

92

36



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

DISEÑO Y CONSTRUCCION DE UN RELOJ PARLANTE UTILIZANDO SINTETIZACION DE VOZ

T E S I S

QUE PARA OBTENER EL TITULO DE:

INGENIERO MECANICO ELECTRICISTA

A R E A E L E C T R O N I C A

P R E S E N T A N :

LUIS FELIPE MENDOZA HERNANDEZ

JORGE EDUARDO PEREZ-JACOME FRISCIONE

Director de Tesis: M. en Ing. José Miguel Martínez Alcaraz

**TESIS CON
FALLA DE ORIGEN**

MEXICO, D. F.

1990



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE TEMATICO:

PAG.

1. INTRODUCCION.....	1
2. ANTECEDENTES DE RELOJ.....	4
2.1 CONCEPTO.....	5
2.2 ¿QUE ES EL TIEMPO?.....	5
2.3 HISTORIA DE LOS RELOJES.....	7
2.4 LOS RELOJES DE CUARZO.....	18
3. NATURALEZA DEL HABLA.....	26
3.1 PRODUCCION Y ANALISIS DEL HABLA... 28	
3.1.1 Análisis de las Vocales....	43
3.1.2 Análisis de las Consonan-	
tes.....	49
3.2 ANALISIS AUDITIVO DEL HABLA.....	64
4. SINTETIZACION DE VOZ.....	70
4.1 GENERALIDADES DE LA SINTETIZACION	
DEL HABLA.....	71
4.2 UNIDADES FONETICAS DE ALMACENA-	
MIENTO DE VOZ.....	76
4.3 TECNICAS DE CODIFICACION DEL	
HABLA.....	78
4.3.1 <i>Vocoders</i>	79
4.3.2 Codificadores por Forma	
de Onda.....	88
4.3.3 Métodos Paramétricos e	
Híbridos.....	99
4.4 SINTETIZACION DEL HABLA	
ALMACENADA.....	109

4.4.1	Entonación.....	109
4.4.2	Modo de operación de los Sintetizadores.....	111
4.4.3	Tecnologías de Hardware para la Sintetización del habla.....	115
5.	DISEÑO DEL RELOJ PARLANTE.....	122
5.1	CARACTERISTICAS DEL RELOJ.....	123
5.2	EL SISTEMA.....	124
5.3	FUNCIONAMIENTO DEL C.I. SP0256...	129
5.3.1	Cómo utilizar el conjunto de <i>Allophones</i>	130
5.4	CONTROL DEL SISTEMA.....	132
5.5	DIAGRAMA DE FLUJO.....	141
5.5.1	Carta ASM.....	142
5.5.2	Reducción de la carta ASM.	142
5.5.3	Tabla de Estado.....	151
5.6	CAPTURA DE DATOS PROVENIENTES DEL C.I. MMS311.....	151
5.6.1	Acoplamiento de Señales...	155
6.	CODIFICACION DE LOS VOCABLOS Y SU DIRECCIONAMIENTO EN MEMORIA.....	157
6.1	CODIFICACION DE LOS VOCABLOS.....	158
6.2	DIRECCIONAMIENTO DE LOS VOCABLOS EN MEMORIA.....	164
6.3	MAPA DE MEMORIA.....	167
7.	CONSTRUCCION Y PRUEBAS DEL RELOJ PARLANTE.....	179

7.1 UTILIZACION DE TABLETAS	
EXPERIMENTALES.....	180
7.2 EL CIRCUITO IMPRESO.....	185
8. CONCLUSIONES.....	198
BIBLIOGRAFIA.....	201
APENDICES.	
A. HOJAS DE DATOS DEL C.I. SP0256....	203
B. HOJAS DE DATOS DEL C.I. MMS311....	224
C. HOJAS DE DATOS DEL C.I. MMS369....	232

I. INTRODUCCION

El objetivo de esta tesis es el de Diseñar y Construir un Reloj Parlante. Este reloj consistirá de un dispositivo que presentará la hora de una manera visible y de un algoritmo capaz de transformar esta hora a una voz en idioma español.

