**UNIVERSIDAD AUSTRAL DE CHILE**
**FACULTAD DE CIENCIAS DE LA INGENIERÍA**
**ESCUELA DE ELECTRICIDAD Y ELECTRÓNICA**

# Estudio analítico y experimental De los circuitos integrados de voz ISD

**Trabajo de titulación para optar al Titulo de Ingeniero Electrónico**

**PROFESOR PATROCINANTE**
**Sr. Julio Zarecht Ortega**

**Julio Sebastián López Vera**
**Valdivia2005**

## COMISIÓN DE TITULACIÓN

**Profesor patrocinante:**

**Julio Zarecht Ortega** _____

**Profesores Informantes:**

**Jorge Morales Vilugrón** _____

**Raúl Urra Ríos** _____

## DEDICATORIA Y AGRADECIMIENTOS

*Hermanita, te dedico este trabajo*
*y espero que te sirva de incentivo*
*para alcanzar tus metas personales.*

*Madre, padre, me imagino que están más felices que yo,*
*Gracias mamá por estar siempre tratando de abrirme los ojos.*

*También quiero agradecer a la comisión evaluadora por aceptar ser quienes me podrán la calificación final de mis estudios de pregrado.*

## ÍNDICE

**RESUMEN** 6
**ABSTRACT** 7
**INTRODUCCIÓN** 8
**OBJETIVOS** 9
**METODOLOGÍA EMPLEADA** 10
**CAPITULO I** 11
**ESTADO DEL ARTE** 11
**1.0 Generalidades** 12
**1.1 Estado del arte en reconocimiento de voz** 13
1.1.1 Principales áreas de trabajo en Reconocimiento de voz 14
**1.2 Estado del arte en conversión texto-voz** 23
1.2.1 Principales áreas de trabajo en conversión texto-voz 24
1.2.2 Principales esquemas de Síntesis de Voz 26
**1.3 Estado del arte en reconocimiento de locutores** 29
1.3.1 Principales áreas de trabajo en Reconocimiento de Locutores 30
1.3.2 Sistemas de Reconocimiento de Locutores 31
**1.4 Estado del arte en codificación de voz** 32
1.4.1 Principales áreas de trabajo en codificación de voz 32
1.4.2 Principales esquemas de codificación de voz 35
**CAPITULO II** 36
**ESTUDIO TEÓRICO DE LOS CIRCUITOS DE VOZ ISD1420 Y ISD4002** 36
**2.0 Generalidades** 37
2.0.1 Funcionamiento general de los ISD 38
**2.1 Descripción del chip de Voz 1420** 39
2.1.1 Arquitectura Interna 40
2.1.2 Descripción de pines 41
**2.2 Funcionamiento del ISD1420** 46
2.2.1 Características generales 46
2.2.2 Métodos de grabado y reproducción del sonido del ISD 1420 47
**2.3 Modos operacionales** 48
2.3.1 Descripción de los Modos operacionales 49
**2.4 Diagramas de tiempo** 51
**2.5 Características eléctricas** 52
**2.6 Descripción del chip de voz ISD4002** 54
2.6.1 Arquitectura interna 55
2.6.2 Configuración de pines 56
2.6.3 Descripción y funcionamiento de cada pin 57
**2.7 Funcionamiento de la serie ISD4002** 61

2.7.1 Características principales 61
2.7.2 Funcionamiento de la interfaz serial periférica SPI 63
2.7.3 Códigos de operación. 65
2.7.4 funcionamiento del ISD4002 66
2.7.5 Secuencia de inicialización del ISD4002 66
2.7.9 Diagramas de tiempo. 68

**2.8 Características eléctricas 70**

**2.9 ISD4002 v/s ISD1420 73**
2.9.1 Diferencias 73
2.9.2 Ventajas 73
2.9.3 Desventajas 74

**2.10 Clasificación de los integrados de Voz ISD 75**

**CAPITULO I I I 77**

**EXPERIENCIAS PRÁCTICAS 77**

**Generalidades 78**

**3.0 Control del ISD 4002 por medio de un microcontrolador 78**
3.0.1 Descripción 78
3.0.2 Etapa de control 78
3.0.3Entrada y salida de audio 87

**3.1Control del ISD4002 por medio de un computador 89**

**3.1Control del ISD4002 por medio de un computador 89**
3.1.1 Descripción 89
3.1.2 Etapa de Control 90
3.1.3Interfaz entre el puerto paralelo y el ISD 91

**3.2 Control del ISD1420 93**
3.2.1Experiencia nº1 93
3.2.2 Experiencia nº2 94
3.3.3 Experiencia nº3 95
3.3.4 Experiencia Nº4 96

**3.3 Ejemplo de aplicación 96**
3.3.1Detector de eventos 96

**Programa para PIC16F84A 98**

**3.5 Descripción de un sistema real empleando chip de voz 101**

**CONCLUSIONES 102**

**Bibliografía 104**

**Anexo 105**

**4.0 Software para tratamiento de señales de audio 105**
4.0.1 Cool Edit 105
4.0.2 pantalla principal del cool edit 105
4.0.3 Como grabar señales de sonido en cool edit 106

## RESUMEN

Se realizo un trabajo de titulación para la carrera de ingeniería electrónica de la Universidad Austral de Chile, orientado a estudiar en forma analítica y experimental los circuitos integrados de voz existentes en el mercado nacional.

Se selecciono este tema ya que, actualmente, en Chile es desconocido por la mayoría de las personas vinculadas a la electrónica, sin embargo es una tecnología emergente que posee un gran potencial aun no explotado a nivel nacional, en Norteamérica se ha incorporado a sistemas electrónicos con aplicaciones como: notas de voz, alertas, menús interactivos y productos de consumo en general.

El trabajo es de carácter teórico-practico, por lo tanto esta dividido en dos partes una enfocada a traducir (la información esta en ingles), comprender y sintetizar la información existente, la cual fue obtenida en un 90% de publicaciones en Internet y la otra parte esta orientada a trabajar en forma experimental en laboratorio, diseñando experiencias que ejemplifican el funcionamiento de estos integrados.

Los integrados seleccionados son los que actualmente comercializa en chile la empresa Victronics, estos son los ISD. Cuya tecnología esta patentada bajo el nombre "chipcorder" por la empresa norteamericana Winbond.

El documento esta estructurado en tres capítulos, Estado del arte, Estudio teórico de los circuitos ISD 4002 y ISD1420, y Experiencias prácticas, en el capitulo III se explica como controlar los circuitos en cuestión por medio de microcontroladores alimentados a 5 volts y por medio del puerto paralelo de un computador.

La parte práctica fue desarrollada en circuitos de carácter experimental, montados en protoboard, sin embargo como resultado final se llego a un hardware orientado a grabar y reproducir sonidos en el ISD4002-240p, este grabador posee entrada de micrófono, entrada de línea, salida de audio de 500mV P-P, y la posibilidad de controlar el ISD con un PIC16F877A, y con el puerto paralelo de un computador.

## ABSTRACT

It is a final work for the career of electronic engineering of the Universidad Austral de Chile, guided to study in analytic and experimental form the existent integrated circuits of voice in the national market.

This topic was selects since, at the moment, in Chile it is ignored by most of people linked to the electronics, however it is an emergent technology that have a great potencial not yet exploded at national level.

In North America it has been incorporated to electronic systems with applications like: voice notes, watchful, interactive menus and consumption products in general.

The work is of character theoretical-practice, therefore this is divided in two parts one focused to translate (the information is in English), to understand and to synthesize the existent information, and the other part is guided to work in experimental form in laboratory, designing experiences that exemplify the operation of these integrated.

The integrated selected are those that at the moment markets in Chile the company Victronics, they are the ISD. Whose technology is patented one under the name "chipcorder".

This document is structured in three chapters, State of the art, theoretical Study of the circuits ISD 4002 and ISD1420, and practical Experiences, in the I chapter III is explained as controlling the circuits in question by means of microcontroladores fed to 5 volts and by means of the parallel port of a computer.

The practical part was developed in circuits of experimental character, mounted in protoboard, however as a result final y arrives to a hardware guided to record and to reproduce sounds in the ISD4002-240p, this engraver possesses microphone entrance, line entrance, exit of audio of 500mV P-P, and the possibility of controlling the ISD with a PIC16F877A, and with the parallel port of a computer.

# INTRODUCCIÓN

En los últimos tiempos se han incorporado al mercado una gran cantidad de artículos electrónicos. Es común observar en los medios de comunicación, avisos publicitando nuevos televisores con mejor sonido, celulares de ultima generación con aplicaciones de mensajes de voz, cámaras fotográficas que graban audio, y una cantidad impresionante de nuevos productos, reivindicando algo tan simple como que *la voz es y seguirá siendo la interfaz más natural de comunicación para las personas.* Esta afirmación que, por una parte, no parece aportar ningún concepto novedoso, por otra, tiene una importancia fundamental en las estrategias y planteamientos que en el mundo de la electrónica, específicamente en el área de las telecomunicaciones, debe realizar cualquier empresa que quiera ofrecer servicios que logren una aceptación mayoritaria por parte de sus clientes.

En el presente trabajo se estudia en particular, una familia de dispositivos electrónicos que surgen como una herramienta para atender las necesidades de tecnologías emergentes enfocadas a trabajar con Voz. Se trata de los circuitos integrados de Voz, ISD, los cuales han sido patentados bajo el nombre de tecnología Chipcorder, orientada a grabar señales de audio en el interior de un circuito integrado y luego poder reproducirlas.

El objetivo del trabajo es realizar un estudio analítico y experimental de los circuitos integrados de voz ISD, comprender su arquitectura interna, experimentar con ellos, y analizar las proyecciones de inserción al mercado. En particular se trabaja con dos circuitos el ISD1420, y el ISD4002, los cuales son estudiados y comparados entre si.

El trabajo esta dividido en tres capítulos, estado del arte, estudio teórico, y estudio experimental.

El presente estudio en su conjunto comprende un documento escrito, y componentes electrónicos, utilizados para el trabajo experimental, los cuales quedan a disposición en la Universidad Astral de Chile.

## OBJETIVOS

### General

Realizar un trabajo de titulación orientado a estudiar en forma analítica y experimental los circuitos integrados de voz de la empresa winbond existentes en el mercado nacional y de esta forma adquirir conocimientos de una tecnología muy utilizada en diferentes áreas de la electrónica (telecomunicaciones, robótica, control, etc.).

### Específicos

- Realizar un análisis del estado del arte.
- Analizar la arquitectura interna de los integrados de voz de la empresa norteamericana winbond específicamente los de la serie ISD 400X.
- Estudiar formas de interacción con tecnologías existentes (controlar el integrado isd4002, por medio de una computadora, y también por medio de un microcontrolador)
- Estudiar los métodos utilizados por estos circuitos para grabación y reproducción del sonido.
- Estudiar software asociados al tema
- Diseñar experiencias practicas para ejemplificar el funcionamiento de estos integrados
- Estudiar como se diferencian las familias de integrados de voz ISD(clasificación)
- Analizar diferencias, ventajas y desventajas de los integrados ISD400X, con integrados de otras series (por ejemplo los isd14xx)
- Analizar la evolución y proyecciones en cuanto a la utilización de esta tecnología.

## METODOLOGÍA EMPLEADA

Para el cumplimiento de los objetivos del trabajo de titulación el desarrollo de este se organizó en tres etapas:

1. Búsqueda de información, en Internet, para esto se utilizo uno de los buscadores mas potentes, ‘‘Google’’, la información encontrada fue 99% en Ingles.

2. Traducción, comprensión y síntesis de la información encontrada (trabajo teórico).

3. Trabajo experimental, orientado a controlar los circuitos integrados de Voz. En esta etapa se repasaron conceptos básicos de electrónica (diodos, transistores, tecnologías CMOS, TTL), y fue necesario obtener orientación de profesores de electrónica de la universidad, además de complementar experiencias con sugerencias de compañeros de carrera.

Fue necesario utilizar las siguientes herramientas.

- Computador personal
- Software traductor de Idiomas ‘‘Power translator’’
- Editor de texto ‘‘Micrsoft Word’’
- Software de desarrollo ‘‘MPLABIDE’’
- Instrumentos de laboratorio; osciloscopio, punta lógica, protoboard, multitester.
- Lenguaje de programación C.
- Lenguaje de programación assembler.

# CAPITULO I

# ESTADO DEL ARTE

## 1.0 Generalidades

En la actualidad contamos con una progresiva proliferación de aplicaciones basadas en el proceso automático del lenguaje hablado. Así, son cada vez más comunes: las interfaces hombre-máquina controladas por voz, los sistemas de respuesta vocal interactiva, y la automatización de sistemas telefónicos, es por esto que grandes empresas del diseño de dispositivos han centrado parte de sus esfuerzos en desarrollar dispositivos capaces de trabajar con señales de voz, existen los procesadores digitales de señales (DSP), también están los DSPIC, estos últimos son microcontroladores que poseen la capacidad de procesar señales, y por su puesto no puede faltar un dispositivo que sea capaz de almacenar señales, en este caso de audio, estos son los ISD . El elevado número de aplicaciones posibles para los próximos años nos lleva a un constante desarrollo tecnológico orientado a trabajar con voz. Así surge una nueva área de desarrollo electrónico, una nueva tecnología que recibe la denominación común de **Tecnología del Habla** y se estructura en cuatro tecnologías básicas principales:

- El Reconocimiento de Voz o Reconocimiento del Habla
- La Conversión Texto-Voz
- El Reconocimiento de Locutores
- La Codificación de Voz

Se pude situar a la Tecnología del Habla como receptora de un amplio conjunto de conocimientos y procedimientos de actuación sobre la información representada en la señal de voz. Conocimientos que se articulan con un alto grado de dificultad y especialización, ya que pertenecen a un marco científico-técnico multidisciplinar, donde se dan cita diferentes ramas del saber como son: fisiología, acústica, lingüística, procesado de señal, inteligencia artificial, teoría de la comunicación y de la información, y ciencia de la computación. Por lo anteriormente descrito, un análisis de "Estado del arte en Tecnología del Habla", no puede pretender abarcar todos y cada uno de los desarrollos y últimas líneas de trabajo abiertas en las diferentes áreas de interés. Por tanto, aún a pesar de introducir importantes simplificaciones, buscando una mayor claridad en la exposición el trabajo esta enfocado a resumir y estructurar el fondo común de los principales desarrollos e innovaciones. Más concretamente, se presentan cuatro secciones diferentes del estado del arte de cada una de las tecnologías básicas.

## 1.1 Estado del arte en reconocimiento de voz

Genéricamente, el principal objetivo que el Reconocimiento de voz persigue es proporcionar una "apropiada" interacción hombre-máquina a través de órdenes habladas. Así, los resultados que esta tecnología proporcione deberán contrastarse con los derivados de otras alternativas como son: teclados, paneles, ratones, etc., en cuanto a si proporcionan un control de procesos de interacción hombre-máquina más o menos "apropiado".

Las principales características que diferencian a los sistemas basados en Reconocimiento de voz frente a otras alternativas son: la naturalidad que supone, utilizar la voz en las operaciones de comando y control, y la precisión y robustez en la comunicación para diferentes usuarios y diferentes entornos. La primera de ellas debería representar la ventaja natural de los sistemas basados en la Tecnología del Habla. Aunque la experiencia nos ha enseñado que, si bien el habla es la forma natural de comunicación entre personas, en el diálogo hombre-máquina esto no parece obvio; por ejemplo, en los diversos estudios que reflejan el elevado número de personas incapaces de responder frente a una máquina. Si bien es cierto que este tipo de rechazos va disminuyendo paulatinamente. Es la segunda de las características anteriores la que se muestra más crítica en las aplicaciones del Reconocimiento del Habla.

El estado actual de la investigación en Reconocimiento voz nos muestra excelentes resultados de sistemas trabajando en entornos controlados de laboratorio. Sin embargo, una aplicación real de esta tecnología exige un funcionamiento en el mundo real donde el grado de dificultad de los problemas es un orden de magnitud mayor.

Bajo esa premisa de buscar una aplicación real, el modelo genérico de comunicación que el Reconocimiento voz propone para el diálogo hombre-máquina puede representarse, de forma simplificada, tal y como muestra el diagrama de la figura 1, para un caso de acceso a una base de datos.

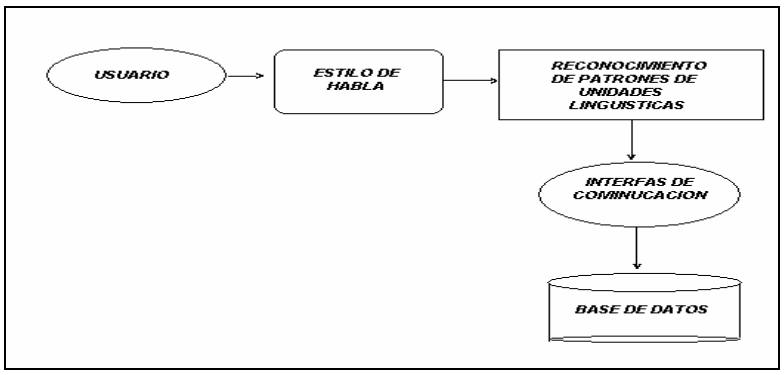



FIGURA 1

En este diagrama, el acceso a la información contenida en una base de datos comienza con la producción de un mensaje hablado por el usuario, pero utilizando una forma o estilo de habla restringido; por ejemplo, utilizando palabras de un vocabulario reducido pronunciadas de forma aislada (como los dígitos), frases tipo, etc. A partir de la señal de voz, un proceso de clasificación, basado en reconocimiento de patrones asociados a diferentes unidades lingüísticas (palabras, fonemas, sílabas, etc.), permite a una interfaz de comunicaciones extraer de la base de datos la información solicitada por el usuario.

### 1.1.1 Principales áreas de trabajo en Reconocimiento de voz

Siguiendo el modelo de la figura 1 podemos presentar las principales áreas de trabajo que intervienen en el diseño y especificación de sistemas de Reconocimiento del Habla actuales. Estas áreas serían las siguientes:

1. Proceso de la señal de voz.
2. Técnicas de reconocimiento de patrones.
3. Diferentes estilos de habla.
4. Dependencia del locutor.
5. Vocabulario de reconocimiento.
6. Tarea de reconocimiento.
7. Bases de datos para entrenamiento y reconocimiento.

## Proceso de la señal de voz

La primera operación que debe realizar un reconocedor es procesar la señal de voz de entrada al sistema, con objeto de extraer la información acústica relevante para la tarea que debemos realizar. En este primer nivel del sistema son dos los interrogantes a resolver:

1. ¿Qué rasgos o características extraer?
2. ¿Qué efectos perturbadores pueden acompañar a la voz? y ¿cómo eliminarlos?

La respuesta a la primera cuestión ha venido precedida de un largo proceso de investigación sobre diferentes procedimientos de parametrización de la voz. Planteándose como solución actual más extendida una parametrización de la envolvente espectral que incluya consideraciones preceptúales a partir del funcionamiento del oído. Para reducir el número de parámetros posibles, la parametrización se combina con la utilización de técnicas discriminativas, seleccionándose el subconjunto con los parámetros más eficientes o distintivos.

En cuanto a la segunda de las preguntas planteadas, la presencia de efectos perturbadores en la señal de entrada, ha generado tres líneas de trabajo principales:

Detección robusta de voz: Apareciendo innumerables procedimientos de discriminación entre voz o ruido (silencio) para diferentes tipos de ruido.

Reducción de ruido: Distinguiéndose procedimientos que actúan directamente sobre la señal de voz y procedimientos que buscan compensar el efecto del ruido sobre la parametrización de la voz.

Cancelación de ecos: Incorporando técnicas de filtrado adaptativo que permitan al usuario comenzar a hablar mientras, desde el terminal remoto, se le está comunicando un mensaje que puede provocar un eco en la voz que entra al reconocedor.

## Técnicas de reconocimiento de patrones

El reconocimiento de patrones es la técnica más específica de todo sistema de reconocimiento. De ahí que muchos reconocedores se identifiquen a partir de la técnica de reconocimiento de patrones que incorporan. A partir de la representación paramétrica de la voz, este módulo realiza un proceso de clasificación utilizando una serie de patrones. Estos patrones se obtienen en una fase de entrenamiento del sistema y son representativos de un conjunto de unidades lingüísticas (palabras, sílabas, sonidos, fonemas). La peculiaridad más característica de este proceso, que marca su dificultad, es la variabilidad temporal que puede presentar una misma unidad lingüística al ser producida por diferentes modos y/o velocidades de habla. Así

pues, las primeras técnicas de reconocimiento de patrones utilizadas fueron las basadas en un Alineamiento Temporal a través de algoritmos de Programación Dinámica, técnicas DTW. Posteriormente se recurrió a la mayor flexibilidad que el modelado de procesos estocásticos permite para representar secuencias de duración variable. Concretamente la alternativa a las técnicas DTW fueron los Modelos Ocultos de Markov, (HMM), que pueden verse como una generalización de algoritmos DTW y han demostrado mejores prestaciones en multitud de sistemas de reconocimiento. También hay que mencionar que, recientemente, la potencia y excelentes capacidades de clasificación mostradas por las denominadas Redes Neuronales Artificiales (RN) las sitúa como posible alternativa frente a los HMM . Hasta el momento las Redes Neuronales han permitido obtener los mejores resultados en Reconocimiento de Locutores, sin embargo en Reconocimiento del Habla encuentran como mayor dificultad la forma de afrontar la variabilidad temporal del habla.

## Modelado dependiente del estilo de habla

Se distinguen tres modos fundamentales de hablar frente a un sistema de reconocimiento:

1. Palabras aisladas: Supone que el usuario pronuncia una sola palabra o comando que el sistema deberá reconocer.
2. Habla conectada: El usuario pronuncia de forma fluida un mensaje utilizando un vocabulario muy restringido; el ejemplo más típico sería la pronunciación de un número telefónico.
3. Habla continua: Corresponde al modo más avanzado de funcionamiento de un reconocedor, y supone la pronunciación de frases de forma natural para un vocabulario amplio de palabras.

Además de los tres modos fundamentales anteriores, los reconocedores de voz tienen que afrontar, para un modelado robusto del habla, los tres aspectos siguientes:

*Reconocimiento en contexto o "Word spotting":*

Técnica especialmente utilizada en reconocimiento de palabras aisladas, encaminada a detectar la presencia de palabras del vocabulario a reconocer en el contexto de otras palabras o pronunciaciones. La mayoría de las veces el contexto es resultado de la dificultad que encuentra el usuario para ceñirse a la pronunciación de una única palabra aislada. En otras

ocasiones, el reconocimiento en contexto es la solución apropiada para robustecer el reconocimiento en ambientes acústicamente hostiles; por ejemplo, cuando la palabra que pronuncia el usuario viene acompañada de ruidos telefónicos, urbanos, etc. En cualquier caso, se trata de una técnica importante para robustecer los sistemas en aplicaciones reales.

*Rechazo*

Otro efecto de la presencia de sonidos indeseados (ruidos, sonidos o palabras fuera del vocabulario), es provocar el reconocimiento de palabras que realmente no han sido pronunciadas. Los procedimientos conocidos como técnicas de rechazo tienen como objetivo permitir incluir entre los resultados de reconocimiento la identificación de esos sonidos indeseados. Nos encontramos ante un problema de gran importancia de cara a la operatividad de un sistema de reconocimiento, que aún hoy por hoy no cuenta con una clara solución.

*Múltiples candidatos*

El proceso de reconocimiento de patrones que realiza un reconocedor se basa en identificar el patrón que ofrezca la puntuación más alta para decidir cuál es la mejor palabra o secuencia de palabras reconocida. Este proceso se basa en información exclusivamente acústica, sin tener en consideración otras posibles fuentes de conocimiento que podrían utilizarse para completar las puntuaciones de las diferentes palabras o secuencias candidatas. En la mayoría de los casos, la aplicación en que se encuentra el reconocedor es la que posee la información necesaria que permitiría seleccionar entre varias hipótesis de reconocimiento. Pensemos, por ejemplo, en una aplicación basada en el reconocimiento de números telefónicos; en esa situación, ante las dos hipótesis mejores de reconocimiento, una compuesta de cinco dígitos y otra de siete, la aplicación seleccionaría esta última independientemente de quién obtuviese la mayor puntuación "acústica" en el proceso de clasificación. Los procedimientos que permiten a un reconocedor disponer de la flexibilidad que supone manejar N hipótesis de reconocimiento se denominan N-best.

