artificial, dentro de cualquier sistema se encuentran tres tipos de unidades, Entrada, salida y ocultas. Las de entrada son aquellas que reciben las señales, las de salida se encargan de enviar la información fuera de la red y las ocultas son las que sus entradas y salidas se encuentran dentro del sistema.

Figura 5: Unidades de Proceso de una RNA



Fuente: Tomado de [22]

Las unidades de salida se caracterizan por tres funciones:

1. **Función de Propagación o excitación:** Que consiste en la sumatoria de cada entrada multiplicada por el peso de su interconexión. Si el peso es positivo, la conexión se denomina excitatoria; si es negativo, se denomina inhibitoria.
2. **Función de Activación:** La cual modifica la función de propagación, y puede o no existir, si no existe la salida es la misma función anterior.
3. **Función de Transferencia:** Es utilizada para limitar la salida de la neurona, y viene generalmente dada por la interpretación que se le quiera dar a la misma.

Entre las principales características o aspectos que describen una RNA, se tiene:

- Topología: es la organización y disposición de las neuronas en la red, formando capas o grupos de neuronas, se tiene en cuenta el numero de neuronas por capa, numero de capas, grado de conectividad y tipo de conexión entre neuronas.

Los tipos de RNA pueden ser Red Monocapa, en los cuales las conexiones de las neuronas son laterales formando una sola capa dentro de la red, y las Redes Multicapa en las que las neuronas son agrupadas por niveles y pueden tener dos tipos de conexiones: feedforward (la conexión es hacia adelante) y feedback (la conexión es hacia atrás).

- Mecanismo de aprendizaje: El aprendizaje es el proceso por el cual las neuronas van modificando sus pesos de acuerdo a una información de entrada. Estos cambios suceden por la modificación, creación o destrucción de las conexiones entre neuronas.

Los pesos se modifican de acuerdo a los criterios de aprendizaje, el cual puede ser supervisado o no supervisado; la diferencia entre estos es la existencia o no de un agente que supervise el proceso.

- Aprendizaje Supervisado: Puede llevarse a cabo de 3 maneras diferentes:
 1. Por corrección de error: En la cual se ajustan los pesos de las conexiones en la red, en función de la diferencia entre los valores deseados y los obtenidos.

2. Por refuerzo: Donde no se indica la salida exacta que se desea, solo se indica mediante una señal si la salida obtenida se ajusta a la deseada y basado en esto se ajustan los pesos por medio de probabilidades.
 3. Estocástico: Consiste en realizar cambios aleatorios a los pesos de las neuronas y evaluar su efecto en la salida basándose en el objetivo deseado.
- Aprendizaje no Supervisado: Puede llevarse a cabo de dos maneras diferentes: Hebbiano: Consiste en extraer características de los datos en entrada, para ajustar los pesos de las conexiones de acuerdo a la relación de las salidas de dos neuronas conectadas. Esto se hace sin tener en cuenta si se deseaba o no obtener estos estados de activación (salidas). El otro método es el Competitivo y cooperativo: Las neuronas compiten entre sí para llevar a cabo una tarea dada, que cuando se cumpla solo una o un grupo de neuronas pueden activarse. El objetivo es clasificar por categorías los datos introducidos a la red, para que las similares se formen activando la misma neurona de salida.
 - Tipo de Asociación de las neuronas de E/S: se emplean dos tipos de asociación:
 1. Redes Heteroasociativas: En el que la red aprende parejas de datos $[(A_1, B_1), (A_2, B_2) \dots (A_n, B_n)]$, de tal forma que cuando se presente cierta información de entrada A_i , la red deberá responder generando la correspondiente salida B_i .
 2. Redes Autoasociativas: Donde la red aprende ciertas informaciones $A_1, A_2 \dots A_n$, de tal forma que cuando se le presente una información de entrada esta realizará una auto correlación, respondiendo con uno de los datos almacenados, en este caso, el más parecido al dato de la entrada.

- Representación de la Información de E/S: Las RNA pueden representar la información de entrada y de salida de manera diferente, con base en lo cual se pueden clasificar:

- Red Continua: Donde los datos de entrada y salida son análogos.
- Red Discreta: Donde los datos de entrada y salida son de tipo binario.
- Red Híbrida: Donde la información de entrada es continua, pero la de salida es binaria.

Ahora bien, entre las principales aplicaciones de las redes neuronales se tiene:

Las Redes Neuronales Artificiales han brindado una alternativa a la computación clásica y ha ayudado a avanzar notablemente en los métodos de investigación, las aplicaciones en las que mayor éxito ha tenido la implementación de dichas redes son:

- Procesamiento de Imágenes y Voz
- Filtrado de señales
- Reconocimiento de Patrones
- Interfaces adaptables para sistemas Hombre/Maquina, entre otros.

También se pueden clasificar de acuerdo al área de aplicación o investigación, entre los cuales podemos encontrar los siguientes:

Finanzas

- Previsión de la evolución de los precios
- Valoración del riesgo de los créditos
- Interpretación de firmas
- Identificación de falsificaciones

Manufacturación

- Robots y sistemas de control (visión artificial y sensores de presión, temperatura, etc)
- Control de producción en líneas de proceso

Medicina

- Analizadores del habla para la ayuda de audición de sordos
- Monitorización en cirugía
- Predicción de reacciones adversas a los medicamentos
- Lectores de rayos X
- Diagnóstico y tratamiento a partir de síntomas y/o datos analíticos

Militares

- Clasificación de las señales de radar
- Creación de armas inteligentes
- Reconocimiento y seguimiento en el tiro al blanco

2.6. SINTESIS DEL ESTADO DEL ARTE

Luego del marco conceptual, referencial e histórico; se plantea una revisión del estado del arte del Reconocimiento de Voz. En este sentido, la investigación se enfoca en los desarrollos más recientes del tratamiento de las señales de voz, con algunas referencias de hace varias décadas¹⁰.

El primer trabajo que se desea destacar es del año 1978 [13], el cual consiste en un sistema de reconocimiento de palabras aisladas y con gramática controlada. Esta prueba fue realizada con palabras independientes presentadas en idioma inglés e italiano. Cada uno de los niveles de reconocimiento son controlador por reglas gramáticas y por variables lingüísticas fuzzy¹¹. El modelo aplicado fue

¹⁰ Principales referencias utilizadas- Bases de datos ELSEVIER, IEEE, entre otras.

¹¹ Metodología de trabajo multivalente, basado en reglas lingüísticas.

interesante para su época a pesar de sus limitantes por la gran cantidad de patrones que requería para una síntesis de vocablos reducidos.

En la segunda parte de la década de los ochentas, se registran trabajos de gran interés para el presente proyecto; tales como un sistema de Reconocimiento de voz de bajo costo [14], el cual fue desarrollado para una aplicación de respuesta automática de operadores telefónicos. Entre los retos tecnológicos de este trabajo se cuenta con exigencias de bajo costo, bajo consumo de potencia y reducido set de componentes. Este proyecto utilizaba un microcomputador de chip sencillo fabricado por Hitachi, y básicamente se fundamentaba en el uso de extracción de patrones con un sistema de cruces por cero, aprovechando una ventana dinámica en su obtención. El modelo fue uno de los primeros ejemplos prácticos e implementados en público para el reconocimiento de palabras aisladas.

Un par de años después (1988), se presenta una mejora al sistema, donde se realiza la expansión y adaptabilidad de la base de conocimiento, aplicado en un sistema microcontrolado [15]. En este caso la cantidad de vocablos manejados fue muy superior y con tasas de precisión bastante interesantes para la época.

Ahora bien, en el contexto de los trabajos encontrados que han realizado mayor énfasis en el filtrado y adecuación de señal; enfocados a ambientes con altas componentes de ruido, es posible referenciar [16], [17] y [18]; los cuales plantean propuestas de aumento de desempeño en condiciones complicadas. El entrenamiento de este tipo de sistemas tiene una alta exigencia matemática, de implementación de software y hardware específico, entre otros. Destacar del trabajo [18], desarrollado en conjunto por institutos de investigación de China y Estados Unidos, la implementación de herramientas con reglas de clasificación, por elementos predictivos Bayesianos, los cuales, según su autor, funcionan

como estrategias anticipativas a las posibles situaciones y afectaciones del ruido sobre la señal de interés.

Aunque para el lector esto suene trivial, se considera necesario destacar que al restringir las condiciones de operación, tanto de registro como de número de vocablos, el sistema reduce su complejidad. En este sentido, si las consideraciones se vuelven menos restrictivas, los sistemas empiezan a exigir mayor complejidad para obtener similares resultados que en el caso anterior.

Continuando con la revisión, entre los trabajos más destacados de los últimos años se encuentran los presentados en [19], [20], [21] y [22]. El primero de estos explica, de manera detallada, el diseño y la ejecución de experimentos con personas discapacitadas y sistemas de reconocimiento de voz. En este caso, se utilizan varios modelos de reconocimiento y se aplican con poblaciones con implantes cocleares¹².

Siguiendo esta misma línea de investigación nos encontramos con un estudio realizado por el Departamento de Ciencias de la computación de la Universidad de Sheffield en el Reino Unido, en el año 2007, en donde se observa claramente una investigación acerca del ASR, y en donde se analizan los obstáculos que impiden un resultado 100% satisfactorio para el usuario, como lo son el ruido,

¹² *“Es un aparato electrónico que se coloca en el oído interno durante una cirugía, diseñado para ayudar a pacientes con sordera neurosensorial severa a profunda, que obtienen muy poco o ningún beneficio de los audífonos convencionales. Los pacientes van a poder detectar sonidos del medio ambiente, la mayoría va a poder entender el lenguaje sin la ayuda de lectura de labios y algunos podrán utilizar el teléfono”. Tomado de [23].*

pocos conocimientos gramaticales y semántica; en el que se concluye la falta de solidez y tolerancia hacia los acentos extranjeros, y se estudia una forma de hacerlos más tolerables y compatibles hacia este tipo de problemas [20].

Como comentario personal cabe resaltar que este tipo de sistemas no siempre se pueden manejar para que sean compatibles con todos los idiomas mundiales, porque sería un trabajo bastante extenuante, sin embargo con aplicaciones, tiempo, dedicación y un estudio más avanzado de las RNA y el Reconocimiento de voz, podría llegar a lograrse.

Seguido a este, se encuentra un proyecto realizado en la Universidad Nacional de Irlanda en el año 2008 [21], en el cual se enfoca en el rendimiento no deseado de algunos sistemas debido al ruido encontrado en las señales y a la poca disminución del mismo. Este proyecto es se enfoca en realizar una comparación entre los sistemas tratados por medio del reconocimiento de voz basados en HMM y a los que no se les realiza ningún tratamiento para la disminución del ruido, obteniendo como resultado que la combinación de tratamiento previo de mejora de señales y el modelo de front-end¹³ ofrecen una mejora en el rendimiento de reconocimiento de la voz en condiciones ruidosas.

Como último proyecto de los cuales se analizó en este proyecto, se encuentra una importante investigación en la que estudiantes del Instituto de automatización de la Academia China de Ciencias en Beijing, investigaron acerca de un método que podría mejorar los sistemas de reconocimiento de voz, ya que ofrece una solución y presenta un novedoso sistema en el cual se pueden diferenciar dos tipos de oradores, por medio de modelos de mezcla Gaussianos (OMG) y vectores cuantificables que recibirían las señales de

¹³ En el cual el tratamiento de disminución de ruido por frecuencias se hace en el momento de recibir la señal de entrada.

agrupación de datos aislados para cada persona, de esta forma podría darse más solidez a los sistemas de Reconocimiento de voz [22].

2.6.1. Aplicaciones En Colombia

En la presente sección se detallan 5 proyectos que se han considerado relevantes y destacables, en el estado del arte de los RAH en Colombia. Aunque se encuentran trabajos adicionales que pueden ser de interés para la investigación, los presentados se consideran de alto impacto para el ejercicio de reconocimiento de habla automático, eje central del presente proyecto:

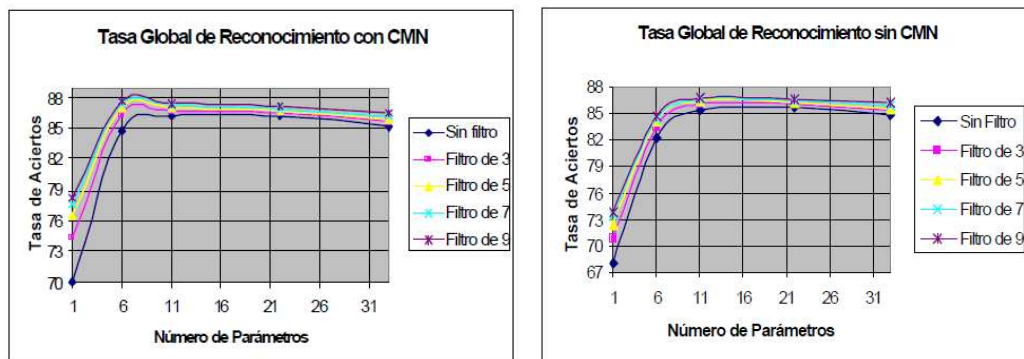
- **Desarrollo de un Clasificador Bayesiano de Ruido y Voz Mediante Estimación por Máxima Verosimilitud:** Ahora bien, es importante para el lector familiarizarse con los casos más destacados que se tienen en Colombia. La totalidad de trabajos presentados acá han sido desarrollados por centros y grupos de investigación de Universidades reconocidas. El primero a destacar fue desarrollado en la Universidad Autónoma de Occidente, en la ciudad de Cali [23]; el cual plantea un estudio de tasas de reconocimiento de voz y ruido, para clasificar las tramas de voz, adecuarlas y estudiar la forma de hacerlas más significativas a la hora de implementarlas en un sistema de educación virtual.

Importante resaltar del trabajo descrito que se ha implementado una técnica clásica que permita normalizar los parámetros espectrales, denominados Cepstral Mean Normalization (CMN), la cual aumenta la tasa de aciertos al ejecutar las pruebas; incrementando la robustez del reconocedor ante el canal de grabación, como se evidencia en la figura 6. Al observar los datos asociados a la figura, se puede determinar

porcentajes de acierto que superan el 80%, además, al implementar la estrategia CMN, se logran mejoras cercanas al 4%.

Como información inicial para el algoritmo, el proyecto en mención ha tomado como una base de datos en español que fue grabada en la Universidad Politécnica de Cataluña, en el cual la mitad de los locutores de dicha base de datos son hombres y la otra mitad mujeres entre los 16 y 30 años de edad, y son personas ubicadas en diferentes lugares de España, por lo cual se cuenta con una gran variedad de dialectos.

Figura 6: Comparación de patrones



Fuente: Tomado de [23]

Finalmente, se puede resumir que este sistema aplica una serie de teoremas y algoritmos matemáticos de normalización, en los cuales el principal objetivo es lograr diferenciar las tasas de voz y ruido, basados en las frecuencias generadas por los mismos, para filtrar los tramos de voz y hacerlos más potentes.

Robot móvil controlado por voz: Cabe también resaltar un proyecto realizado por el Politécnico Colombiano Jaime Isaza Cadavid en la ciudad de Medellín [24]. En el cual se presentó un robot que obedece comandos de voz para moverse en distintas direcciones. La implementación de este tipo de robot se dio gracias a la investigación y trabajo realizado en base a redes neuronales y procesamiento digital de la voz; este avance podría convertirse en un dispositivo de gran ayuda para facilitar la movilidad de las personas con algún tipo de discapacidad. La imagen de dicho robot se presenta en la figura 7.

Figura 7: Robot controlado por voz



Fuente: Politécnico Colombiano Jaime Isaza Cadavid.

Para este móvil, y en termino de valores cuantitativos, se puede indicar que sus autores referenciar que el mecanismo de reconocimiento tiene una efectividad cercana al 94%, además la implementación de filtrado digital avanzado, hace de gran interés el trabajo de este grupo de investigación.

Como apreciación personal del autor, aunque el móvil desarrollado cuenta con aplicaciones de movilidad; también puede ser considerado para robótica lúdica, educativa y de competición, generando nuevos tipos de

concursos que secunden a los ya clásicos seguidores de línea, luchadores de sumo, entre otros. Este tipo de trabajos son de gran interés para los desarrollos de los semilleros de investigación de las universidades, como es el caso del existente en la Universidad Pontificia Bolivariana de Bucaramanga; a la cual está adscrito el autor, en calidad de estudiante, del presente documento.

- **Automatización de una Silla de Ruedas Controlada por Comandos de Voz:** Siguiendo la línea de investigación en robótica, se puede encontrar un proyecto realizado por la Universidad de Pamplona [25]; en el cual se implementa la automatización de una silla de ruedas controlada por comandos de voz simples, tales como adelante, atrás, derecha, izquierda y alto. De igual forma desde la silla de ruedas puede realizar varias funciones como manejo de puertas y luces de la casa.

