Diseño de un Sistema de Codificación de Predicción Lineal (LPC)

M. en C. Pablo Manrique Ramírez
Profesor e Investigador del CIC-IPN
Ing. Miguel A. Meléndez Velázquez
Alumno del CIC-IPN

El área de procesamiento de señales de voz, desde hace tiempo ha sido tema de investigación en el desarrollo de proyectos. Sin embargo tras años de investigación y comenzando con los prototipos basados en resonadores acústicos hasta llegar a los modernos programas para computadora que sintetizan voz, todos estos productos han sido diseñados para el idioma inglés primordialmente; mientras que para el idioma español todavía se trabaja en una etapa de laboratorio. Además, muchos de los productos relacionados con el procesamiento de señales que ya son comercializados, solamente están implementados en software, el hardware empleado corresponde generalmente a tarjetas de sonido de las que ya se dispone en la mayoría de computadoras personales.

El objetivo principal de este proyecto es realizar el procesamiento de señales de voz utilizando el método de Codificación de Predicción Lineal (LPC) y enfocándose en el idioma español; no solo aprovechando el desarrollo de software sino también explotar el uso de dispositivos digitales para el diseño de una interfaz, por lo que no se empleará una tarjeta de sonido de propósito general sino una tarjeta sintetizadora de voz con un microprocesador de propósito específico, esto con el fin de lograr una calidad aceptable en la señal de voz sintetizada.

Introducción

El modelo de Codificación de Predicción Lineal es considerado uno de los modelos más próximos (analógicamente hablando) al sistema vocal humano. Por esta característica se ha seleccionado como base para el diseño de una interfaz digital que sea capaz de sintetizar señales de voz. Para esto, es importante considerar cómo está constituido el Sistema Vocal Humano para realizar una analogía con un Sistema Digital.

Generalidades del Sistema Vocal Humano

La voz es un sonido producido en la laringe debido a la salida del aire que, al atravesar las cuerdas vocales, las hace vibrar. Uno de los parámetros que definen a la voz es su tono. El tono depende de cada individuo y está determinado por la longitud y masa de las cuerdas vocales. Por lo tanto, el tono puede alterarse variando la presión del aire exhalado y la tensión sobre las cuerdas vocales. Esta combinación determina la frecuencia a la que vibran las cuerdas: a mayor frecuencia de vibración, más alto es el tono.

Otro aspecto de la voz es la resonancia. Una vez que ésta se origina, resuena en el pecho, garganta y cavidad bucal. Finalmente, otro parámetro importante de la voz es su intensidad o fuerza, que depende de la resonancia y de la fuerza de vibración de las cuerdas vocales.

Para que la voz sirva como parte de un sistema de comunicación, se requiere también la articulación. La articulación se refiere a los sonidos del habla que se producen y combinan para formar las palabras del lenguaje. Los instrumentos de la articulación son: los labios, la lengua, los dientes, las mandíbulas y el paladar. El habla se articula mediante la interrupción o modelación de los flujos de aire, vocalizados y no vocalizados, a través del movimiento de la lengua, los labios la mandíbula inferior y el paladar. Los dientes se usan para producir algunos sonidos específicos.

Otro elemento para conformar un sistema de comunicación por medio del habla es el lenguaje. El lenguaje es un sistema de símbolos abstractos reconocido por un grupo de personas que sirve para comunicar sus pensamientos y sentimientos. Los símbolos pueden ser verbales o no verbales, es decir, hablados o escritos, además, los símbolos no verbales pueden ser gestos y movimientos