## Dependencia del locutor

El grado de dependencia del locutor define si el sistema incorpora patrones de unidades lingüísticas adaptados a un locutor determinado, y, por tanto, sólo funcionará correctamente para él, o si los patrones pretenden ser válidos para cualquier hablante. En el primer caso se

habla de reconocimiento dependiente del locutor, mientras que en el segundo de reconocimiento independiente del locutor. A parte de las actividades específicas que se desarrollan para sistemas dependientes e independientes del locutor, existe un importante número de esfuerzos dirigidos a conseguir la adaptación de un reconocedor a un locutor específico con la menor cantidad de voz posible.

## Dependencia del vocabulario

Las prestaciones de un reconocedor dependen fuertemente del tamaño y grado de dificultad del vocabulario. Es decir, del número de palabras que el sistema es capaz de reconocer, y de la mayor o menor dificultad de su reconocimiento en base a las relaciones de similitud fonética entre palabras. En la actualidad se diseñan sistemas tanto para vocabularios pequeños (menos de 50 palabras) y medios (entre 50 y 500 palabras), como para grandes vocabularios (más de 500 palabras), llegándose hasta 50.000 palabras para aplicaciones de dictado o acceso a bases de datos mediante lenguaje natural.

Otra importante dimensión, en relación con el vocabulario, es la que afecta a la distinción entre vocabularios fijos y flexibles. Una determinada aplicación, cuando esté reconociendo, siempre actuará sobre un vocabulario fijo. Pero en muchos casos ese vocabulario deberá variarse o actualizarse para eliminar y/o dar cabida a nuevas palabras. Tradicionalmente, una variación del vocabulario suponía comenzar un largo y costoso proceso de recogida de una nueva base de datos y re-entrenamiento de los patrones del sistema. En la actualidad hay diversas aproximaciones para conseguir un sistema con vocabulario flexible, que no necesite re-entrenarse para cada nuevo vocabulario.

## Gramáticas de reconocimiento

Según aumenta el número de palabras del vocabulario, el número de posibles combinaciones crece exponencialmente. Por tanto, se hace imprescindible la incorporación de restricciones, en cuanto al número de combinaciones válidas, según la tarea en que se inserte el sistema. Restricciones que suelen incorporarse en forma de gramáticas basadas en reglas sintácticas y/o semánticas destinadas a reducir el número de palabras susceptibles de ser reconocidas en cada momento. La medida utilizada para definir el grado de dificultad que supone una determinada tarea es la denominada perplejidad, de modo que un nivel de perplejidad bajo supone que en cada momento el número de posibles palabras candidatas es bajo, mientras que una perplejidad alta supone que ese número es alto, y consiguientemente el reconocimiento será más difícil.

## Principales sistemas de reconocimiento del habla (bases de datos)

A principios de los años 80 más de 10 compañías de Estados Unidos ofrecían reconocedores de palabras aisladas dependientes del locutor con un vocabulario de hasta 300 palabras. Sólo las firmas VERBEX y NEC ofrecían un sistema independiente del locutor con posibilidades de reconocimiento de palabras conectadas. En ese momento, la situación del Reconocimiento del Habla podría resumirse como:

- Reconocedores de palabras aisladas dependientes del locutor como tecnología asentada.
- Reconocedores independientes del locutor y reconocedores de palabras conectadas como tecnologías nacientes.

Por otro lado, debido a las limitaciones en el ancho de banda y la sensibilidad frente al ruido, sólo un número muy reducido de estos reconocedores trabajaban sobre la línea telefónica. En esta época sólo se encuentra en la literatura referencia a tres aplicaciones del Reconocimiento del Habla dentro del ámbito de las telecomunicaciones:

Dos prototipos de reconocedores de palabras aisladas independiente del locutor para:

1. Marcación por voz en la red privada.
2. Reconocimiento de letras.

- Un reconocedor de palabras aisladas dependiente del locutor aplicado a la marcación de números de teléfono por voz.

Como consecuencia de las investigaciones en Reconocimiento del Habla llevadas a cabo durante los últimos diez años, actualmente son muchas las compañías que cuentan con

reconocedores de palabras aisladas (dígitos mas un número reducido de comandos) independiente del locutor. Sistemas diseñados, en su mayor parte, para incorporarse en aplicaciones de telecomunicación.

Las prestaciones obtenidas para palabras aisladas, vocabularios con un número de palabras inferior a 200, e independencia del locutor, dependen en gran medida de las características acústicas de las palabras del vocabulario. Así, mientras que el reconocimiento de los diez dígitos puede presentar una tasa de error de palabra inferior al 2 por 100, el reconocimiento de 39 caracteres alfanuméricos (dígitos y letras) en inglés supone un 7% de error, y el de 129 palabras dentro del ámbito de las compañías aéreas un 2,9 por 100.

El reconocimiento de dígitos conectados es otra de las tareas con mayores posibilidades de utilización en diversas aplicaciones. Los resultados que proporcionan los mejores sistemas desarrollados para el inglés por los Laboratorios Bell de AT&T y por el Centro de Investigación Informática de Montreal (CRIM), suponen una tasa de error de palabra inferior al 1 por 100 cuando trabajan en condiciones de laboratorio. Sin embargo, sobre la red telefónica las prestaciones se reducen de forma importante hasta tasas de error de palabra cercanas al 4 por 100.

La evolución de los sistemas de reconocimiento con mayor proyección de futuro: reconocimiento para grandes vocabularios y habla continua, tiene como mejores representantes a los siguientes sistemas experimentales:

BYBLOS

Desarrollado por BBN. Byblos es el nombre de una ciudad fenicia donde se descubrió la primera muestra de escritura fonética. Este detalle marca el énfasis que se pone actualmente en desarrollar sistemas sobre una base fonética. Aunque se trata de un sistema dependiente de locutor, este sistema ha aportado un nuevo y eficiente procedimiento de reconocimiento rápido (búsqueda rápida) basado en algoritmos N-best.

TANGORA

Desarrollado en IBM. También se trata de un sistema dependiente de locutor para grandes vocabularios. Su principal interés es un proceso de adaptación a un nuevo locutor que require 20 minutos para leer 100 frases de 1.200 palabras, 700 de las cuales son distintas.

SPHINX-II

Desarrollado en la Universidad de Carnegie-Mellon . Es un sistema pionero en reconocimiento independiente de locutor para grandes vocabularios. Su más reciente innovación es el procedimiento VOCIND para hacer al sistema independiente del vocabulario.

LINCOLN

Desarrollado en el laboratorio del mismo nombre. Su principal aportación es el modelado de voz rápida, con emoción, tensión, etc.

DECIPHER

Desarrollado en SRI Internacional. Su principal novedad fue la representación detallada de aspectos fonéticos importantes, tales como la coarticulación entre palabras.

ATR HMM-LR

Sistema japonés desarrollado en ATR . Está basado en procedimientos específicos de modelado de sonidos que no utilizan estructuras intermedias de modelos de fonema o palabra.

CSELT

Desarrollado en el centro italiano del mismo nombre. Su principal innovación es un sistema de búsqueda rápida basada en un primer descifrado fonético simple y rápido, seguido por una búsqueda más detallada.

PHILIPS

Desarrollado por la empresa del mismo nombre. Es un sistema pionero en procesos de reconocimiento rápidos para habla continua y vocabularios de hasta 10.000 palabras.

Sistemas telefónicos de AT&T y Bell Northern Research (BNR).

Ambos sistemas incorporan procedimientos específicos para aplicaciones de automatización de servicios telefónicos.

Para disponer de una idea general de las prestaciones proporcionadas por los sistemas anteriores, presentamos algunos datos orientativos sobre evaluaciones realizadas para tres bases de datos correspondientes a tres tareas diferentes dentro del programa DARPA (Defenece Advanced Research Projects Agency) de Estados Unidos. Hay que dejar claro que las tres bases de datos se componen de voz grabada en condiciones de laboratorio. Una breve descripción de cada base de datos sería:

Base de datos correspondiente a la gestión de recursos navales (Naval Resource Management), con un vocabulario de 991 palabras.

Base de datos de información sobre vuelos de líneas aéreas ATIS (Air Travel Information System), con vocabulario de 1.800 palabras.

Base de datos leída del Wall Street Journal, con un vocabulario de 20.000 palabras.

Cada una de las tareas de reconocimiento cuenta con un nivel de restricción gramatical diferente, siendo mayor para la base de datos ATIS, algo inferior para la de recursos navales, y un orden de magnitud inferior para la del Wall Street Journal.

Entre 1987 y 1999 se consiguió reducir la tasa de error de reconocimiento a nivel de palabra del 20 por 100 al 4,5 por 100, sobre la base de datos de gestión de recursos navales. Para la base de datos ATIS los mejores resultados suponen una tasa de error a nivel de palabra del 4 por 100. Desde 1990 se está trabajando con la base de datos del Wall Steet Journal, que presenta un nivel de dificultad muy superior a las otras dos, consiguiéndose una tasa de error próxima al 13 por 100.

Hay que tener presente que, aún en los casos de un error de palabra del 4 por 100, el error a nivel de frase será próximo al 20 por 100. Por tanto, cara a la incorporación de esta tecnología en aplicaciones reales se hace imprescindible complementar el reconocimiento con técnicas de proceso de lenguaje natural, que permitan extraer información del mensaje hablado aún a partir de frases con errores de reconocimiento. A pesar de esto son muchos los desarrollos que a nivel de prototipo muestran las posibilidades futuras del Reconocimiento del Habla.

## Trabajos futuros en reconocimientos del habla

Después de revisar las principales innovaciones y resultados obtenidos para sistemas de Reconocimiento del Habla actuales, pasamos a presentar los principales cauces que guiarán los trabajos que se desarrollen en el futuro. Desde la óptica de la problemática específica del reconocimiento, los objetivos principales de avance planteados pueden resumirse en los siguientes:

Robustez

Frente a la diversidad de situaciones de ruido ambiental, variedades dialectales, habla espontánea y entornos de trabajo: línea telefónica, tipos de micrófono, etc.

Grandes vocabularios

Adecuados para manejar aplicaciones avanzadas en las que intervenga un léxico superior a 1.000 palabras.

*Vocabularios flexibles*

Permitiendo extender el reconocimiento a cualquier vocabulario.

*Habla continua*

Permitiendo diálogos más naturales y espontáneos en el proceso de comunicación hombre-máquina.

Cada uno de los objetivos anteriores tiene asociado un conjunto de líneas de avance entre las que destacan las siguientes:

Incorporación de técnicas eficientes de compensación en los procesos de acondicionamiento de la voz y extracción de características, que permitan adaptarse a las condiciones variantes debidas a diferentes locutores, canales de transmisión, micrófonos, ruido ambiente, etc.

Definición y modelado de un conjunto de unidades inferiores a la palabra, imprescindibles para el desarrollo de sistemas de grandes vocabularios. Buscándose dos objetivos principales:

Representar de forma precisa los principales sonidos y contextos del habla, Permitir un funcionamiento independiente del vocabulario que evite el tedioso y costoso proceso de grabar una base de datos cada vez que varíe el vocabulario a reconocer.

Desarrollo de procedimientos de búsqueda eficiente. Imprescindibles para poder disponer de realizaciones en tiempo real cuando el conjunto de palabras del vocabulario crece y la red de reconocimiento es compleja.

Diseño de procedimientos automáticos o semi-automáticos dirigidos a definir la gramática más apropiada para una determinada tarea de reconocimiento, que sea lo suficientemente flexible para permitir construcciones no totalmente correctas y habla espontánea.

Integración de algoritmos de reconocimiento y de proceso de lenguaje natural, tanto para mejorar las prestaciones de los sistemas de reconocimiento como para permitir el desarrollo de sistemas avanzados de diálogo hombre-máquina.

Junto a estas líneas de avance, que concentrarán los esfuerzos futuros en la División de Tecnologías del Habla, no hay que olvidar la tarea básica que supone la grabación, mantenimiento y actualización de bases de datos cada vez más completas.

## 1.2 Estado del arte en conversión texto-voz

Dentro de la tecnología conocida como Síntesis de Voz existen dos aproximaciones diferenciadas: la síntesis a partir de concepto y la síntesis a partir de texto. En el primer caso la voz es generada a partir de una representación simbólica del mensaje que queremos emitir y, por lo tanto, requiere un proceso de generación de lenguaje natural. En el segundo caso, no disponemos del significado que queremos representar mediante la señal de voz, sino que disponemos de un texto que deberemos leer. Nos centraremos en este último tipo de sistemas, que denominaremos de conversión texto-voz.

Para la aceptación por el público de los sistemas de conversión texto-voz, es necesaria una gran calidad en la voz sintética. Esta calidad, hoy en día, debe ir más allá de la simple inteligibilidad, se trata de conseguir la naturalidad de un hablante humano. Un lector humano ante un texto introduce una información que denominaremos prosodia. La prosodia relaciona los diferentes sonidos del mensaje hablado, y permite reflejar tanto elementos lingüísticos, imprescindibles para el sentido de la oración (modalidad, énfasis, etc.), como elementos no lingüísticos (características del locutor, estado de ánimo, etc.). La habilidad del lector para reflejar acústicamente el contenido o sentido del mensaje mide la calidad del proceso de lectura que realiza.

Un sistema de síntesis se compone de dos módulos claramente diferenciados, que requieren para su realización una metodología y conocimientos de base radicalmente distintos: el proceso lingüístico-prosódico y el proceso acústico (figura 4).

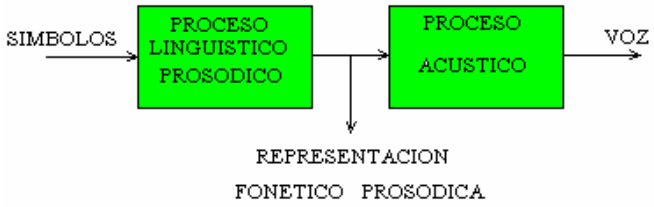



Figura 4. Diagrama de bloques general de un sistema de conversión texto a habla.

Debe añadirse, que existen puntos de conexión entre el proceso lingüístico y el proceso acústico. El punto de encuentro entre ellos será la representación fonética y prosódica del texto.

### 1.2.1 Principales áreas de trabajo en conversión texto-voz

Siguiendo el diagrama de la figura 4, las principales áreas de actividad en conversión texto-voz, se articulan en dos líneas diferenciadas: proceso lingüístico y proceso acústico.

#### Proceso lingüístico

El objetivo general del proceso lingüístico es determinar, a partir de un texto, dos tipos de información necesarios para proporcionar al proceso acústico datos para generar voz natural. Estos dos tipos de información se conocen como información segmental e información suprasegmental.

La información segmental es la asociada a la cadena de sonidos que componen el mensaje. Los sonidos que se pueden producir con los órganos de articulación son variadísimos, aún considerando un único locutor. Pero cada idioma ha seleccionado una serie limitada de "sonidos ideales" aceptados por todos para su uso en el habla. Tenemos así conciencia de una serie limitada de representaciones abstractas fáciles de aprender y percibir. Estas representaciones abstractas se denominan fonemas y su número depende del idioma considerado. Sin embargo, si abandonamos el plano abstracto de la lengua y nos situamos en el plano del habla, debido al carácter continuo del habla hay variaciones en el punto, modo y carácter sordo/sonoro de cada "sonido ideal". Estas condiciones dan lugar a los diferentes alófonos de un fonema, así, por ejemplo, en español los fonemas oclusivos sonoros se fricatizan en contextos vocálicos (compárese el sonido de /g/ en venga y en vega), lo que complica aún más el diseño de sistemas de síntesis de voz de alta calidad. Establecer mecanismos para determinar la variación concreta para cada sonido es, por tanto, una de las principales actividades de investigación relacionadas con la información segmental.

Representación de la información suprasegmental

La información suprasegmental es aquella que queda asociada a la prosodia. En consecuencia, refleja tanto elementos lingüísticos (carácter de la frase, pausas, acentos, agrupación en elementos de significado, etc.), como elementos no lingüísticos (características personales del locutor, estado de ánimo, etc.). Esta información es la clave para conseguir una alta naturalidad en los sistemas de síntesis de voz.

Es por ello por lo que se destinan muchos esfuerzos de investigación a determinar su representación a través de tres parámetros acústicos de la señal de voz:

La evolución temporal de la frecuencia fundamental, que es el correlato físico más importante desde el punto de vista perceptivo, La duración de los segmentos o sonidos que componen la frase, La curva de energía de la señal acústica (lo menos importante desde el punto de vista preceptivo).

## Proceso acústico

El objetivo general que el proceso acústico persigue es convertir la cadena fonética y las variables de control prosódico en la forma de onda asociada a la voz sintetizada. Un diagrama de bloques típico para el proceso acústico es el representado en la figura 5.

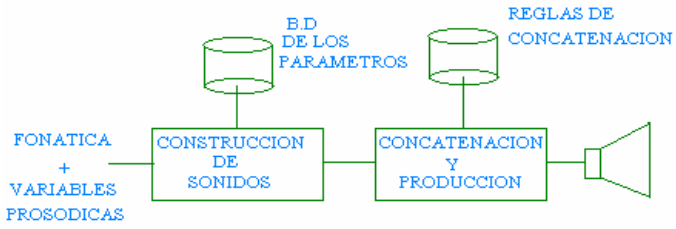



Figura 5. Proceso acústico**.**

Existen dos actividades involucradas en el proceso acústico: La construcción o generación de sonidos, y la concatenación y producción de la cadena hablada. La construcción de sonidos supone la obtención de la información que define cada sonido a sintetizar, a partir de una representación, en general, paramétrica de los mismos. La concatenación y producción será el proceso de unión de los diferentes sonidos, así como la generación y modificación de sus parámetros acústicos. El primer aspecto a resaltar es la existencia de un compromiso entre, el número de reglas de parametrización y concatenación (reglas destinadas a evitar transiciones bruscas desagradables para el oído) y, el tamaño de la base de datos de parámetros. De manera que, atendiendo al proceso acústico, podemos establecer una amplia gama de sistemas de conversión de texto a habla que abarca desde los sistemas dirigidos por regla a los dirigidos por datos. Dicho de una manera concisa, en un sistema "puro" dirigido por reglas, éstas generan la representación paramétrica que alimentará un sintetizador de voz. En uno dirigido por datos, éstos representan directamente segmentos de voz. Entre estos dos polos podemos Encontrar multitud de casos intermedios. Así, por ejemplo, los segmentos de voz pueden estar parametrizados según un determinado modelo de codificación de voz.

### 1.2.2 Principales esquemas de Síntesis de Voz

Existe un gran número de sistemas de conversión texto-voz que incorporan una amplia variedad de aproximaciones diferentes para el proceso lingüístico, difíciles de estructurar de una manera sistemática. En consecuencia, los sistemas se suelen catalogar atendiendo a la técnica de proceso acústico que incorporen. Actualmente se utilizan métodos de síntesis que pueden

clasificarse en tres grupos principales: los sintetizadores de formantes, los sintetizadores mediante modelos articulatorios y los sintetizadores basados en concatenación de unidades.

## Sintetizadores de formantes

En éstos la cadena fonética y la prosodia controlan las frecuencias, anchos de banda y la excitación de un conjunto de resonadores y antirresonadores. Un exponente claro de este tipo de sistemas lo constituye el sintetizador de Klatt. El sintetizador de formantes consiste en una composición de filtros que modelan las resonancias y antirresonancias de las cavidades vocal y nasal. Para este modelado se usan filtros que en la configuración más general están conectados en serie y en paralelo. Es un procedimiento de enorme flexibilidad que se pone de manifiesto en la alta calidad de la voz sintética que se puede obtener mediante ajuste manual de los parámetros del sintetizador. Sin embargo, se necesita un número enorme de reglas en las síntesis automáticas, lo que requiere compiladores cada vez más sofisticados capaces de integrar todo el conocimiento que se adquiere a base de experimentar con el sistema.

## Sintetizadores mediante modelos articulatorios

En estos, se trata de simular la propagación de las ondas acústicas en el tracto vocal. Los segmentos y las variables prosódicas se traducen en parámetros de un modelo simplificado del aparato fonador humano, que implícitamente restringen la dinámica del sistema, pudiendo producir voz de la más alta calidad. Surgieron para tratar de hacer corresponder explícitamente los sintetizadores de formantes con un modelo más explícito del tracto vocal. Su interés se centra en que las restricciones implícitas en este modelo permiten ver el habla como un continuo acústico, por lo que se solventan los problemas de concatenación de segmentos. Sin embargo, la dificultad principal de estos tipos de sistemas es que todavía no se conoce totalmente el proceso de producción del habla humana.

## Sintetizadores basados en concatenación de unidades

En estos sintetizadores, como su propio nombre indica, se concatena un conjunto de unidades extraídas de producción humana. El representante más clásico de este tipo de sintetizadores es el conocido como PSOLA (Pitch Synchronous Overlap Add). En este tipo de sintetizadores debe estar presente un algoritmo que permita, además de la concatenación de unidades, modificar prosódicamente los segmentos a concatenar. Adicionalmente, se pueden

usar técnicas de codificación de voz para reducir las necesidades de almacenamiento en la base de unidades acústicas. También existe la posibilidad de incluir en el modelo de codificación de voz las tareas de concatenación y modificación prosódica, siempre que el codificador parametrice la señal de voz con la suficiente flexibilidad para el modelado prosódico de las unidades.

## 1.3 Estado del arte en reconocimiento de locutores

El Reconocimiento de Locutores plantea un problema, en principio, próximo al de Reconocimiento de Voz. Los dos modos básicos de operación del Reconocimiento de Locutores son la Identificación de Locutores (IL) y la Verificación de Locutores (VL). La Identificación de Locutores es el proceso de determinar la identidad de un hablante perteneciente a una población de hablantes pre-establecida. La Verificación de Locutores supone determinar si una persona es quien dice ser a través del análisis de su voz.

IL y VL suponen por tanto el reconocimiento de quién habla en lugar de qué es lo que se habla. En consecuencia, estos sistemas buscan explotar las diferencias en la forma y estilo de habla, aspectos que se pretende eliminar o normalizar en los sistemas de Reconocimiento de Voz. A pesar de ello, los sistemas de reconocimiento de locutores presentan una arquitectura muy similar a la presentada para los sistemas de Reconocimiento de Habla. El reconocimiento de un locutor se basa en la aplicación de una determinada técnica de clasificación de patrones a partir de un conjunto de parámetros o características obtenidas de la voz del hablante.

El grado de dificultad de IL es función directa del tamaño de la población de locutores, ya que el número de comparaciones a realizar es igual al número de locutores posibles. Por el contrario, la dificultad en VL es relativamente independiente del número de locutores, puesto que supone una única comparación con el patrón asociado a la identidad presentada por el hablante (por ejemplo a través de una clave previamente proporcionada al sistema). El resultado de la VL será simplemente sí o no, es decir, aceptar o rechazar la identidad presentada.

Las prestaciones de un reconocedor de locutores dependen también del tipo de discurso que pronuncie el locutor. Así se habla de sistemas dependientes de texto, cuando el locutor debe obligatoriamente pronunciar un texto fijo pre-establecido, y sistemas independientes de texto, cuando el locutor tiene total libertad en su pronunciación. Habitualmente el reconocimiento dependiente de texto, se utiliza en aplicaciones de control de accesos con usuarios cooperativos, mientras que el independiente de texto es habitual en aplicaciones forenses (por ejemplo, investigaciones policiales) con usuarios no cooperativos. Las prestaciones de los sistemas dependientes de texto son superiores a las de los independientes de texto.

### 1.3.1 Principales áreas de trabajo en Reconocimiento de Locutores

El factor más crítico para el diseño de sistemas de Reconocimiento de Locutores es la determinación de los parámetros o características sobre los que el sistema basará su funcionamiento. Idealmente deberán elegirse características fáciles de medir, estables con el paso del tiempo, robustas frente a los diferentes entornos de trabajo del sistema y altamente discriminativas frente a posibles impostores.

En muchos sistemas se utilizan conjuntos de parámetros similares a los utilizados en reconocimiento de voz: coeficientes de predicción lineal, parámetros diferenciales, etc. También suele considerarse, al menos teóricamente, información explícita sobre formantes o resonancias del tracto vocal y frecuencia fundamental de vibración de las cuerdas vocales. Sin embargo, esa información es difícil de obtener de forma precisa, especialmente en entornos adversos. Adicionalmente, la evaluación de la frecuencia fundamental presenta como problema principal la facilidad de ser imitado, si bien su valor medio se ha utilizado con éxito en diversos sistemas. Otras características que han proporcionado buenos resultados han sido: el grado de co-articulación en la pronunciación de sonidos nasales, la pendiente de los formantes en diptongos y el tiempo de arranque de las oclusivas sonoras.