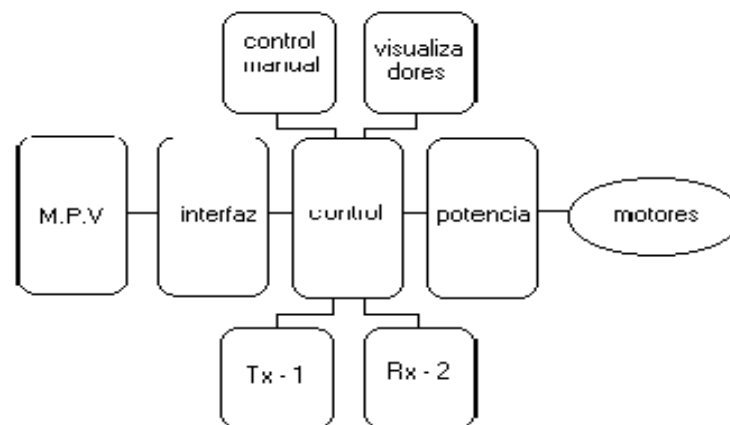
El proceso de diseñar la silla se comenzó teniendo conocimientos en las áreas de control, electrónica digital y tratamiento de señales, esto con el fin de implementar la tecnología no solo con fines lucrativos, sino que por el contrario adquirir un mayor compromiso con la sociedad.

En cuanto al procedimiento, la señal de la voz ingresa al sistema por medio de un micrófono, y es capturada por un circuito integrado que incluye un DSP¹⁴, con el cual se realiza el procesamiento de la señal y es convertida en información digital para que a su vez pueda ser reconocida por una red neuronal que se encarga de la identificación de la palabra, para dar paso a una salida digital que es utilizada para dar marcha al sistema.

¹⁴ Procesador digital de señales, el cual se considera un sistema de cómputo que posee un juego de instrucciones Hardware y Software; el cual puede trabajar con varios datos en paralelo y un diseño e instrucciones específicas para el tratamiento de señales.

Al incluirse dentro de la silla dos módulos de radio frecuencia se le permite al usuario que al pronunciar una palabra de control domótico, el circuito dentro de la silla genera una señal al circuito de control de la casa y se puede ejecutar la acción. En la figura 8 se presenta el modelo general utilizado en el sistema de control, el cual está constituido por la etapa de reconocimiento de voz, la interfaz al microcontrolador, el sistema microcontrolado encargado de la gestión de visualización, control manual, y comunicación de información con las componentes domóticas; a su vez se cuenta con el sistema de potencia para el control de los motores que se encargan del movimiento del equipo.

Figura 8: Sistema de control en la silla

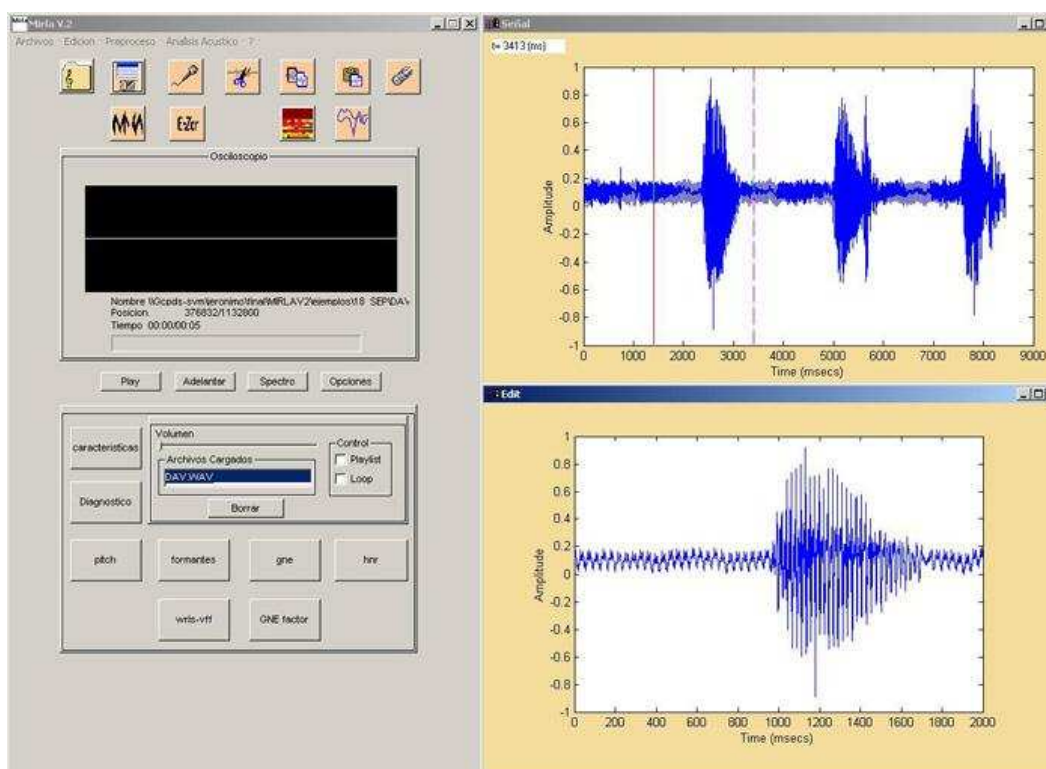


Fuente: Automatización de una silla de ruedas controlada por comandos de voz. [25]

- **Prototipo de análisis de patologías de voz:** En el área de la medicina y ayuda a la sociedad, recientemente fue creado un prototipo que permite hacer estudios y análisis acústicos computarizado de la voz [26] para diagnosticar patologías de la voz, y por medio del cual los especialistas y

expertos de la medicina pueden observar el grado de calidad de la misma para descubrir enfermedades y dar tratamientos oportunos a las mismas. En la figura 9 se puede apreciar el panel frontal del software diseñado.

Figura 9: Prototipo software de análisis de patologías de voz



Fuente: Análisis acústico de voz, control y procesamiento digital de señales [26].

Dicho software permite medir pérdidas de potencia en la voz, pérdida de claridad, inicio de ronquera entre otros. En ese orden de ideas, la calidad de la voz es una propiedad diferente a la del reconocimiento, sin embargo se consideró importante y relevantes presentar este trabajo por el impacto social y en la salud del desarrollo.

Finalmente, este trabajo ha sido destacado, pues a nivel mundial no se cuentan con software comerciales ampliamente desarrollados para cumplir este apoyo diagnóstico.

- **Aplicación de reconocimiento de voz utilizando HTK:** Continuando con la revisión, se encuentra un proyecto desarrollado en la Universidad Javeriana; el cual utiliza un reconocido software denominad HTK [27] para la obtención de un patrón de reconocimiento de voz. Es pertinente indicar que dicho aplicativo utiliza o trabaja con modelos HMM¹⁵, sobre los cuales ya se ha descrito en la sección del marco referencial del presente trabajo de grado. El aplicativo fue diseñado principalmente para reconocimiento de voz, pero se ha utilizado en muchas aplicaciones de reconocimiento de patrones que emplean a otros modelos de Markov.

Para la implementación de este modelo fue importante concretar fases de entrenamiento y reconocimiento de los parámetros de los HMM, para observar claramente las secuencias del modelo y calcular la disminución de las frecuencias del ruido causado en la señal que al momento de su entrada es convertida en una serie de vectores acústicos que son procesados y obtener una señal más limpia y clara.

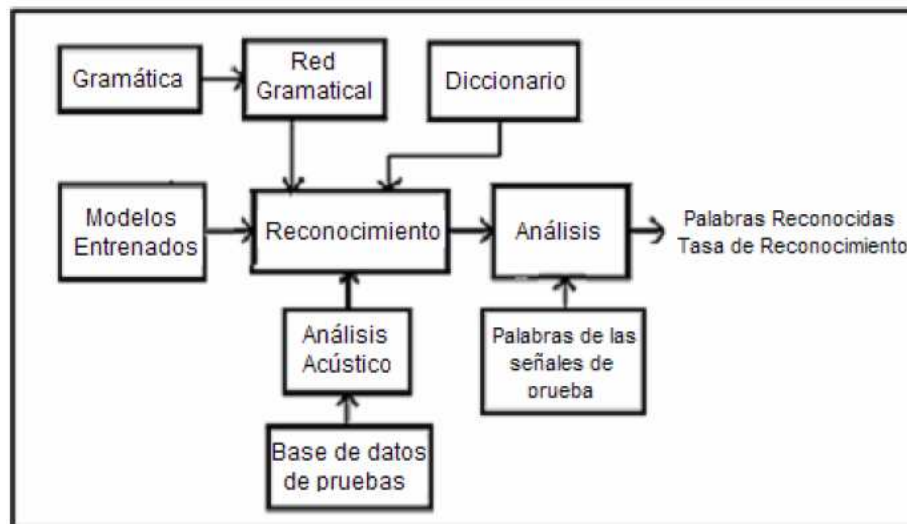
En la figura 10 se destaca el proceso de reconocimiento basado en los Modelos de Markov.

- **Sistema inteligente de reconocimiento de voz para la traducción del lenguaje verbal a la lengua de señas:** Complementario a los anteriores trabajos, otro proyecto de alto impacto se encuentra presentado en [26].

¹⁵ Hidden Markov Model ó Modelo Oculto de Markov, es un modelo estadístico que determina parámetros desconocidos u ocultos, de una cadena a partir de los parámetros observables, se considera como una red bayesiana dinámica simple.

En el cual se implementa una herramienta desarrollada en MATLAB™¹⁶, que permite la creación y modificación de redes neuronales que se aplican al reconocimiento de palabras, para luego presentar grafica y visualmente la palabra en lengua de señas colombiana y a su vez el deletreo grafico de dicha palabra.

Figura 10: Modelo de reconocimiento por modelos ocultos de MARKOV



Fuente: Tomado del documento presentado en [27].

En la figura 11 se muestra el sistema de tratamiento de señal de voz realizado en el proyecto de procesamiento de vocablos para generación de lenguaje por señas.

El desarrollo de la aplicación comienza con la grabación de las señales de voz, esta es guardada como un archivo de audio. La señal que ha sido

¹⁶ Es un ambiente de computo, de alta ejecución numérica y de visualización, que integra el análisis numérico, calculo de matrices, procesamiento de señales, y graficación, en un ambiente sencillo de utilizar, donde los problemas y sus soluciones son expresadas justamente como están escritas; a diferencia de la programación tradicional

almacenada posee un ruido generado por el ambiente; el cual está rodeado de frecuencias cercanas a los 60 Hz, por tanto se debe aplicar un filtro que elimine estas frecuencias.

Figura 11: Síntesis del tratamiento de señal proyecto lenguaje por señas



Fuente: Autor del proyecto

En el sistema inteligente de reconocimiento de voz para la traducción del lenguaje verbal a la lengua de señas colombiana (VLSC), a partir de las palabras a reconocer por medio de procesamiento digital de las señales, procede a realizar una extracción de características específicas de las palabras para posteriormente implementar una red neuronal

probabilística, que se encarga de comparar las matrices características de cada palabra con la señal de voz que está entrando.

De esta forma se permite finalmente adquirir la señal de entrada y visualizarla en el lenguaje de señas colombiano, a través de una interfaz grafica que me integra la representación, el deletreo y el texto de dicha seña.

De acuerdo a procesos actuales de la Universidad Pedagógica Nacional en el cual se integran estudiantes sordos, se realizó un estudio de viabilidad implementando una serie de pruebas con 14 docentes de la institución, y se llegó a la conclusión de que el sistema VLSC es una herramienta de apoyo eficiente para el docente, ya que posee una interfaz útil, clara y coherente, que mejora el rendimiento de la educación para la comunidad sorda.

- **Reconocimiento Automático del Habla Utilizando la Transformada De Fourier y Redes Neuronales**

Este último trabajo de la revisión, estudia el reconocimiento de señales o vocablos utilizando dos herramientas matemáticas reconocidas: Transformada de Fourier y Redes Neuronales. Este proyecto se ha encargado de generar diferentes patrones de comparación con el dominio de la frecuencia y se ha entrenado una red neuronal para ejecutar dicha tarea.

La adquisición de señales se realiza a través de una tarjeta de sonido de los sistemas de cómputo, cumpliendo el Teorema de Nyquist, con tasas de muestreo superiores a 8000 muestras/segundo. Se utiliza el algoritmo

de la Transformada Rápida de Fourier (FFT) para posteriormente promediar dicha señal para la reducción del costo computacional [29].

Entre las conclusiones de este trabajo se demuestra, junto a los resultados, que la FFT es una herramienta útil para la extracción de las características de la voz. Se recomienda el uso de pre-procesamiento de las señales.

Finalmente, luego de la revisión bibliográfica realizada (objetivo inicial del presente proyecto), se abre paso al tercer capítulo del presente documento, donde se detallan las consideraciones de diseño del sistema de reconocimiento de voz a través de circuitos integrados y soluciones hardware.

3. ELECCIÓN E IMPLEMENTACIÓN DEL SISTEMA DE RECONOCIMIENTO

Luego de un marco referencial claro y una revisión del estado del arte de lo que se tiene, a nivel mundial y local, se procede a describir el diseño del sistema de reconocimiento de voz en el presente ejercicio académico. Se ha elegido desarrollar un proyecto de aplicación efectiva, fundamentado en hardware de reconocimiento que aplique un sistema de **identificación dependiente**¹⁷ del hablante y con detección de **palabras aisladas**¹⁸. Se eligen estos criterios directamente por el tipo de elementos que se desean controlar (brazo robótico), en el cual solamente se realizará el direccionamiento y posicionamiento en el plano X-Y.

Para llevar a cabo esta tarea, fue importante determinar los comandos que se van a necesitar para controlar el brazo robótico, asociados a palabras o frases comúnmente asociadas con estos comandos. También, es importante elegir un set o conjunto de palabras que su sonido sea diferente, para evitar las confusiones del equipo de cómputo (ej: mesa/pesa, casa/taza, entre otros). En la tabla 1 se muestra un típico ejemplo para el desplazamiento de un sistema robótico, el cual es presentado en el trabajo de la referencia [28]; dicho ejemplo describe un conjunto de palabras utilizado para la ejecución de un número específico y reducido de acciones.

Ahora bien, en una profunda revisión de opciones de circuitos integrados existentes para cumplir la tarea, se han acotado a los siguientes:

¹⁷ Speaker-dependent

¹⁸ Isolated-word

- HM2007 (HMC semiconductors)
- MSM6679 (OKI semiconductors)
- RSC-164 (Interactive Speech)
- RSC-300 (Interactive Speech)
- RSC-484 (Interactive Speech)
- Hello IC (Philips Semiconductor)

Los mas destacados serán descritos en la sección 3.1 de manera sucinta, para al final de la misma presentar una tabla comparativa y las razones para la elección de uno de estos.

Tabla 1. Ejemplo de listado de vocabulario y acciones de un RAH (sistema robótico).

CODIGO	PALABRA	ACCION
0	Arriba	Mover en eje Z+
1	Abajo	Mover en eje Z-
2	Derecha	Mover en eje Y+
3	Izquierda	Mover en eje Y-
4	Frente	Mover en eje X+
5	Atrás	Mover en eje X-
6	Abrir	Abrir pinza
7	Cerrar	Cerrar Pinza
8	Rápido	Aumentar el tiempo de paso
9	Lento	Disminuir el tiempo de pazo
10	Salir	Salir del Sistema

Fuente: Autor del proyecto basado en referencia [28]

3.1 SISTEMAS RAH POR HARDWARE COMPARADOS

Los siguientes elementos fueron elegidos por ser comúnmente utilizados en las referencias encontradas en el estado del arte. Entre los criterios de elección se cuenta: facilidad de implementación, alta precisión y bajo coste, entre otros.

3.1.1 HM2007 [30]

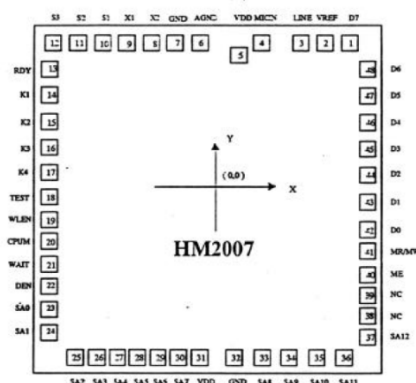
El fabricante de este dispositivo es la empresa norteamericana HMC. Su desarrollador lo describe como un chip simple de reconocimiento de voz basado en tecnología CMOS que cuenta entre sus propiedades con análisis de voz, procesos de reconocimiento y funciones para sistemas de control. Los principales elementos que requiere el sistema para su funcionamiento es: micrófono externo, teclado de configuración, memoria SRAM de 64kB, entre otros. El HM2007 cuenta con la posibilidad de gestionar hasta 40 palabras y se puede combinar con elementos más complejos, como microprocesadores que implementen redes neuronales, algoritmos genéticos, lógica fuzzy, entre otros.