El motivo principal que nos condujo a la realización de este proyecto fué el de ayudar a los ciegos a percibir la hora. Este aspecto es importante, puesto que representa una ayuda a una parte de la sociedad, que aunque no es representativa de ésta (pues hay mucho menos invidentes que videntes), sí tiene derecho a ciertos medios que le permitan substituir algunas de las carencias de sus integrantes por sistemas artificiales de ayuda.

Por medio del trabajo y de las características creativas de este diseño, podemos asomarnos a la vida profesional, es decir, planteamos una posibilidad de lo que podría ser nuestra vida ingenieril: El diseño de sistemas electrónicos.

Este diseño puede servir como base para sistemas de control y monitoreo en donde los operadores eviten el cansancio visual que

produce el hecho de tener que estar viendo tableros.

La organización formal de esta tesis es la siguiente. El capítulo 2 presenta algunos conceptos sobre la medición del tiempo, y contiene una breve historia de los relojes, haciendo énfasis en relojes electrónicos. El capítulo 3 presenta y analiza las características del habla, introduciendo conceptos que se utilizan en el capítulo 4; en el cual se describen las técnicas de codificación y sintetización de voz más utilizadas. El capítulo 5 contiene en detalle todos los pasos que se siguieron en el diseño del Reloj Parlante. En el capítulo 6 se tiene lo referente a la codificación de la voz correspondiente a las diferentes horas. El capítulo 7 describe los pasos que se siguieron hasta obtener el proyecto en circuito impreso funcionando. En el capítulo 8 se presentan las conclusiones.

En el presente escrito, nos hemos permitido reproducir figuras de textos originales pues varias de éstas fueron hechas con graficadores especiales. Por esta razón, algunas figuras muestran indicaciones en inglés.

2. ANTECEDENTES DE RELOJ.

2.1 CONCEPTO.

El reloj es una máquina ideada para registrar e indicar el paso del tiempo en forma que permite dividir el día en horas, minutos y segundos. El único modo de construir una máquina tal es el de conectar un mecanismo que ejecute movimientos regulares, a intervalos isócronos, a un dispositivo contador que registre el número de tales movimientos. Todos los relojes, de cualquier forma que fueren, se guían por este principio.

2.2 ¿QUE ES EL TIEMPO?

Podemos aquí tratar de definir lo que es tiempo, un concepto difícil de comprender. En cuanto a esto, escribió Agustín de Hipona en el siglo V : *"Todos sabemos lo que es el tiempo hasta que nos vemos obligados a explicarlo"*. El tiempo es algo extraordinariamente importante para nuestra existencia, pero ¿Qué es en realidad?

Los pensadores antiguos equiparaban el tiempo con el concepto de cambio. Las primeras colectividades humanas se

encontraban atrapadas por el ritmo constante de las estaciones del año, y su idea del tiempo era más bien la de un ciclo. Esto era así incluso para los antiguos griegos, a pesar de que poseían ya unas leyes matemáticas exactas y un avanzado sistema de geometría. La idea del tiempo como una magnitud que puede medirse con toda precisión es todavía muy moderna.

Un paso importante en este campo lo dió por primera vez Galileo en el siglo XVI, al descubrir que los trayectos recorridos por los cuerpos que caen, se comportan entre sí como los respectivos cuadrados de los tiempos transcurridos durante la caída. Con ello, ya podía considerarse al tiempo como un fenómeno situado dentro del campo físico.

A principios de este siglo, Einstein descubrió que el tiempo es algo que depende del observador, del grado de sus movimientos o de la gravitación que actúa en ese momento sobre él, borrando así la opinión que había dado Newton de un tiempo universal y absoluto. Estas nuevas teorías quedan fuera del contexto de nuestros objetivos para esta tesis. Sobre el tiempo, nos permitimos

repetir el primer párrafo que menciona la Micropaedia Britannica: "Una medida o período medible, una continuidad que carece de dimensiones espaciales. El tiempo es de interés filosófico y también es tema de investigaciones matemáticas y científicas..."

En algún momento de su evolución, quizás cuando descubrió la agricultura y se hizo sedentario, el hombre empezó a requerir de alguna forma de medir el tiempo.