En cuanto al proceso de reconocimiento de patrones, muchas veces comparte las mismas técnicas que se utilizan en reconocimiento de voz: DTW, HMM y RN. Sin embargo, la diferencia fundamental que puede plantearse es la utilización de parámetros o características obtenidas a partir de estadísticas a largo plazo sobre la voz, en lugar de secuencias de parámetros obtenidos en tramos de tiempo restringido.

Otro aspecto crucial característico del Reconocimiento de Locutores es la necesidad de mantener una actualización continua de los patrones de cada locutor. Esta actualización continua permitirá garantizar el correcto funcionamiento del sistema durante largos períodos de tiempo adaptándose a las naturales variaciones de la forma de hablar de los diferentes locutores.

Como ya mencionamos anteriormente, es en el Reconocimiento de Locutores donde la técnica de clasificación basada en Redes Neuronales parece mostrar posibilidades importantes de éxito. Cabe destacar los trabajos sobre el diseño de una red neuronal entre cada par de locutores presenta una tasa de reconocimiento del 100 por 100 para una población de 47 locutores, en un sistema de VL independiente de texto.

### 1.3.2 Sistemas de Reconocimiento de Locutores

La VL encuentra en nuestros días aplicación en sistemas reales, normalmente complementando algún otro sistema de identificación (por ejemplo tarjetas magnéticas, huellas dactilares, etc.). Sin embargo, la aplicación de sistemas de IL es problemática por las menores prestaciones que proporcionan dado su mayor nivel de complejidad, pues se exigen tasas de error muy próximas al 100 por 100. Existen varios sistemas de VL dependientes de texto que consiguen tasas de rechazo de locutores correctos y de aceptación de impostores inferiores al 1 por 100 en aplicaciones de control de acceso para poblaciones de hasta 200 locutores. Típicamente estos sistemas requieren un tiempo de entrenamiento de dos a tres minutos, y un tiempo de verificación típico de 10 s.

Una de las áreas de aplicación importante para la VL es a través de la red telefónica (por ejemplo, para operaciones de tele-banco y acceso restringido a sistemas cerrados). Desafortunadamente las prestaciones de los actuales sistemas de reconocimiento de locutores se degradan en entornos telefónicos de manera importante. Esto es debido, principalmente, a la reducción del ancho de banda, la presencia de ruidos y los diferentes tipos de micrófonos y canales, que añaden una variabilidad adicional al habla de un locutor determinado.

Las mejores tasas de rechazo y falsa aceptación de locutores, para VL a través de la línea telefónica, son del orden del 5 por 100 y 1 por 100, respectivamente. Claramente es necesaria una actividad investigadora intensa dirigida a mejorar las capacidades de discriminación y los procedimientos de extracción de información característica de la identidad de un locutor.

## 1.4 Estado del arte en codificación de voz

La tecnología dedicada a la compresión de voz telefónica (ancho de banda 300 a 3.400 Hz) ha sido objeto de una intensa actividad investigadora durante décadas. En los últimos años, y como consecuencia del alto grado de madurez alcanzado, ha surgido una extensa actividad orientada al desarrollo de numerosas aplicaciones tanto en el ámbito de la transmisión (telecomunicaciones) como en el del almacenamiento (informática). Durante ese proceso han nacido varios estándares tanto nacionales como internacionales, siendo el sector de Estandarización de Telecomunicaciones de la Unión Internacional de Telecomunicación, UIT-T (el sucesor del CCITT), el organismo más involucrado en esta actividad. Especialmente importante en nuestros días es la aplicación de la codificación de voz a la transmisión radio digital, sobre todo en los futuros Sistemas de Comunicaciones Personales (PCS).

Junto a la codificación de voz ha surgido también un importante crecimiento en investigación y desarrollo de procedimientos de codificación de señales de audio (anchos de banda de 20 Khz.) para transmisión y almacenamiento con calidad de Compact Disc (CD), y voz de banda ancha (7 Khz.) para aplicaciones de telé conferencia. Dentro del proceso de estandarización en este ámbito destaca en la actualidad el algoritmo de codificación de audio incluido en el estándar MPEG de la ISO.

El objetivo básico de la codificación de voz y audio es lograr un compromiso entre velocidad binaria y degradación introducida por el proceso de codificación. Con unas exigencias muy inferiores en la codificación de voz telefónica, donde los usuarios, hoy por hoy, demandan relativamente poca calidad, frente a la codificación de audio, que requiere una calidad próxima a la proporcionada por el CD que se toma como referencia.

### 1.4.1 Principales áreas de trabajo en codificación de voz

Se presenta un resumen de las principales áreas de trabajo en este ámbito:

- Codificación a velocidad binaria variable.
- Codificación de bajo retardo.
- Relación Codificación de fuente-Codificación de canal.
- Enmascaramiento del ruido de Cuantificación.
- Procedimientos objetivos de evaluación de la calidad.

## Codificación a velocidad binaria variable

Disponer de una velocidad binaria variable comienza a ser imprescindible en muchas aplicaciones de telecomunicaciones, especialmente para sistemas de telefonía celular digital que incorporan Acceso Múltiple por División de Códigos (CDMA) y transmisión digital con Modo de Transferencia Asíncrono (ATM). Obtener una velocidad variable puede lograrse explotando la presencia de silencios en la voz y codificando de forma diferenciada los diferentes tipos de sonidos.

Típicamente, el cambio de velocidad de un codificador puede hacerse en un intervalo de 10 ms, controlándose, bien internamente (por ejemplo atendiendo a los canales de la voz entrante), o bien externamente (por ejemplo en función del nivel de tráfico).

## Codificación de bajo retardo

La mayor parte de los esquemas de codificación de voz más evolucionados introducen retardos en el proceso de codificación/decodificación del orden de 60 a 100 ms. En muchos sistemas de comunicación este retardo es crítico para el correcto funcionamiento del mismo, por este motivo en 1988 el CCITT (ahora UIT-T) estableció para su estándar a 16 kbit/s unos requerimientos de retardo máximo de cinco milisegundos, con un objetivo deseable de dos milisegundos. Estas exigencias han motivado el diseño de procedimientos que eviten retardos y estén basados en la obtención de los parámetros, de un codificador, de una forma recurrente a partir de la señal que se va sintetizando.

## Relación Codificación de fuente-Codificación de canal

Un aspecto clave en la incorporación de codificadores de voz sobre aplicaciones de comunicación radio, es la interrelación entre el proceso de codificación de fuente, que realiza el codificador, y la codificación de canal, que emplea el sistema de comunicación. El punto más importante a tener en cuenta es determinar cómo los posibles errores que se produzcan en el proceso de transmisión degradan la calidad que proporciona el codificador. Puede suceder que un codificador, que proporciona una mejor calidad que otro en ausencia de errores de transmisión, se degrade más que el segundo en presencia de errores. Por tanto, son actividades importantes: el estudio de la robustez de los codificadores frente a errores de transmisión, el diseño de técnicas de protección de los bits que representan los parámetros más sensibles del codificador y la incorporación de procesos de suavizado de parámetros cuando se detecten errores en los mismos.

## Enmascaramiento del ruido de Cuantificación

Cualquier codificador de voz o de audio introduce ruido en el proceso de compresión. Este ruido puede controlarse atendiendo a las diferencias entre la señal de entrada y la señal sintetizada, pero lo realmente importante es que el receptor, el oyente, perciba las dos señales con las menores diferencias posibles. Entramos, por tanto, en el ámbito de conocimiento del proceso de percepción de sonidos, que presenta fenómenos importantes como el conocido como enmascaramiento: nuestro oído, en determinadas condiciones, es incapaz de percibir un sonido en presencia de otro; diremos que este segundo sonido enmascara al primero. Concluimos entonces que, si un codificador siempre introduce ruido, su éxito se basará en conseguir que ese ruido no se oiga, que esté enmascarado. La investigación sobre los procesos acústicos y psicológicos que influyen en el proceso de percepción de sonidos es hoy un aspecto incorporado en algunos codificadores de voz, y clave en la codificación de ancho a través de los denominados modelos psicoacústicos.

## Procedimientos objetivos de evaluación de la calidad

En estrecha relación con el punto anterior, puede comprenderse fácilmente que la evaluación de la calidad de un codificador no podrá hacerse a través de medidas clásicas tales como la relación señal a ruido. Es preciso incluir procedimientos dirigidos a replicar los mecanismos propios del proceso de audición. Disponer de este tipo de procedimientos evitaría el tedioso y costoso procedimiento que supone realizar una evaluación subjetiva a partir de una determinada población de oyentes.

### 1.4.2 Principales esquemas de codificación de voz

La señal de entrada de los diferentes esquemas de codificación de voz actuales, puede pertenecer a una de las siguientes categorías:

- Voz Telefónica: Ancho de banda aproximado 3,2 kHz y frecuencia de muestreo 8 kHz.
- Voz de Banda Ancha: ancho de banda 20 kHz, frecuencia de muestreo 16 kHz.
- Ancho de Alta Calidad: Ancho de banda 20 kHz, frecuencia de muestreo 44,1 kHz o 42 kHz.
- Ancho de Calidad Media: Ancho de banda 15 kHz, frecuencia de muestreo 32 kHz.

La familia de esquemas de codificación para voz telefónica más extendida actualmente es la denominada Codificación por Predicción Lineal con Excitación por Código (CELP). La codificación CELP cuenta como principales representantes:

- El estándar federal USA FS-1016 para 4,8 kbit/s.
- El estándar americano y japonés para telefonía digital TDMA denominado VCELP (recientemente PSI-CELP para el estándar japonés de velocidad media).
- El estándar CCITT (ahora UIT-T) para codificación a bajo retardo a 16 kbit/s: LD-CELP (Low Delay CELP).
- La mayoría de los candidatos al estándar de velocidad media para telefonía móvil GSM en Europa y la Asociación de Industria Telefónica (TIA) de Estados Unidos.
- Estándar TIA para telefonía celular CDMA: QCELP de velocidad variable. Para velocidades en torno a 2 kbit/s, tras el clásico estándar FS-1015 conocido como vocoder LPC-10, es objeto de una investigación activa el denominado Vocoder con Excitación Multibanda (MBE) adoptado como estándar INMARSAT (comunicaciones marítimas vía satélite).

Finalmente, para voz de banda ancha y audio, son destacables los siguientes esquemas:

- El estándar UIT-T G722, esquema de Codificación en Sub-bandas ADPCM para sonidos de ancho de 7 kHz.

Codificador de audio para la norma ISO/MPEG, basado en modelos psicoacústicos. Permite señales de entrada muestreadas a 32, 44,1 y 48 Khz., y proporciona velocidades entre 32 y 448 kbit/s monocanal (un canal estéreo).

# CAPITULO II

# ESTUDIO TEÓRICO DE LOS CIRCUITOS DE VOZ ISD1420 Y ISD4002

## 2.0 Generalidades

La tecnología chipcorder patentada por ISD se caracteriza por guardar información analógica, en el mundo de las memorias semiconductoras, esto se traduce en un almacenamiento de información sin la necesidad de conversores análogo digital ni digital análogo.

En el presente capitulo se estudian los integrados de voz en cuestión (ISD), se presentan diagramas de bloques internos de 2 familias de integrados, los ISD 1400, y los ISD4002.

Para el correcto funcionamiento de un chip de Voz se requieren algunos dispositivos externos: micrófonos, parlantes, interruptores, algunas resistencias, condensadores y alimentación. Con estos dispositivos es suficiente para formar un sistema de grabación y reproducción del sonido, los demás elementos: preamplificadores, filtros, control automático de ganancia, amplificadores de poder, control lógico y almacenamiento analógico, están disponibles en el interior del integrado. Estas son las familias ISD1100, ISD1200, ISD1400 y ISD2500 las cuales se controlan en forma paralela, y pueden interactuar directamente con sistemas alimentados a 5 Volts.

Una nueva serie de dispositivos chipcorder surgen con la aparición de la familia ISD33000, los cuales funcionan con 3 Volts, un puerto de control serial y además se proyectan para ser incorporados en mercados específicos ya en operación. Por ejemplo no traen incorporados en el chip el amplificador de micrófono ni el control para parlantes, lo que los hace una solución más rentable para esas aplicaciones que ya tienen incorporadas estas funciones en otra parte del circuito.

**2.0.1 Funcionamiento general de los ISD**

Durante grabación los dispositivos ISD ejecutan diversas etapas de acondicionamiento de la señal antes de grabarla.

Por ejemplo en aquellos productos que incluyen en el interior del chip, control automático de ganancia, la primera etapa se compone de un preamplificador, amplificador, y bloques de control automático de ganancia. El preamplificador es conectado al micrófono a través de un condensador de acoplamiento, el cual filtra los niveles bajos de la señal alterna (2- 20mv). La amplificación se realiza en dos etapas, inicialmente por el preamplificador de entrada y luego por el amplificador de ganancia. El camino de la señal se completa conectando un condensador entre los pines ANA OUT y ANA IN.

El circuito AGC monitorea en forma dinámica el nivel de señal del amplificador de salida y envía un voltaje de control de ganancia al preamplificador, la ganancia del preamplificador se ajusta automáticamente para mantener una señal de entrada óptima al filtro.

Las características del AGC se describen por dos constantes de tiempo, el tiempo de carga y el tiempo de descarga.

- El tiempo de carga es el requerido por el AGC para responder a un incremento en la señal de entrada.
- El tiempo de descarga es la constante de tiempo del incremento de ganancia en presencia de una señal decreciente.

El usuario puede ajustar estos tiempos seleccionando los valores de los componentes conectados al pin AGC.

Todos los otros dispositivos (los que no tienen AGC) poseen una etapa de acondicionamiento de la señal proporcionada por un filtro de entrada. A pesar de que el almacenamiento se realiza en forma analógica sin alterar la señal es necesario emplear técnicas de muestreo. Por esto es necesario un filtro antialiasing, para filtrar los componentes de la frecuencia de entrada que están por encima de la mitad de la

frecuencia de muestreo. Esto es para satisfacer el ya conocido criterio de nyquist que se aplica a todos los sistemas de muestreo de datos.

Una calidad de voz mejor que la de telefonía, se alcanza con una frecuencia de muestreo de 8Khz, la frecuencia de corte del filtro pasa bajo es seleccionada a 3.4Khz. Satisfaciendo el criterio de nyquist.

Ahora el acondicionamiento de la señal esta completo y es pasada a través de tranceivers análogos para ser guardada en el interior de arreglos analógicos de almacenaje.

Las muestras son tomadas con una frecuencia de reloj 8.0Khz. Estas muestras son sometidas a un proceso de cambio de nivel de voltaje para el procedimiento de escritura en la memoria no volátil.

Durante el proceso de reproducción los voltajes analógicos guardados son leídos secuencial mente bajo el control de la frecuencia de reloj reconstruyendo así la forma de onda guardada. El filtro smooting sirve para eliminar los componentes de la frecuencia de muestreo de la forma de onda final.

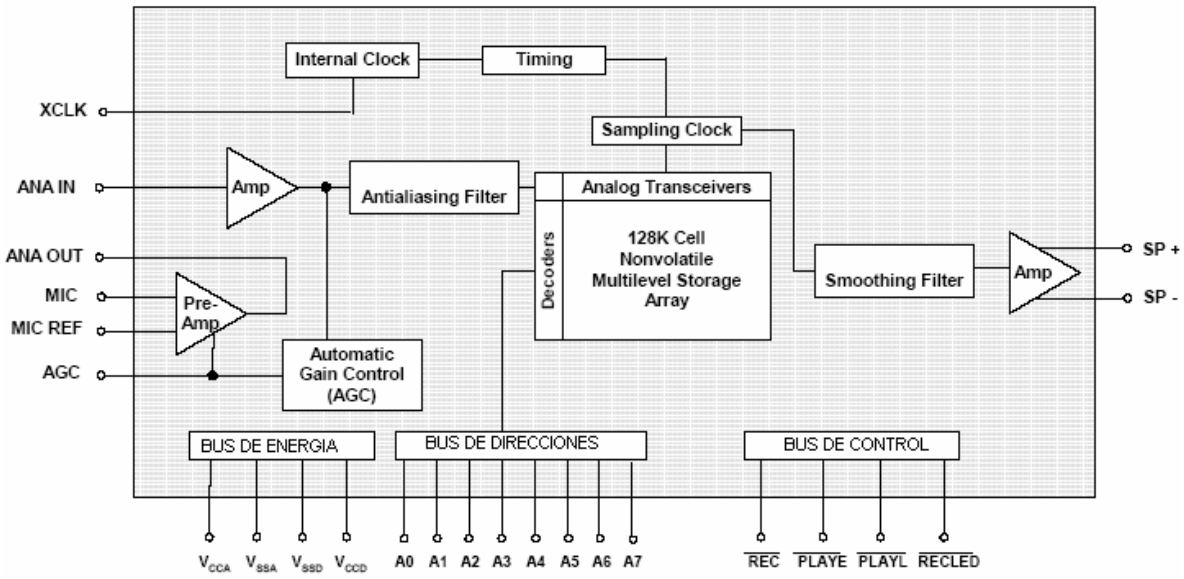
## 2.1 Descripción del chip de Voz 1420

El chip de voz ISD1420, pertenece a la familia de circuitos ISD1400, que fue una de las primeras en ser lanzada al mercado por la compañía Winbond, este circuito es capaz de grabar en su interior 20 segundos de sonido, y posee interfaz de comunicación paralela. Es un dispositivo construido con tecnología CMOS, y en su interior cuenta con un oscilador, preamplifcador de micrófono, control automático de ganancia, filtros smooting y antilaliasing y amplificador de parlante.

Para que este circuito opere se necesita agregar en forma externa un subsistema formado por un parlante, micrófono y algunos componentes pasivos. Las grabaciones son almacenadas en la memoria flash, con la característica de cero poder de consumo de almacenamiento, las señales de voz y audio son almacenadas directamente en su forma natural.

### 2.1.1 Arquitectura Interna

**Diagrama de bloques**



### 2.1.2 Descripción de pines

El integrado ISD1420, tiene 20 pines de entrada, 4 de salida y cuatro que no se utilizan, ya que no están habilitados.

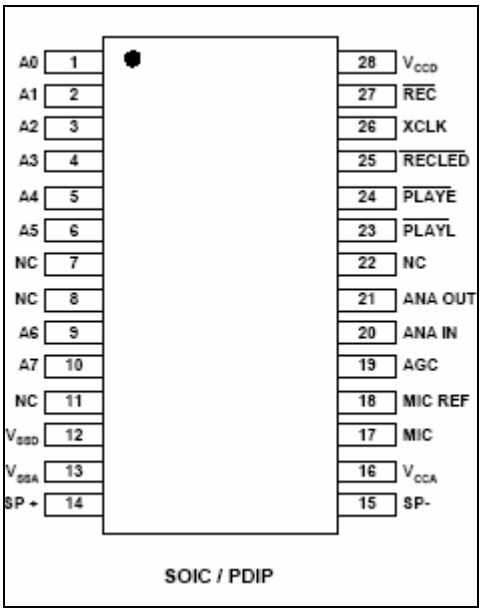
Las entradas son las siguientes:

1. Tiene 8 pines de direccionamiento que son desde A0 hasta A7.
2. Cuatro pines de alimentación dos analógicos y dos digitales con sus respectivas tierras (vccd, vcca, vssd, vssa).
3. Tres pines de control con los cuales se indica si debe grabar o reproducir sonidos (RECLED, PLAYE, PLAYL).
4. Dos pines para conectar el micrófono (MIC, MIC REF)
5. Un pin para entrada analógica, uno para conectar el oscilador externo y otro pin de control automático de ganancia.

Los pines de salida son:

dos para conectar el parlante (SP+, SP-), uno de salida analógica (ANA OUT), y un pin indicador

(REC).



| Nombre del pin | Numero | Función |
|---|---|---|
| A0 hasta A7 | 1, 2, 3, 4, 5, 6, 9, 10 | **Bus de direcciones:** Los pines de direccionamiento tienen dos funciones, dependiendo del nivel lógico en que se encuentren los dos bits más significativos MSB.<br>Si cualquiera de los dos MSB de direcciones esta en nivel bajo, todos los pines son interpretados como pines de direccionamiento, y la dirección que ellos indiquen será interpretada como la dirección inicial para comenzar un ciclo de grabado o reproducción de sonido.<br>El bus de direcciones es solamente entrada y no proporciona información del progreso interno de la operación. El bus de direcciones se activa con el flanco de bajada de los 3 pines de control (23, 24, 25).<br>Si los pines A6 y A7 están en nivel lógico alto, el chip esta en un modo de funcionamiento especial (ver modos de operación). |
| **Vssd y Vssa,** | 12 y 13 | **Tierra:** Los circuitos digitales y análogos internos de la serie ISD1400, poseen conexiones a tierra separadas para minimizar el ruido. Estos pines deben ser conectados juntos, tan cerca del encapsulado como sea posible. |
| **SP+ Y SP-** | 14 Y 15. | **Salidas de parlante:**<br>Estos pines proporcionan conexión directa a altavoces<br>con impedancias tan bajas como 16Ω. Una |

| | | |
|---|---|---|
| | | única salida puede ser utilizada, pero, para una conexión directa de los altavoces, las dos polaridades opuestas proporcionan una mejora en el poder de salida, este aumenta cuatro veces en comparación con la salida unida. Cuando sp+ y sp- son usados no se requiere conectar los parlantes a través de un condensador de adaptación. Una salida única requiere ser conectada a través de un condensador. Las salidas de parlante están en un estado de alta impedancia durante el ciclo de grabado y se conectan a Vssa cuando esta apagado. |
| **Vcca, Vccd** | 16 y 28 | **Voltaje de alimentación:** Estos pines deben ser conectados juntos tan cerca del encapsulado como sea posible. Y desacoplados de la itera con un condensador. |
| **MIC** | 17 | **Entrada de micrófono:** La entrada de micrófono transfiere esta señal al preamplificador interno del chip, cuya ganancia es controlada entre -15 Y 24 DB por un control automático de ganancia (CAG) interno al chip. Un micrófono externo debe ser conectado a este in a través de un condensador e serie. El valor de este condensador junto con la resistencia interna de 10K Ω de este pin determina la frecuencia de corte pasa banda para la SERIE ISD1400. |
| **MIC REF**, | | **Referencia de micrófono:** Corresponde al pin 18. Esta es la entrada inversa al preamplificador. Este proporciona una |

| | | |
|---|---|---|
| | | cancelación del ruido o un modo común de rechazo a la entrada del chip cuando se conecta un micrófono diferencial. |
| **AGC** | 19 | **Control automático de ganancia:** Ajusta la ganancia del preamplificador para compensar debido al amplio rango de variación de la señal de entrada de micrófono, este control permite trabajar con un amplio rango de sonidos, desde susurros hasta sonidos fuertes, los cuales son grabados con un mínimo rango de distorsión. El tiempo de ‘‘ataque’’ es determinado por la constante de tiempo formada por una resistencia interna de 5K Ω y un condensador externo (C6 en la figura esquemática) conectado desde el pin AGC y Vssa. El tiempo de descarga esta determinado por el circuito formado entre la resistencia R5 y el condensador C6 ambos conectados en paralelo entre el pin AGC Y Vssa. Los valores nominales de 470k Ω Y 4,7µf dan excelentes resultados. |
| **ANA IN** | 20 | **Entrada analógica:** Transfiere la señal a ser grabada al chip para entradas de micrófono el pin ANA OUT debe ser conectado con el ANA IN a través de un condensador. El valor de este condensador junto con la impedancia de entrada de 3kohm de ANA IN es seleccionada para dar un corte adicional a la frecuencia pasa banda de la voz. Si la señal deseada de entrada proviene de una fuente diferente a un micrófono, puede ser ingresada al chip directamente a través |

| | | |
|---|---|---|
| | | del pin ANA IN acoplada capacitivamente. |
| **ANA OUT** | 21 | **Salida analógica**. Este pin proporciona el preamplificador de salida para el usuario. El voltaje de ganancia para este preamplificador es determinado por el nivel de voltaje en el pin AGC. |
| $\overline{PLAYL}$ | 23 | **Reproducción (Activación por nivel)** Cuando este pin es llevado a un nivel lógico bajo un ciclo de reproducción comienza. La reproducción continúa hasta que $\overline{PLAYL}$ es llevado a nivel alto. |
| $\overline{PLAYE}$ | 24 | **Reproducción (edge activation)** Cuando este pin es llevado a un nivel bajo la reproducción comienza y continúa aun cuando el pin sea llevado a nivel alto. |
| $\overline{RECLED}$ | 25 | **Led de grabado** Es un pin de salida que entrega un nivel bajo cuando, el circuito se encuentra en proceso de grabación, además, entrega un pulso de nivel bajo cuando un fin de mensaje es encontrado en un proceso de reproducción. |
| XCLK | 26 | **Reloj externo** La entrada tiene un dispositivo interno desplegable El ISD1400 se configura en la fábrica con una frecuencia interna de reloj que garantiza el tiempo nominal mínimo de grabado/reproducción. Por ejemplo, un ISD1420 que opera dentro de la especificación se observará que siempre tiene un mínimo de 20 segundos de grabación La frecuencia de muestreo se mantiene Entonces |

| | | |
|---|---|---|
| | | a una variación de+/- 2. 25 por ciento encima de la temperatura y voltaje que operación comercial. El reloj interior tiene una tolerancia de ±5 por ciento encima de la temperatura industrial y rango de voltaje. UN suministro de poder regulado se recomienda para temperaturas industriales Si se requiere una precisión mayor, el dispositivo puede cronometrarse a través de los pines de XCLK. |

## 2.2 Funcionamiento del ISD1420

La serie ISD1400 de Winbond ofrece frecuencias de muestreo de 6.4 y 8.0 Khz. Las muestras del discurso se guardan directamente en la memoria no volátil sin digitalizar, ni comprimir. El almacenamiento analógico directo proporciona a una muy verdadera y natural reproducción sonora de voz, música, tonos.