A continuación se listan las principales características de este circuito integrado:

- Sistemas LSI de reconocimiento de voz.
- Sistema de reconocimiento dependiente del hablante.
- Posibilidad de expansión con memoria SRAM de 64KB.
- Máximo de 40 palabras reconocibles por chip.
- 1.92 segundos para reconocer una palabra.
- Posibilidad de conexión del sistema en paralelo.
- Conexión directa a micrófono para registro de señal.
- Tiempos de respuesta menor a 300ms.
- Fuente de alimentación de 5V.

- **Modo Manual:** En este modo de operación, se recurre a un teclado externo, a una memoria SRAM y otros componentes para construir un sistema de reconocimiento simple. La gestión de pines es bastante sencilla y se utiliza un standard de bus de datos sencillo de gestionar.
- **Modo de control CPU:** El modo de operación CPU provee funciones adicionales de operación que serán descritos a continuación. En este modo, el bus puede ser utilizado de manera bidireccional, para así poder operar interconectado a un microcontrolador o a un DSP. Existen tres registros en el chip, uno encargado de almacenar, un registro de status y uno buffer de salida.

Figura 12: Circuito esquemático HM2007



53

3.1.2 MSM6679 (OKI semiconductors) [31]

El presente circuito integrado se considera un VRP, ósea un Procesador de Reconocimiento de Voz; el cual actúa como un dispositivo esclavo que permite varias funciones, entre las que se destaca:

- Reconocimiento de voz independiente del hablante.
- Reconocimiento de voz dependiente del hablante.
- Grabador de sonido de estado-sólido.
- Reproductor de sonido.
- Sintetizador de audio.

Otras de las características a destacar es su conversor de datos, así como una salida de PWM para monitoreo del correcto funcionamiento del sistema. Este chip permite establecer un conjunto de vocablos grabados en memoria externa o establecidos en el microchip. Los requerimientos de memoria están directamente ligados al tamaño del vocabulario.

Entre los elementos más interesantes de este tipo de aplicación, se cuenta con una alta tolerancia del ruido del ambiente y una precisión considerable que supera el 95%. Se pueden configurar hasta 61 palabras predefinidas, y se puede expandir con memoria SD. Sintetizando los parámetros principales se cuenta con:

- Soporte de vocabulario múltiple con más de 20 palabras (independiente del hablante).
- Soporte de vocabulario múltiple con más de 60 palabras (dependiente del hablante).
- Síntesis de palabras y conversiones cercanas al 2.3 segundos. El sistema de síntesis está embebido en el chip del controlador.

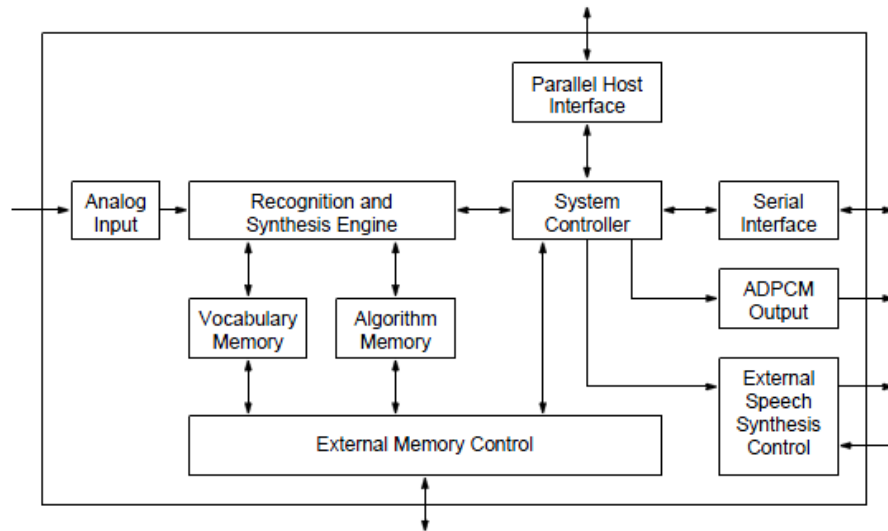
- Cuenta con una tasa máxima de muestreo de 10 kHz.
- Latencia de reconocimiento cercano a 200ms.
- Reloj de operación de 32MHz.
- Empaquetado de 84 pines superficial.

En la figura 13, se presenta un diagrama de bloques descriptivo del funcionamiento del integrado descrito. Interesante destacar algunos elementos tales como:

- **Analog Input:** Entrada de conversión analógica. Conversión de señales y digitalización.
- **Recognition and Synthesis Engine:** Este motor de reconocimiento y de síntesis de resultados permite construir la comparación de las señales con los patrones establecidos. Dicha componente se encuentra asociada a las **Memorias de vocabulario y de algoritmo**, en las cuales se encuentra registradas las palabras establecidas y el algoritmo de comparación.
- **System Controller:** Etapa de gestión de comunicación y conversión de señales con las partes exteriores del integrado.

Un punto clave al momento de realizar la elección de este dispositivo, es que el consumo de corriente supera los 120 mA en el peor de los casos; lo cual dista del criterio planteado al inicio del texto de bajo consumo de potencia.

Figura 13. Esquema de funcionamiento circuito integrado MSM6679



Fuente: Hoja de Datos del fabricante.

3.1.3 RSC-164/300/464 [33]

El conjunto de circuitos integrados de la línea RSC-164/300/464 describe una familia de microcontroladores de propósito general con características de reconocimiento de voz, verificación de habla y capacidad para grabar y reproducir audio. Estos elementos son desarrollados por Interactive Speech TM y han sido aplicados en buena cuenta en proyectos sobre reconocimiento de voz.

Se consideran microcontroladores de bajo costo, cuentan con una resolución de 8 bits (indicar que en el estado del arte ya se encuentran microcontroladores de 32 bits), y que cuenta con un sistema de reconocimiento independiente/dependiente del hablante. El dispositivo cuenta con una potente **red neuronal artificial**, encargada del aprendizaje y la clasificación de los sonidos.

En el conjunto de datos suministrados por el fabricante, se encuentra una efectividad superior al 96% en el caso de modelos independiente del hablante y de un 99% para reconocimiento dependiente del hablante. El uso de las redes neuronales en este dispositivo, reduce la necesidad de utilizar más circuitos integrados para cumplir una tarea específica.

El autor del presente documento considera necesario resaltar que dicho integrado cuenta con un gran nivel de integración, lo que ha reducido el uso de elementos externos para su funcionamiento. Al igual que los ejemplos anteriores, se puede realizar conexión directa a batería (alimentación), micrófono (registro de datos), speaker(reproducción de audio), entre otros. Es destacable el bajo consumo de potencia que lo hace interesante para aplicaciones autónomas alimentadas por batería o dispositivos portables.

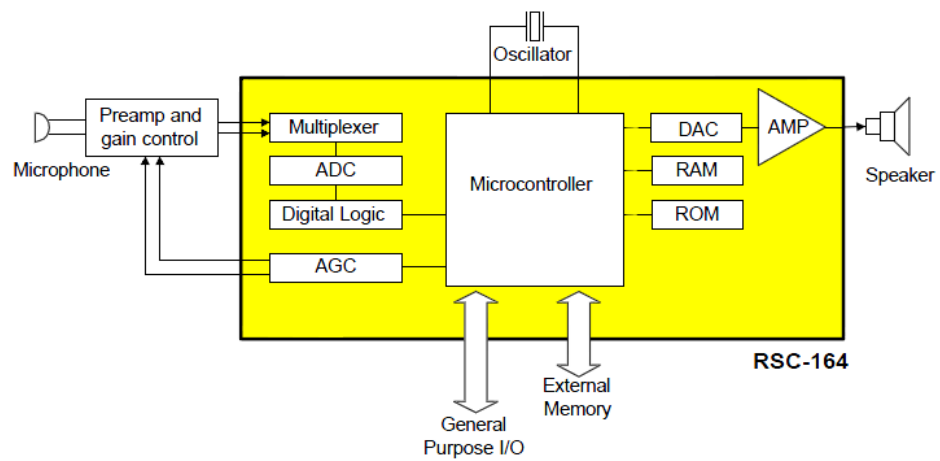
En la figura 13, se describe de manera puntual las partes que componen el circuito integrado. Por destacar en este trabajo, el sistema de reconocimiento y la red neuronal se encuentra instalado en la sección denominada microcontrolador, así como mostrar que existe la posibilidad de expansión de su capacidad con memorias externas. En dicho gráfico se observa que, como elementos externos de funcionamiento, se requiere simplemente un oscilador para la generación de señal de reloj, etapas de preamplificación que sean necesarias para el correcto funcionamiento del circuito y las entradas y salidas de propósito general.

Finalmente, sintetizando sus características, se tiene:

- Capacidad de reconocimiento dependiente o independiente del hablante.
- Verificación de speaker y síntesis de sonidos de gran calidad.
- Microcontrolador de 8-bits con 4 MIPS.
- Conversores de datos internos, tanto D/A como A/D; cuenta con posibilidad de filtrado digital.

- Capacidad de grabación de información y reproducción de sonido.
- Memoria interna de 64kB y posibilidad de expansión externa.
- 16 líneas de entrada y salida de propósito general.
- Voltaje de 3.5 V a 5 V
- Consumo de corriente promedio de: 25 mA

Figura 13. Diagrama de bloques del circuito RSC-164



Fuente: Hoja de datos del fabricante

3.2 TABLA COMPARATIVA DE CIRCUITOS INTEGRADOS REVISADOS

Los criterios elegidos para esta comparación se encuentran asociados a características técnicas, detalles eléctricos, facilidad de consecución y coste.

Tabla 2. Cuadro comparativo de tres opciones de sistemas de reconocimiento de voz

C.I	S.D/S.I *	Frecuencia de Procesamiento [Hz]	Memor. interna [kB]	Número de palabras max.	Latencia de conver. [s]	Síntesis y reprod. de audio	Consumo de potencia I[A]-P[W]	Capacidad de expansión de memoria	Costo promedio [USD\$]	Lugar de compra **
HM2007	SI/SI	8 MHz	64	40	1.9	SI/NO	6 mA – 30mW	SI – Memoria SRAM	Chip: 16 Demo board: 180	Colombi a /USA
MSM6679	SI/SI	32 MHZ	64- 1024***	60	1.2	SI/SI	>150mA - 1300mW	SI- Memoria SD	Chip: 25.5 Demo board: 270- 300	USA
RSC-164	SI/SI	14.32 MHz****	128	52	2.1	SI/SI	12 mA – 65 mW	SI- Memoria SD	Chip: 27.5 Demo board: 500- 700	USA

* Speaker Dependent /Speaker Independent

** Capacidad de consecución en el mercado electrónico colombiano.

*** Dependiendo del modelo se tiene hasta 1024 kB de memoria interna.

**** En la línea más baja. Este valor se incrementa en el RSC – 300/464.

Basado en la tabla 2 y considerando los siguientes puntos:

- Aplicación sencilla de control que no requiere más de 15 palabras o direcciones.
- Bajo consumo de potencia para el control del brazo robótica en modo stand-alone o autónomo.
- Simplicidad en programación.
- Fácil consecución y bajo coste.

Se ha decidido utilizar la opción del HM2007, ya que cuenta con los requerimientos básicos para el funcionamiento del sistema, así como el más bajo costo y consumo de potencia de las opciones planteadas. Importante resaltar que en ningún momento se afirma que en las especificaciones técnicas generales de reconocimiento, las propiedades de expansión y las metodologías de comunicación no sean más eficientes las otras dos opciones; simplemente que el HM2007 es el más cercano a los requerimientos establecidos para el control del brazo robótico de prueba.

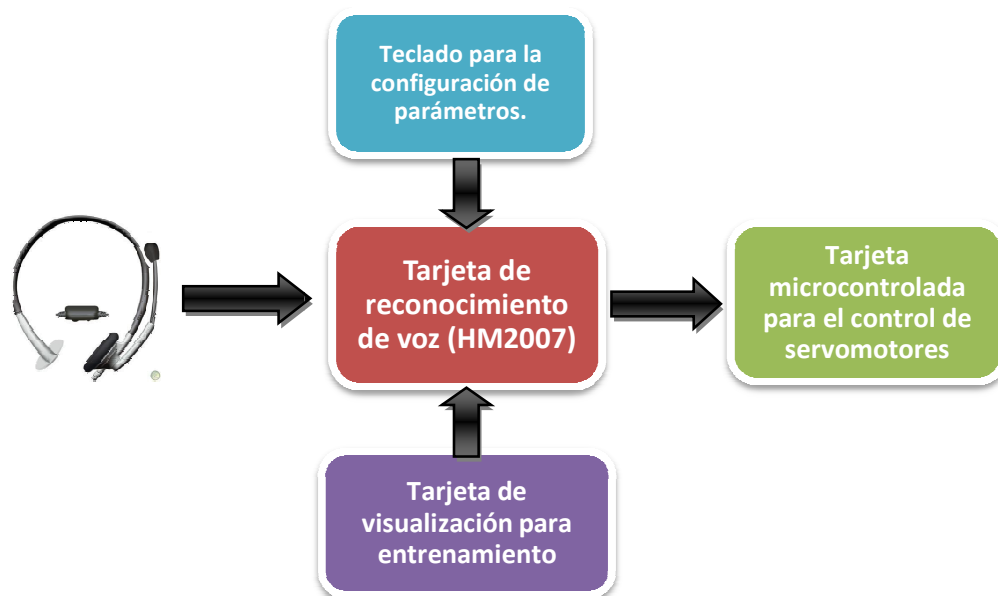
3.3 IMPLEMENTACIÓN DEL SISTEMA CON HM2007

Se ha decidió entonces trabajar, de manera efectiva, con un sistema de reconocimiento de voz HM2007 aplicado a un brazo robótico. Para dicho fin, se ha realizado el diseño y construcción de una tarjeta de reconocimiento y control, la cual debe ser previamente entrenada para el cumplimiento de las tareas específicas. Aunque la concepción de la misma es entregada por el fabricante, y puede ser adquirida como un Kit asociado al circuito integrado, se ha decidido construir un prototipo propio para validar el funcionamiento del mismo.

El sistema implementado cuenta con los siguientes elementos:

- Tarjeta de reconocimiento de voz (programación del HM2007).
- Tarjeta de visualización de datos (Display 7 segmentos).
- Teclado para ingreso de información.
- Tarjeta microcontrolada para la gestión del movimiento de los servomotores del brazo robótico.

Figura 14. Diagrama de bloques del circuito RSC-164



Fuente: Autor del Proyecto

Además, a la salida de la última etapa se conecta directamente al sistema de potencia del brazo robótico. Aunque dicho brazo no es más que un hardware de validación del funcionamiento del prototipo creado, se realizará en las siguientes líneas una breve descripción del mismo.

3.2.1 Brazo robótico (periférico final)

Se ha elegido un brazo robótico muy utilizado para las aplicaciones académicas y de trabajo en semilleros y grupos de investigación asociados al tema de robótica. El robot elegido para estas pruebas es un OWI007 [34], el cual es una gran herramienta para la enseñanza de la robótica básica y los principios de locomoción, para probar las capacidades de los motores y la capacidad de construir y controlar un brazo. En la figura 15 se muestran algunas ilustraciones del brazo utilizado. Es importante destacar que la unidad viene por defecto con un control cableado que permite mover la pinza, subir, rotar, entre otras funciones que puede realizar el brazo robótico elegido. Entre las principales características se tiene:

- Sistema con 5 grados de libertad.
- Base con rotación de 350°.
- Articulación del hombro con 120°.
- Articulación del codo con 135°.
- Rotación CW y CCW de 340°.
- Abertura y cierre de la pinza 50mm.
- Alimentación con 4 baterías.

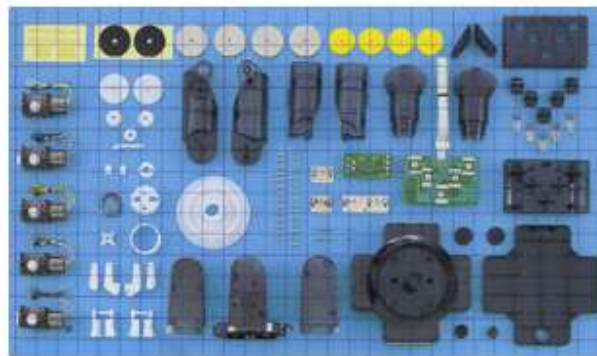
Figura 15. Imágenes asociadas al brazo utilizado.