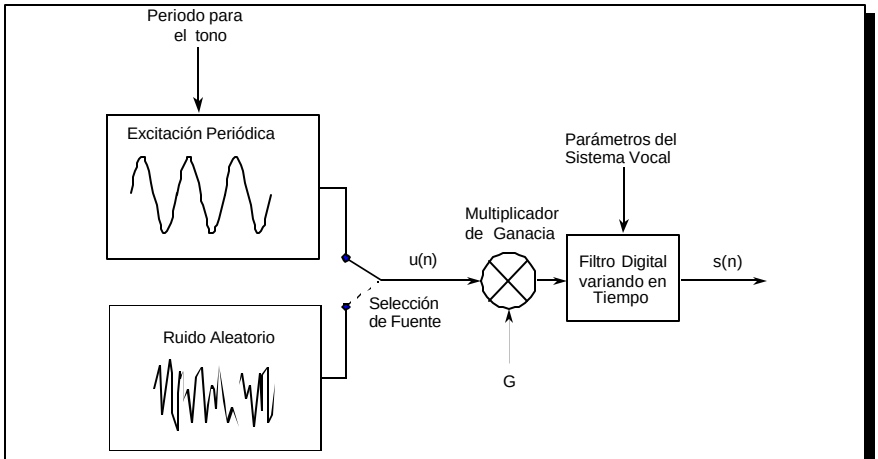

Figura 1. Modelo LPC para síntesis de voz

corporales. En el lenguaje hablado se utiliza la capacidad de articular sonidos y en el lenguaje escrito ésta se sustituye por la ortografía. Las capacidades auditivas y visuales son esenciales para la comprensión y expresión del lenguaje.

El Modelo LPC

El modelo LPC para señales de voz es una buena aproximación al sistema vocal humano. El LPC, matemáticamente es preciso [1] y además no es muy complejo llevarlo a la práctica a través de software y/o hardware, comparado con otros métodos digitales.

La idea fundamental del modelo LPC es representar a la señal de voz como una función de excitación constituida por un tren de pulsos cuasiperiódicos (para sonidos vocalizados) o una fuente de ruido aleatorio (para sonidos no vocalizados) [1]; el modelo de síntesis para voz con LPC [3] es mostrado en la figura 1.

En este modelo la fuente de excitación es seleccionada por un interruptor de posición controlado por un caracter vocalizado / no vocalizado de la voz. La ganancia G apropiada es estimada de la señal de voz, lo que es utilizado como entrada para un filtro digital que tiene la función de transferencia H(z). Este filtro digital es controlado por los parámetros característicos del sistema vocal de la voz que se está sintetizando. De esta manera, los parámetros para este modelo son:

- **Selección de la fuente (periódica o de ruido aleatorio).**
- **Periodo de tono (para sonidos vocalizados).**
- **Parámetro G de ganancia.**
- **Los coeficientes $\{a_i\}$ para el filtro digital.**

Estos parámetros varían lentamente en el tiempo y son los que hay que considerar en el diseño de un sintetizador digital de voz que utilice el modelo LPC.

El Sintetizador de Voz

En lo que corresponde al hardware, el sistema digital está basado en el microprocesador TSP53C30, de Texas Instruments [3]; que es un microprocesador de propósito específico que puede ser programado para seguir el modelo LPC. El TSP53C30 trabaja como un dispositivo esclavo a otro microprocesador, por lo que puede ser parte de la interfaz de una tarjeta de expansión o como un sistema independiente de una PC si se incluye en el diseño un microprocesador que realice todo el control, así como bancos de memoria para datos de voz.

El funcionamiento general del TSP53C30 es análogo al modelo LPC, ya mencionado anteriormente. Este modelo incorpora elementos análogos a los existentes en el sistema vocal humano. Tiene un generador de funciones periódicas y aleatorias (para producir sonido vocalizado y no vocalizado, respectivamente), un multiplicador de ganancia (realiza la función de la presión de aire que define la intensidad de la voz) y un

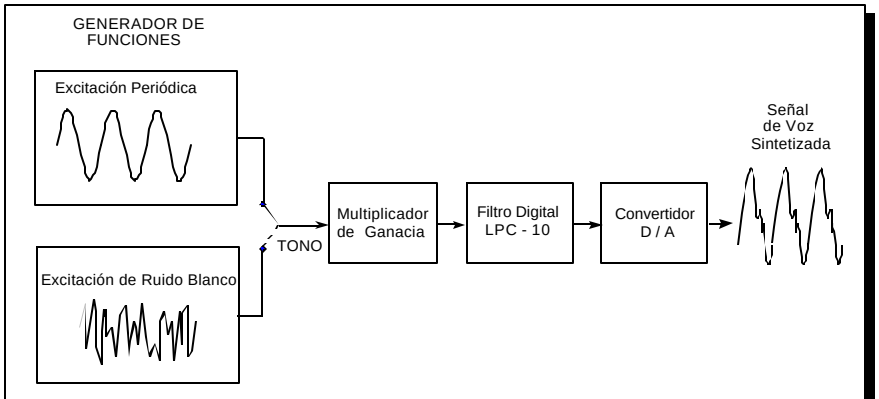

Figura 2. Esquema del modelo LPC basado en el TSP53C30

filtro digital de 10 polos (modela la resonancia de la cavidad oral).