### 2.2.1 Características generales

**Duración**

La serie ISD1400 es capaz de almacenar información que va desde 16 segundos (ISD1416) hasta 20 segundos (ISD1420).

**Almacenamiento en la memoria EEPROM**

Uno de los beneficios de la tecnología ChipCorder® de Winbond es incorporar en el interior del circuito la memoria no volátil, además no se gasta energía en el almacenamiento de mensaje. El mensaje se retiene típicamente 100 años sin consumir energía. Además, el dispositivo puede re-grabarse encima 100.000 veces.

**Modo de apagado Automático**

Al final de una reproducción o ciclo de grabado, la serie de ISD1400 automáticamente entra en un modo de bajo consumo de energía, consumiendo 0.5 µA. Después de un ciclo de reproducción o grabado, el dispositivo entra en este modo automáticamente.

**Frecuencia para reloj externo**

Éstos estas frecuencias de reloj no deben variarse ya que pueden ocurrir problemas de aliasing si la proporción de la muestra difiere del lo recomendado. Si el XCLK no se usa, esta entrada debe conectarse a tierra.

| Dispositivo | Frecuencia de muestreo | Reloj requerido |
|---|---|---|
| ISD1416 | 8.0Khz | 1024Khz |
| ISD1420 | 6.4Khz | 819.2Khz |

**2.2.2 Métodos de grabado y reproducción del sonido del ISD 1420**

**Como grabar y reproducir un mensaje**

Para grabar mensajes en el interior del chip es necesario armar la configuración mínima necesaria para el funcionamiento de este circuito ver figura 3.2 (capitulo3), un ciclo de grabación comienza cada ves que el pin Rec. es llevado a nivel bajo, mientras este pin permanezca en bajo la grabación continuara, hasta que la memoria se llene o el pin sea llevado a nivel alto cuando esto ocurre el chip pone una marca de fin de menaje (EOM).

La grabación comenzará desde la dirección de memoria que indique el bus externo de direcciones, A7. . . A0, y esta dirección va entre los valores, 00000000 y 10100000.

Para reproducir los sonidos basta con indicar la dirección desde la cual se quiere comenzar la reproducción, y presionar el botan play (figura 3.2, capitulo 3).

También existen otros métodos de grabación y reproducción los cuales trabajan en conjunto con los modos operacionales.

Por ejemplo si se desea grabar un mensaje a continuación de otro es necesario utilizar el modo A4 “direccionamiento consecutivo”

En el capitulo 3 se realizan experiencias practicas que ayudan a comprender mejor los métodos de grabado y reproducción de sonido en el ISD1420.

## 2.3 Modos operacionales

La serie de ISD1400 tiene varios modos operacionales incorporados proporcionando máxima funcionalidad con componentes adicionales mínimos. Los modos operacionales usan los pines de direccionamiento, pero están mapeados fuera del rango de dirección normal. Cuando los dos bits más significativos del bus de direcciones (MSB), A6 y A7, están en alto, los demás pines se interpretan como pines de modo de funcionamiento y no como pines de direccionamiento. Por consiguiente, los modos operacionales y el direccionamiento directo no son compatibles y no pueden usarse simultáneamente.

Hay dos consideraciones importantes para usar los modos operacionales. Primeramente, todas las operaciones empiezan en la dirección 0, que es la dirección del principio. Las operaciones restantes pueden comenzar en otra dirección, dependiendo del modo de funcionamiento seleccionado. Además, el puntero de dirección se restablece a 0 cuando el dispositivo se cambia de modo de grabado a reproducción pero no de reproducción a grabado cuando A4 es alto en el Modo Operacional.

En segundo lugar, un Modo Operacional se ejecuta cuando cualquiera de las entradas de mando, $\overline{PLAYE}$, $\overline{PLAYL}$, o $\overline{REC}$, reciben un pulso bajo y los dos MSB están en alto. Esta operación se mantiene hasta que alguna de las señales de control se lleva a

nivel bajo, en este punto la dirección a ejecutar será la que indiquen en ese momento el puntero interno de direccionamiento.

### 2.3.1 Descripción de los Modos operacionales

**AO - Mensaje Cueing**

El mensaje Cueing le permite al usuario saltar a través de los mensajes, sin saber las direcciones físicas reales de cada mensaje. Cada pulso de nivel BAJO que llegue al pin PLAYL, cuando el pin A0 se encuentra en nivel alto causa que el puntero de direcciones salte al próximo mensaje. Para poder escuchar el mensaje es necesario poner el pin A0 en nivel bajo nuevamente Este modo sólo se usa para reproducir, y es utilizado en conjunto con el modo operacional A4.

**A1- borrar las marcas de fin de mensaje (EOM)**

Este modo de operación, permite grabar mensajes en forma secuencial y reproducirlos como un solo mensaje con solo una (EOM) al final del último mensaje. El modo de funcionamiento A1 funciona solamente en conjunto con el modo A4

**A3- repetición de mensaje**

Este modo permite repetir un mensaje como una reproducción sin fin. La reproducción de un mensaje desde el principio de la memoria. Presionando $\overline{PLAYE}$ , comenzara la reproducción, y presionando $\overline{PLAYL}$ , la reproducción se detiene.

**A4-Direccionamiento consecutivo**

En funcionamiento normal, el puntero de direccionamiento se reseteara cuando ha pasado a través de un (EOM). El modo A4 inhibe este receteo permitiendo que la grabación o reproducción se realice en forma consecutiva. Cuando el dispositivo esta en un estado estático, ni grabando ni reproduciendo, es posible recetear el contador de direcciones a cero, llevando este pin momentáneamente a cero. Este modo de

funcionamiento sirve tanto para grabar como para reproducir un mensaje. Ambos procesos son tal como se indica en la sección descripción de pines

El puntero de direcciones se recetea al pasar de un ciclo de reproducción a grabado y viceversa.

## 2.4 Diagramas de tiempo

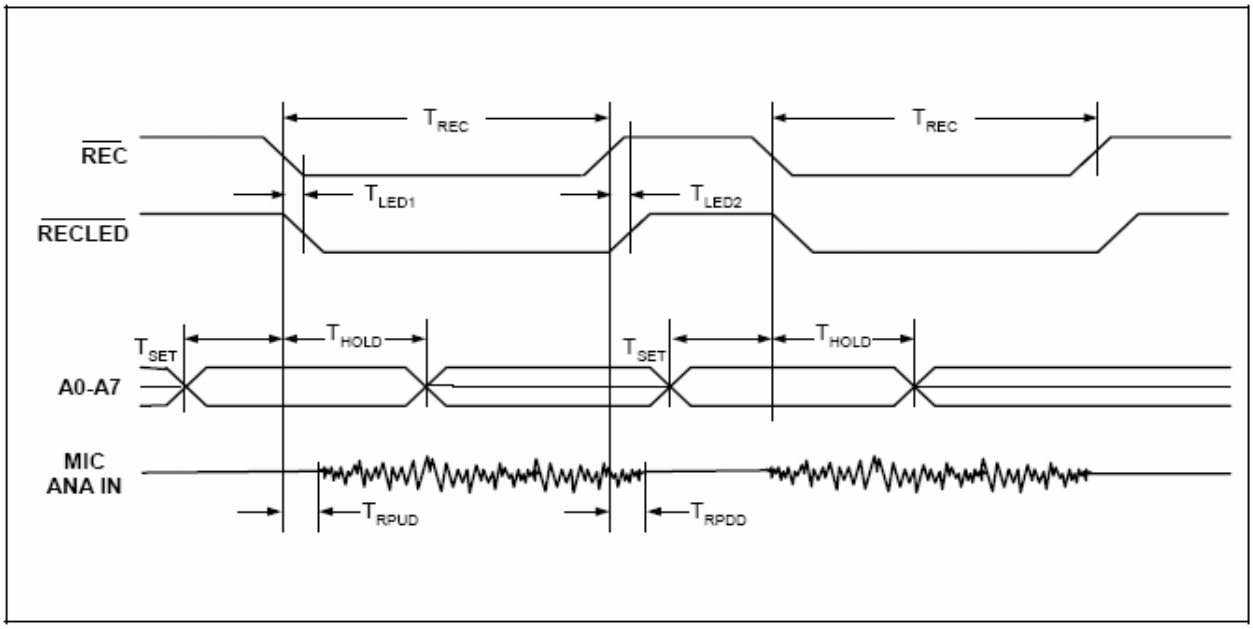



Diagrama de grabación

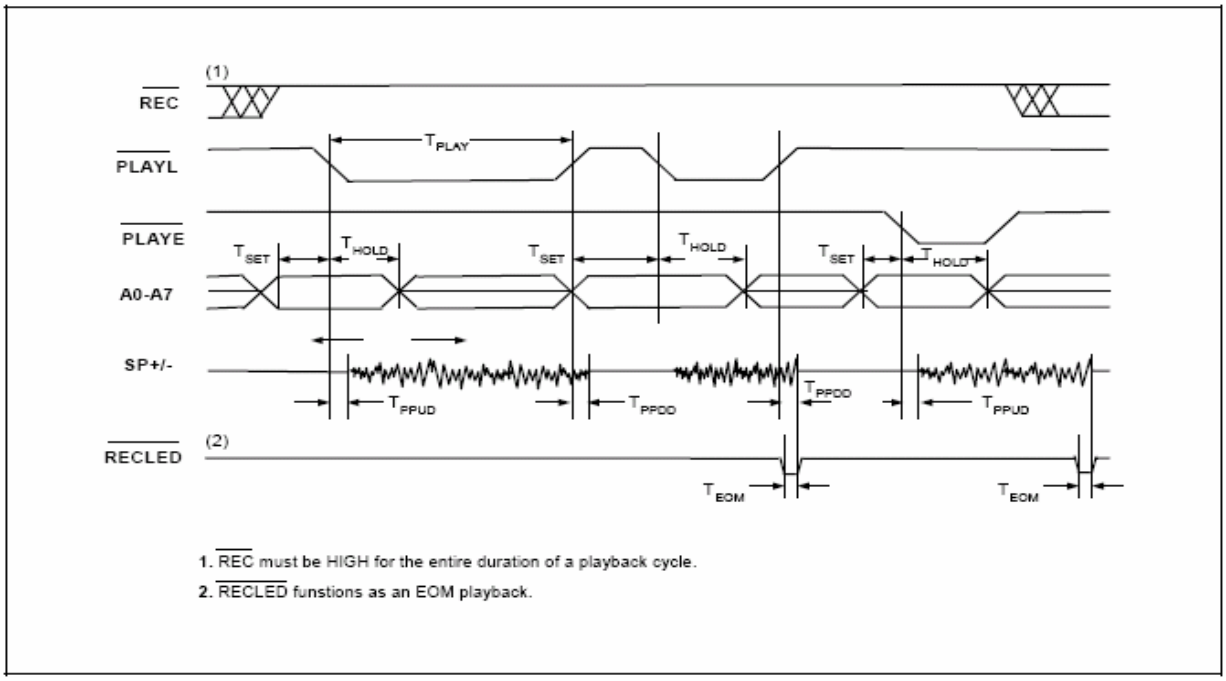



Diagrama de reproducción

## 2.5 Características eléctricas

| Condición | Valor |
|---|---|
| Temperatura de almacenamiento | -65ºC a +150ºC |
| Voltaje aplicado a cada pin | (Vss - 0.3V) a (Vcc + 0.3V) |
| Voltaje aplicado a cada pin con corriente 20mA | (Vss - 1.0V) a (Vcc + 1.0V) |
| Temperatura al soldar (durante 10 segundos) | 300ºC |
| Voltaje de alimentación | +4.5 a +6.5 Volts |

## Características en funcionamiento

| Condición | Valor |
|---|---|
| Temperatura (versión comercial) | 0ºC a +70ºC |
| Temperatura (versión industrial) | -40ºC a +85ºC |
| Voltaje de alimentación | +4.5V a +5.5V |

## Parámetros AC

| Característica | Símbolo | Min. | Tip. | Max. | Unidad |
|---|---|---|---|---|---|
| Frecuencia de muestreo<br>ISD1416<br>ISD1420 | Fs | | | 8.0<br>6.4 | KHZ<br>KHZ |
| Filtro pasa banda<br>ISD1416<br>ISD1420 | Fcf | | 3.3<br>2.6 | | KHZ<br>KHZ |
| Duración de grabado<br>ISD1416<br>ISD1420 | Trec | 16<br>20 | | | segundos |
| tiempo de reproducción | Tplay | 16 | | | Segundos |

| | | | | | |
|---|---|---|---|---|---|
| ISD1416<br>ISD1420 | | 20 | | | |
| $\overline{RECLED}$ ON | Tled1 | | 5 | | mseg. |
| $\overline{RECLED}$ OFF<br>ISD1416<br>ISD1420 | Tled2 | 30<br>40 | 38.9<br>48.6 | 95<br>100 | mseg. |
| Tiempo de direcciones | Tset | 300 | | | ns. |
| | Thold | 0 | | | ns. |
| Tiempo power up<br>ISD1416<br>ISD1420 | Trpud | | 26<br>32 | | mseg. |
| PD pulse width (record)<br>ISD1416<br>ISD1420 | Trpud | | 26<br>32 | | mseg. |
| PD pulse width (play)<br>ISD1416<br>ISD1420 | Tppud | | 26<br>32 | | mseg. |
| Trpdd<br>ISD1416<br>ISD1420 | Trpdd | | 26<br>32 | | mseg. |
| Tiempo Power up de play<br>ISD1416<br>ISD1420 | Tppud | | 26<br>32 | | mseg. |
| Tiempo de power down<br>ISD1416<br>ISD1420 | Tppdd | | 6.5<br>8.1 | | mseg. |

| | | | | | |
|---|---|---|---|---|---|
| Ancho del pulso $\overline{EOM}$<br>ISD1416<br>ISD1420 | Teom | | 12.5<br>15.625 | | mseg. |
| Distorsión armónica | THD | | 1 | 3 | % |
| Potencia de salida de parlante | Pout | | 12.2 | | mW |
| Voltaje entre los pines del parlante | Vout | | 1.25 | 2.5 | Vp-p |
| Entrada de voltaje de MIC | Vin1 | | | 20 | mVp-p |
| ANA IN voltaje de entrada | VIN2 | | | 50 | mVp-p |

## 2.6 Descripción del chip de voz ISD4002

Este circuito pertenece a la familia de circuitos de voz ISD con interfaz de control serial, con protocolo de comunicación SPI.

La familia de circuitos ISD 4002 se expone en la siguiente tabla.

| Nombre del dispositivo | Duración en segundos | Frecuencia de muestreo (Khz.) | Filtro pasa banda(Khz.) |
|---|---|---|---|
| ISD4002-120 | 120 | 8.0 | 3.4 |
| ISD4002-150 | 150 | 6.4 | 2.7 |
| ISD4002-180 | 180 | 5.3 | 2.3 |
| ISD4002-240 | 240 | 4.0 | 1.7 |

Tabla.4002A

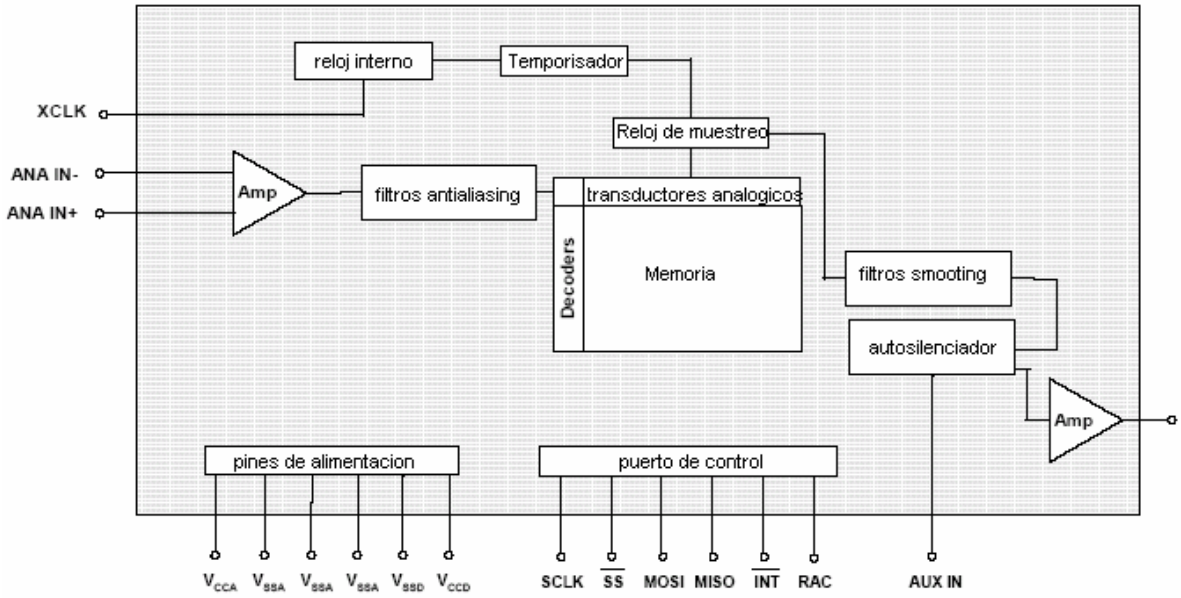
La serie ISD 4002, es una alternativa para aplicaciones electrónicas que requieran trabajar con voz, estos circuitos son útiles tanto para grabar como para reproducir sonidos, funcionan con una alimentación de tres volts, y el tiempo de almacenamiento de sonido esta entre 120 y 240 segundos dependiendo de la frecuencia de muestro (ver tabla 4002A).

Estos dispositivos están basados en tecnología CMOS, e internamente poseen un oscilador interno, filtros antialiasing, filtros smooting, características de auto silenciador, y memorias flash.

Estos son dispositivos de tipo esclavo los cuales fueron diseñados para ser utilizados en sistemas basados en microcontroladores o microprocesadores. El direccionamiento y el control se realizan a través de una interfaz serial periférica (SPI).

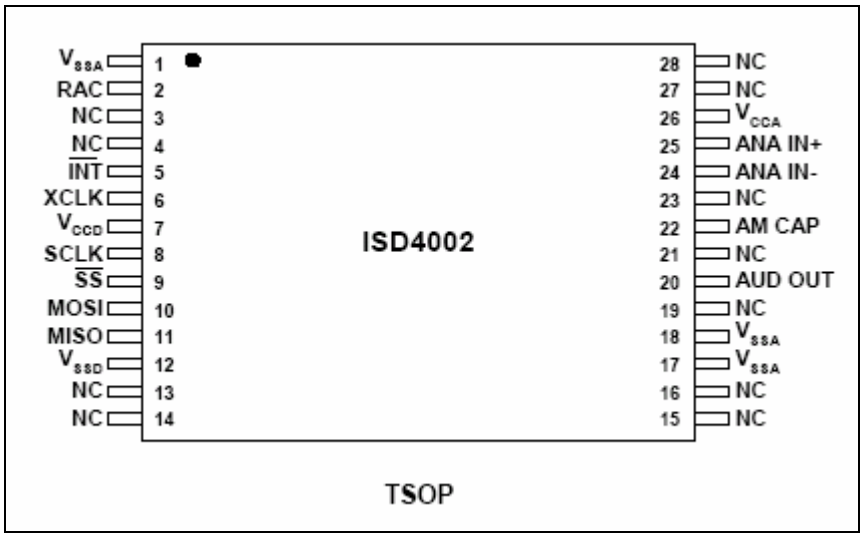
Las grabaciones son almacenadas en la memoria flash, con la característica de cero poder de consumo de almacenamiento, las señales de voz y audio son almacenadas directamente en su forma natural.

### 2.6.1 Arquitectura interna

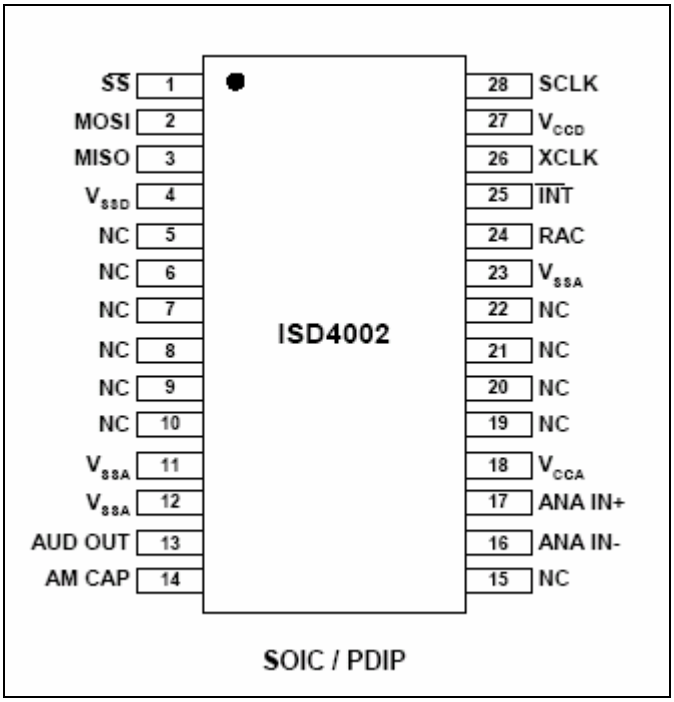


### 2.6.2 Configuración de pines

**Encapsulado TSOP**




**Encapsulado SOIC / PDIP**



### 2.6.3 Descripción y funcionamiento de cada pin

| Nombre | Pin numero PDIP/SOIC | Pin numero PDIP/SOIC | Función |
|---|---|---|---|
| SS | 1 | 9 | **Selección de esclavo**: al llevar este pin a nivel bajo , se selecciona el dispositivo |
| MOSI | 2 | 10 | **Salida de maestro entrada de esclavo:** Esta es la entrada serial al dispositivo ISD4002, el circuito maestro, (microcontrolador o microprocesador), pone el bit a ser enviado en el pin MISO medio ciclo entes del flanco de subida del reloj (SCLK) que sincroniza el intercambio de información entre el maestro y el esclavo. |
| MISO | 3 | 11 | **Entrada de maestro salida de esclavo:** este pin es la salida serial del dispositivo ISD4002 (drenaje abierto), este pin entrara en un estado de alta impedancia si el dispositivo no es seleccionado |
| Vssa/Vssd | 11,12,23,4 | 1,17,18,12 | **Tierras:** El ISD4002, utiliza buses de tierra separados para, los circuitos analógicos y digitales que posee internamente, las tierras analógicas (Vssa) deben ser conectadas juntas y a través de una baja impedancia a tierra.Los pines de tierra analógica (Vssd) serán conectados a la tierra a través de un camino de baja impedancia. Es importante asegurarse que la impedancia entre Vssa y Vssd sea menor que 3 ohm. |
| NC | 5-10, 15, | 3, 4, 13-16, | No conectar |

| | | | |
|---|---|---|---|
| | 19-22 | 19, 21, 23, 27, 28 | |
| AUD out | 13 | 20 | **Salida de audio:** este pin es la salida del audio almacenado en el circuito, y es capaz de manejar impedancias de 5Kohm, debe ser acoplado a la salida a través de un condensador. |

| Nombre | Pin numero PDIP/SOIC | Pin numero PDIP/SOIC | Función |
|---|---|---|---|
| AMP CAP | 14 | 22 | **Auto mute: La** característica de auto silenciador solo se aplica en operaciones de reproducción y aluda a minimizar el ruido (con 6db de atenuación). Se recomienda conectar este pin a tierra a través de un condensador de 1uf. Este condensador pasa a formar parte de un detector interno de amplitud de señal. Este pin puede ser deshabilitado conectándolo directamente a Vcca. |
| ANA IN- | 16 | 24 | **Entrada analógica invertida:** Este pin en la entrada analógica invertida del circuito y se utiliza en modo diferencial. La señal de entrada a este pin no debe superar los 16mvpp, y debe ser acoplada a través de un condensador, el cual debe ser del mismo valor que el que se conecte a ANA IN+ la impedancia de entrada del pin ANA IN- normalmente es de 56Kohm. En modo de funcionamiento single ended, el pin ANA IN- debe ser acoplado a tierra a través de un condensador de igual magnitud al |

| | | | |
|---|---|---|---|
| | | | utilizado en el pin ANA IN+. |
| ANA IN+ | 17 | 25 | **Entrada analógica no invertida:** Este pin transfiere la señal a ser grabada al interior del dispositivo. Para mas detalles ver la figura ANA IN MODES. |
| Vcca/Vcc d | 18/27 | 26/7 | **Alimentación:** Para minimizar el ruido los circuitos analógicos y digitales utilizan buses de alimentación separados. La alimentación es de +3V. |
| RAC | 24 | 2 | **Row addres clock:** esta es una salida de drenaje abierto (es necesario conectar una resistencia pull up para conectarla con otro dispositivo.) y entrega una señal que representa a una única línea de memoria, es decir representa el final y el fin de una línea. Para detalles de diagramas de tiempo de esta señal ver tabla de parámetros AC. |
| $\overline{INT}$ | 25 | 5 | **Interrupción:** También es una salida drenaje abierto por lo tanto es necesario conectar una resistencia pull up para conectarla con otro dispositivo. Esta señal se pondrá en nivel bajo y permanecerá en nivel bajo cuando se detecta un desborde de memoria o una marca de fin de mensaje. Esta señal se limpiara la próxima ves que comience un ciclo SPI. |
| XCLK | 26 | 6 | **Entrada de reloj externo:** Este pin tiene un dispositivo interno pull down. La serie ISD4002 esta configurada por defecto con una |

| | | | |
|---|---|---|---|
| | | | frecuencia de muestreo interna centrada a +/- 1% de la especificación. El reloj interno tiene una tolerancia entre -6/+4 % sobre excesos de temperatura, temperaturas industriales y rangos de voltaje. Si se requiere un reloj con mayor precisión se puede utilizar una configuración externa, como sigue:<br>Si la entrada de reloj externa no se utiliza debe ser conectada a tierra. |
| SCLK | 28 | 8 | **Reloj serial:** Este es el reloj de entrada al dispositivo y se encarga de sincronizar la entrada y salida de información de los pines MOSI/MISO respectivamente. Los datos ingresan el flanco de subida y salen en el flanco de bajada. |