Fuente: Autor del proyecto

Finalmente, queda por destacar que el brazo robótico está compuesto por elementos mecánicos (poleas, piñones, engranajes, entre otros), así como de partes electrónicas de control, motores y componentes de acrílico para su estructura. Cada una de estas partes se muestra en la figura 16 suministrada por el autor del proyecto.

Figura 16. Partes constitutivas del brazo OWI007



Fuente: Hoja de datos del fabricante

3.2.2 Tarjeta de Reconocimiento de voz.

En este punto se revisa la tarjeta de mayor interés del desarrollo del Reconocimiento de Voz. A este punto ya se ha comentado que el circuito integrado que se ha elegido es el HM2007, pero además de esto es necesario destacar los elementos adicionales que constituyen la tarjeta.

La tarjeta principal está compuesta principalmente por el integrado de 48 pines HM2007, además de eso se adjunta una memoria RAM adicional de 8kB que permite almacenar la información de las palabras a almacenar. La hoja de datos del HM2007 recomienda el uso de un integrado compuesto por varios circuitos de latch (unidades de memoria) que se conecta a la salida de la memoria RAM.

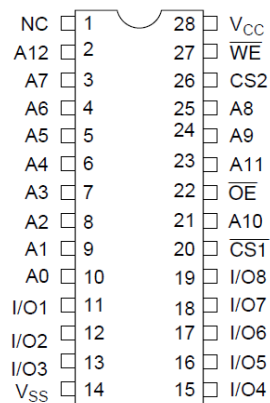
Se adicionan a la tarjeta elementos pasivos adicionales para el correcto funcionamiento del sistema, entre los que se destacan cuatro resistores, dos diodos y cuatro condensadores; sin embargo estos son elementos descritos por la hoja de datos del fabricante y no se considera relevante ahondar en ellos.

Finalmente, se cuenta con un regulador de voltaje LM7805, un switch de reset, un jack para la conexión del micrófono y una conexión para batería de 9V.

En las siguientes líneas se resaltan algunos datos particulares de los elementos más importantes de esta tarjeta (exceptuando el HM2007 que ya fue descrito):

- **Memoria HM6264A [35]:**
 - o Memoria RAM Estática de tecnología CMOS de 8-bits.
 - o Cuenta con una capacidad nominal de almacenamiento de 8 bits.
 - o Bajo consumo de potencia, típicamente de 0.1 mW.
 - o Tiempo de acceso máximo de 150ns.
 - o Fuente de alimentación de 5V.
 - o Compatibel con niveles TTL en todas las entradas y las salidas.

Figura 17. Diagrama de pines de la Memoria utilizada (HM6264A)



Fuente: Hoja de datos del fabricante

- **Latch 74LS373 [36]:**

- o Empaquetado que cuenta con 8 latches internos.
- o Interfaz de salida de 3 estados.
- o Histeresis en la Activación del Latch.
- o Diodos de control de alta corriente en las entradas.

En la tabla 3 se muestra la relación entre las salidas y las entradas de cada uno de los latch incluidos en el integrado.

Tabla 3. Tabla de verdad LS373

LS373			
D _n	LE	OE	O _n
H	H	L	H
L	H	L	L
X	L	L	Q ₀
X	X	H	Z*

H = HIGH Voltage Level

L = LOW Voltage Level

X = Immaterial

Z = High Impedance

Fuente: Hoja de datos del fabricante

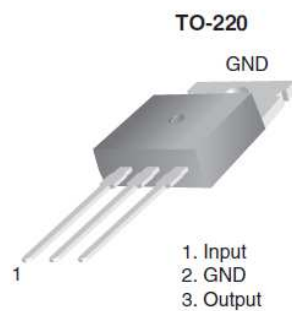
- **Regulador LM7805 [37]**

Regulador reconocido para la generación de señal constante de 5 V, sus principales características son:

- o Elemento de tres terminales: Entrada, Referencia y Salida.
- o Corriente de salida hasta 1 A
- o Voltaje de salida 5 V.
- o Tolerancia del 2% - 5% según modelo.

En la figura 18 se muestra el diagrama de pines de este integrado. Es importante destacar la importancia de utilizar condensadores a la entrada y la salida del mismo, para reducir la influencia del ruido eléctrico.

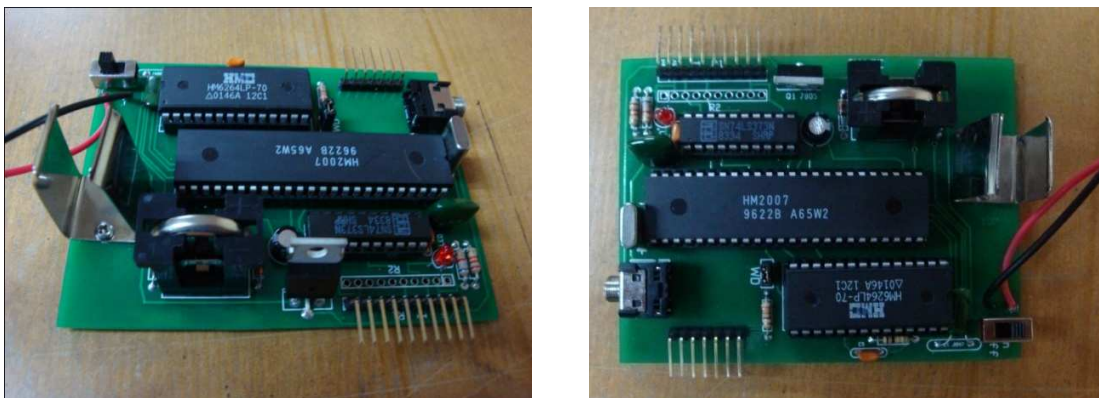
Figura 18. Diagrama de pines del regulador LM7805



Fuente: Hoja de datos del fabricante

Cabe resaltar que la tarjeta cuenta con una batería adicional para suplir como back-up a la memoria y evitar que se borren los patrones de las palabras almacenados cuando el circuito se apaga. En la figura 19 se ilustra la tarjeta reconocimiento, además el bosquejo del circuito impreso (PCB) es presentado en el Anexo 1-A del presente documento de trabajo de grado.

Figura 19. Tarjeta de reconocimiento de voz

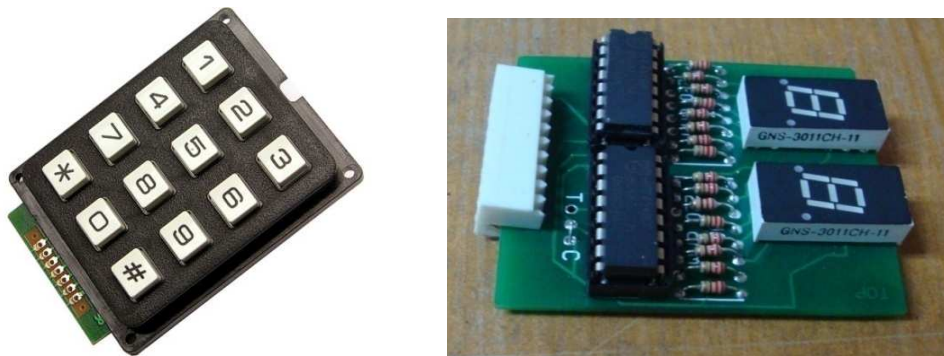


Fuente: Autor del Proyecto

3.2.3 Tarjeta de Teclado- Configuración de parámetros

Basado en las especificaciones de la hoja de datos del fabricante del HM2007, se adiciona un teclado para la configuración del mismo. Dicho elemento está construido con contactos sencillos, específicamente 12 switches normalmente abiertos dispuestos en un PCB convencional (Anexo 1-B). Lo anterior se evidencia en la figura 20.

Figura 20. Ejemplo de teclado para configuración



Fuente: Autor del Proyecto

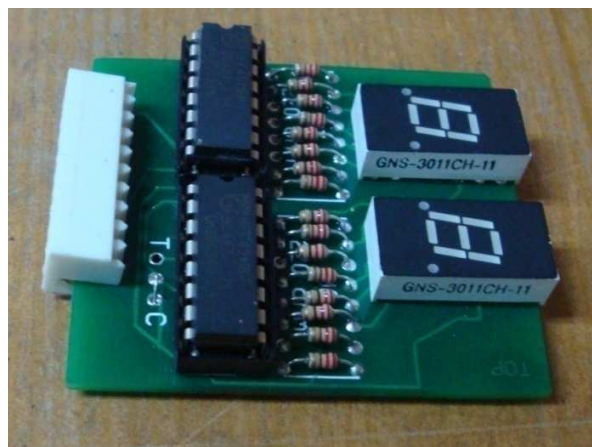
3.2.4 Tarjeta de Display o visualización

Esta tarjeta permite visualizar la configuración y almacenamiento de los datos. La idea es utilizar dos displays de siete segmentos convencionales para visualizar los datos que son enviados por el HM2007, tanto de configuración, como de almacenamiento y error. En las siguientes líneas se presenta la descripción de los principales elementos de esta tarjeta:

- **Display de 7 segmentos:** Este elemento permite la visualización de dígitos (valores de 0-9) que son transmitidos por un circuito integrado, en este caso por el HM2007. En la figura 21 se observa la tarjeta real con la

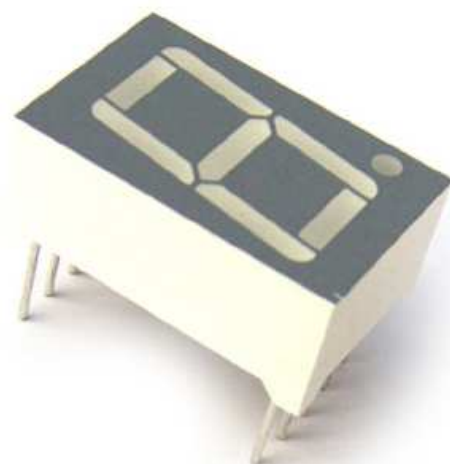
que se han trabajado las pruebas. En la literatura se encuentra displays de ánodo común y de cátodo común; donde su diferencia radica en la tensión de excitación para la iluminación (en un caso se activa con un valor de entrada alto y en el otro caso con un valor de entrada bajo) [38]. En la figura 22 se ilustra un ejemplo de dicho display.

Figura 21. Tarjeta de visualización - Displays



Fuente: Autor del Proyecto

Figura 22. Ejemplo de display de 7 segmentos



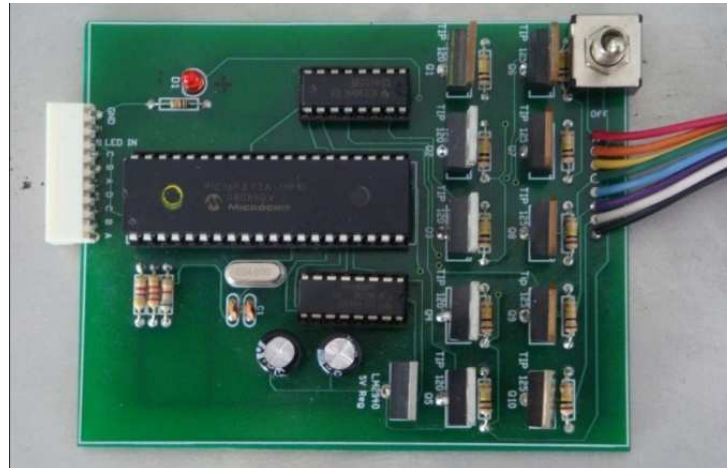
Fuente: Hoja de datos del fabricante

- **Circuito integrado 4511:** Este circuito se utiliza convencionalmente para realizar la decodificación de código BCD (utilizado por el integrado HM2007) a 7 segmentos. En ese sentido, su uso es meramente de adecuación de señal y no cumple ninguna función de procesamiento o reconocimiento de patrones [39]. Entre los detalles técnicos, se cuenta con un circuito integrado de 16 pines, tiene bajo consumo de potencia, pueden manejar hasta 25 mA de salida, tiene un pin de entrada para borrado entre otras características.

3.2.5 Tarjeta Microcontrolada para la manipulación de motores.

Aunque no hace parte directa del sistema de reconocimiento de voz, se ha adquirido una tarjeta adicional para el acople entre la etapa de reconocimiento y la manipulación de los motores del robot. Esta tarjeta hace parte de un kit que provee el fabricante del brazo robótico, y está vinculado directamente al sistema de reconocimiento del HM2007, así que el acople se convierte en algo relativamente sencillo de realizar. La tarjeta de control de motores se muestra en la figura 23, donde se destaca el set de transistores de potencia, el microcontrolador y las etapas de comunicación de datos

Figura 23. Tarjeta de control de motores (microcontrolada)

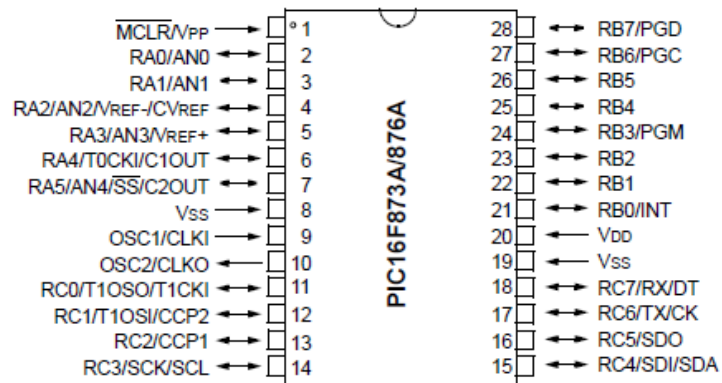


Fuente. Autor del Proyecto

De manera descriptiva, el sistema utiliza un microcontrolador central PIC16F877A, el cual actúa como receptor de la información suministrada por la tarjeta de reconocimiento y como controlador de las señales que son enviadas a los servomotores del brazo. Se cuenta con un conjunto de 10 transistores de potencia, cinco del tipo NPN TIP120 y cinco del tipo PNP TIP125; los cuales entregan la corriente necesaria para el control de los motores. En la figura 24 se ilustra el diagrama de pines del microcontrolador utilizado. Ahora bien, los principales elementos utilizados en esta tarjeta son:

- Microcontrolador PIC16F877A [40]:
 - o Frecuencia de operación de 20 MHz.
 - o Memoria RAM de 2.8 kB y 16kB para memoria Flash.
 - o Cuenta con temporizador de 8-16 bits.
 - o Módulos de comunicación SSP y SPI.
 - o Conversor Análogo a Digital A/D de 10 bits.
 - o Bajo consumo de potencia – Voltaje de operación de 3-5 V.

Figura 24. Diagrama de pines del microcontrolador PIC16F877A



Fuente: Hoja de datos del fabricante

- Transistor TIP120-TIP125 [41]:
 - o Transistor MOS de mediana potencia.
 - o Configuración NPN-PNP para gestión de la dirección de los motores DC del brazo robótico.
 - o Capacidad de manejo de corriente de colector de 5 A (suficientes para los 200 mA que consume cada servo-motor del robot).

En la figura 24 se ilustra el diagrama de pines del transistor TUP-120.

Figura 24. Diagrama de pines del transistor TIP120/TIP125



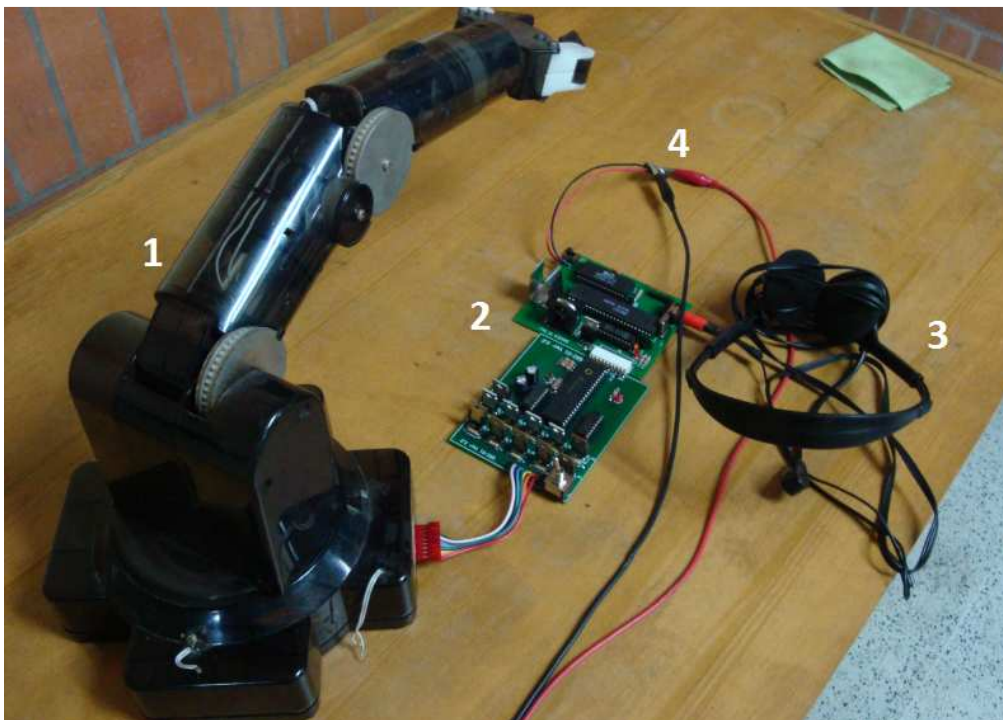
Fuente: Hoja de datos del fabricante

Hasta este punto se han descrito los elementos más importantes del hardware utilizado para el reconocimiento de voz, tanto en la adquisición, como en el procesamiento y control de las señales dirigidas al brazo robótico. En el siguiente capítulo se presentarán las pruebas realizadas al circuito, la metodología de funcionamiento y los resultados finales del presente ejercicio académico.