En la figura 2 se muestra un esquema del modelo LPC basado en el TSP53C30. En dicho modelo el generador de funciones requiere como parámetro el periodo para producir un tono similar al obtenido con la vibración de las cuerdas vocales (sonido vocalizado). Este generador también produce ruido blanco, correspondiente al sonido no vocalizado. La función de salida del generador, es entonces multiplicada por un factor de energía que representa la presión de los pulmones. Finalmente, la señal es pasada a un filtro digital que modela la forma de la cavidad oral. Este filtro tiene 10 polos, razón por la cual la síntesis es referida como LPC-10.

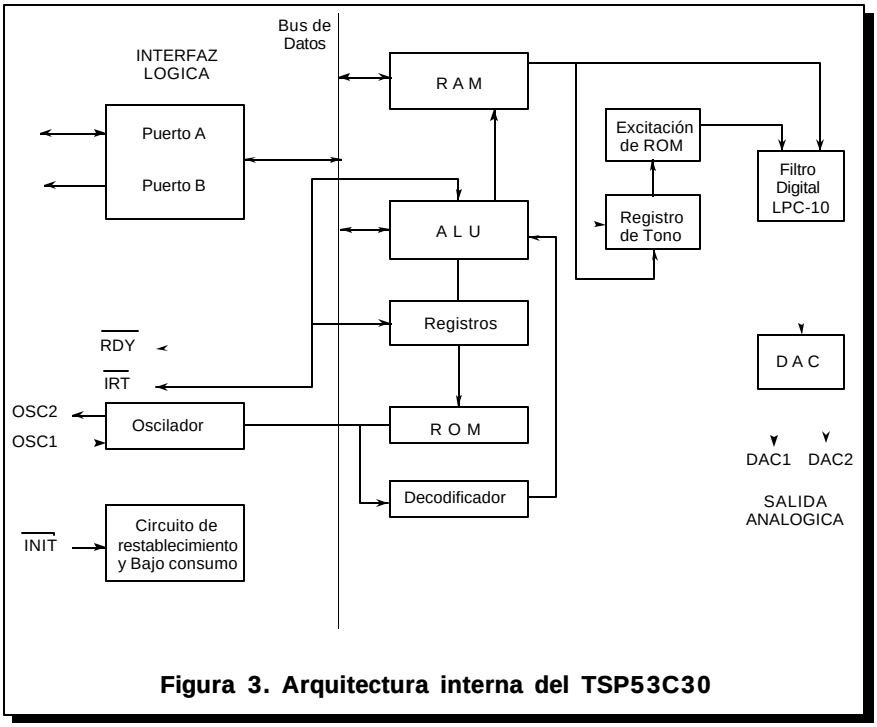

Figura 3. Arquitectura interna del TSP53C30

Arquitectura del TSP53C30

En la figura 3 se muestran a bloques los componentes internos del TSP53C30.

Básicamente se tienen los siguientes componentes: Un microprocesador de 8 bits, una ROM interna de 8K y una interfaz lógica de E/S. Las instrucciones son buscadas en la ROM cada 9 ms y se usan para controlar la acción que desarrolla el TSP53C30. Para producir voz, el TSP53C30 accesa datos de voz de una memoria externa o los que recibe del procesador al cual está conectado como esclavo. Una vez que el dato es leído, el procesador debe "descompactar" los parámetros individuales de voz y guardar los resultados en una sección de la RAM interna. De esta RAM interna se obtienen los parámetros de voz cada vez que son requeridos.

La sección de E/S, está formada por un puerto A bidireccional de 8 bits y un puerto B de 8 bits como interfaz a una memoria externa.

Buffer de datos de entrada

Todos los datos de entrada para la síntesis de voz son almacenados en una sección de la memoria RAM del TSP53C30, excepto cuando es inicializado para usar el formato PCM. En este caso, los datos son pasados directamente a la sección del sintetizador.

Software de Control

Para que el TSP53C30 realice sus funciones, es necesario que un programa controle los diferentes bloques de su arquitectura. Las instrucciones no son detalladas, pero el diagrama de flujo general del software de control se muestra en la figura 4.

Selección del modo de operación

El procesador debe inicializar al TSP53C30 para seleccionar un modo de operación del dispositivo y el formato de datos a manejar. El modo de operación determina si el TSP53C30 aceptará datos para la sintesis de voz o las direcciones en donde residen esos datos. Una vez inicializado el TSP53C30, este opera en uno de los siguientes tres modos de operación:

Modo "host": En este modo, el procesador envía directamente al TSP53C30 los datos para la síntesis de voz.