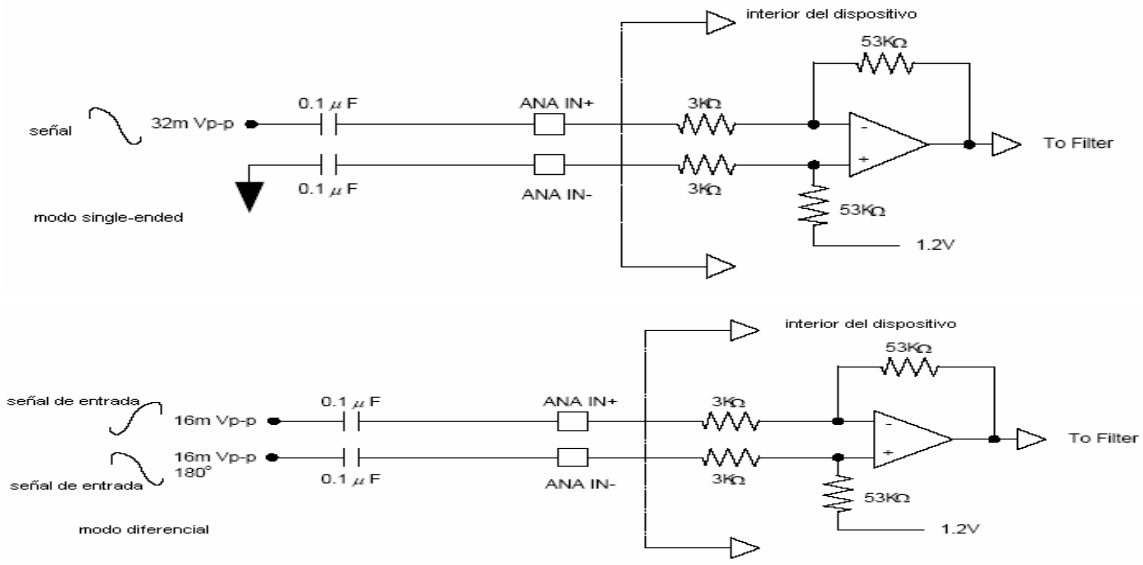
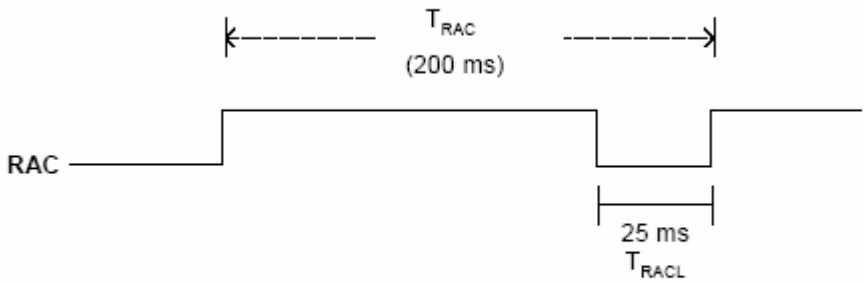



FIGURA ANA IN MODES




forma de onda de la señal RAC durante operacion normal
(ejemplo para una frecuencia de muestreo de 8KHz)

## 2.7 Funcionamiento de la serie ISD4002

### 2.7.1 Características principales

- **Calidad de audio:** La serie ISD4002 de Winbond, ofrece frecuencias de muestreo de 8.0, 6.4, 5.3, y 4.0 KHz. lo cual permite al usuario escoger la frecuencia que mejor le acomode considerando que a mayor frecuencia de muestreo es mejor la calidad del sonido almacenado pero disminuye el tiempo de grabación. Las muestras del sonido son grabadas directamente en su forma natural. Sin digitalizar ni comprimir, lo cual permite obtener una mayor calidad de sonido reproducido.
- **Duración:** La serie ISD 4002 nos ofrece 120, 150, 180 y 240 segundos de grabación.
- **Almacenamiento flash:** La serie ISD 4002 utiliza memorias flash, lo cual se traduce cero consumo de almacenamiento de mensaje, estos pueden ser retenidos típicamente por más de 100 años sin la necesidad de energizar el dispositivo. Además esta memoria puede ser regrabada 100.000 veces.

- **Interfaz con microcontrolador**: Este dispositivo utiliza una interfaz SPI, para sus funciones de direccionamiento, y esta diseñado para trabajar como un circuito esclavo. Solo cuatro líneas son necesarias para controlar el funcionamiento del dispositivo (MOSI, MISO, SCLK, $\overline{SS}$ ).

- **Programación**: La serie ISD 4002 también es ideal solo para funciones de reproducción de sonidos pregrabados, una ves que la configuración de sonidos fue creada, es posible reproducir los segmentos deseados con un programa adecuado.

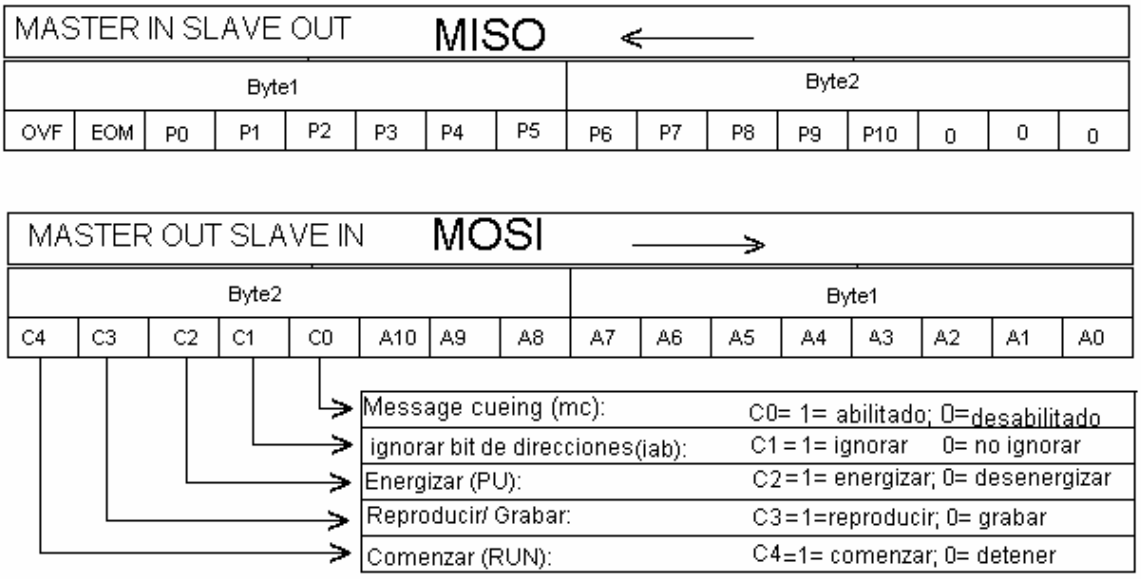
| Dispositivo | Frecuencia de muestreo | Reloj requerido |
|---|---|---|
| ISD4002-120 | 8.0 kHz | 1024 kHz |
| ISD4002-150 | 6.4 kHz | 819.2 kHz |
| ISD4002-180 | 5.3 kHz | 682.7 kHz |
| ISD4002-240 | 4.0 KHz | 512 KHz |

### 2.7.2 Funcionamiento de la interfaz serial periférica SPI

En primer lugar este protocolo de transferencia de datos asume que los registros del microcontrolador son cloqueados en el flanco de bajada de SCLK, nuestro dispositivo funciona de la siguiente manera.

1. Toda comunicación serial comienza con el flanco de bajada del pin selección de esclavo $\overline{SS}$.
2. $\overline{SS}$ permanece en nivel bajo durante cualquier comunicación serial y debe cambiarse a nivel alto entre instrucciones.
3. Un dato (BIT) ingresa al ISD en el flanco de subida del reloj, y un dato (BIT) sale en el flanco de bajada del reloj (SCLK).
4. Una operación de grabado o reproducción comienza llevando el $\overline{SS}$ a nivel bajo y enviando un código de operación.
5. Los códigos de operación contienen 11 bits de direcciones y 5 bits de control.
6. Cada operación que finalice con una marca de fin de mensaje o de un desborde de memoria generara una interrupción, la cual será limpiada la próxima vez que se inicie un nuevo ciclo SPI.
7. Como el dato de interrupción sale a través del pin MISO, y simultáneamente se esta enviando información. Se debe tomar precauciones como enviar datos compatibles con la operación en curso del sistema, ya que es posible leer un dato de interrupción y comenzar un nuevo ciclo SPI.
8. una operación comienza llevando a uno el bit RUN y finaliza llevando a cero este bit.
9. todas las operaciones comienzan después del flanco de subida de $\overline{SS}$.

PUERTO SPI DEL CHIP DE VOZ ISD4002




El puerto SPI esta conformado por dos registros de desplazamiento uno es el MISO (master in slave out), entrada al maestro y salida del esclavo, este registró entrega información al circuito de control maestro, en primer lugar entrega el bit de desborde de memoria, seguido de el bit de fin de mensaje y luego entrega la dirección a la cual quedo apuntando el puntero interno de direcciones del ISD (P0-P10). El segundo registro el MOSI (master out slave in) salida del maestro y entrada al esclavo, es el utilizado para entregar al ISD la dirección desde la cual debe comenzar a trabajar (A0-A9), A10 toma la condición sin cuidado(X), y los bit de control (C0-C4). El primer bit a enviar es el A0, y el último es el C4.

### 2.7.3 Códigos de operación.

En la siguiente tabla se presenta un resumen de los códigos de operación para el ISD4002

| Instrucción | direcciones (A0-A9,0) | Bits de control ,5 bits C0 C1 C2 C3 C4 | Función |
|---|---|---|---|
| POWER UP | (X-X,0) | 0 0 1 0 0 | Energizar el dispositivo |
| SET PLAY | (A0-A9,0) | 0 0 1 1 1 | Comienza un ciclo de reproducción desde la dirección especificada |
| PLAY | | 0 1 1 1 1 | Reproduce desde la dirección actual hasta una EOM, OVF |
| SET REC | (A0-A9,0) | 0 0 1 0 1 | Comienza un ciclo de grabación desde la dirección;A0-A9 |
| REC | | 0 1 1 0 1 | Reproduce desde la dirección actual hasta un OVF o un comando de stop |
| SETMC | (A0-A9,0) | 1 0 1 1 1 | Inicia un salto de mensajes desde la dirección (A0-A9,0) |
| MC[2] | | 1 1 1 1 1 | Ejecuta salto de mensaje desde la dirección actual procede al final de un mensaje o entra en estado de OVF si no se encuentran mas mensajes. |
| STOP | (X-X,0) | 0 1 1 X 0 | Detiene la operación en curso. |
| STOPP WRDN | (X-X,0) | X 1 0 X 0 | Detiene la operación e curso y entra en estado de reposo. |
| RINT | (X-X,0) | 0 1 1 X 0 | Lee el estado de los bits de interrupción, OVF y EOM. |

Nota:

C0 = Salto de mensaje (message cueing)

C1 = Ignorar bits de direcciones (Ignore address bits)

C2 = Control de encendido o apagado (Master power control)

C3 = Selección de grabado o reproducción (Record or play operation)

C4 = habilitar o deshabilitar una operación

[2] la opción de salto de mensajes solo puede ser seleccionada al principio de una operación de reproducción.

### 2.7.4 funcionamiento del ISD4002

Salto de mensajes

(MC) permite al usuario saltar a través de mensajes. Sin conocer la localización física actual de ese mensaje. La reproducción se detendrá cuando se encuentre una marca de fin de mensaje. En este momento el puntero de líneas de mensaje se incrementara y quedara apuntando al siguiente mensaje. Al encontrarse con el final de la memoria el dispositivo entrara en la condición de OVF.

### 2.7.5 Secuencia de inicialización del ISD4002

El ISD4002 estará listo para funcionar después que se le envíe la secuencia de inicialización. Se recomienda utililizar la siguiente secuencia para optimizar las operaciones de grabado y reproducción.

Modo de grabación.

1. Enviar el comando POWERUP.
2. Esperar el tiempo Tpud.
3. Enviar el comando POWERUP.
4. Esperar 2x Tpud.
5. a.) Enviar el comando SETREC con la dirección deseada o

b.) Enviar el comando REC

6. Enviar el comando STOP para detener la grabación.
7. Esperar el tiempo Tstop/pause

Para 5.a.) El dispositivo comenzara a grabar desde la dirección deseada y generara una interrupción cuando encuentre el final de la memoria, si no se envía un comando STOP antes de ese evento, el dispositivo se detendrá automáticamente.

Modo de reproducción

1. Enviar el comando POWERUP.
2. Esperar el tiempo Tpud.
3. a.)Enviar el comando SETPLAY con la dirección deseada, o
   b.)Enviar el comando PLAY (reproducción desde la dirección actual)
4. a.) Enviar el comando STOP
   b.)Esperar que la reproducción finalice al encontrar una marca de fin de mensaje o un fin de memoria.
5. Esperar Tstop/pause.

Para 3.a.) El dispositivo comenzara a reproducir en la dirección deseada y generara una interrupción al encontrar un EOM o un OVF.

## 2.7.9 Diagramas de tiempo.

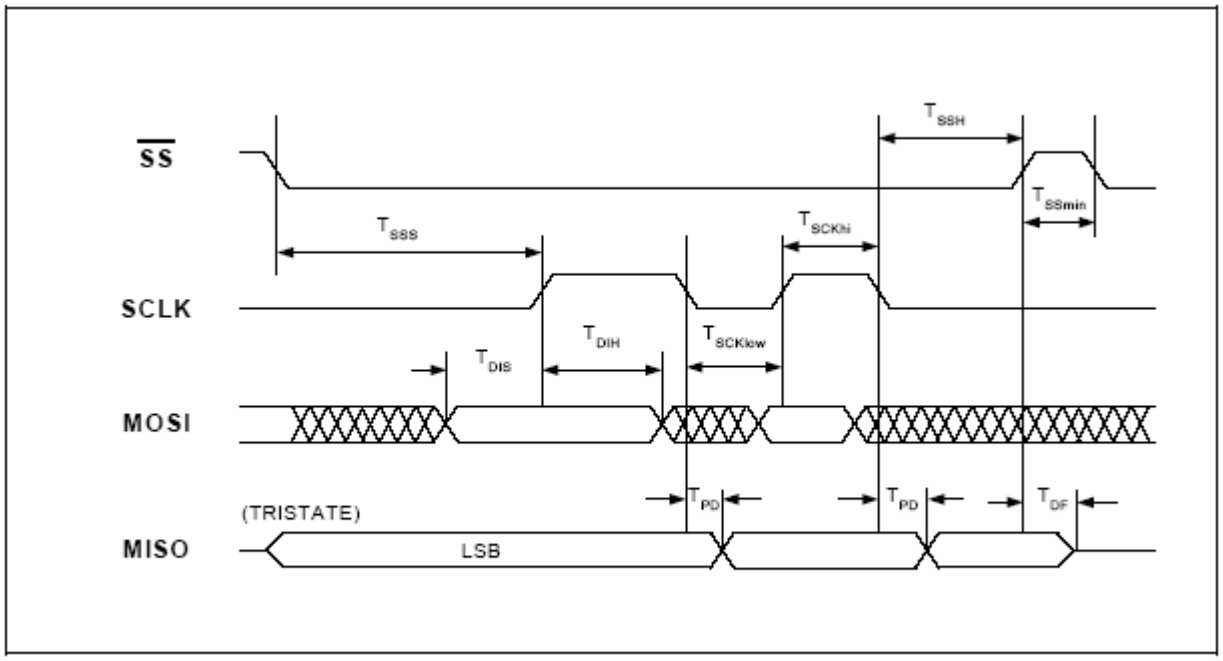
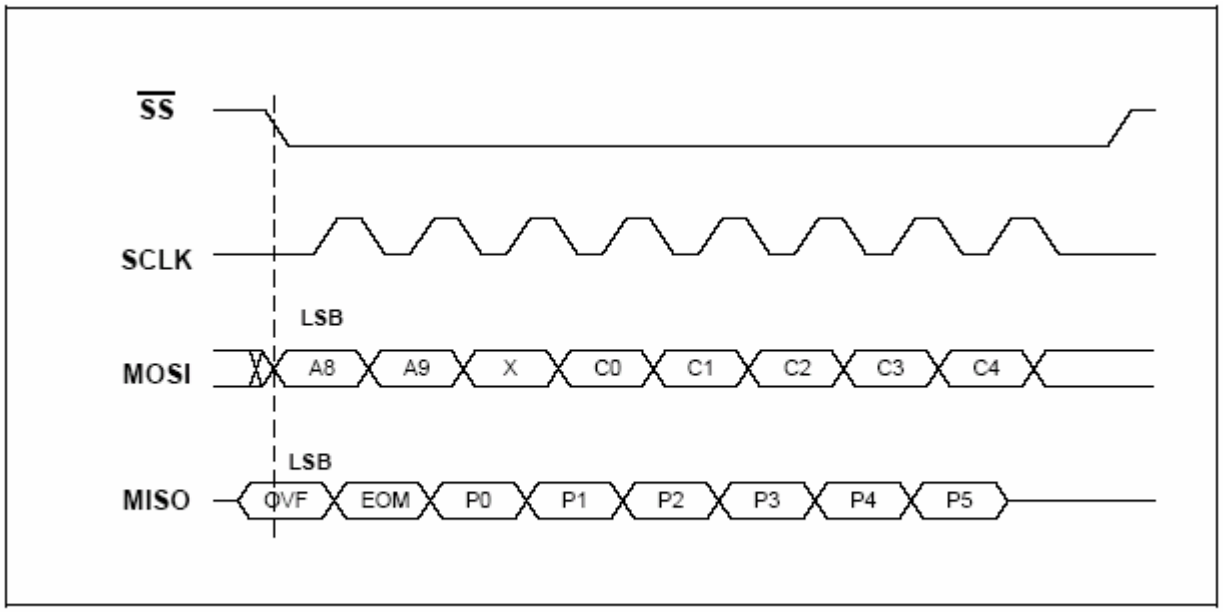
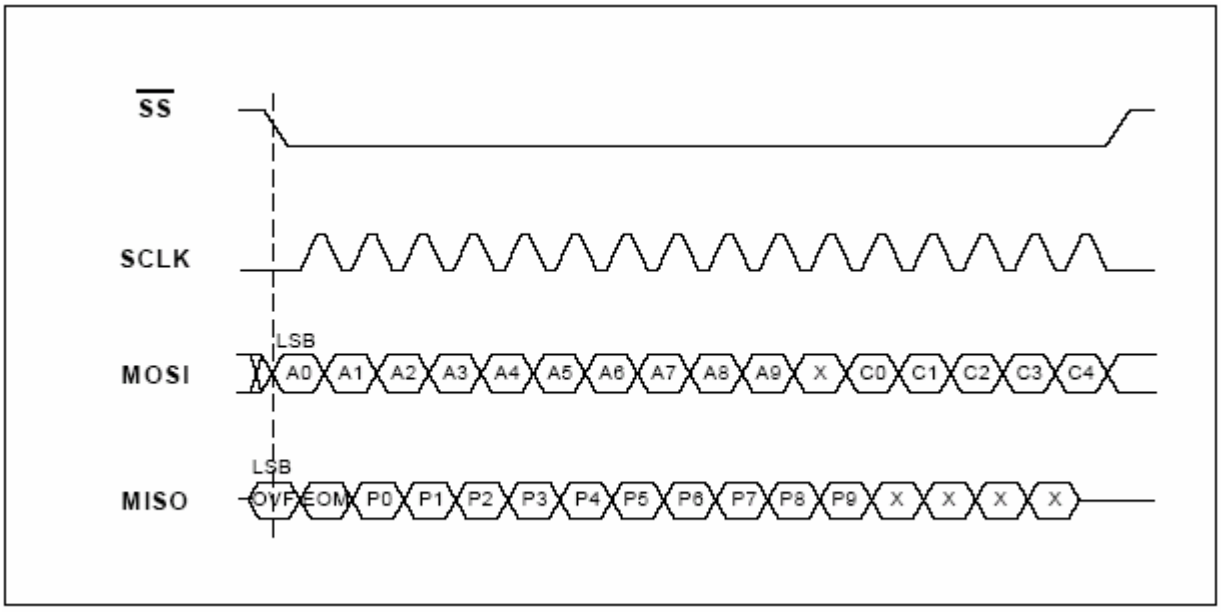



Diagrama de tiempo

| Símbolo | Mínimo | Máximo | Unidad |
|---|---|---|---|
| Tsss | 500 | | n seg. |
| TssH | 500 | | n seg. |
| TDIS | 200 | | n seg. |
| TDIH | 200 | | n seg. |
| TPD | | 500 | n seg. |
| TDF | | 500 | n seg. |
| TSSmin | 1 | | u seg. |
| TSCKHI | 400 | | n seg. |




Formato de comandos de 8 bit




Formato de comandos de 16 bits

## 2.8 Características eléctricas

| Condición | Valor |
|---|---|
| Temperatura de almacenamiento | -65ºC a 150ºC |
| Voltaje aplicado a cada pin | (Vss - 0,3V) a (Vcc + 0,3V) |
| Voltaje aplicado a cada pin con la corriente de entrada limitada a ± 20mA | (Vss – 1V) a (Vcc + 1V) |
| Voltaje aplicado a los pines MOSI, SCLK, y $\overline{SS}$ , con la corriente de entrada limitada a ± 20mA. | (Vss- 1.0 V) a 5.5V |
| Temperatura durante periodos de soldadura. (10 segundos) | 300ºC |
| Vcc – Vss | -0.3V a + 7.0V |

## Características en funcionamiento

| Condición, en funcionamiento | Valor |
|---|---|
| Temperatura del encapsulado | 0ºC a 70ºC |
| Temperatura del encapsulado en versión extendida | -20ºC a 70ºC |
| Temperatura de funcionamiento para versiones industriales | -40ºC a +85ºC |
| Voltaje de alimentación (Vcc)[1] | +2,7V a 3,3 V |
| Votaje de tierra (Vss)[2] | 0V |