4. PRUEBAS DE FUNCIONAMIENTO DEL HARDWARE

En esta etapa final del documento, se describen las pruebas de funcionamiento realizadas con cada una de las tarjetas mostradas en el capítulo anterior. Es importante iniciar remitiéndose a la figura 25, donde se muestra el montaje total, en laboratorio, del sistema.

Figura 25. Montaje completo en laboratorio



Fuente: Autor del proyecto

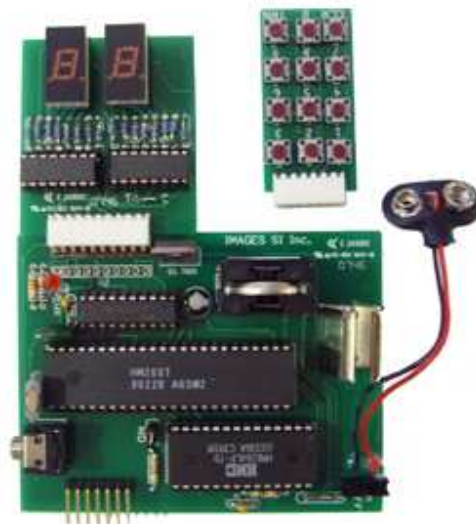
De la figura 25 se puede destacar: 1. Brazo robótico con 5 grados de libertad, 2. Tarjetas de reconocimiento de voz y control de motores, 3. Diadema con micrófono para pruebas (elemento de registro) y 4. Conector a fuente de 9 V (en este caso se está utilizando una fuente regulada de DC).

4.1 METODOLOGÍA DE CONFIGURACIÓN DEL SISTEMA

En esta sección se describe la metodología de configuración y uso para realizar la puesta en marcha del sistema. Los pasos recomendados son:

1. Se conectan el teclado y el display a la tarjeta de reconocimiento. Este es el **modo de entrenamiento**. En este punto se van a grabar las palabras para cada una de las acciones. El teclado cuenta con 12 teclas, la numeración del 1 hasta el 9, adicionalmente el 0. Se cuenta con las teclas CLR (Borrar) y TRN (Entrenar). Un ejemplo de este sistema se ilustra en la figura 26.

Figura 26. Modo de entrenamiento



Fuente: Autor del Proyecto

2. Para realizar el entrenamiento basta con digitar el número de la palabra que se desea programar, por ejemplo **01** y luego, al darle la tecla **TRN**; grabar la palabra que se desea asignar en esta posición. Es importante destacar que para realizar la grabación se requiere conectar el **micrófono**. El entrenamiento se puede realizar de manera continua hasta llegar al número **40**.
3. Después de grabar cada palabra es importante verificar la correcta operación de la misma; para esto basta hablarle a través del micrófono y comprobar que en el display aparece el valor registrado, El fabricante del HM2007 indica en su hoja de datos, que el chip provee los siguientes códigos de error:
 - Palabra muy larga = **55**.
 - Palabra muy corta = **66**
 - No hay concordancia con dato anterior = **77**
4. Para realizar la limpieza de la memoria basta con digitar el código 99 y posteriormente el botón CLR.

NOTA: El fabricante del HM2007 ha indicado en su hoja de datos que el sistema opera adecuadamente, tanto en modelo dependiente del hablante como independiente del hablante; sin embargo cuando se revisa con detalle el manual de instrucciones se recomienda que para que la operación **independiente del hablante** (la más compleja) funcione de manera adecuada es necesario grabar las palabras en posiciones de memoria alejadas. Por ejemplo grabar en 1, 5, 10, 15, entre otros.

5. Para el ejercicio académico realizado se ha grabado un set de palabras en español para ejecutar las tareas de manera eficiente. Esta configuración se puede realizar, en teoría, con cualquier palabra; sin

embargo el autor del presente documentno recomienda vocablos con sonidos diferentes. Las palabras registradas son mostradas en la tabla 3.

Tabla 3. Palabras configuradas en las pruebas

Núm. Registrado	Función	Palabra grabada
1	Cierre de la Pinza	Agarrar
2	Descenso del “Codo”	Bajar
3	Giro de la base a izquierda - CCW ¹⁹	Izquierda
4	Subir “Hombro”	Elevar
5	Giro CCW “Muñeca”	Rotar
6	Giro CW ²⁰ “Muñeca”	Giro
7	Descenso del “Hombro”	Caer
8	Giro de la base a derecha – CW	Derecha
9	Subir “Codo”	Subir
10	Abrir Pinza	Abrir
11	Detener	Stop

Fuente: Autor del Proyecto

¹⁹ Acrónimo de Counter ClockWise → Giro en sentido opuesto a las manecillas del reloj.

²⁰ Acrónimo de ClockWise → giro en sentido de las manecillas del reloj.

4.2 PRUEBAS ELÉCTRICAS Y DE FUNCIONAMIENTO

Para verificar el correcto comportamiento eléctrico hay que verificar cada una de las etapas del sistema. Las principales recomendaciones o pruebas para comprobar funcionamiento eléctrico son:

- Tarjeta de teclado: Verificar que el teclado este correctamente conectado y que los switches que lo conforman no se encuentren en corto. Se puede utilizar el multímetro (midiendo continuidad) para verificar este funcionamiento.
- Tarjeta de Display: Importante elegir de manera adecuada los displays de siete segmentos de cátodo común. También se puede verificar a través de la aplicación de diferenciales de tensión en los terminales de prueba para comprobar encendido de cada uno de los ocho leds (7 del número y 1 del punto).
- Tarjeta de Reconocimiento: Se han adicionado leds para verificar funcionamiento del HM2007. El led indicador permite detectar cuando graba, cuando almacena y cuando recibe una señal de voz. Es importante considerar el uso y buen estado de la batería de Ión para poder garantizar el almacenamiento de la información.
- Tarjeta de Control de Motores: Esta tarjeta provista por el fabricante puede ser evaluada de manera sencilla conectando la tarjeta de reconocimiento a la entrada. El sistema emite señales bipolares referidas a tierra, una gestionada por los transistores NPN y otros transistores PNP. Estas señales se pueden registrar sin ningún problema a través de los terminales de la tarjeta y un multímetro.

4.2.1 Pruebas de funcionamiento con el mismo hablante

Se realiza la verificación, en varios experimentos, de la eficiencia del sistema. Para esto se ejecutan barridos con diferente cantidad de palabras, y con cada una se ejecutan 15 pruebas. Estas pruebas se realizan con palabras **fonéticamente muy diferentes**, con terminaciones y longitudes diferentes, obteniendo los resultados verificables en la tabla 4 los resultados.

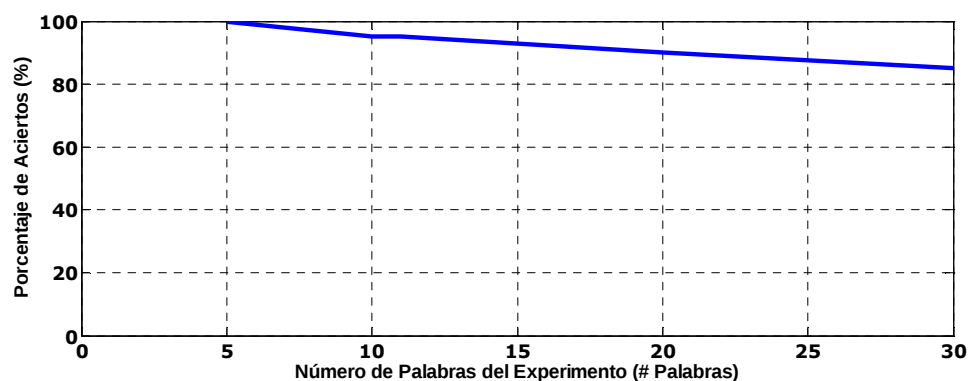
Tabla 4. Pruebas con palabras considerablemente diferentes (fonética) – 20 intentos

Número de palabras	Aciertos	Aciertos [%]
5	20	100
10	19	95
11	19	95
20	18	90
30	17	85

Fuente: Autores del Proyecto

En la figura 27 se ilustra los porcentajes de error en estas pruebas.

Figura 27. Prueba de eficiencia con palabras fonéticamente diferentes



Fuente: Autor del Proyecto

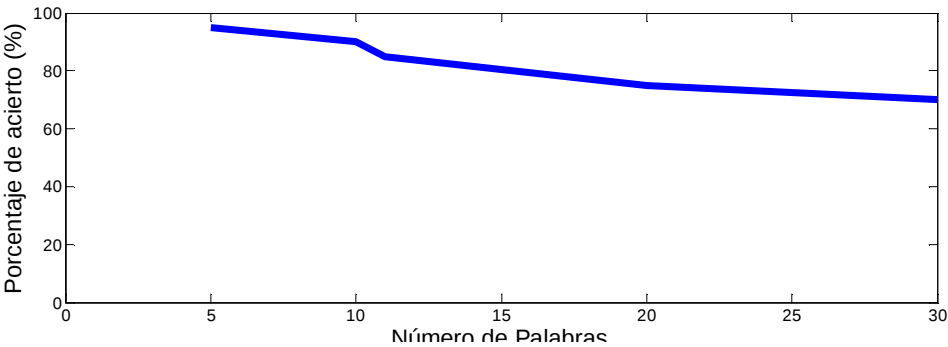
En la siguiente prueba se realiza una experimentación con 20 intentos, pero esta vez se han grabado 20 palabras relativamente similares. Se utilizaron verbos con terminaciones similares como: bajar, rotar, girar, elevar, entre otros; los datos tabulados de dicha prueba se presenta en la tabla 5 y se grafican en la figura 28.

Tabla 5. Pruebas con palabras considerablemente similares (fonética) – 20 intentos

Número de palabras	Aciertos	Aciertos [%]
5	19	95
10	17	85
11	16	80
20	16	80
30	15	75

Fuente: Autor del Proyecto

Figura 28. Prueba de eficiencia con palabras fonéticamente similares



Fuente: Autor del Proyecto

En la figura 28 es evidente el descenso del porcentaje de acierto en más del 10%, sin embargo para el caso de interés (11 palabras) se tiene un porcentaje

cerca al 85%, que sin ser una excelente referencia, tiene un funcionamiento aceptable.

4.2.2 Pruebas de funcionamiento con diferente hablante

En esta sección se presenta el resultado de probar el funcionamiento del sistema con once (11) palabras y tres hablantes. La programación es realizada por el sujeto A (hombre) y se ejecutan 20 pruebas por tres (3) sujetos mas, B (hombre), C(mujer) y D (mujer). En la tabla 6 se muestra la síntesis de los resultados.

Tabla 6. Eficiencia en el registro de resultados de cuatro (4) personas.

PERSONA	ACIERTOS	Aciertos [%]
A	19	95
B	17	85
C	16	80
D	15	75

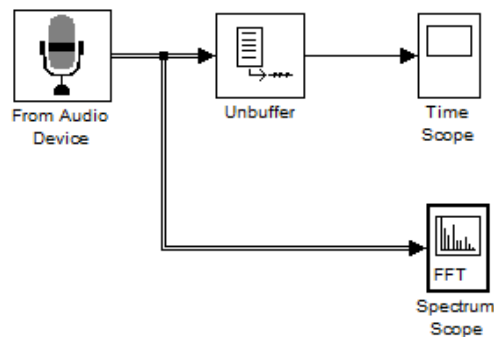
Fuente: Autor del Proyecto

Se observa que existe una dependencia del hablante considerable, sobre todo en las voces de los hablantes femeninos. Se recomienda reprogramar el hardware HM2007 si existe un cambio de género.

Ahora bien, la robustez del sistema está fuertemente ligada a la capacidad de procesamiento del hardware de reconocimiento, así como al nivel de complejidad y eficiencia de los algoritmos utilizados. Se ha realizado el registro de voz de los cuatro hablantes comparados a través de la herramienta MATLAB®, recurriendo

a la caja de herramientas Signal Processing Blockset²¹; esto para examinar las características frecuencial de la voz de cada uno de los participantes de la prueba. Para dicho objetivo se implemento, en Simulink, el sistema mostrado en la figura 28.

Figura 28. Sistema de grabación y visualización de señales



Fuente: Autor del proyecto

El modelo implementado está compuesto de cuatro bloques principalmente, los cual son:

- From Audio Device: Este bloque permite grabar, a partir de la tarjeta de sonido, una señal de audio proveniente de una fuente externa. En el presente ejemplo se ha registrado a través de un micrófono convencional la señal de interés.
- Unbuffer: Este bloque permite la gestión del array de datos proveniente del primer proceso, esto para permitir visualizar la señal en la siguiente etapa de forma correcta.

²¹ Caja de herramientas de Procesamiento de Señales, la cual permite adquirir, procesar y visualizar señales; permitiendo obtener características frecuencia-temporales, filtrado de señal, entre otros.

- Time Scope: Visualizador de la señal temporal. Este bloque se actualiza automáticamente a medida que se va realizando el registro.
- Spectrum Scope – FFT: Bloque principal del análisis. Dicho bloque permite visualizar la respuesta en frecuencia de la señal grabada, recurriendo al reconocido algoritmo de Transformada Rápida de Fourier, el cual ejecuta la DFT (Transformada Discreta de Fourier).

Al realizar estas mediciones se observa que los sujetos masculino, al menos los verificados, generan una señal de mayor potencia al realizar el registro, así como un ancho de banda en frecuencias más reducido que los sujetos femeninos. En la figura 29 se ilustra cada uno de los registros. Estos resultados permiten estimar, así como las pruebas sobre el hardware, que es altamente recomendable trabajar en modalidad dependiente del hablante, para de esta forma reducir los problemas de fallos en la manipulación del brazo robótico.

4.2.3 Ejecución de prueba con desplazamiento de objeto

Esta prueba se ha realizado de forma cualitativa, buscando probar el correcto funcionamiento del brazo robótico a través de los comandos de voz. Para esto se ha decidió utilizar una botella plástica para ser desplazada por el brazo, combinando las opciones y utilizando únicamente ordenes de voz previamente configuradas. En la figura 30 se muestran varias imágenes de la ejecución de este procedimiento, el cual es muy ilustrativo y genera un resultado evidenciable del funcionamiento de todo el sistema.

Figura 29. Imágenes de experimento de desplazamiento de botella plástica



Fuente: Autor del proyecto.

CONCLUSIONES

Luego de una revisión exhaustiva del estado del arte y de colocar a prueba uno de los diversos sistemas de Reconocimiento Automático de Voz por hardware, el autor del presente proyecto concluye que:

- La complejidad de los sistemas de reconocimiento de voz se encuentra asociada a varios parámetros propios del proceso, entre los que se destacan el número de palabras a procesar, la dependencia o independencia del hablante, la aplicación del sistema en ambientes con alto ruido sonoro, entre otros. Cada uno de estos parámetros limita la funcionalidad de los sistemas convencionales y exige el uso de metodologías más eficientes.
- La comparación de patrones de voz, los modelos ocultos de Markov y las Redes Neuronales Artificiales; han sido las metodologías de reconocimiento de habla más utilizadas en los proyectos que integran esta estrategia de control. La complejidad en los algoritmos que emplean cada una de estas metodologías permiten generar sistemas de mayor robustez, rechazo al ruido y capacidad de vocablos.
- Esta claramente diferenciada la solución del problema meramente resuelta con dispositivos hardware y aquella que integra estos dispositivos con herramientas software. Los proyectos revisados muestran desde desarrollo de reconocimiento básico hasta aplicativos que permiten apoyar diagnósticos médicos complejos.
- El integrado HM2007, elegido en este proyecto como base para el reconocimiento, presenta características de bajo consumo de potencia, baja latencia en la conversión y un costo reducido. Cabe resaltar que si se desea desarrollar una aplicación compleja que reconozca más de 50

palabras y que tenga buenas cualidades de filtrado, el integrado utilizado no es la mejor opción, y se recomienda el uso de integrados con mayor capacidad de procesamiento, memoria y calidad de filtrado de la señal.