2.3 HISTORIA DE LOS RELOJES.

2.3.1 Reloj de Sol.

El reloj de sol, que es el primer indicador de tiempo que inventó el hombre, nació de la observación de que las sombras que proyectan los objetos iluminados por dicho astro, cambian de longitud en el curso del día. Los primeros fueron, probablemente piedras colocadas en derredor de un poste vertical, marcando las posiciones de la sombra de éste conforme el sol avanzaba en la bóveda celeste. Allá por el año 300 a. de J.C., los caldeos perfeccionaron un reloj de sol que tenía la forma de un cuenco, en el que la sombra trazada por el estilo o gnomon, marcaba 12 horas del día.

2.3.2 Reloj de Fuego.

Para medir el tiempo de noche, o en días sin sol, o en interiores, el hombre tuvo que inventar otros métodos, y, entre los primeros que discurrió figuran los relojes de fuego, en los cuales se podía precisar el tiempo que tardaba en consumirse determinada substancia combustible. Solían hacerse arder con este fin cuerdas de cáñamo que llevaban nudos a intervalos regulares, bujías marcadas con anillos, o bien, aceite que se colocaba en lámparas en que se podía medir la cantidad de combustible.

2.3.3 Reloj de Arena.

El reloj de arena, de muy antiguo origen, ha sobrevivido hasta el día de hoy. En Inglaterra, en la cámara de los comunes se emplea aún un reloj de arena, de 2 minutos, con que se mide el tiempo asignado para que se reúnan los miembros antes de una votación. A menudo, las amas de casa usan relojes de arena en las cocinas, al preparar los alimentos.

2.3.4 Reloj de Agua.

Los relojes de agua fueron los verdaderos

precursores de los relojes modernos. En un antiquísimo reloj chino, el agua goteaba de un vaso a otro que contenía un flotador de madera, el que se elevaba con el agua, indicando así el tiempo y, para mayor exactitud, se usaban hasta 4 vasos, vaciándose de nuevo el agua del último en el vaso superior y el ciclo se repetía. En la Grecia antigua, se perfeccionó una clepsidra, en que el nivel del agua, al elevarse, hacía girar una aguja que indicaba las horas.

Durante la edad media se introdujeron las clepsidras a rodaje, provistas de pesas y ruedas. Conforme el agua elevaba al flotador, hacía descender una pesa, y esto hacía girar un husillo al que se afianzaba la aguja que indicaba las horas.

2.3.5 Reloj de Pesa.

El primer reloj movido exclusivamente por la acción de una pesa, sobre el que existe una descripción completa, fué hecho en 1364 por Enrique de Vick, un relojero alemán, por encargo de Carlos V de Francia. Hacía funcionar dicho reloj el descenso de una pesa de 200 kg; suspendida de una cuerda enrollada en un tambor giratorio. Al girar este,

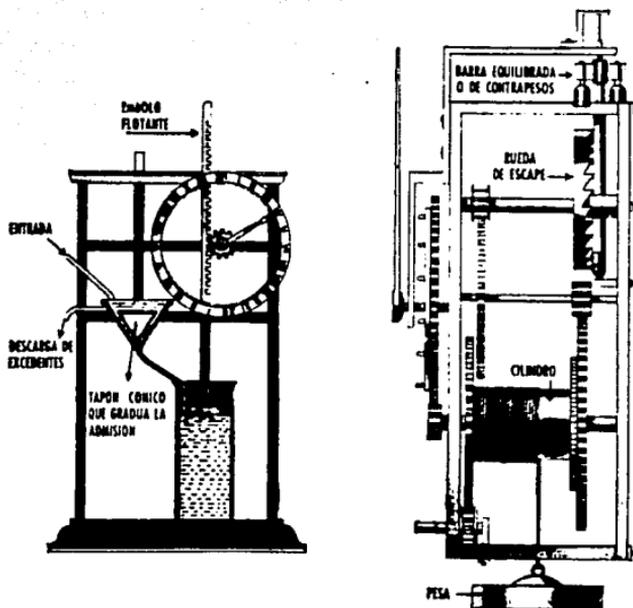


Fig. 2-1. Los relojes medievales dieron origen a los relojes modernos. Izquierda: Al gotear el agua ascendía el émbolo que movía la manecilla. Derecha: La misma idea se aplicó más tarde al aprovechar el descenso lento de una pesa.