[1] VCC = VCCA = VCCD

[2] VSS = VSSA = VSSD

## Parametros AC

| Característica | Símbolo | Típico | Máximo | Unidad |
|---|---|---|---|---|
| Frecuencia de muestreo<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | FS | 8.0<br>6.4<br>5.3<br>4.0 | | KHz<br>KHz<br>KHz<br>KHz |
| Filtro pasa banda<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | FCF | 3.4<br>2.7<br>2.3<br>1.7 | | KHz<br>KHz<br>KHz<br>KHz |
| Tiempo de grabado<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TREC | 120<br>150<br>180<br>240 | | sec<br>sec<br>sec<br>sec |
| Tiempo de reproduccion<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TPLAY | 120<br>150<br>180<br>240 | | sec<br>sec<br>sec<br>sec |
| Tiempo power up<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TPUD | 25<br>31.25<br>37.5<br>50 | | msec<br>msec<br>msec<br>msec |
| Tiempo de stop<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TSTOP or TPAUSE | 50<br>62.5<br>75<br>100 | | msec<br>msec<br>msec<br>msec |
| Periodo de rac<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TRAC | 200<br>250<br>300<br>400 | | msec<br>msec<br>msec<br>msec |

| | | | | |
|---|---|---|---|---|
| Tiempo de RAC en nivel bajo<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TRACL | 25<br>31.25<br>37.5<br>50 | | msec<br>msec<br>msec<br>msec |
| Periodo del RAC en mensaje cueing<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TRACM | 125<br>156.3<br>187.5<br>250 | | µsec<br>µsec<br>µsec<br>µsec |
| Tiempo del RAC en nivel bajo para MC<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | TRACML | 15.63<br>19.53<br>23.44<br>31.25 | | µsec<br>µsec<br>µsec<br>µsec |
| Distorsion armonica<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | THD | 1 | 2 | % |
| Voltaje de entrada a el pin ANA IN +<br>ISD4002-120<br>ISD4002-150<br>ISD4002-180<br>ISD4002-240 | VIN | | 32 | mV |

## 2.9 ISD4002 v/s ISD1420

### 2.9.1 Diferencias

| Característica | ISD1420 | ISD4002 |
|---|---|---|
| Tiempo de grabación | 20 segundos | 120 segundos<br>150 segundos<br>180 segundos<br>240 segundos |
| Voltaje de alimentación | Vcc 5 volts<br>Vss 0 volts | Vcc 3 volts<br>Vss 0 volts |
| Formas de direccionamiento | Las direcciones de memoria pueden ser indicadas en forma manual o a través de un microcontrolador o microprocesador. | Las direcciones de memoria solo pueden ser indicadas por medio de un microprocesador o microcontrolador. |
| Métodos de comunicación | Posee interfaz paralela | Posees interfaz serial con protocolo SPI |
| Encapsulado | SOIC/PDIP | SOIC/PDIP / TSOP |

### 2.9.2 Ventajas

| Característica | ISD4002 | ISD1420 |
|---|---|---|
| Tiempo de grabación | Posee mas memoria lo cual se traduce en mayor tiempo de grabación. | Posee menos memoria, lo cual se refleja en menor tiempo de grabación. |
| Métodos de comunicación | El hecho de tener un bus de comunicación | Para funciones de direccionamiento y |

| | | |
|---|---|---|
| | serial, para interactuar con el dispositivo maestro, se traduce en una ventaja a la hora de diseñar el hardware ya que se ahorran una cantidad significativa de pines, destinados a las funciones de direccionamiento, y control, solo se utilizan 4 pines. | control se utililizan un total de 12 pines |

### 2.9.3 Desventajas

| Característica | ISD4002 | ISD1420 |
|---|---|---|
| Costos | La alimentación de 3 Volts, nos obliga a diseñar interfaces de adaptación de niveles, ya que la mayoría de los microcontroladores que se utilizan en chile funcionan con 5 volts. Esto se traduce en mayores costos asociados a nuestros diseños. Además el valor actual del circuito es de $9000. | Compatible en forma directa con microcontroladores alimentados a 5 Volts. El valor actual del circuito es de $6000. |
| Software | Se requiere un software sofisticado para poder controlar el circuito, | Es comunicación paralela, solo basta con poner la palabra binaria |

| | | |
|---|---|---|
| | como se trata de Comunicación serial el mínimo error en sincronía provoca que el dispositivo no funcione. | correcta en el bus de control. |

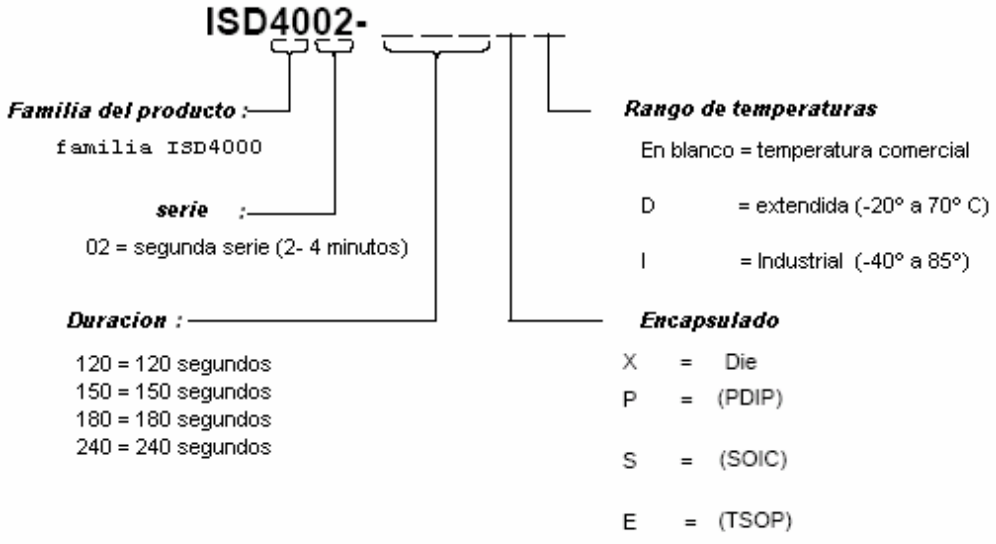
## 2.10 Clasificación de los integrados de Voz ISD

Estos dispositivos están separados en dos grupos, y estos grupos a la vez en sub. grupos, según el tiempo de grabación.

1. Según la interfaz de comunicación (paralela o serial)
2. Según el tiempo de grabación (capacidad de almacenaje.)

Los de interfaz paralela son los de las series 1XXX y 2XXX, y los de interfaz serial son los de la serie 3XXX en adelante.

A continuación se grafica un ejemplo de cómo se identifica un dispositivo de Voz ISD, se toma como ejemplo la serie ISD 4002.



## Familias de circuitos de Voz ISD

| Serie | Dispositivo | Duración | Memoria (K celdas) | Frecuencia de muestreo Khz. | Filtro pasa bandas Khz | DIP | SOIC | TSOP |
|---|---|---|---|---|---|---|---|---|
| ISD1100 | ISD1110 | 10 sec. | 64 | 6.4 | 2.6 | X | X | |
| | ISD1112 | 12 sec. | 64 | 5.3 | 2.2 | X | X | |
| ISD1200 | ISD1210 | 10 sec. | 64 | 6.4 | 2.6 | X | X | |
| | ISD1212 | 12 sec. | 64 | 5.3 | 2.2 | X | X | |
| ISD1400 | ISD1416 | 16 sec. | 128 | 8.0 | 3.3 | X | X | |
| | ISD1420 | 20 sec. | 128 | 6.4 | 2.6 | X | X | |
| ISD1800 | ISD1806 | 6-12 sec. | 48 | 4 ~ 8 | 2.9 @ 8KHz | S[1] | X | |
| | ISD1810 | 8 -16 sec. | 68 | 4 ~ 8 | 2.2 @ 6.4KHz | S[1] | X | |
| ISD2500 | ISD2532 | 32 sec. | 256 | 8.0 | 3.4 | X | X | |
| | ISD2540 | 40 sec. | 256 | 6.4 | 2.7 | X | X | |
| | ISD2548 | 48 sec. | 256 | 5.3 | 2.3 | X | X | |
| | ISD2564 | 64 sec. | 256 | 4.0 | 1.7 | X | | |
| | ISD2560 | 60 sec. | 480 | 8.0 | 3.4 | X | X | X |
| | ISD2575 | 75 sec. | 480 | 6.4 | 2.7 | X | X | X |
| | ISD2590 | 90 sec. | 480 | 5.3 | 2.3 | X | X | X |
| | ISD25120 | 120 sec. | 480 | 4.0 | 1.7 | X | X | |
| ISD4002 | ISD4002-120 | 2.0 min. | 960 | 8.0 | 3.4 | X | X | X |
| | ISD4002-150 | 2.5 min. | 960 | 6.4 | 2.7 | X | X | X |
| | ISD4002-180 | 3.0 min. | 960 | 5.3 | 2.3 | X | X | X |
| | ISD4002-240 | 4.0 min. | 960 | 4.0 | 1.7 | X | X | X |
| ISD4003 | ISD4003-04M | 4.0 min. | 1920 | 8.0 | 3.4 | X | X | X |
| | ISD4003-05M | 5.0 min. | 1920 | 6.4 | 2.7 | X | X | X |
| | ISD4003-06M | 6.0 min. | 1920 | 5.3 | 2.3 | X | X | X |
| | ISD4003-08M | 8.0 min. | 1920 | 4.0 | 1.7 | X | X | X |
| ISD4004 | ISD4004-08M | 8.0 min. | 3840 | 8.0 | 3.4 | X | X | X |
| | ISD4004-10M | 10.0 min. | 3840 | 6.4 | 2.7 | X | X | X |
| | ISD4004-12M | 12.0 min. | 3840 | 5.3 | 2.3 | X | X | X |
| | ISD4004-16M | 16.0 min. | 3840 | 4.0 | 1.7 | X | X | X |
| ISD5008 | ISD5008 | 4 - 8 min. | 1920 | 4.0/5.3/6.4/8.0 | 1.7/2.3/2.7/3.4 | X | X | X |
| ISD5100 | ISD5104[2] | 2 - 4 min. | 1048 | 4.0/5.3/6.4/8.0 | 1.7/2.3/2.7/3.4 | | | X |
| | ISD5108 | 4 - 8 min. | 2096 | 4.0/5.3/6.4/8.0 | 1.7/2.3/2.7/3.4 | | X | X |
| | ISD5116 | 8 - 16 min. | 4194 | 4.0/5.3/6.4/8.0 | 1.7/2.3/2.7/3.4 | X | X | X |
| ISD5216 | ISD5216 | 8 - 16 min. | 3866 | 4.0/5.3/6.4/8.0 | 1.8/2.5/2.9/3.7 | X | X | X |

# CAPITULO I I I
# EXPERIENCIAS PRÁCTICAS

## Generalidades

En el presente capitulo se explica la realización de una serie experiencias practicas con los integrados de voz orientadas a explicar su funcionamiento.

### 3.0 Control del ISD 4002 por medio de un microcontrolador

#### 3.0.1 Descripción

Para cumplir uno de los objetivos del trabajo de titulación. ‘‘**control del integrado de voz ISD4002 por medio de un microcontrolador**’’ Se diseño un circuito capaz de operar el chip de voz utilizando las seis funciones básicas que este posee:

1. SETPLAY
2. PLAY
3. SETREC
4. REC
5. STOP
6. MC

El sistema esta dividido en etapa de control y entrada y salida de audio.

#### 3.0.2 Etapa de control

La cual esta a cargo del microcontrolador PIC16F877A, se selecciona este integrado ya que posee un modulo serial denominado MASTER SYNCHRONOUS SERIAL PORT (MSSP). Este modulo puede operar en dos modos:

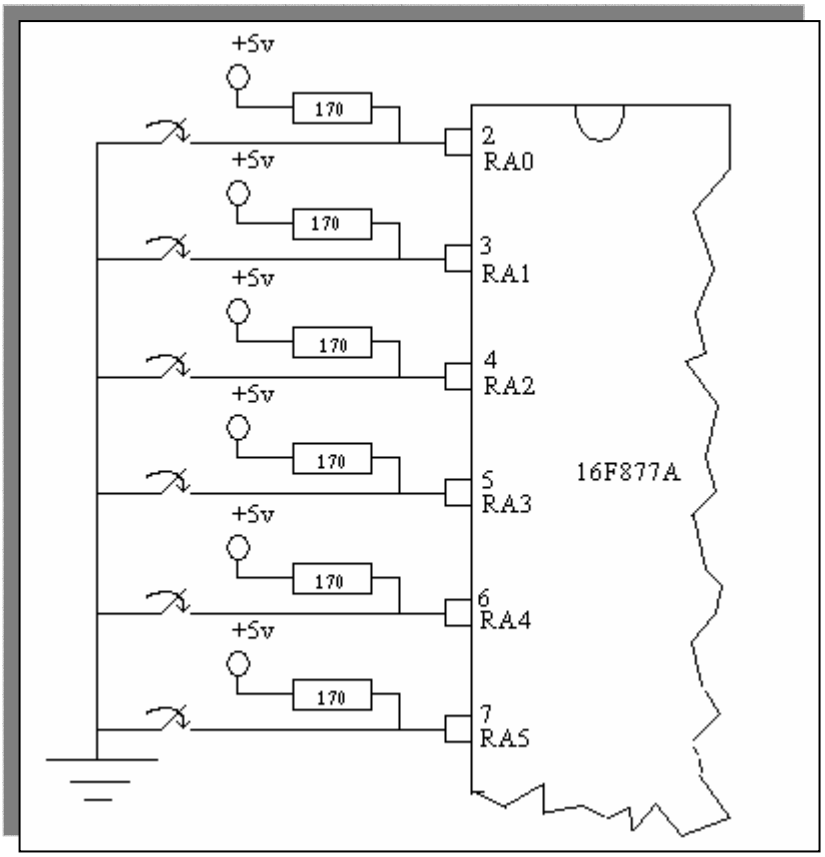
- Serial Peripheral Interface (SPI).
- Inter Integrated Circuit (I2C)

Como podemos observar, el PIC16F877A posee el modulo SPI que necesitamos para controlar nuestro integrado de voz. Esta es la razón principal por la cual se seleccionó este microcontrolador.

La etapa de control esta conformada por tres bloques, entrada, proceso y salida hacia el ISD.

**Entrada**

Se utiliza el Puerto A, como entradas digitales, a este puerto se conectan micro pulsadores, los cuales servirán para manipular los niveles de entrada a los pines del puerto, se trabaja con el estado normalmente en alto.




En el dibujo anterior se observa la conexión realizada al puerto A del PIC 16F877A, cada pin se configura como entrada digital (esto se realiza por software), y se conectan a 5 volts a través de una resistencia de 170Ω, por otro lado se conectan a tierra a través del micro pulsador.

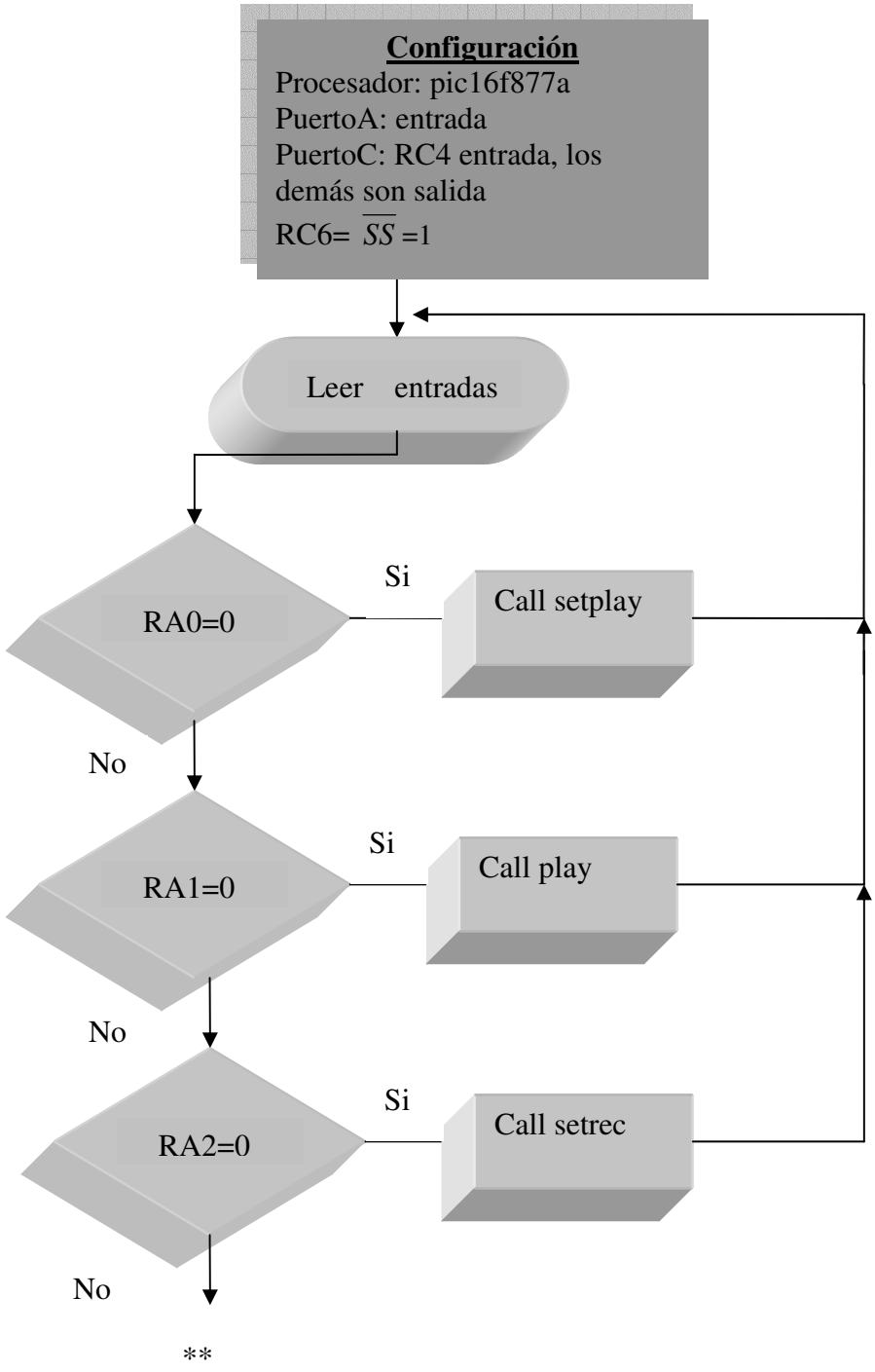
De esta forma nuestras entradas estarán normalmente en nivel alto, y cuando el micro pulsador opere el PIC detectara el cambio de nivel en su entrada (de alto a bajo) y realizara la tarea que se le haya programado.

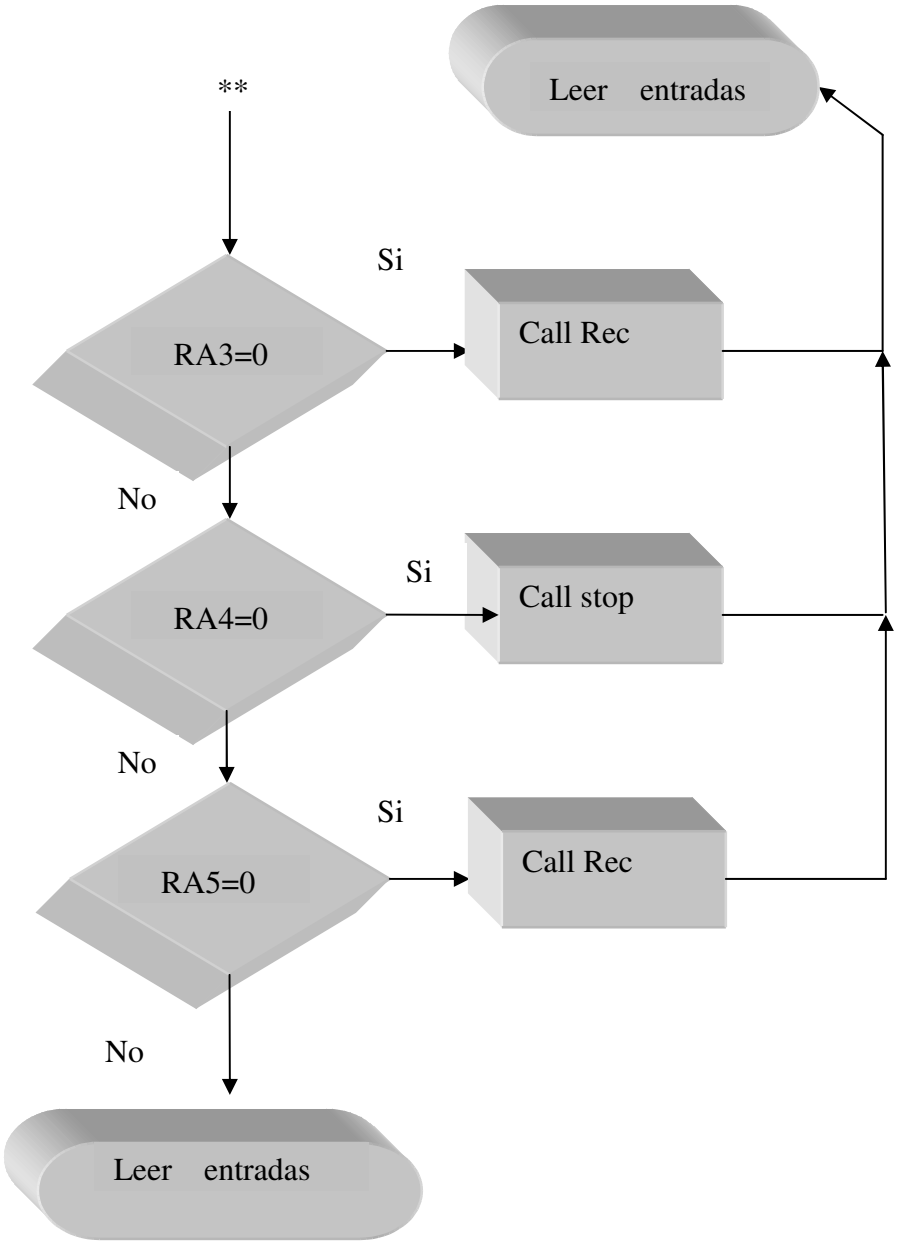
**Ejemplo de inicialización del puerto a**

```
BCF     STATUS, RP0     ; CON RP0, RP1 NOS CAMBIAMOS DE BANCO
BCF     STATUS, RP1     ; AHORA NOS ENCONTRAMOS EN EL BANCO 0
CLRF    PORTA           ; INICIALIZACION DEL PUERTOA LIMPIANDO LA
                        ; SALIDA
BSF     STATUS, RP0     ; SELECIONAMOS EL BANCO 1
MOVLW   0x06            ; CONFIGURAMOS TODOS LOS PINES DEL PUERTO
MOVWF   ADCON1          ; COMO ENTRADAS DIGITALES
MOVLW   0xFF            ; VALOR PARA ENVIAR A TRISA
MOVWF   TRISA           ; SET RA<5:0> COMO ENTRADAS
BCF     STATUS, RP0     ; REGRESAMOS AL BANCO 0
```

**PROCESO**

Esta etapa es software y consiste en generar los comandos de control del ISD. El diagrama de flujo del software generado es el siguiente:




**
Leer entradas
Si
RA3=0
Call Rec
No
Si
RA4=0
Call stop
No
Si
RA5=0
Call Rec
No
Leer entradas

## Subrutinas Call

Estructuralmente son todas iguales, la diferencia esta en el dato que se envía, estos son palabras de 16 bits, en el caso donde se ignoran los bits de direcciones pueden ser palabras de 1 byte, como se muestra en el comando stop. El primer bit enviado es A0 y el último C4, A10 no esta implementado, por lo tanto se recomienda dejarlo en cero.

| **Bits de control ( C4:C0), A10=0, Bits de direcciones (A9:A8)** | **Bits de direcciones (A7:A0)** |
|---|---|

| Bit15 | Bit14 | Bit13 | Bit12 | Bit11 | Bit10 | Bit9 | Bit8 | Bit7 | Bit6 | Bit5 | Bit4 | Bit3 | Bit2 | Bit1 | Bit0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C4 | C3 | C2 | C1 | C0 | A10 | A9 | A8 | A7 | A6 | A5 | A4 | A3 | A2 | A1 | A0 |