Modo Directo: El TSP53C30 recibe las direcciones de memoria en donde se encuentran los datos para la síntesis de voz.

Modo Indirecto: En este modo, el TSP53C30 recibe la dirección de una tabla de búsqueda que contiene la dirección de los datos para la síntesis.

En cuanto al formato de los datos utilizados por el TSP53C30, estos formatos pueden ser:

5220: Este formato es utilizado desde el sintetizador TSP5220C.

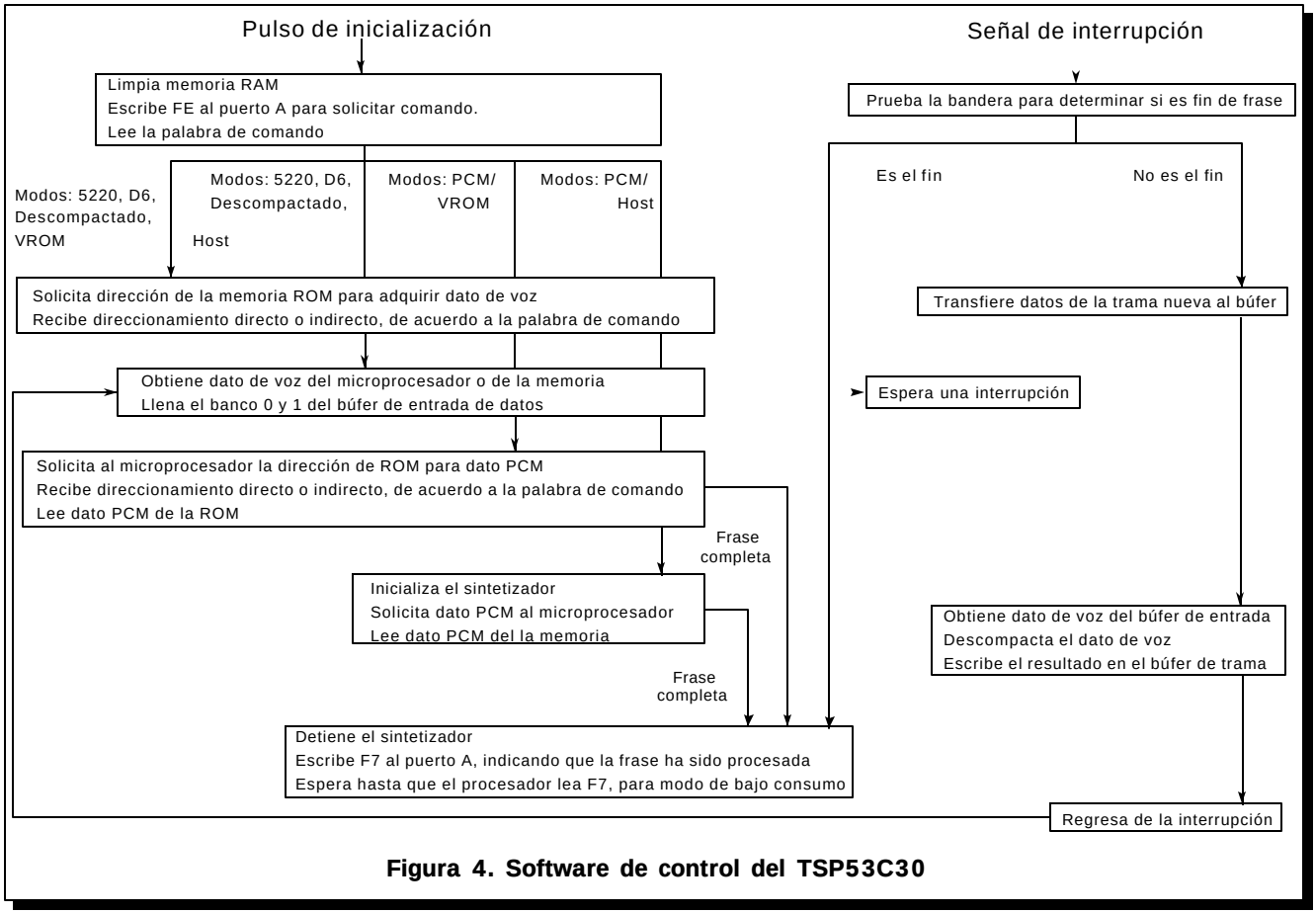

Figura 4. Software de control del TSP53C30

D6: Con este formato, los datos tienen ligeramente una mayor velocidad de transferencia y un control de tono más fino que el formato 5220.