impulsaba una serie de ruedas dentadas que movían la aguja de las horas. Una manivela grande permitía volver a enrollar la cuerda. Para hacer que la pesa descendiese lentamente y a un ritmo constante, se recurría al escape de báscula para lo cual se utilizaba una corona o rueda dentada de escape como la final del tren de engranaje. La función de la rueda de escape era aprovechar los movimientos de un balancín de contrapesos para regular la rotación del mecanismo de ruedas dentadas, manteniéndola a un ritmo parejo y constante. La rueda de escape movía un árbol giratorio que llevaba una barra con contrapesos en su parte superior; en la parte inferior del árbol había 2 aletas longitudinales, normales entre sí, colocadas frente a la rueda de escape. Las aletas eran trozos metálicos que sobresalían y enganchaban en los dientes de la rueda. Al hacer contacto una de las aletas, giraba un poco el árbol hasta que la otra pegaba en los dientes. La barra de los contrapesos mantenía el árbol en constante oscilación y ello hacía que la rueda de escape girase en cada movimiento sólo la distancia correspondiente a un diente.

2.3.6 Reloj de Péndulo.

El siguiente adelanto de importancia en la construcción de relojes fué la introducción del péndulo con el que se substituyó el poco práctico balancín de contrapesos que hacía las veces de oscilador en el reloj de Vick. La teoría del péndulo había sido dada a conocer por Galileo desde 1581. Consiste el péndulo en una pesa o plomada que se afianza en el extremo inferior de una barra vertical cuya extremidad superior está suspendida de tal modo que la barra y su plomada pueden oscilar libremente por la acción de la gravedad.

La gran virtud de un péndulo como dispositivo medidor de tiempo estriba en que, dentro de ciertas limitaciones, el espacio de tiempo de su oscilación no se altera por las variaciones en la longitud del arco que describe al oscilar. La duración de la oscilación de un péndulo varía en proporción con su longitud. Se atribuye al astrónomo holandés Christián Huygens la aplicación del principio del péndulo a los relojes, por haber hecho un reloj de péndulo en 1657.

Dado que la duración de la oscilación de

un péndulo depende de un modo tan preciso de su longitud, se deduce que cualquier expansión o contracción de la barra, ocasionadas por cambios de temperatura, afectan la exactitud del reloj en que se emplee. Cuando se requiere una medición exacta del tiempo, hay que compensar tales cambios de longitud del péndulo mediante un dispositivo que la mantenga lo más constante posible en todo tiempo. Un inglés, llamado George Graham, inventó un péndulo con un recipiente de azogue en el extremo inferior. Al aumentar la longitud de la barra metálica por efecto del calor, subía simultáneamente el azogue dentro de un tubo y en esta forma se mantenía constante el centro de gravedad del péndulo. John Harrison, otro inglés, construyó en 1726, un péndulo compensador del tipo de parrilla, basado igualmente en la expansión de los metales. Estaba formado por 9 barras, 5 de acero y 4 de latón, las que se disponían de modo que el efecto del calor se contrarrestaba por la acción de éste sobre los dos metales diferentes. Otro tipo de péndulo compensador en que también se recurre a la diferencia de expansión de metales diversos es uno en que se emplean tubos concéntricos de hierro y cinc.

2.3.7 Reloj con Muelle Real.

Un método más eficaz consiste en fabricar la barra del péndulo con una aleación especial de acero y níquel, llamada *invar*, cuyo coeficiente de expansión es tan pequeño, que el efecto de los cambios de temperatura resulta insignificante.

El muelle real, en la forma de una angosta cinta de acero templado, apretadamente arrollada en espiral, reemplazó pronto a la pesa como el dispositivo que ponía en movimiento a los relojes. Se acredita este adelanto a Peter Henlein, un cerrajero de Nuremberg, y se cree que lo dió a conocer por el año de 1500.

El ritmo de movimiento de un reloj, del que depende la exactitud con que mide el tiempo, se rige por la forma en que funciona su mecanismo oscilante, desde fines del siglo XVII, este mecanismo ha consistido de un muelle y un volante. La fuerza necesaria para mantener el volante en oscilación se conserva en el muelle real, reponiéndose cuando se le dá cuerda al reloj.