**SET PALY**

| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**PALY**

| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**SET REC**

| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**REC**

| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**STOP**

| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

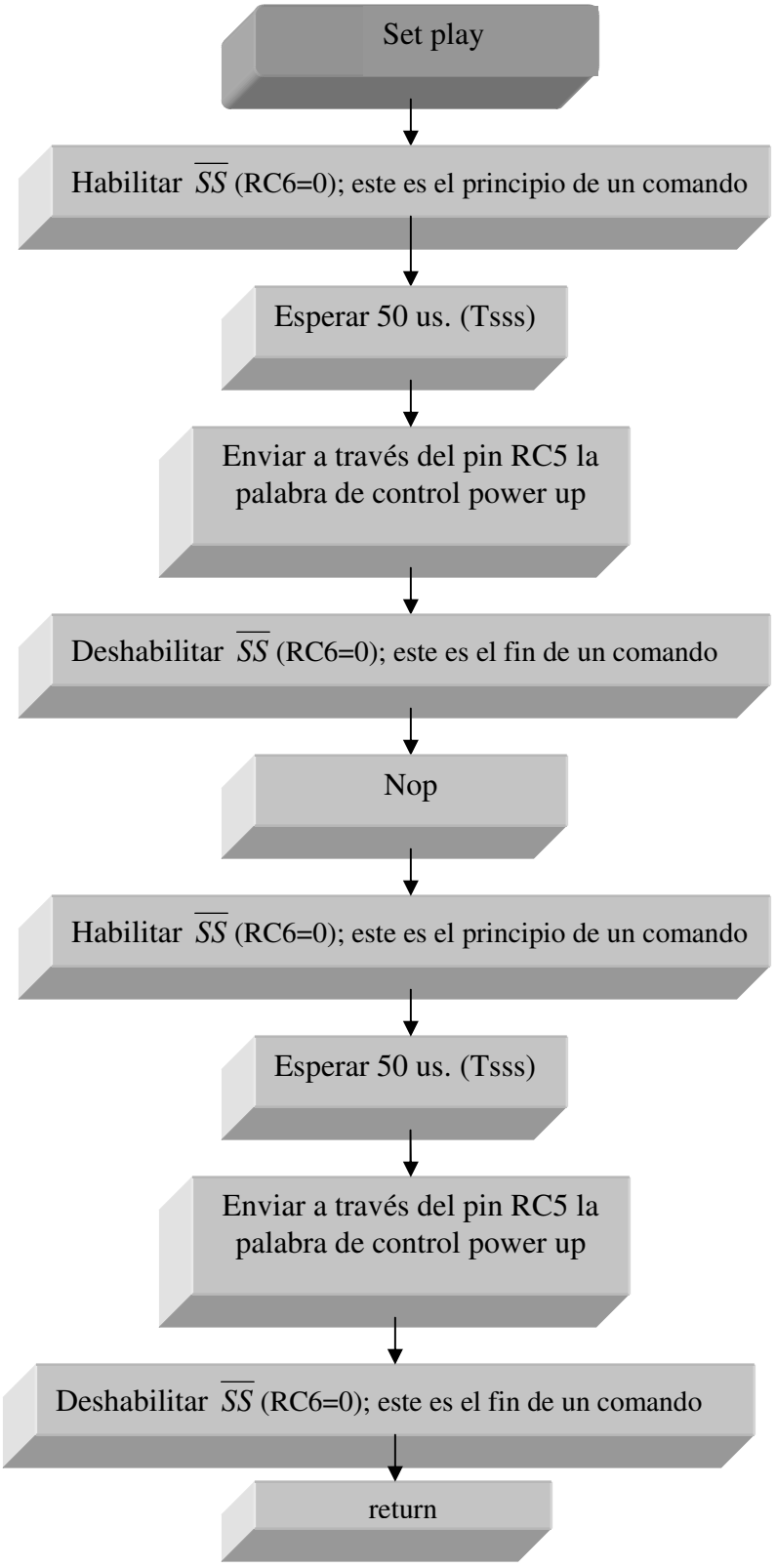
**MC**

| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**POWER UP**

| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

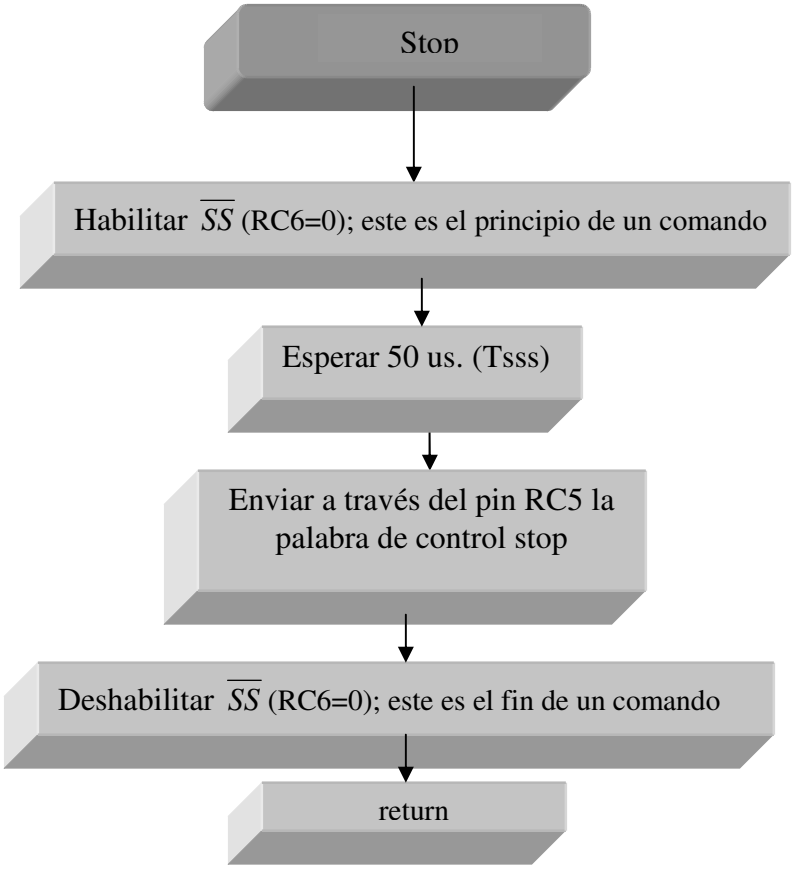
## Diagramas de flujo de subrutinas




Esta es una rutina genérica para el caso en que se envían comandos de 16 bits, en primer lugar se envía la rutina de power up del dispositivo, y luego el comando a ejecutar, con su dirección de comienzo en el ejemplo se envía set play.

Este diagrama es valido para los comandos play, set Rec., Rec., MC.

Para el comando stop se enviaron solamente los ocho bits correspondientes al byte de control.



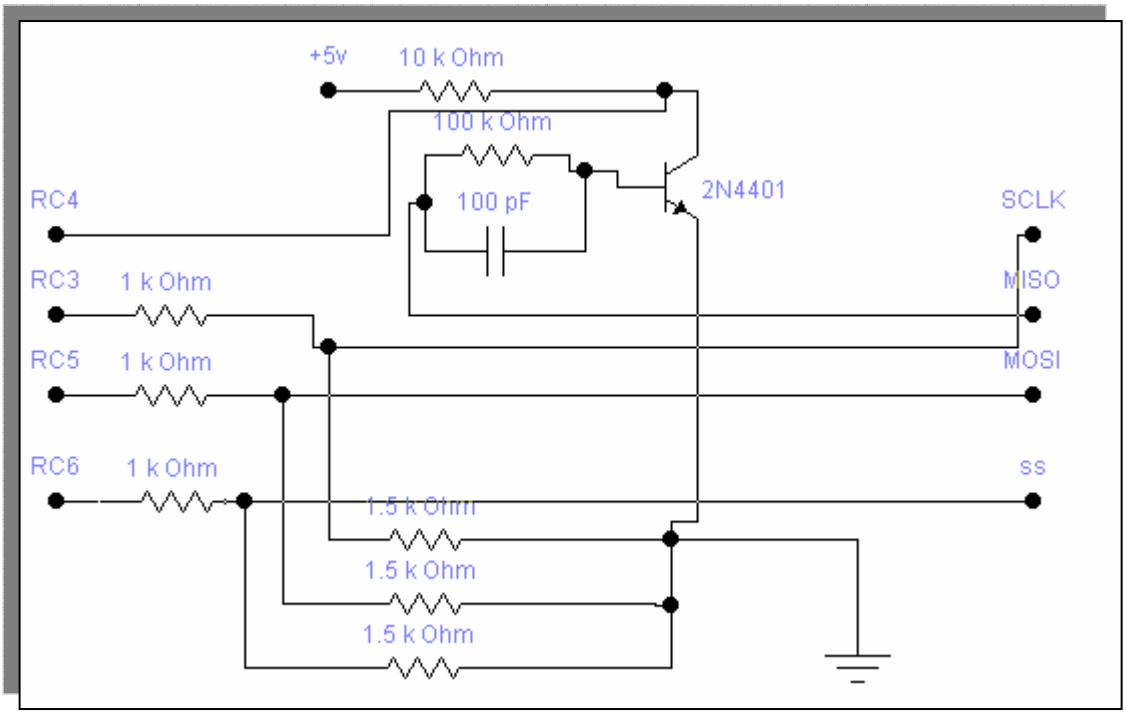

**Configuración del modo SPI**

```
BSF     STATUS,5        ;INGRESAMOS AL BANCO 1
MOVLW   B'10000000'     ; BIT 7 INGRESO DE DATOS AL FINAL DEL DATO DE SALIDA
MOVWF   SSPSTAT         ; BIT 6 TX EN EL FLANCO DE SUBIDA DE SCK
MOVLW   B'00010000'     ; PUERTO C QUEDA COMO SALIDA EXEPTO
MOVWF   TRISC                   ; RC4 EL CUAL ES SDI (MISO).
BCF      STATUS,5               ;regresamos al banco 0
MOVLW   B'00100000'       ; BIT 5 SSPCON ABILITA EL PUERTO SPI, BIT4 NIVEL
                          ; INACTIVO DEL RELOJ EN  BAJO, BITS 3,2,1 SPI
MODO
                          ; MAESTRO
MOVWF   SSPCON          ; CLOCK = FOCS/4
CLRF     PORTC          ; INICIALIZAMOS EL PUERTO C
BSF     PORTC, 6        ; SS  DESACTIVADO
```

**Salida**

El PIC esta alimentado con 5 volts, por lo tanto sus salidas son de 5 volts. El ISD trabaja con 3 volts.

La interfaz entre el PIC y el isd, necesaria para controlar las señales MISO, MOSI, SCLK Y $\overline{SS}$ (recordar que MISO, es una entrada al microcontrolador) es la siguiente.




El circuito anterior es la interfaz entre el microcontrolador funcionando a 5 volts y el ISD funcionando a 3 volts. El juego de resistencias forma un divisor de voltaje que asegura 3 volts a la entrada del ISD, el circuito del transistor es un elevador-inversor de voltaje de 3 a 5 volts para poder ingresar señales desde el ISD al PIC.

## 3.0.3Entrada y salida de audio

### Entrada de audio

El ISD dispone de dos pines ANAIN+ y ANAIN- para entrada de audio, si la señal de audio se inyecta por ambos pines(modo diferencial) esta debe ser de 16mvp-p en cada pin, y si el audio ingresa por un solo pin (modo single ended) este debe ser ANAIN+ con una amplitud de 32mvp-p, ANAIN- es llevado a tierra.

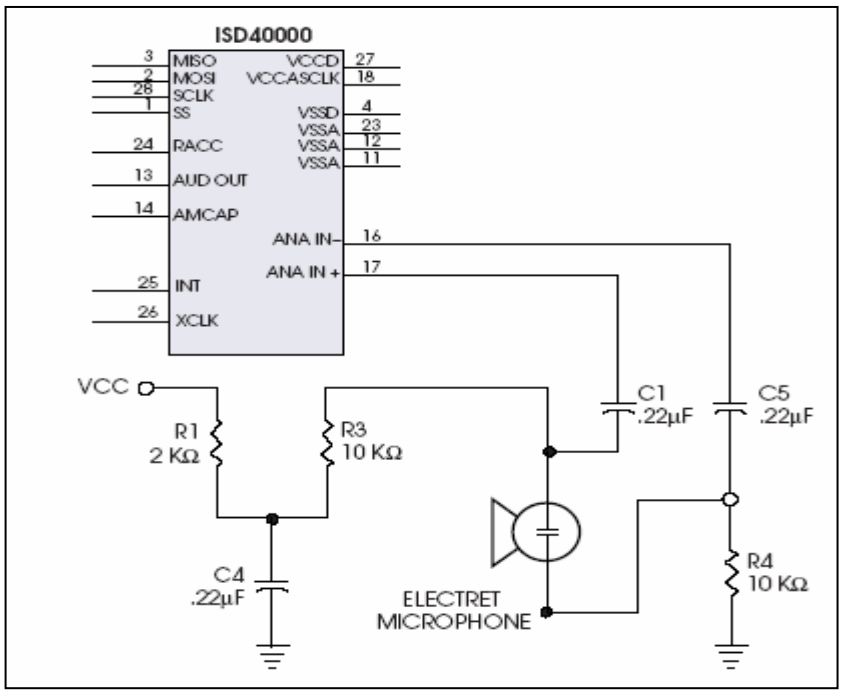
Se habilitaron entrada de micrófono y entrada de línea.

### Entrada de micrófono

Características generales:

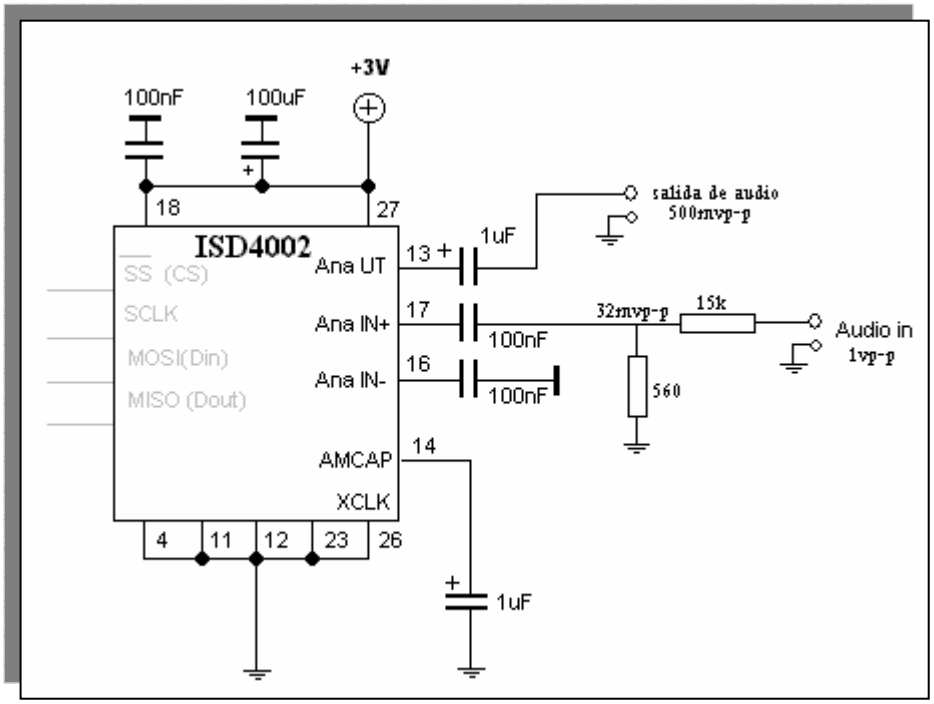
- Funciona con modo diferencial
- Utiliza un micrófono electret
- El circuito se alimenta con tres volts
- La amplitud de la señal es de 16mvp-p en cada pin

Circuito:



**Entrada de línea:**

La entrada de línea de un computador entrega 1Vp-p, para reducir esta señal a 32 mvp-p se conecto un divisor de tensión formado por resistencias, tal como se aprecia en la siguiente figura.




**Salida de audio**

La salida de audio entrega 500mvp-p, si se desea un mayor amplitud se puede amplificar, en laboratorio se realizaron pruebas con el amplificador operacional lm386 y se logro amplificar hasta 10 veces la señal, para efectos de prueba se puede conectar un parlante de computador con el cual se consigue una amplificación de hasta 10 veces la señal original.

## 3.1Control del ISD4002 por medio de un computador

### 3.1.1 Descripción

Para demostrar la compatibilidad de los ISD, con sistemas actuales, se controlo el integrado de voz por medio de un microcontrolador, ahora se controlara con la ayuda de un computador personal de escritorio.

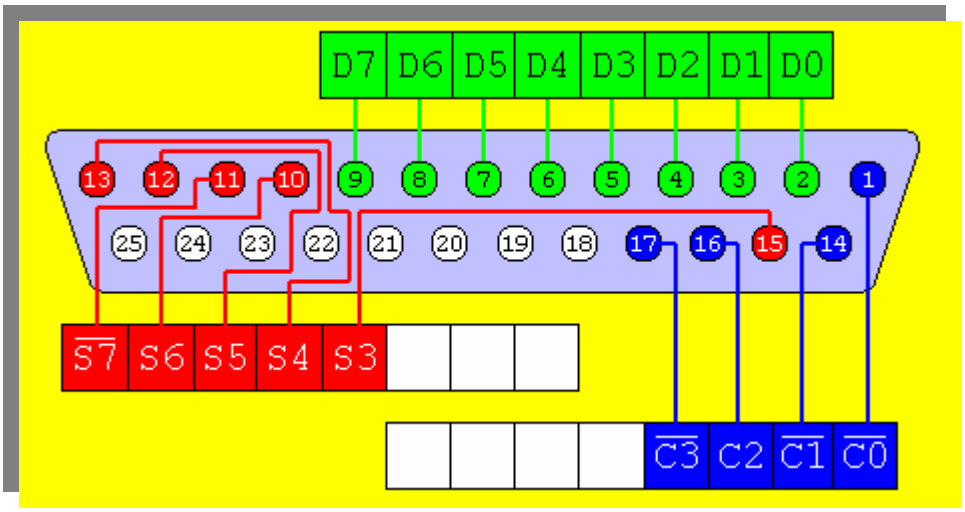
El control se realizara por medio del puerto paralelo. Y se controla el chip con tres funciones 64 funciones básicas: play, setplay, setRec, rec, MC. , y stop. Además no se utiliza el pin MISO del ISD, esto para comprobar que es posible controlar el chip con solo 3 pines de comunicación.

### El puerto paralelo

Los niveles de tensión y corriente presentes en el puerto paralelo del PC responden a los estándares de la familia lógica TTL, siendo un estado alto representado por 5V de corriente continua y un estado bajo indicado por la ausencia de tensión (0V). Hablando de corriente, cada pin del puerto paralelo puede proporcionar hasta 10mA en forma segura y 20mA como límite por breves períodos de tiempo. Para ingresar en circuitos integrados lógicos TTL ó CMOS no es necesario "potenciar" la señal, incluso es posible encender un diodo LED. Pero para manejar elementos como relays, motores, lámparas o cargas mayores se hace preciso reforzar la señal y, de ser posible, aislar el sistema. Este puerto es accesible a nivel software por medio de la BIOS, la cual asigna a éste una dirección de memoria base y dos direcciones suplementarias. La dirección base (0x3789 corresponde al bus de datos del puerto, esto quiere decir que lo que enviemos a esta dirección será exteriorizado por los terminales D0 a D7. El bus de estado se encuentra en la posición base + 1. Cabe aclarar que este bus sólo tiene implementadas cinco posiciones S3 a S7, esta última con estado lógico invertido (cuando no presenta voltaje el estado del bit es uno, mientras que cuando presenta voltaje el estado del bit es cero) y que las posiciones inferiores S0, S1 y S2 no tienen función alguna. Por último, el bus de control dispone de cuatro señales implementadas, en este caso las mas bajas C0 a C3 de las cuales sólo C2 está en estado lógico normal.

Las tres restantes (C0, C1 y C3) presentan estados lógicos invertidos, lo que implica que si se pone a uno el bit correspondiente a C0 el terminal no presentará tensión alguna, mientras que si se pone el bit a cero el terminal presentará tensión.

En el dibujo de abajo los terminales 2 al 9 (simbolizados por el color verde) representan el canal de datos. Los azules el de control y los rojos el de estados (los pines de estado son entrada).




Los pines seleccionados para trabajar son el 2(SS), 3 (MOSI) y el 5 (SCLK), en esta ocasión no se utiliza el pin MISO, para demostrar que es posible controlar el chip de voz sin este pin.

### 3.1.2 Etapa de Control

Esta es software, el lenguaje de programación fue C, el diagrama de flujo del software es el mismo presentado en la sección 3.0.2 con la diferencia que ahora se utiliza un computador y no un PIC, la entrada del sistema es el teclado, y como salida los pines 2, 3, y 5 del puerto paralelo.

### 3.1.3Interfaz entre el puerto paralelo y el ISD

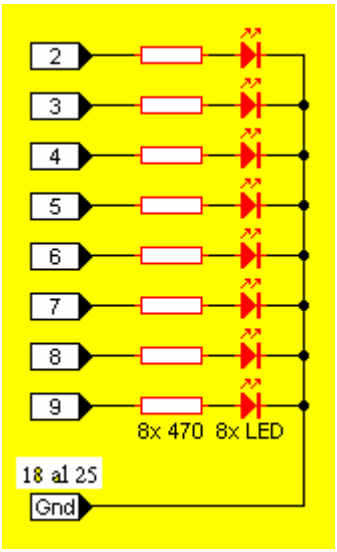
Se realizaron pruebas con el puerto del PC orientadas a medir las tensiones de salida y estas estaban entre 4,5 y 4,7 Volts en niveles altos, y 0,3 volts en nivel bajo para controlar los niveles de salida del puerto paralelo se diseño un código fuente en C.

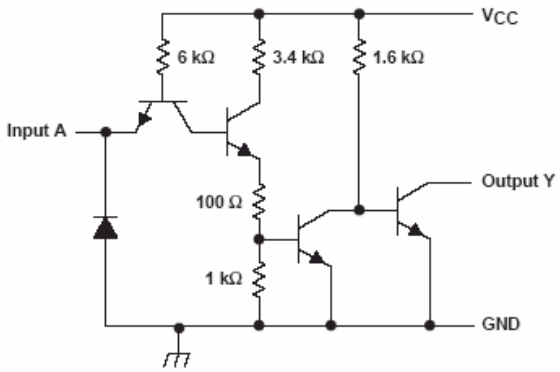
```
#include <stdio.h>
void main()
{
outport (0x378,0x01);
}
```

Este programa permite manipular los niveles de salida de los pines 2 al 8 del puerto paralelo. La instrucción outport(0x378,0x01) se lee:

Saque por el bus de datos del puerto paralelo (0x378), el dato exadecimal 0x01(00000001, en binario), con la ayuda de un multitester digital se puede medir los niveles de salida del puerto, considerando que las tierras van desde los pines 18 al 25.

Al lado el circuito con diodos led que se utilizo de monitor para el bus de datos del puerto paralelo, además se conecto un búfer de salida a los pines 2, 3, y 5 para mejorar sus niveles ttl y asegurar que la salida de datos del puerto sea unidireccional. El búfer utilizado fue el SN7417, este es con salida open colector, y se alienta con 5 volts. Abajo se muestra el esquemático del bufer.



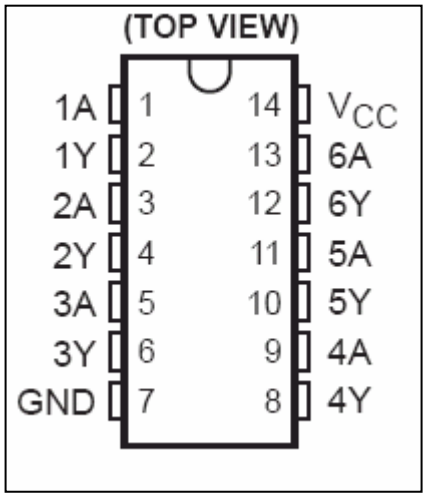



Figura 3.1.3 Configuración de pines del buffer SN7417

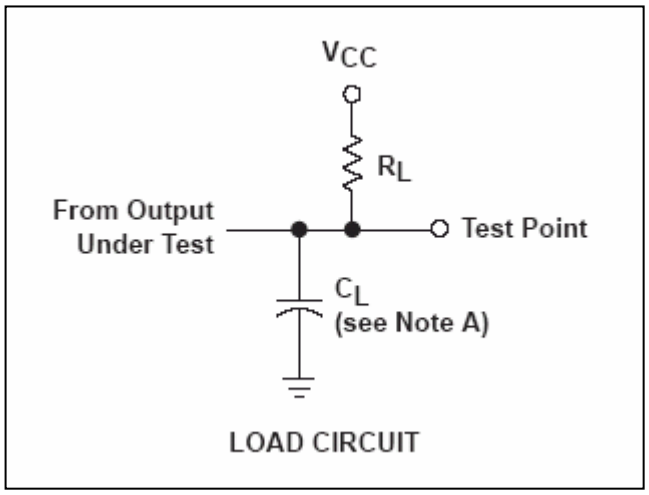



Figura3.1.3.1 circuito de carga para la salida open colector del buffer.
RL 51OHM, CL 104.