- Aunque el sistema desarrollado en el presente proyecto se fundamenta en un circuito que tiene características de poder manejar sistemas independientes del hablante, el entrenamiento del mismo llega a ser complejo, sobre todo si lo trata de utilizar una persona con tonos y timbres de voz diferente. Se lograron obtener aciertos superiores al 90% si el hablante es el mismo que entrenó el sistema y el que ejecuta el control del robot.

RECOMENDACIONES

Entre las principales recomendaciones que realiza el autor de este proyecto a los futuros trabajos de investigación y desarrollo en el tema del Reconocimiento Automático de Voz son:

1. El circuito integrado HM2007 ha mostrado un comportamiento eficiente y su configuración es sencilla; sin embargo, si se desea trabajar un sistema con más de 3 hablantes y con un set de palabras considerables, es necesario recurrir a dispositivos de mayor capacidad, mayor velocidad y con mejores algoritmos de reconocimiento.
2. El apoyo del sistema hardware con herramientas de visualización en software es una interesante idea para una segunda fase, es decir, se puede integrar el hardware desarrollado con una solución de apoyo, tanto en el procesamiento como en la visualización, en un equipo de computo convencional.

BIBLIOGRAFIA

- [1] Intel Co. La innovación de Intel sigue convirtiendo en realidad la Ley de Moore. [en línea]. <http://www.intel.com/cd/corporate/techtrends/EMEA/spa/209840.htm> [citado el 12 de Enero de 2010].
- [2] RAMIREZ, José David. GRAJALES, Carlos Andrés. La robótica en la medicina. Universidad de Manizales. Enero de 2003. [en línea]. <http://www.umanizales.edu.co/programs/ingenieria/ventana/ventana11/RoboticaenMedicina.pdf> [citado el 12 de Enero de 2010].
- [3] PREHN, Ricardo. Xplorer- Detector de minas antipersonales. EUSS- Escola Universitaria Salesiana de Sarriá. 2005. [en línea]. <http://www.depeca.uah.es/alcabot/alcabot2002/documentos/ponencias/lib001.pdf> [citado el 15 de Enero de 2010].
- [4] Sistemas de Locomoción de robots móviles. Automatización y Robótica Industrial. [en línea]. http://www.esi2.us.es/~vivas/ayr2iaei/LOC_MOV.pdf . [citado el 15 de Enero de 2010].
- [5] COLOMBIA ACCESIBLE. Colombia no está acondicionada para facilitar la movilidad de las personas con discapacidad [en línea]. www.ladisapacidad.com/colombiaaccesible/colombia_no_esta_acondicionada_para_facilitar_la_movilidad_d.html [citado el 12 Enero de 2010].
- [6] WIKIPEDIA. Reconocimiento Automático de Habla. [en línea]. http://es.wikipedia.org/wiki/Reconocimiento_del_habla . [citado el 15 de diciembre de 2009].
- [7] OROPEZA, José Luis. Reconocimiento de voz aplicado a la robótica. Centro de Innovación y Desarrollo Tecnológico en Computación. Instituto Politécnico Nacional. Mexico. 2006.
- [8] THOMAS, William. Remote Control of a Robotic Arm Trough speaker-dependent, isolated-word speech recognition. UMI. Estados Unidos. 2001
- [9] SAMPER, Juan José. Introducción a los Sistemas Expertos [en línea]. Red científica/Ciencia, Tecnología y Pensamiento. <http://www.redcientifica.com/doc/doc199908210001.html>. [citado el 12 de Enero de 2010].

- [10] GRUPO PAS. Reconocimiento de la voz [en línea]. reve Historia de las redes Neuronales [en línea]. Universidad de Málaga. Universidad DEUSTO. Mayo 2006. <http://www.redcientifica.com/doc/doc199908210001.html>. [citado el 14 de Febrero de 2010].
- [11] Neural Network Framework. Breve Historia de las redes Neuronales [en línea]. Universidad de Málaga. <http://www.redes-neuronales.netfirms.com/tutorial-redes-neuronales/historia-de-las-redes-neuronales.htm> . [citado el 12 de Enero de 2010].
- [12] Reconocimiento de la voz [en línea]. reve Historia de las redes Neuronales [en línea]. Universidad de Málaga. Universidad DEUSTO. Mayo 2006. <http://www.redcientifica.com/doc/doc199908210001.html>. [citado el 14 de Febrero de 2010].
- [13] RIVORA, S y TORASSO, P. An isolated-word recognizer based on grammar-controlled classification processes. CENS, Politécnico de Torino. Universidad e Torino. Revista Pattern Recognition, Volume 10, Issue 2, Pages 73-84, 1978.
- [14] CHOUKIR, K. y CHOLLET, G. Adaptation of automatic speech recognizers to new speakers using canonical correlation analysis techniques Computer Speech & Language, Volume 1, Issue 2, , Pages 95-107, 1986.
- [15] ANGUS, james y WHITAKER, Martin. Low-cost speech recognizer. Microprocessors and Microsystems, Volume 11, Issue 2, March 1987, Pages 94-98
- [16] GONG, Yifan. Speech recognition in noisy environments: A survey. CRIN/CNRS- INRIA-Lorraine,France. Speech Communication, Volume 16, Issue 3, April 1995, Pages 261-291.
- [17] YAO. Kaisheng, PALIWAL. Kuldeep. Noise adaptive speech recognition based on sequential noise parameter estimation. ATR Spoken Language Translation Research Labs, Kyoto, Japan. Speech Communication, Volume 42, Issue 1, January 2004, Pages 5-23.
- [18] HUO, Qiang. HUI, Chin. Robust speech recognition based on adaptative classification and decision strategies. *Speech Communication*, Volume 34, Issues 1-2, April 2001, Pages 175-194.
- [20] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvett, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi and C. Wellekens. Automatic Speech Recognition and Speech Variability: A Review. *Speech Communication*, Volume 49, Issues 10-11, October-November 2007

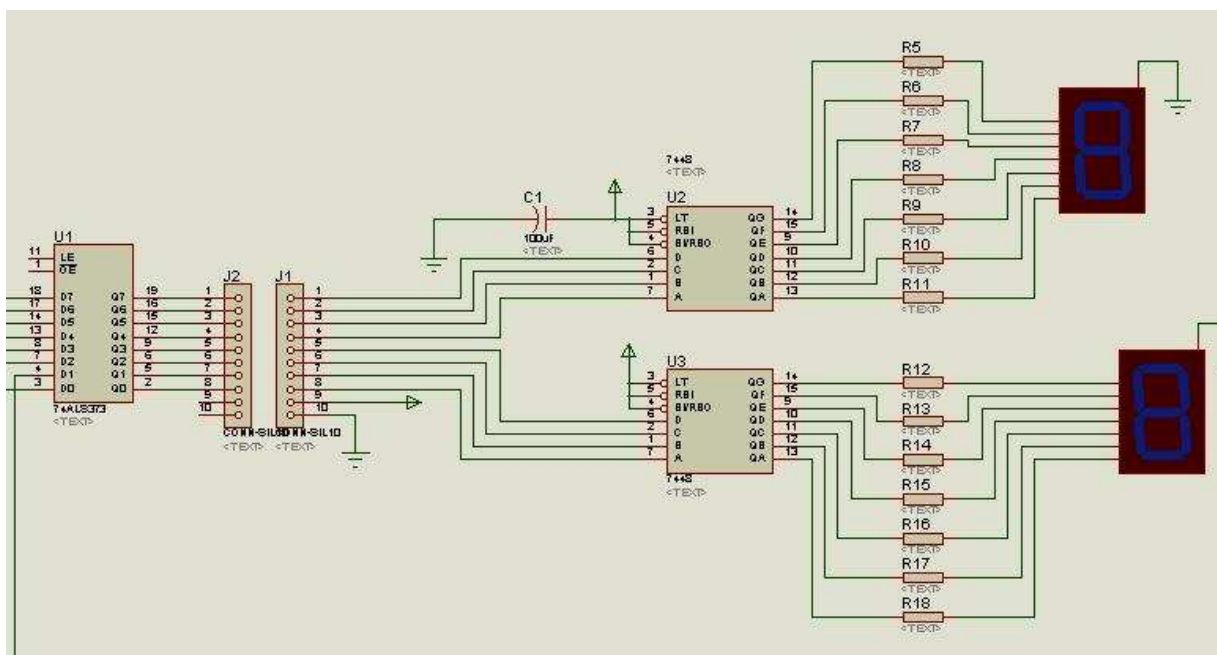
- [21] FLYNN, Ronan; Jones, Edward. Combined speech enhancement and auditory modelling for robust distributed speech recognition. *Speech Communication*, Volume 50, Issue 10, October 2008, Pages 797-809
- [22] Peng Li, Yong Guan, Shijin Wang , Bo Xu, and Wenju Liu. Monaural speech separation based on MAXVQ and CASA for robust speech recognition. *Computer Speech & Language*, Volume 24, Issue 1, January 2010, Pages 30-44
- [23] Grupo Médico Otológico. Implante Coclear. [en línea]. http://www.susmedicos.com/articulos_otologia_coclear1.htm. 2005. [citado el 16 de Enero de 2010]
- [24] D'HARO Luis Fernando, Desarrollo de un clasificador Bayesiano de ruido y voz mediante estimación por máxima verisimilitud. http://www-gth.die.upm.es/~lfdharo/Papers/Clasificador_STSIVA_2002.pdf. [citado el 16 de Enero de 2010]
- [25] Politécnico Colombiano Jaime Isaza Cadavid, (2006). "Robot controlado por voz fue presentado en el politécnico Jaime Isaza Cadavid" Red de Universidades-Universia, 2007. [en línea] <http://www.universia.net.co/investigacion/destacado/robot-controlado-por-voz-fue-presentado-por-el-politecnico-jaime-isaza-cadavid.html>. [citado el 16 de Enero de 2010]
- [26] PAEZ, John Jairo. Sistema inteligente de reconocimiento de voz para la traducción del lenguaje verbal a la lengua de señas. Universidad Politécnica Nacional. [en línea]. <http://www.ribiecol.org/nueve/ponencias/35.pdf>. [citado el 16 de Enero de 2010]
- [27] PARDO Garcia Aldo, RUBIANO Calderon Zolanyi Marcela, MONTAGURT Ferizzola Yeison Javier. Automatización de una silla de ruedas, controlada por comandos de voz. Universidad de Pamplona.[en línea] . http://www.upbmonteria.edu.co/electronica/gapa/descargas/Tercer%20Congreso%20Unipamplona/automatizacion_de_una_silla_de_ruedas_controlada_por.pdf
- [28] ATKINSON, William Thomas. Remote control of robotics amr through speaker-dependent, isolated-word speech recognition. UMI (University Microfilms International)- A Beel & Howell Information Company. 2001
- [29] PARDO, Jose D. CASTRO, Jose. Mattos, Lorenzo. Reconocimiento automático del habla utilizando la transformada de Fourier y Redes Neuronales. Universidad Popular del Cesar. Noviembre 20 de 2006.

- [30] HMC Semiconductors. HM2007 Speech Recognition. Hoja de Datos. [en línea]. http://www.datasheetcatalog.org/datasheets/2300/499674_DS.pdf. [citado: 25 de Febrero de 2010].
- [31] OKI Semiconductors. MSM6679 Voice Recognition Processor. Hoja de Datos. [en línea]. <http://datasheet.digchip.com/000/000-1-MSM6679.pdf> . [citado: 25 de Febrero de 2010].
- [33] INTERACTIVE Speech. RSC-164. Hoja de Datos [en línea]. <http://www.imagesco.com/speech/rsc-164.pdf> . [citado: 25 de Febrero de 2010]
- [34] OWI007 - Robotic Arm Trainer Kit. [en línea]. <http://www.electronickits.com/robot/owi007.htm>. [citado: 10 de Marzo de 2010].
- [35] HM6264A Series – Memoria de almacenamiento. Hoja de Datos [en línea]. <http://faculty.kfupm.edu.sa/COE/masud/RichText/RAM6264.pdf>. [citado: 10 de Marzo de 2010].
- [36] MOTOROLA. Latch 74LS373. Hoja de Datos. [en línea]. <http://ecee.colorado.edu/~mcclurel/sn74ls373rev5.pdf>. [citado: 10 de Marzo de 2010].
- [37] FAIRCHILD Semiconductor. Regulador LM7805. Hoja de Datos [en línea]. <http://www.fairchildsemi.com/ds/LM/LM7805.pdf> . [citado: 10 de Marzo de 2010].
- [38] Electrónica Umicron. Display de 7 segmentos. [En línea] http://www.unicrom.com/Tut_display-7-segmentos.asp. [citado: 10 de Marzo de 2010]
- [39] TEXAS Instruments. CD4511B (BCD-7 Segmentos). [En línea]. <http://focus.ti.com/lit/ds/symlink/cd4511b.pdf>. [citado: 10 de Marzo de 2010]
- [40] MICROCHIP. PIC16F87XA (Microcontrolador). [En línea]. <http://ww1.microchip.com/downloads/en/DeviceDoc/39582b.pdf> . [citado: 10 de Marzo de 2010]

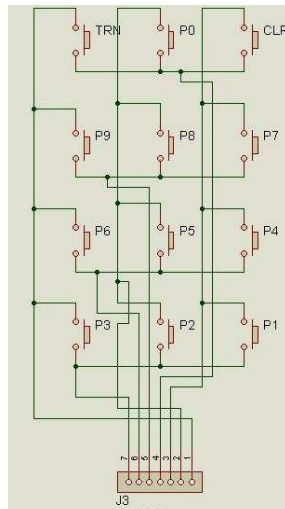
ANEXOS

ANEXO A – Circuitos esquemáticos en herramienta Proteus

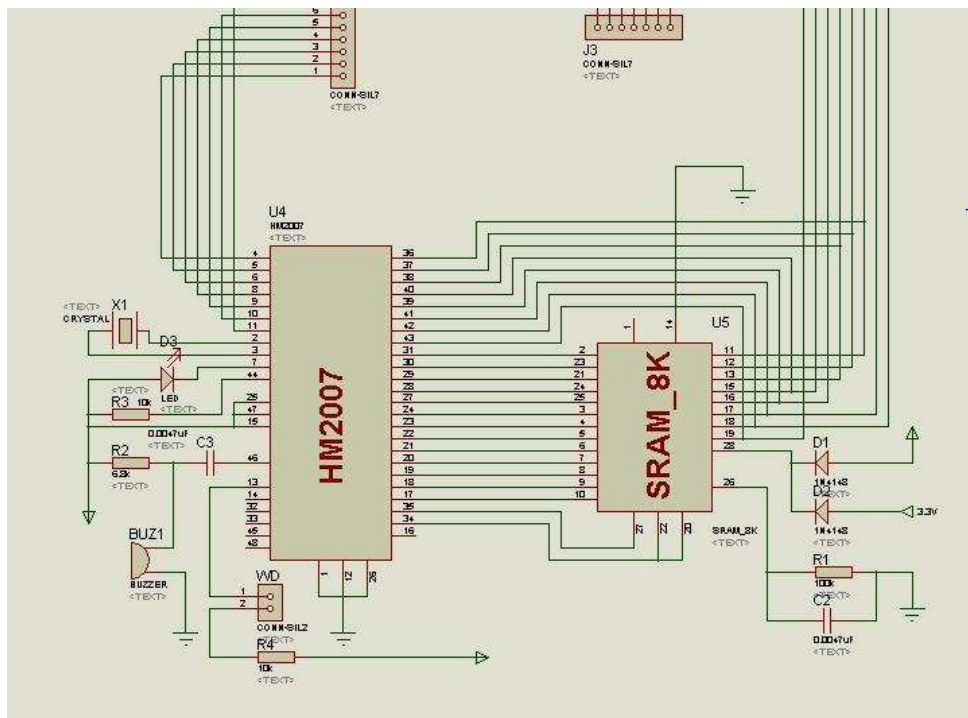
Circuito Impreso 1- Tarjeta de Display (Visualización)



Circuito Impreso 2 – Tarjeta de Teclado



Circuito Impreso 3 – Tarjeta de Reconocimiento



ANEXO B- SINTESIS DE HOJAS DE DATOS DEL FABRICANTE

B.1 HM2007 – Sistema para reconocimiento de voz



HM2007
SPEECH RECOGNITION

GENERAL DESCRIPTION

HM2007 is a single chip CMOS voice recognition LSI circuit with the on-chip analog front end, voice analysis, recognition process and system control functions. A 40 isolated-word voice recognition system can be composed of external microphone, keyboard, 64K SRAM and some other components. Combined with the microprocessor, an intelligent recognition system can be built.