PCM: Debido a la codificación, provee una mayor velocidad de transferencia para los datos.

Descompactados: A diferencia de los formatos anteriores, los datos no son "compactados", lo que permite un control mucho más fino sobre los parámetros de voz. Este es el formato ideal para el modelo LPC.

Las velocidades de Transferencia para estos formatos son:

Formato de Datos	*Bytes por Segundo*
5220	225
D6	250
Descompactado	1000
PCM	10 000

Procedimiento de operación

La tabla 1 muestra los comandos del TSP53C30 para sus diferentes modos de operación.

Para establecer un protocolo de comunicación entre el TSP53C30 y el procesador principal, también es necesario un conjunto de palabras de estado reportadas por el TSP53C30 hacia el procesador. Esto puede apreciarse en la tabla 2.

El TSP52C30 es inicializado al principio de cada palabra o frase, llevando /INIT a estado bajo. El dispositivo permanece en modo de bajo consumo mientras esta terminal permanezca en estado bajo. Una vez que /INIT se vaya a estado alto, el protocolo de comunicación es esta-

Tabla 1. Comandos del TSP53C30

Entrada del puerto P1 durante la inicialización								Formato de Datos	Fuente de Datos de Voz
7	6	5	4	3	2	1	0		
0	0	0	0	0	0	0	0	PCM	Procesador principal
n1	n0	s	0	0	1	1	0	5220	Procesador principal
n2	n0	s	0	0	1	0	0	D6	Procesador principal
p2	p1	p0	0	0	0	1	0	Descompactado	Procesador principal
0	0	0	0	i	0	0	1	PCM	Memoria Externa
0	0	s	0	i	1	1	1	5220	Memoria Externa
0	0	s	0	i	1	0	1	D6	Memoria Externa
0	0	0	0	i	0	1	1	Descompactado	Memoria Externa

Tabla 2. Palabras de estado para el TSP53C30

Palabra de Estado								Código	Interpretación
7	6	5	4	3	2	1	0	Hexadecimal	del Estado
1	1	1	1	1	1	1	0	FE	Esperando comando de modo de operación.
1	1	1	1	1	1	0	1	FD	Esperando dirección de dato de voz.
1	1	1	1	1	0	1	1	FB	Esperando dato de voz.
1	1	1	1	0	1	1	1	F7	Fin de síntesis. Se detectó código final.
1	1	1	0	1	1	1	1	EF	Terminación anormal de síntesis.

blecido entre el microprocesador y el sintetizador. Este protocolo se muestra en la tabla 3.

MODO "HOST" CON DATOS DESCOMPRIMIDOS

En este modo, el microprocesador principal envía los comandos y los datos de síntesis, los cuales son:

No. de byte	Dato
1	Longitud de la trama
2	Tono
3*	tono S/Fraccional
4	Energía
5	Energía fraccional
6	Parámetro K1
7	Parámetro K2
8	Parámetro K3
9	Parámetro K4
10	Parámetro K5
11	Parámetro K6
12	Parámetro K7
13	Parámetro K8
14	Parámetro K9
15	Parámetro K10
16	Parámetro K1 fraccional
17	Parámetro K2 fraccional
18	Parámetro K3 fraccional
19	Parámetro K4 fraccional
20	Parámetro K5 fraccional
21	Parámetro K6 fraccional

*El bit más significativo corresponde al código de parada. Los cuatro bits menos significativos son el dato de tono.

La secuencia de la operación a seguir se enlista a continuación:

1. **Permitir a /INIT realizar una transición a estado bajo.**
2. **El TSP53C30 lleva /RDY a estado bajo.**
3. **Lectura de dato del puerto A. Debe ser "FE" (Esperando comando de modo de operación).**
4. **Escribir uno de los siguientes comandos al puerto A (/RDY será puesta en estado alto para escribir el dato de operación).**

Comando	Longitud de la trama	Coeficiente de Reflexión Fraccional
02	15 bytes	Ninguno
22	16 bytes	K1
42	17 bytes	K1, K2
62	18 bytes	K1, K2, K3
82	19 bytes	K1, K2, K3, K4
A2	20 bytes	K1, K2, K3, K4, K5
C2	21 bytes	K1, K2, K3, K4, K5, K6