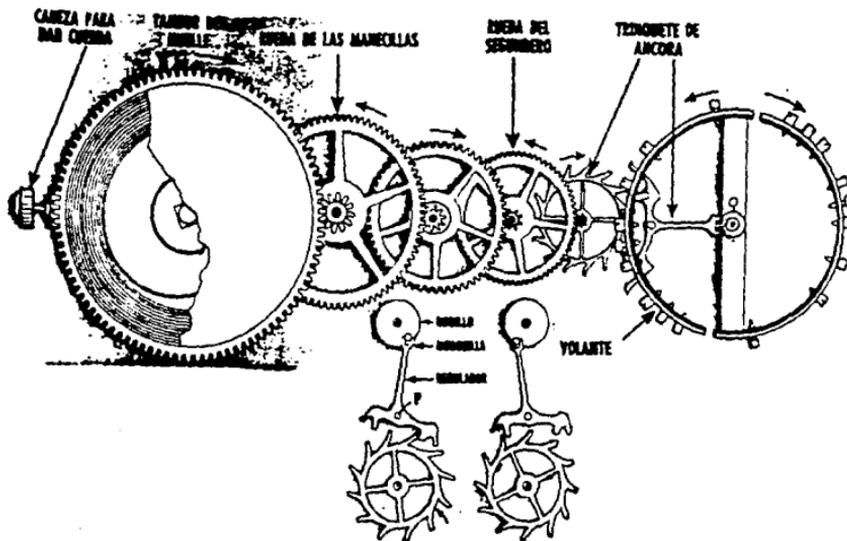


Fig. 2-2. Diagrama del mecanismo de un reloj de bolsillo. Para mantener el muelle, se encierra éste en un tambor dentado. El tambor transmite el movimiento al tren de ruedas dentadas y piñones que finalmente lo hacen llegar a la rueda de escape. La marcha de esta rueda depende de las oscilaciones del volante. A medida que éste llega a su posición extrema y gira en sentido contrario, el tope del rodillo mueve el trinquete (con pivote en P) y deja escapar un diente de la rueda de escape. A cada oscilación se repite el ciclo. Para contrarrestar la dilatación motivada por los cambios de temperatura, el volante está hecho con dos metales distintos.

La fuerza retenida en el muelle real se transmite al volante por medio del tren de las ruedas dentadas y la de escape. Las manecillas se mueven debido a la acción de una rueda dentada que gira con un ritmo adecuado.

2.3.8 Reloj de Impulsos Eléctricos:

Un tipo de reloj mas moderno lo constituyen unos sistemas de designación unitaria del tiempo en los que un reloj maestro hace funcionar automáticamente, y a distancia, varias esferas subordinadas a las que se hacen llegar impulsos eléctricos. Por su confiabilidad, bajo costo de mantenimiento, y también por su precisión (si se les regula debidamente), se extendió el uso de tales relojes. Consisten, por lo general, en un péndulo que marca un segundo en cada oscilación y hace girar una pequeña rueda de 15 dientes una revolución por cada 30 segundos. Conforme dicha rueda completa su rotación, suelta un retén y deja caer una palanca de gravedad sobre el péndulo, dando a este un impulso de continuidad. Al llegar la palanca de gravedad a su punto mas bajo, cierra un contacto eléctrico y un electroimán vuelve inmediatamente a colocarla en su

retén, donde permanece hasta el fin del siguiente medio minuto, cuando se repite el ciclo.

Las esferas subordinadas, de las que puede haber un gran número, son mecanismos sencillos en que una rueda de 120 dientes se desplaza hacia adelante a cada medio minuto por la acción de los mismos impulsos eléctricos que han vuelto a poner en su lugar la placa de gravedad del reloj maestro. En las esferas subordinadas, dichos impulsos eléctricos hacen avanzar un garfio de trinquete hasta el diente siguiente de la rueda de 120, moviendo así las manecillas. La ventaja principal de este sistema es su perfección teórica, demostrada por el hecho de que se utiliza en forma modificada en el reloj de péndulo libre.

2.3.9 Reloj Síncrono.

Posteriormente aparecieron los relojes eléctricos síncronos que consisten en un motor pequeño del tipo llamado síncrono acoplado a un engranaje reductor y, como lo indica el nombre, el rotor del motor gira al ritmo exacto de la frecuencia de la corriente alterna. Si la frecuencia se mantiene al

número de ciclos por segundo de antemano determinado y, si se provee un engranaje reductor, de diseño adecuado que conecte el motor con las manecillas de reloj, indicarán éstas la hora correcta. Los relojes síncronos no son, en realidad, relojes en el sentido verdadero de la expresión; son sencillamente medidores de frecuencia que repiten el tiempo que les transmite la planta de fuerza eléctrica.