FEATURES

- Single chip voice recognition CMOS LSI.
- Speaker-dependent isolated-word recognition system.
- External 64K SRAM can be connected directly.
- Maximum 40 words can be recognized for one chip.
- Maximum 1.92 sec of word can be recognized.
- Multiple-chip configuration is possible.
- A microphone can be connected directly.
- Two control mode is supported: Manual mode and CPU mode.
- Response time : less than 300 ms.
- 5V single power supply.
- 48-pin PDIP, 52 pin PLCC, 48 pad bare chip.

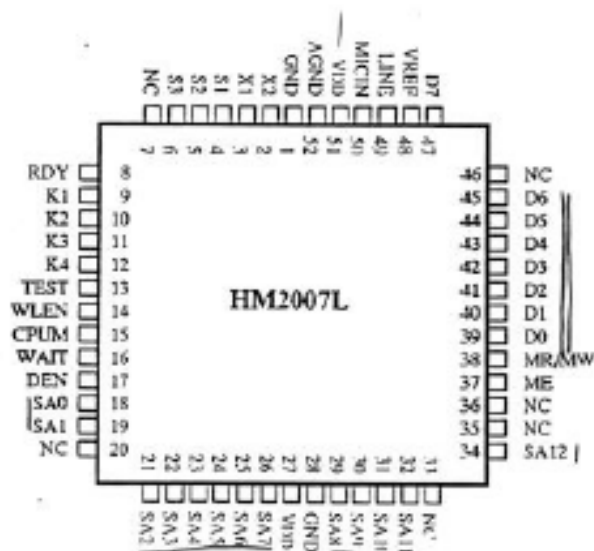


HM2007
SPEECH RECOGNITION

PIN CONFIGURATIONS

HM2007P

GND	1	48	AGND
X2	2	47	VDD
X1	3	46	MICIN
S1	4	45	LINE
S2	5	44	VREF
S3	6	43	D7
RDY	7	42	D6
K1	8	41	D5
K2	9	40	D4
K3	10	39	D3
K4	11	38	D2
TEST	12	37	D1
WLEN	13	36	D0
CPUM	14	35	MR/MW
WAIT	15	34	ME
DEN	16	33	NC
SA0	17	32	NC
SA1	18	31	SA12
SA2	19	30	SA11
SA3	20	29	SA10
SA4	21	28	SA9
SA5	22	27	SA8
SA6	23	26	GND
SA7	24	25	VDD



PIN DESCRIPTIONS

Symbol	Pin No.		I/O	Function
	PDIP	PLCC		
	48L	52L		
Vref	44	48	I	The voltage reference input of internal ADC. Supply the reference voltage of the internal A/D converter.
LINE	45	49	O	For testing only.
MICIN	46	50	I	Microphone connect pin. A microphone should be connected via a coupling capacitor and resistor.
V _{DD}	47	51		Positive power supply.
AGND	48	52		Analog Ground.
GND	1	1		Negative power supply.
X2,X1	2,3	2,3	I	Crystal connect pin. A 3.58 MHz crystal is connected to these pin.
S1,S2	4,5	4,5	I/O	Keypad scanning pin for manual mode and the read/write control Pins in the CPU mode.
S3	6	6		
RDY	7	8	O	Voice input ready indicator. Active low output. When HM2007 is ready for the voice input in training or recognition mode, a low signal is sent. If the chip is busy, a high signal is sent.
K1,K2	8-11	9-12	I/O	The keypad input pin in the manual mode and the bidirectional data bus (K-bus) in the CPU mode. In the manual mode, the four pins combined with S1 to S3 form the keypad scanning circuit. Maximum 12 keys can be scanned. In the CPU mode, the data bus direction is determined by the S2 and S3. A high level signal that appears in the pin S2 will place the content of internal register onto to the data bus (K-bus). The data may be come from the status register or the output buffer which is selected by the pin S1. If S1 is high, output buffer is selected, otherwise, the status register is selected. A high level signal that appears in the pin S3 will place the content of K-bus into the input register. Note that user can not place high level signal on S2 and S3 simultaneously.
K3,K4				
TEST	12	13	I	"H": test mode. "L": Normal mode.
WLEN	13	14	I	Word length select pin. Selecting the voice length to be recognized. When set to high, 1.92 sec is selected. Internally pull low for 0.9 sec is selected. Note that when 1.92 sec is selected, only 20 words maximum can be recognized if 8K-byte memory is used.
CPUM	14	15	I	CPU mode select pin. Internally pull low for manual mode. When set to high, CPU mode is selected.

Symbol	Pin No.		I/O	Function
	PDIP	PLCC		
	48L	52L		
WAIT	15	16	I	Waiting control input. Active low input. When this pin is set to "L" and manual mode is selected, HM2007 will enter the waiting state and do not accept voice input until this pin back to "H". For CPU mode, when HM2007 is ready to get voice input, if this pin is set to "L", HM2007 will skip the voice input process and enter the get-command process.
DEN	16	17	O	Data enable signal. When the recognition or training process is complete, the chip will place its response on the data bus D0 to D7 and which can be latched onto external devices by this pin.
SA0-SA1	17-24	18,19	O	External memory address bus. The bus is used as an external memory address when ME pin is active.
SA2-SA7		21-26		
SA8-SA11	27-31	29-32	O	Positive power supply.
SA12		34		
V _{DD}	25	27		
GND	26	28	O	Negative power supply.
NC	32,33	35,36		
		7,20		
		33,46	O	Memory enable pin. Active low output. This pin will send the memory enable signal to the external SRAM. This pin can be connected directly to the CE pin of 6264 SRAM.
ME	34	37		
MR/MW	35	38	O	Memory read/write select pin. Read/write control signal of the external SRAM. This pin can be connected directly to the R/W pin of 6264 SRAM.
D0-D6	36-42	39-45	I/O	External memory data bus(D-bus). The bus is used as an external memory I/O bus when ME pin is active and used as output response bus when DEN pin is active.
D7	43	47		

FUNCTION DESCRIPTIONS

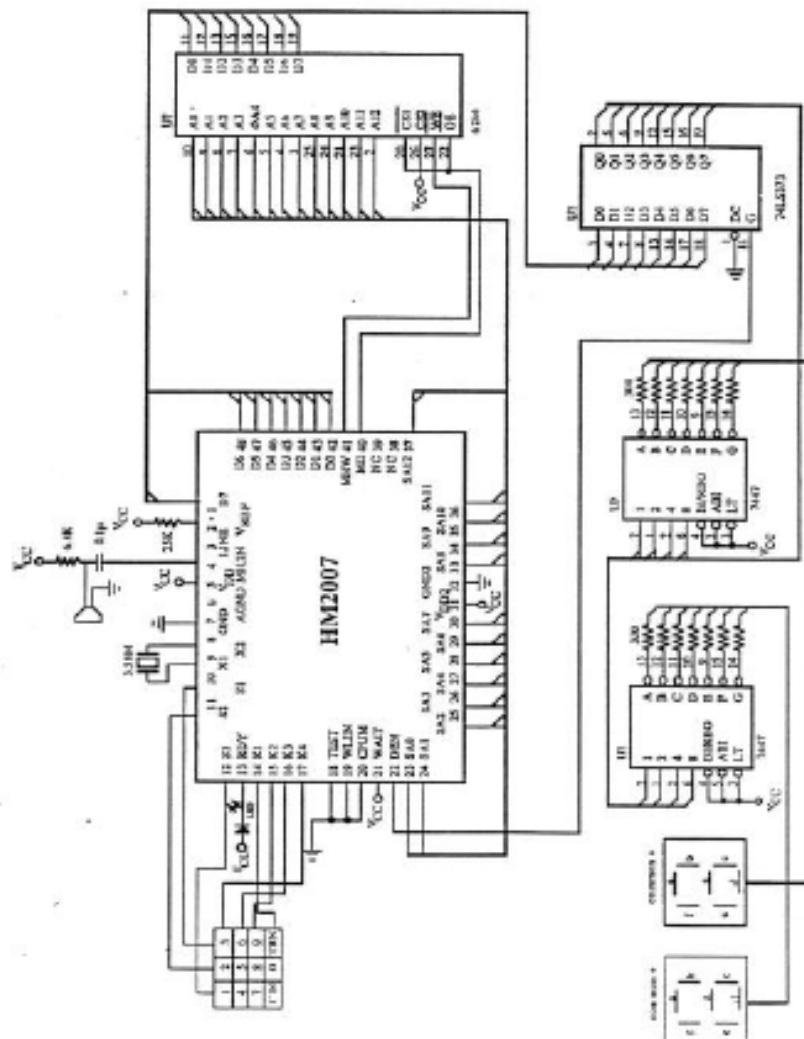
There are two operation mode which are provided by HM2007.

A). Manual mode.

In this operation mode, a keypad, a SRAM and other components may be connected to HM2007 to build a simple recognition system (See application circuit).
The type of SRAM can be used is a 8K-byte memory.

APPLICATION CIRCUIT

A). Manual mode



B.2 MSM669 – Procesador para reconocimiento de voz

OKI Semiconductor

MSM6679 Voice Recognition Processor

SI/SD Voice Recognizer, Recorder/Player, and Speech Synthesizer

DESCRIPTION

The MSM6679-XXXJS Voice Recognition Processor (VRP) is a slave-mode device that performs five functions: speaker-independent (SI) voice recognition, speaker-dependent (SD) voice recognition, solid-state sound recording, sound playback, and speech synthesis. The highly integrated device also provides an on-chip memory controller, Flash memory interface, analog data conversion, OKI speech synthesizer interface, and PWM sound output.

For SI recognition, the MSM6679 contains a vocabulary template in on-chip or external memory. Pre-trained SI vocabularies eliminate the need for laborious training, as usually required by SD products. The memory requirements are dependent on the size of the vocabulary. The MSM6679 can tolerate background noise, while providing high recognition accuracy. In its designated operating environment, the device achieves a typical recognition accuracy of >95% (using an OKI-defined test procedure).

For SD recognition, the MSM6679 stores speaker-dependent vocabulary templates, as defined by the user, in external SRAM. The MSM6679 can create SD vocabularies of up to 61 words each, with each word using approximately 50 bytes.

As well as providing voice recognition capabilities, the MSM6679 integrates a solid-state recorder/player, speech synthesis functions, and a tone generator. ADPCM recording/playback provides high quality sound and efficient memory utilization. The MSM6679 can respond to spoken commands, verbally or with tones, via an on-chip speech synthesizer and tone generator. For larger speech-synthesis requirements, the MSM6679 also provides a glueless MSM665x control interface for off-chip speech synthesis.

The MSM6679 can interface to any application or personal computer via a parallel or serial interface through an open, device-independent serial mode API (SMAPI). To accelerate code development, OKI supplies an evaluation kit, and assembly and C language programs for this product.

FEATURES

- Speaker-independent recognition
 - Up to 20~25 words in each vocabulary
 - Multiple vocabulary support
- Speaker-dependent recognition
 - Up to 61 words in each vocabulary
 - Multiple vocabulary support
- Speech synthesis
 - Up to 2.3sec internal and 27.6sec external speech synthesis on-chip; sample looping and concatenation allows even longer phrases.
 - On-chip controller for MSM665x speech synthesizer
 - Standard beep tone outputs
 - PCM and ADPCM voice or sound-effect output
- Speech capture and playback
 - 28kbps ADPCM speech compression
- Serial ASCII and parallel command interface
- 6944Hz audio input sample rate for record and playback
- 10kHz sample rate for voice recognition
- 200msec recognition latency
- Flexible memory mapping for EPROM, FLASH, and SRAM
- 32MHz operation
- Package: 84-pin PLCC (QFJ84-P-S115)

FUNCTIONAL AND I/O DIAGRAMS

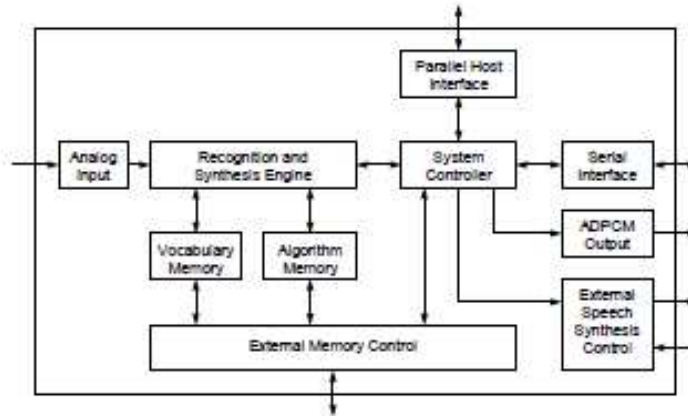


Figure 1. MSM6679 Block Diagram

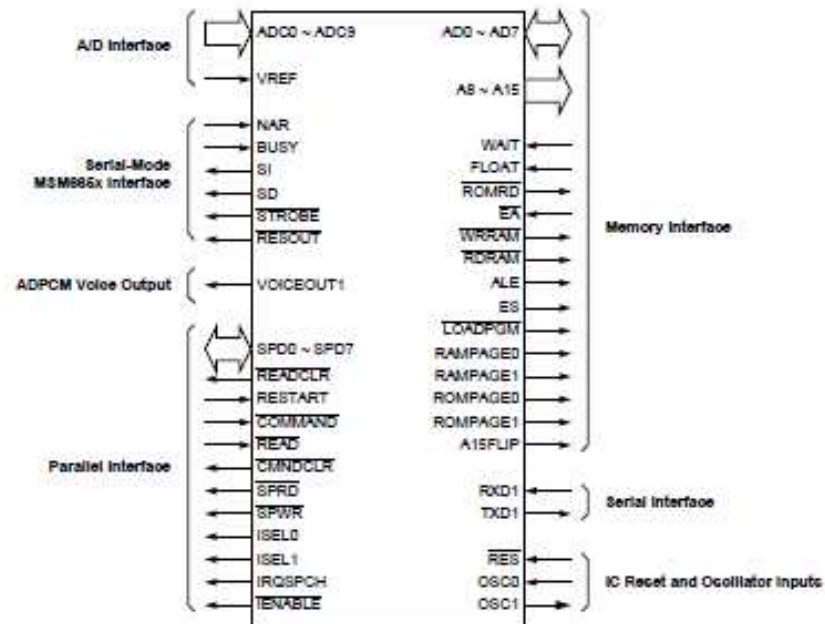


Figure 2. MSM6679 Logic Symbol

DC Characteristics (VDD = 4.5 ~ 5.5 V, Ta = -40 ~ 70 °C)

Parameter	Symbol	Condition	Rated Value			Unit
			Min	Typ ⁽¹⁾	Max	
High-level input voltage	V _{IH}	Applied to AD0-AD7	2.2	—	V _{DD} + 0.3	V
		Applied to OSC0	0.85 x V _{DD}	—	V _{DD} + 0.3	
		Applied to all other I/O	0.80 x V _{DD}	—	V _{DD} + 0.3	
Low-level input voltage	V _{IL}	Applied to AD0-AD7	-0.3	—	0.8	
		Applied to OSC0	-0.3	—	0.15 x V _{DD}	
		Applied to all other I/O	-0.3	—	0.2 x V _{DD}	
High-level output voltage	V _{OH}	Output current = 400 μA, applied to AD0-AD7, ALE, and ROMRD	V _{DD} - 0.4	—	—	
		Output current = 200 μA, for all other I/O	V _{DD} - 0.4	—	—	
Low-level output voltage	V _{OL}	Output current = 3.2 mA, applied to AD0-AD7, ALE, and ROMRD	—	—	0.4	
		Output current = 1.6 mA, for all other I/O	—	—	0.4	
Input leak current	I _{IH} , I _{IL}	V _I = V _{DD} /10 V, applied to AIn, EA, FLOAT, and RESTART	—	—	1/-1	μA
Input current		V _I = V _{DD} /10 V, applied to RES	—	—	1/-250	
		V _I = V _{DD} /10 V, applied to OSC0	—	—	15/-15	
High-level output current	I _{OH}	V _O = 2.4 V, applied to AD0-AD7	-2	—	—	mA
		V _O = 2.4 V, applied to all other I/O	-1	—	—	
Low-level output current	I _{OL}	V _O = 2.4 V, applied to AD0-AD7	10	—	—	
		V _O = 2.4 V, applied to all other I/O	5	—	—	
Output leakage current	I _{LO}	V _O = V _{DD} /10 V	—	—	±2	μA
Input capacitance	C _I	f = 1 MHz, T _a = 25 °C	—	5	—	pF
Output capacitance	C _O		—	7	—	
Analog reference power supply voltage	I _{REF}	During voice input	—	—	4	mA
		When voice input is halted	—	—	10	
Power consumption in STOP mode	I _{DDs}	V _{DD} = 2 V, T _a = 25 °C	—	0.2	10	μA
		Ports for input pins are V _{DD} or GND, otherwise no load	—	1	100	
Power consumption in HALT mode	I _{DDH}	f _{OSC} = 32 MHz, no load	—	—	60	mA
Power consumption	I _{DD}		—	—	144	

1. Typical condition is 5 V 25 °C.

B. 2- RSC-164- Procesador para reconocimiento de voz



RSC-164/ RSC-164i Data Book

GENERAL DESCRIPTION

The RSC-164 and RSC-164i are low-cost speech recognition ICs designed for use in consumer electronics. They combine an 8-bit processor with neural-net algorithms to provide high-quality speaker-independent speech recognition, speaker-dependent speech recognition, and speaker verification. The chips also support speech synthesis, voice record/playback, 4-voice music synthesis, speaker verification, and system control. These CMOS device includes on-chip RAM, ROM, 16 general-purpose I/O lines, A/D and D/A converters, and a 4-MIPS dedicated processor.