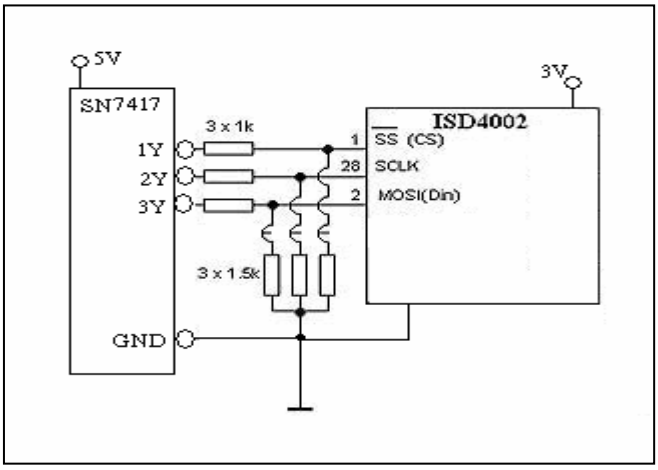



Figura3.1.3.2 divisor de tensión entre el bufer y el isd.

## 3.2 Control del ISD1420

Se realizaron una serie de experiencias orientadas a controlar el integrado ISD1420, utilizando los comandos de control qué este ofrece al circuito master. Se controlo al esclavo (ISD1420) en forma manual y con un microcontrolador PIC16F84A, el cual fue seleccionado en forma aleatoria , la entrada de audio fue a través de un micrófono electret..

### 3.2.1Experiencia nº1

### Grabado y reproducción de un mensaje

Para esto se monto en protoboard el circuito de la figura3.2, y en el bus de direcciones conectados al deep switch se puso la dirección B'00000000', al presionar el botón REC el dispositivo comienza a grabar los sonidos que ingresen por el micrófono y continua grabando hasta que la memoria se llene o hasta que el botón RECLED regrese a su posición normal, para reproducir el mensaje grabado basta con presionar el botón PLAYE, o mantener presionado el botón PLAYL.

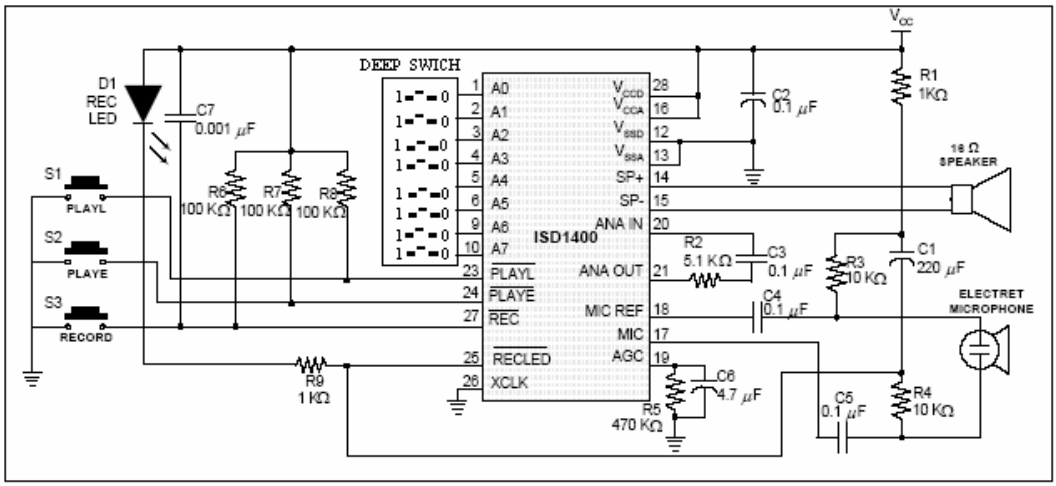



Figura 3.2

### 3.2.2 Experiencia nº2

**Direccionamiento consecutivo**

Con el circuito de la figura3.2 se graban diversos mensajes uno a continuación de otro y luego se reproducen uno a continuación de otro, para esto se utiliza el modo de funcionamiento A4.

1. El bus de direcciones se configuro de la siguiente manera

| A7 | A6 | A5 | A4 | A3 | A2 | A1 | A0 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | x | 1 | 0 | x | 0 | 0 |

Donde:

A7 y A6 = 11; se selecciona la opción “modos de funcionamiento”.

A4 = 1 ; selección del modo A4.

2. Se energizo el circuito

3. Se grabaron 4 mensajes distintos de 5 segundos de grabación cada uno, para esto basto con mantener presionado el botón RECLED durante el tiempo deseado, al presionar nuevamente el botón RECLED, el puntero interno de direcciones del ISD se incrementa, lo cual permite grabar un segundo mensaje a continuación del anterior.

4. una vez finalizada la grabacion de los mensajes comienza la reproducción, llevando momentáneamente el pin PLAYE a nivel bajo empieza la reproducción del primer mensaje hasta encontrar un marca de fin de mensaje donde la reproducción se detiene.

5. Al repetir el proceso el punto 4 los mensajes restantes se reproducen de igual forma que el primero. Al llegar al final del último mensaje, es necesario llevar momentáneamente el pin A4 del bus de direcciones a cero para poder realizar un nuevo proceso de reproducción, o grabado.

6.- Se llevaron los pines que están en la condición de sin cuidado a niveles alto y bajo y se comprobó el funcionamiento normal del circuito bajo el modo A4.

### 3.3.3 Experiencia nº3

**Borrar las marcas de fin de mensaje**

Con el circuito de la figura 3.2 se probara el modo de funcionamiento A1. Este modo de operación, permite grabar mensajes en forma secuencial y reproducirlos como un solo mensaje con solo una (EOM) al final del último mensaje.

El modo de funcionamiento A1 funciona solamente en conjunto con el modo A4

1. Se configuro el bus de direcciones de la siguiente manera:

| A7 | A6 | A5 | A4 | A3 | A2 | A1 | A0 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | x | 0 | 0 | x | 1 | 0 |

2. Comenzó un ciclo de grabado, intentando grabar diferentes mensajes, pero a la hora de reproducirlos, solamente se conseguía reproducir el último mensaje grabado, se desconecto la fuente de alimentación del sistema, con el objetivo de recetear el puntero interno de direcciones y se inicio un ciclo de reproducción, el resultado fue el mismo, solo se escuchaba el ultimo mensaje grabado

4. Con la siguiente configuración del bus de direcciones

| A7 | A6 | A5 | A4 | A3 | A2 | A1 | A0 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | x | 1 | 0 | x | 1 | 0 |

Se realizaron las mismas pruebas que en el punto 3 y ahora si se consiguió escuchar todos los mensajes grabados sin marcas de fin de mensaje entre ellos notar que ahora se activo A4 y A1.

### 3.3.4 Experiencia Nº4

### Repetir un mensaje indefinidamente

Con el circuito de la figura 3.2 se experimento lo siguiente

1. Poner en el bus de direcciones la siguiente configuración

| A7 | A6 | A5 | A4 | A3 | A2 | A1 | A0 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | x | 0 | 1 | x | 0 | 0 |

2. Con los mensajes previamente grabados comienza un ciclo de reproducción. Notara que el mensaje se reproduce en forma continua, es decir, la reproducción comienza desde el lugar que indique el puntero y al encontrar una marca de fin de mensaje la reproducción se repite.

### 3.3 Ejemplo de aplicación

### 3.3.1Detector de eventos

Se diseño un sistema que sea capaz de detectar el cambio de nivel lógico, en tres pines de entrada, cuando un evento ocurra el sistema avisara reproduciendo un mensaje previamente gravado.

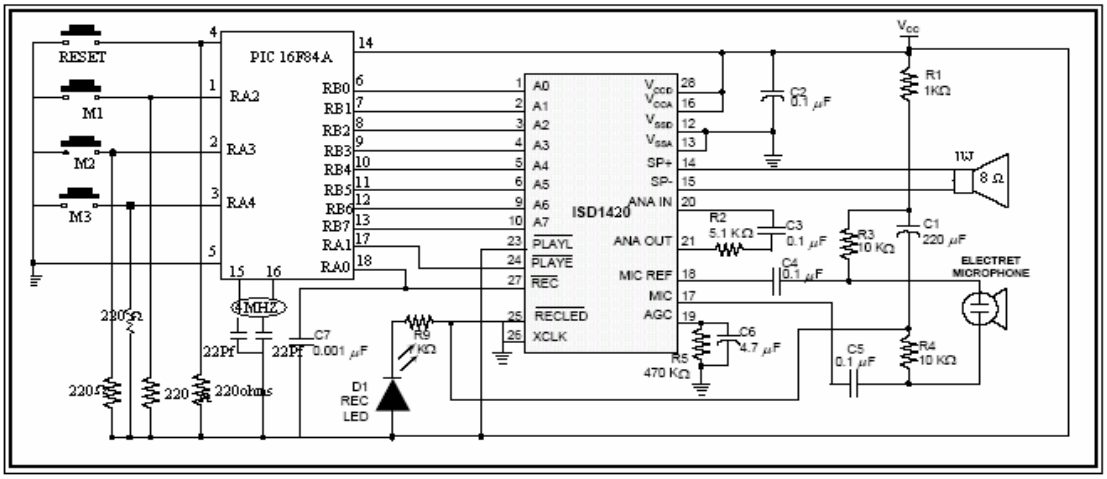



**Figura3.3.1**

**Descripción:**

Los pines de entrada se encontraran normalmente en alto, el sistema estará controlado por un microcontrolador PIC16F84A, el almacenamiento del sonido será a través del integrado ISD1420P. Además los mensajes a reproducir están previamente grabados, esto se realizó mediante el modo de funcionamiento A4.

Para el logro de la experiencia se siguieron los siguientes pasos

1.- Se grabaron tres mensajes distintos con sus respectivas marcas de fin de mensaje

2.- Se monto en protoboard el circuito de la figura 2

3.- Se cargo el microcontrolador con el código fuente para PIC 16f84A.

Cada vez que se detecta un evento se activan los modos operacionales A0 y A4 De esta forma si se presiona el pulsador M2 el microcontrolador pondrá la palabra de control, 11000000, en el bus redirecciones, para recetear el puntero de direcciones, luego se pone la palabra, 11010001, de esta forma activamos el modo de funcionamiento A0, que permite incrementar el puntero interno de direcciones según la cantidad de pulsos de nivel bajo enviados a el pin PLAYL, en este caso se enviara un pulso ya que M2 es el segundo mensaje, después de enviar este pulso la palabra de control debe ser, 11010000, para activar el modo A4, y enviar un nuevo pulso al pin PLAYL, para que comience la reproducción.

**Programa para PIC16F84A**

; Este software controla al detector de eventos con el isd 1420.

```
Processor=p16f84a
Include  <p16f84.inc>
__config  _XT_OSC & _WDT_OFF & _PWRTE_ON


                cblock 0x20
PDel0
PDel1
PDel2
CONTADOR
MENSAJE1
MENSAJE2
                endc
                org     0
INICIO
        BSF             STATUS,5        ;  ingresamos al banco 1
        movlw           h'00'           ;  puertob queda como salidas
        movwf           TRISB
        movlw           b'11100'
        movwf           TRISA
        bcf             STATUS,5        ;   regresamos al banco 0
        clrf            PORTB           ;   inicializamos el puerto b
         CLRF           PORTA
         CLRF           CONTADOR
         BSF            PORTA,0
         BSF            PORTA,1
```

```
COMIENZO
A2          BTFSC   PORTA,2
            GOTO        A3
            GOTO    M1
A3          BTFSC   PORTA,3
            GOTO    A4
            GOTO    M2
A4          BTFSC   PORTA,4
            GOTO    COMIENZO
            GOTO    M3
M1          CALL    RECETEO
            MOVLW   B'11010000'
            MOVWF   PORTB
            CALL    PULSO
            GOTO    COMIENZO
M2          CALL    RECETEO
            MOVLW   B'11010001'
            MOVWF   PORTB
            CALL    PULSO
            MOVLW   B'11010000'
            MOVWF   PORTB
            CALL    PULSO
            GOTO    COMIENZO
M3          CALL    RECETEO
            MOVLW   B'11010001'
            MOVWF   PORTB
            CALL    PULSO
            CALL    PULSO
            MOVLW   B'11010000'
            MOVWF   PORTB
```

```
                CALL    PULSO
                GOTO    COMIENZO
RECETEO
                MOVLW   B'11000000'
                MOVWF   PORTB
                CALL    R125mS
                RETURN

PULSO           BCF     PORTA,1
                CALL    R125mS
                BSF     PORTA,1
                CALL    R125mS
                RETURN
R125mS

                movlw    .124       ; 1 set numero de repeticion  (B)
                 movwf   PDel0      ; 1 |
PLoop1           movlw    .251      ; 1 set numero de repeticion  (A)
                movwf    PDel1      ; 1 |
PLoop2          clrwdt              ;  1 clear watchdog
                decfsz  PDel1, 1    ; 1 + (1) es el tiempo 0  ? (A)
                 goto   PLoop2      ; 2 no, loop
                decfsz  PDel0,  1   ; 1 + (1) es el tiempo 0  ? (B)
                goto    PLoop1      ; 2 no, loop
PDelL1          goto    PDelL2       ; 2 ciclos delay
PDelL2          clrwdt              ; 1 ciclo delay
                return                      ; 2+2 Fin.
                 END
```

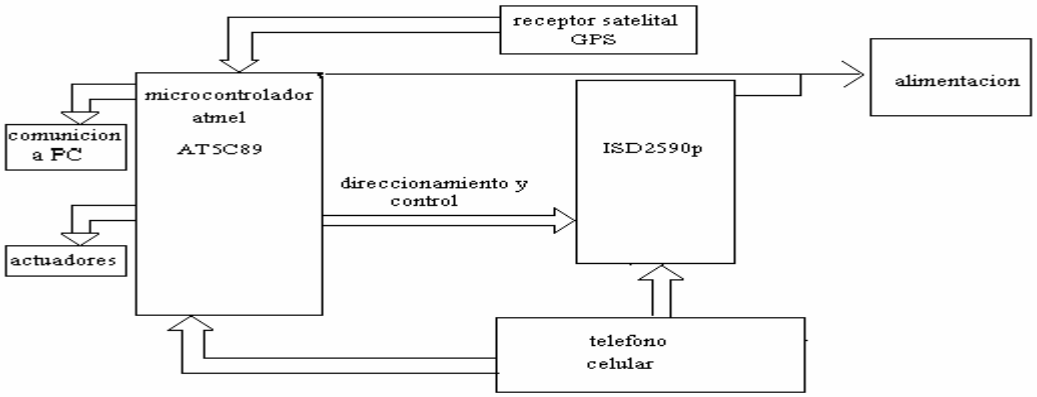
## 3.4 Descripción de un sistema real empleando chip de voz

CVX-R (control vehicular décima región) es una empresa valdiviana dedicada al diseño de sistemas de control electrónico. La especialidad de cvx-r es el control vehicular, es decir, registrar la posición de un vehículo, y poder monitorearla en tiempo real.

Uno de los últimos productos lanzados al mercado es el denominado" TRUK FINDER", el cual posee las siguientes características:

1. Se puede instalar en cualquier tipo de vehículos
2. Permite monitorear en forma remota vía Internet:
   a. la posición actual del vehículo, este este en movimiento o no.
   b. Temperatura
   c. Velocidad de desplazamiento
3. Permite vía telefónica ingresar a un menú de voz el cual esta grabado en un chip ISD, y realizar las siguientes operaciones
   a. Ingresar y cambiar contraseñas de usuario
   b. Detener o arrancar el motor del vehículo
   c. Controlar algún equipo eléctrico a elección del usuario.

## Diagrama de bloques del sistema Truk Finder



## CONCLUSIONES

1. En la actualidad todo avance tecnológico relacionado con señales de voz, esta inmerso dentro de un área denominada Tecnologías de voz o tecnologías del Habla.

Las principales áreas de investigación y desarrollo en tecnologías de voz son:

- Reconocimiento de voz.
- Conversión texto voz.
- Reconocimiento de locutores.
- Codificación de la voz.

El objetivo principal de las tecnologías de voz es potenciar el método más común y natural de comunicación humana (la voz) incorporándolo a sistemas de telecomunicaciones y a interfases hombre maquina. De esta forma los circuitos integrados de voz ISD, capaces de grabar y reproducir señales de voz son una herramienta importante dentro de las tecnologías del habla.

2. Los circuitos integrados de voz, son grabadores/ reproductores de audio que en su interior cuentan con los elementos necesarios para desarrollar sus funciones, filtros de entrada y salida, preamplificadores, amplificadores, memorias, transductores.

3. La interacción de los ISD con circuitos master, actualmente utilizados por los ingenieros, es perfectamente viable, si se cuenta con la interfaz correcta de comunicación.

4. Para utilizar los ISD, basta con conectar externamente algunos componentes, resistencias, condensadores, micrófono, parlantes y enviar el comando de control adecuado.

5. Respecto de software asociado a este tema, se pueden dividir en dos grupos: software editores de audio y software de control para el circuito master.

6. Cuando se de desarrollan experiencias practicas, ya sea para un trabajo de titulacion, o para otros fines, es fundamental seguir una pauta de trabajo, y trabajar sobre esta, incluso al extremo de la exageración, y lo mas importante, todas las pruebas son importantes por mas obvias que estas sean, es necesario hacerlas todas, incluso las que no estaban consideradas ya que al menor cabo suelto el proyecto puede fracasar.

7. los integrados de voz ISD se clasifican en dos grupos:

- interfaz de comunicación paralela
- interfaz de comunicación serial

Y estos grupos a su ves se diferencian según el tiempo de grabación que soporta cada integrado.

8. Se cuenta con circuitos integrados capaces de grabar y reproducir señales de audio, sin embargo tienen sus diferencias entres si, diferentes capacidades de almacenamiento, interfaz de comunicación serial, interfaz de comunicación paralela, alimentación desde 2,7 volts a 5.5 volts, en fin, hay gran variedad y el diseñador debe ser capaz de seleccionar el integrado que mas se adecue a los requerimientos del sistema.

9. Los circuitos integrados de voz son una herramienta disponible para los diseñadores de sistemas electrónicos desde principios de los años 90, sus aplicaciones principales son: menús interactivos, contestadores telefónicos, e interfaz hombre maquina en general, su inserción al mercado ha ido en aumento y esto se ve reflejado con la evolución sufrida por estos circuitos, los primeros solo tenían capacidad para almacenar 16 segundos de señal de voz el ultimo circuito lanzado al mercado es capaz de guardar hasta 16 minutos de sonido.

## Bibliografía

### Libros


- Muewer, Manual de actualización y reparación de PC's, Prentice Hall, octava edición, 1992
- Javier Ceballos, curso de programación e C, Macrobit, 1989.


### Hojas de datos


- PIC 16F87XA data sheet, Microchip Technologies , 2003
- PIC 16F8XA data sheet, Microchip Technologies, 2003
- ISD 1400 series SINGLE-CHIP, Voice Record/Playback devices,16-and- 20 seconds duration, march 2004 revision 1.0
- ISD4002 Series SINGLE-CHIP, multiple messages, Voice Record/Playback devices,120, 150, 180 and 240 seconds duration, september 2003, revision 1.0, windbond electronics corp.
- Sn5407, SN5417, SN7407, SN7417 HEX BUFFERS DRIVERS, whith open collector higth voltaje outpt, Revised may 2004, Texas Instruments.


### Sitios Web


- http://www.melabs.com/products/labxt.htm
- http://www.winbond-usa.com
- http://www.microchip.com
- http://www.monogafias.com/El Puerto Paralelo IBM.htm
- http://www.pablin.com.ar/ Monitor para Puerto Paralelo.htm
- http://hem.passagen.se/comunication/speach.htm
- http://www.telefonicaI+D.com/tecnologíasdelhabla.htm
- http://www.victronics.cl

# Anexo

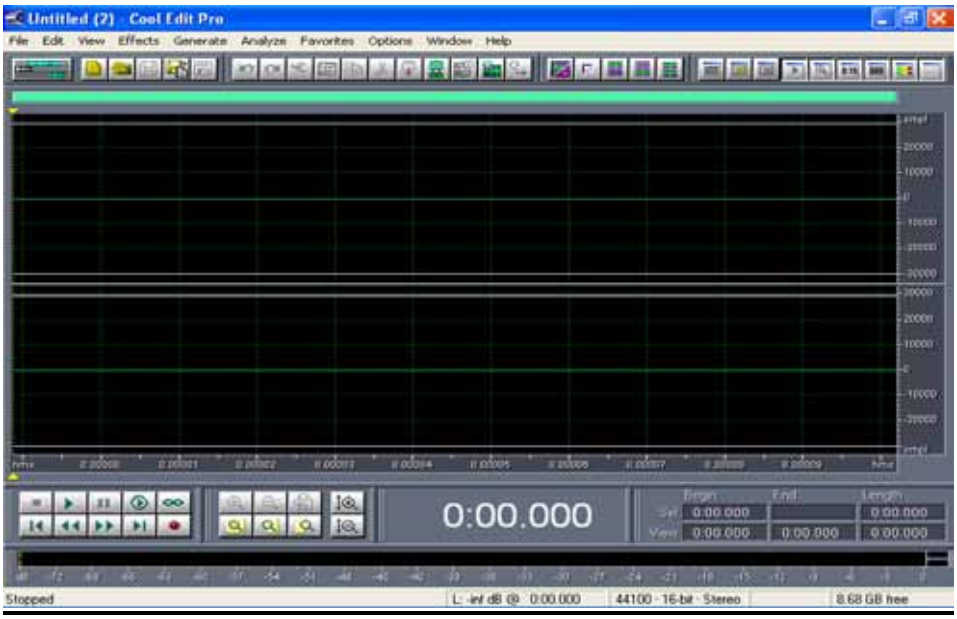
## 4.0 Software para tratamiento de señales de audio

### 4.0.1 Cool Edit

Cool edit es un software diseñado para trabajar con señales de audio en un PC, ofrece las siguientes prestaciones:

- Grabar señales de audio en un PC
- Editar el audio grabado
- Reproducir las señales de audio

### 4.0.2 pantalla principal del cool edit



### 4.0.3 Como grabar señales de sonido en cool edit

En la pantalla principal esquina inferior izquierda se aprecia una barra de comandos el botón rojo es REC, (en esa misma ventana tenemos los botones play, pause y stop) al presionar este botón comenzará la grabación del sonido entrante al computador por la tarjeta de audio, ya sea entrada de micrófono o entrada de línea. Para guardar un archivo hay que seleccionar la opción file→ save as, tal como en cualquier programa de Windows. Una ves guardado nuestro archivo lo podemos reproducir, por ejemplo con Windows media player u otro reproductor, y pasar nuestro sonido al chip de voz.

Figura 4.3, se aprecia la señal Que entra por medio de un micrófono a la tarjeta de sonido de un PC, esta siendo capturada por cool edit en formato stereo

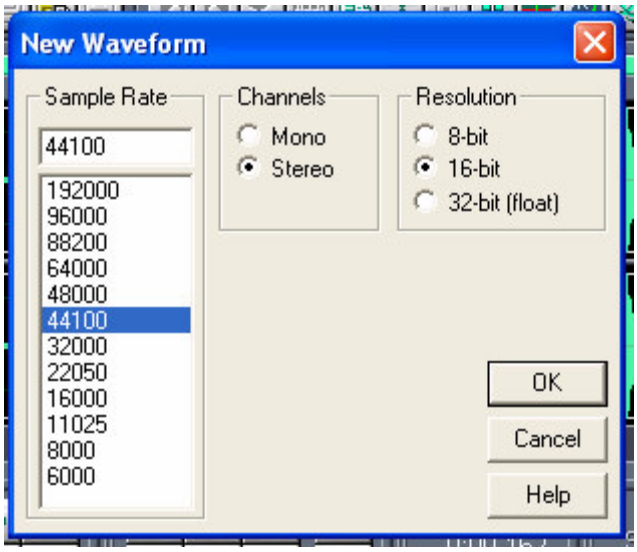



Figura 4.4, es la pantalla que aparece cuando queremos comenzar un nuevo archivo, aquí se puede seleccionar las características de nuestra grabación, frecuencia de muestreo, mono o stereo, y la codificación