5. **El TSP53C30 lleva la terminal /RDY a estado bajo.**
6. **Lectura del dato en el puerto A. Si es "FB" (Esperando dato de voz), entonces se realiza el siguiente paso, de otra manera con "F7" (Fin de síntesis) se detecta el código final.**
7. **Se escribe el dato de síntesis de voz al puerto A. /RDY se lleva de estado bajo a estado alto. Se repite desde el paso 5.**

TSP53C30	Procesador Principal
a) Pone "FE" en el latch del puerto A y la linea /RDY en estado bajo	b) Detecta /RDY en estado bajo. Lee el código de petición. escribe el comando de operaciónal puerto A. (Cuando /ENA2 es puesto en estado bajo, /RDY se va a estado alto)
c) Pone "FD"(petición de dirección de dato) o "FB" (petición de dato de voz) en el latch del puerto A y la líne/RDY en estado bajo	d) Detecta /RDY en estado bajo. Lee el código de petición. Escribe la dirección o el dato de voz en el puerto A
e) Se repiten los pasos (c) y (d) hasta que el TSP53C30 detecte el código de finalización al término de una palabra o frase	
f) Código de finalización detectado. Pone "F7" en el latch del puerto A y la línea /RDY en estado bajo	g) Detecta /RDY en estado bajo. Lee el código de estado
h) Permanece en estado de bajo consumo mientras el procesador lee el código de estado "F7"	

Tabla 3. Protocolo de comunicación

COMPRESIÓN DE DATOS EN EL MODELO *LPC*

El modelo LPC toma ventajas de las características de la voz para "ahorrar" información redundante. La señal de voz cambia lentamente y la cavidad oral tiende a caer dentro de ciertas áreas de resonancia más que en otras. La voz es analizada en periodos de 10 a 25 milisegundos. En este periodo considerado, la señal de voz es interpolada de tal manera que no existan cambios abruptos respecto a la siguiente muestra. Además, con el modelo LPC los coeficientes del filtro [5] son constantes (para un periodo repetido de señal de voz) por lo que solamente son requeridos los valores del tono y el factor de energía. Los coeficientes del filtro son mantenidos con los valores anteriores. Sumado a esto, todos los coeficientes son codificados de 7 a 3

bits por cada coeficiente. Esta codificación está hecha de tal forma que se tomen los valores de los coeficientes que ocurren con más frecuencia.

Resultados

Actualmente, el sistema aún se encuentra en una etapa de pruebas de laboratorio, en lo que respecta a la programación de las rutinas de síntesis. Sin embargo, la construcción del prototipo de la tarjeta en su versión para expansión de PC ha sido completada. En la imagen 1 se muestra el prototipo de la tarjeta vista desde la cara de componentes.

Conclusiones

El desarrollo de este proyecto muestra la aplicación de los Sistemas Digitales dentro de la tecnología de voz. Como la mayoría de las aplicaciones de procesamiento de señales, la producción de voz consume muchos recursos informáticos, sin embargo, el uso de un procesador dedicado (el TSP53C30) a la síntesis de voz "libera" de mucho trabajo al sistema, pudiendo enfocarse el diseño y la implementación en el software de control para mejorar la calidad de la voz producida, así como rutinas para edición y reproducción de voz.

El prototipo continuará como auxiliar experimental para la síntesis de voz, permitiendo la investigación más a fondo dentro de esta área. Esto es muy importante, ya que sirve como base para el desarrollo de aplicaciones que exploten la tecnología de la síntesis de voz a través de un sistema dedicado a este fin.

Bibliografía

[1] **Lawrence Rabiner, Biing-Hwang J.** *"Fundamentals of Speech Recognition"*. **Prentice Hall, 1993.**

[3] *"TSP53C30 Speech Synthesizer"*. **Texas Instruments, 1990.**

[4] *"TSP53C0x Family Speech Synthesizer, Design Manual".* **Texas Instruments, 1994.**

[5] **Grice, Donald G. Rensselaer.** *"Adaptive bandpass filtering and its relationship to techniques used in speech compression, synthesis, and recognition"*. **Polytechnic Inst.**

[6] **F. J. Owens.** *"Signal Processing of Speech".* **McGraw-Hill, 1993.**

Imágen 1. Cara de componentes de la tarjeta para Síntesis de Voz con el TSP53C30