In addition to providing the horsepower needed to perform speech recognition and speech synthesis, the processor has sufficient cycles available for general-purpose product control. The RSC Development Kit allows developers to create custom applications for the RSC chips. The Development Kit includes an assembler, linker, simulator, in-circuit emulator, and library of Sensory technology object code.

The highly-integrated nature of these chips reduce external parts count. A complete system may be built with few additional parts other than a battery, speaker, microphone, and audio input support circuitry. Low power requirements make the RSC chips an ideal solution for battery-powered and hand-held devices.

The RSC chips use a pre-trained neural network to perform speaker-independent speech recognition, while high-quality speech synthesis is achieved using a time-domain compression scheme that improves on conventional ADPCM. Four-voice music synthesis allows multiple, simultaneous instruments for harmonizing. Dynamic AGC control can compensate for people not optimally positioned with respect to the microphone or for people who speak too softly or loudly.

FEATURES

High-Performance Processor

- 4-MIPS performance at 14.32 MHz
- 12-16 general-purpose I/O lines
- Interrupts, timers and counters
- Fully static operation; clock speed from DC to 14.32 MHz

Highly-Integrated Single-Chip Solution

- Internal 64K ROM
- 384 bytes RAM
- 12-bit A/D (Analog to Digital converter)
- Pulse width modulator for direct speaker drive

Low Power Requirements

- Requires single 3.5 to 5.0 volt supply.
- Typical current drain is 10 mA
- Low-power 32,768 Hz oscillator for clock applications

High-Quality Recognition and Synthesis

- Recognition accuracy better than 96% (Speaker Independent) and 99% (Speaker Dependent).
- Synthesis data rates from 5,000-15,000 bits per second
- 4-voice music synthesis capabilities
- Optional AGC control compensates for variations in input signal

Easily Expanded to larger-scale systems

- Separate 16-bit Address and 8-bit Data buses compatible with common memory components
- Separate Code and Data address spaces and memory strobes
- Instruction set compatibility between RSC-164 and RSC-164i



RSC-164/ RSC-164i Data Book

GENERAL DESCRIPTION

The RSC-164 and RSC-164i are low-cost speech recognition ICs designed for use in consumer electronics. They combine an 8-bit processor with neural-net algorithms to provide high-quality speaker-independent speech recognition, speaker-dependent speech recognition, and speaker verification. The chips also support speech synthesis, voice record/playback, 4-voice music synthesis, speaker verification, and system control. These CMOS device includes on-chip RAM, ROM, 16 general-purpose I/O lines, A/D and D/A converters, and a 4-MIPS dedicated processor.

In addition to providing the horsepower needed to perform speech recognition and speech synthesis, the processor has sufficient cycles available for general-purpose product control. The RSC Development Kit allows developers to create custom applications for the RSC chips. The Development Kit includes an assembler, linker, simulator, in-circuit emulator, and library of Sensory technology object code.

The highly-integrated nature of these chips reduce external parts count. A complete system may be built with few additional parts other than a battery, speaker, microphone, and audio input support circuitry. Low power requirements make the RSC chips an ideal solution for battery-powered and hand-held devices.

The RSC chips use a pre-trained neural network to perform speaker-independent speech recognition, while high-quality speech synthesis is achieved using a time-domain compression scheme that improves on conventional ADPCM. Four-voice music synthesis allows multiple, simultaneous instruments for harmonizing. Dynamic AGC control can compensate for people not optimally positioned with respect to the microphone or for people who speak too softly or loudly.

FEATURES

High-Performance Processor

- 4-MIPS performance at 14.32 MHz
- 12-16 general-purpose I/O lines
- Interrupts, timers and counters
- Fully static operation; clock speed from DC to 14.32 MHz

Highly-Integrated Single-Chip Solution

- Internal 64K ROM
- 384 bytes RAM
- 12-bit A/D (Analog to Digital converter)
- Pulse width modulator for direct speaker drive

Low Power Requirements

- Requires single 3.5 to 5.0 volt supply.
- Typical current drain is 10 mA
- Low-power 32,768 Hz oscillator for clock applications

High-Quality Recognition and Synthesis

- Recognition accuracy better than 96% (Speaker Independent) and 99% (Speaker Dependent).
- Synthesis data rates from 5,000-15,000 bits per second
- 4-voice music synthesis capabilities
- Optional AGC control compensates for variations in input signal

Easily Expanded to larger-scale systems

- Separate 16-bit Address and 8-bit Data buses compatible with common memory components
- Separate Code and Data address spaces and memory strobes
- Instruction set compatibility between RSC-164 and RSC-164i

2. RSC-164 Architecture

The RSC-164 is a highly integrated device that combines:

- An 8-bit RISC microprocessor.
- On-chip ROM (64 Kbytes) and RAM (384 bytes), and the ability to address off-chip RAM or ROM.
- An analog-to-digital converter, a digital-to-analog converter, and a pulse width modulator.

The RSC-164 has an external memory interface for accessing external RAMs or ROMs. It also has an internal ROM that can be enabled or disabled (partially or fully) by pin inputs (signals -XMH, -XML; see page 4). The RSC-164i does not access external parallel memory and relies solely on the internal ROM for its program storage needs.

The 8-bit processor can directly access 384 on-chip registers (RAM), of which 352 are general purpose registers and 32 are Special Functions Registers (SFRs). The instruction set accessing these registers is completely symmetrical, allowing *movs*, arithmetic, and logical operations with any register as the destination. Two bi-directional ports provide 16 general purpose I/O pins to communicate with external devices (see page 6). The RSC-164 has a high frequency (14.32 MHz) oscillator as well as a low frequency (32,768 Hz) oscillator suitable for timekeeping applications. The processor clock can be selected from either source, with a selectable divider value. Sensory's technology code requires the use of the 14.32 MHz clock. The device performs speech recognition when running at 14.32 MHz, with an optional divide-by-2 CPU clock (see page 10). There are two programmable 8-bit counters / timers, one derived from each oscillator.

A microphone with an external preamp converts sound into two audio signals that are fed into the AIN0 and AIN1 inputs of the RSC-164. The gain of the external preamp may be controlled by the RSC-164 in some designs. The RSC-164 uses an ADC (Analog-to-Digital Converter) and a Sample and Hold (SH) circuit to convert the incoming analog speech signal into digital data. The output audio signal of the RSC-164 is derived either from a DAC (Digital-to-Analog Converter) or a PWM (Pulse Width Modulator).

In addition to its on-chip ROM and RAM, the RSC-164 has 8 data lines (D[7:0]) and 16 address lines (A[15:0]), along with associated control signals (-RDC, -RDD, -WRC, -WRD, -XML, -XMH) for interfacing to external memory. The memory control signals on the RSC-164 and the processor instruction set provide independent Code and Data spaces, allowing configuration of systems up to 192 Kbytes with no additional hardware decoding. The RSC-164i does not contain these address, data, and control signals. The RSC-164 features 16 general-purpose I/O pins (Px.y) for product and memory bank control; the RSC-164i has 12 general purpose I/O lines.

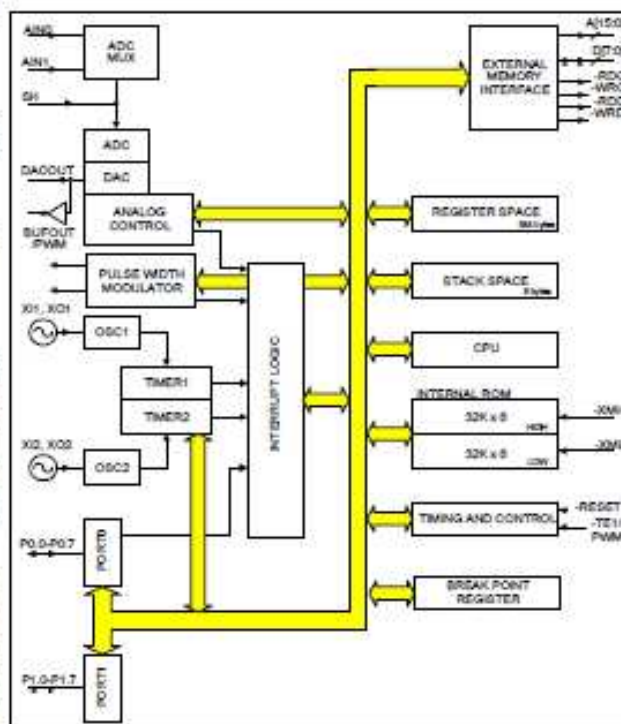


Figure 1 - RSC-164 Block Diagram

B.4- HM6264A – Memoria de apoyo al sistema HM2007

HM6264A Series

8192-word $\times$ 8-bit High Speed CMOS Static RAM

Features

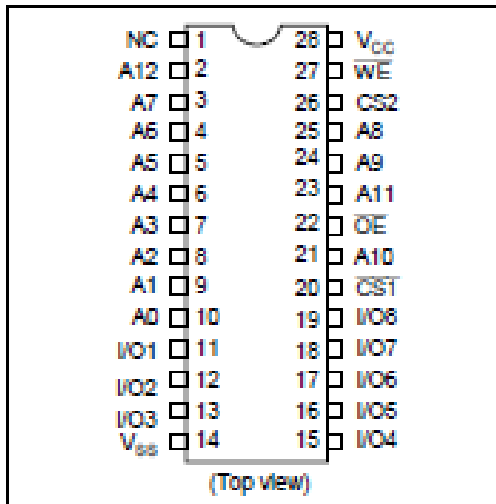
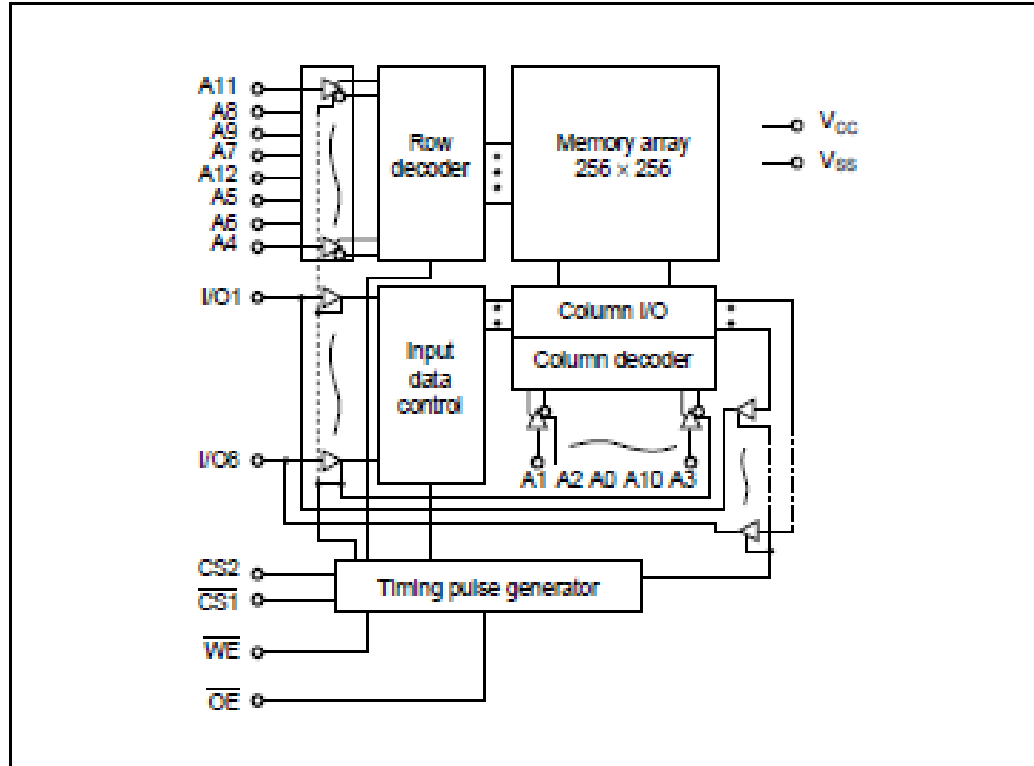
- Low-power standby
 - 0.1 mW (typ)
 - 10 μ W (typ) L-/LL-version
- Low power operation
 - 15 mW/MHz (typ)
- Fast access time
 - 100/120/150 ns (max)
- Single +5 V supply
- Completely static memory
 - No clock or timing strobe required
- Equal access and cycle time
- Common data input and output, three-state output
- Directly TTL compatible
 - All inputs and outputs
- Battery back up operation capability (L-/LL-version)

Ordering Information

Type No.	Access time	Package
HM6264AP-10	100 ns	600-mil, 28-pin plastic DIP (DP-28)
HM6264AP-12	120 ns	
HM6264AP-15	150 ns	
HM6264ALP-10	100 ns	
HM6264ALP-12	120 ns	
HM6264ALP-15	150 ns	
HM6264ALP-10L	100 ns	
HM6264ALP-12L	120 ns	
HM6264ALP-15L	150 ns	

Type No.	Access time	Package
HM6264ASP-10	100 ns	300-mil, 28-pin plastic DIP (DP-28N)
HM6264ASP-12	120 ns	
HM6264ASP-15	150 ns	
HM6264ALSP-10	100 ns	
HM6264ALSP-12	120 ns	
HM6264ALSP-15	150 ns	
HM6264ALSP-10L	100 ns	
HM6264ALSP-12L	120 ns	
HM6264ALSP-15L	150 ns	
HM6264AFP-10	100 ns	28-pin plastic SOP *1 (FP-28D/DA)
HM6264AFP-12	120 ns	
HM6264AFP-15	150 ns	
HM6264ALFP-10	100 ns	
HM6264ALFP-12	120 ns	
HM6264ALFP-15	150 ns	
HM6264ALFP-10L	100 ns	
HM6264ALFP-12L	120 ns	
HM6264ALFP-15L	150 ns	

Note: 1. T is added to the end of the type number for a SOP of 3.00 mm (max) thickness.

Pin Arrangement**Block Diagram**

HM6264A Series

HM6264A Series

Truth Table

$\overline{WE}$	CST	CS2	$\overline{OE}$	Mode	IO pin	V_{CC} current	Note
x	H	x	x	Not selected (power down)	High Z	I_{SS}, I_{SB1}	
x	x	L	x		High Z	I_{SS}, I_{SB1}	
H	L	H	H	Output disabled	High Z	I_{CC}	
H	L	H	L	Read	Dout	I_{CC}	Read cycle
L	L	H	H	Write	Din	I_{CC}	Write cycle 1
L	L	H	L	Write	Din	I_{CC}	Write cycle 2

Note: x: Don't care.

Absolute Maximum Ratings

Parameter	Symbol	Rating	Unit
Terminal voltage ¹⁾	V_T	-0.5 ²⁾ to +7.0	V
Power dissipation	P_T	1.0	W
Operating temperature	T_{opr}	0 to +70	°C
Storage temperature	T_{stg}	-55 to +125	°C
Storage temperature (under bias)	T_{bias}	-10 to +85	°C

Notes: 1. With respect to V_{SS} .
2. -3.0 V for pulse width $\leq$ 50 ns

Recommended DC Operating Conditions ($T_a = 0$ to +70°C)

Parameter	Symbol	Min	Typ	Max	Unit
Supply voltage	V_{CC}	4.5	5.0	5.5	V
	V_{SS}	0	0	0	V
Input voltage	V_{IH}	2.2	—	6.0	V
	V_{IL}	-0.3 ¹⁾	—	0.8	V

Note: 1. -3.0 V for pulse width $\leq$ 50 ns