2.4 LOS RELOJES DE CUARZO.

Paralelamente a la aplicación comercial de los relojes eléctricos, se estaba desarrollando un tipo de reloj que cambiaría el curso de la relojería en el mundo. Fue aproximadamente en el año 1930 que se empezó a dar un importante paso en la medición del tiempo, al descubrir que los cristales de cuarzo podían ser unos extraordinarios metrónomos para los relojes eléctricos. Aparecieron entonces en el mercado los relojes de cuarzo, que resultaron ser con mucho los más exactos concebidos hasta el momento. Estos nuevos medidores constataron que la rotación terrestre como punto de partida para medir el tiempo no era demasiado

fiable; debido a los desplazamientos en las masas internas de la tierra y a los cambios en la correlación de energías sobre la corteza terrestre, en cada milenio se produciría un error de unos 5 segundos más o menos.

El cuarzo es uno de los componentes más comunes de las rocas. Cuando es puro y de estructura homogénea, posee unas propiedades denominadas piezoeléctricas: Bajo la influencia de una corriente alterna, vibra con extrema regularidad; y esta frecuencia, asombrosamente estable, ha sabido utilizarse para medir el tiempo. Los cientos de millones de vibraciones de las partículas de cuarzo, han llevado a cabo una minirevolución. La hora ya no necesita ningún elemento mecánico.

Dentro de las características principales de los relojes de cuarzo está la enorme precisión: En un año, es decir, en 525800 minutos, sólo observan un margen de error de un minuto, contra el de 2 a la semana que suelen tener los relojes con sistema mecánico. El secreto de tal precisión reside en un segmento de cristal de cuarzo de 8 mm de longitud por 2 de anchura, una barrita

minúscula que ha hecho pasar a la historia el mecanismo complicado del péndulo. Las últimas investigaciones están introduciendo relojes sofisticados capaces de conseguir un margen de error de 30 segundos al siglo.

La técnica relojera acaparó durante 2 siglos el protagonismo de la medida del tiempo, debido a la intervención de 3 elementos fundamentales: Un oscilador, el péndulo; una fuente de energía, el resorte; y un dispositivo de máquina contable, que a través del juego de una rueda dentada, mueve las manecillas. Hoy, estas piezas que siempre están en movimiento, han sido reemplazadas por otras inmóviles; es decir, una fuente de energía eléctrica como una pila, un circuito impreso, y el cristal de cuarzo. El interior de la caja de este nuevo reloj no contiene ningún elemento de relojería tradicional.

El físico francés Pierre Curie fué el primero en descubrir las propiedades eléctricas de los cristales. En 1883, en el curso de una de sus investigaciones, comprobó que una ligera sacudida hacía aparecer cargas eléctricas en la superficie de los cristales. Entonces procedió a la experiencia inversa y

sometió un trozo de cuarzo a una corriente alterna. De repente, los cristales empezaron a emitir oscilaciones de gran rapidez (de 50000 Hertz a varios millones, según fuera la dimensión del cristal) y de una extraordinaria regularidad, hasta el punto de que la frecuencia de estas vibraciones tenía la particularidad de ser constante. Ahí residía toda la importancia de este descubrimiento que él denominó *efecto piezoeléctrico*.

Este hallazgo quedó largo tiempo en vía muerta, no se sabía que hacer con él. Fué preciso esperar a las investigaciones de los científicos alemanes Adelsberger y Sheibe, en 1933, para que la teoría se aplicase a la relojería. Debido a su gran margen de inexactitud, los relojes mecánicos no estaban capacitados para colaborar en sus trabajos científicos. Entonces recordaron los descubrimientos de Pierre Curie y dedujeron que si los cristales de cuarzo podían emitir durante algunos años unas vibraciones de frecuencia constante (regularidad que no conocían los péndulos), se debían poder utilizar en lugar de los péndulos.

Quedaba, sin embargo, un problema: Cómo transformar las vibraciones de cuarzo en un contador visible de unidades de tiempo. Dicho de otra forma, cómo reducir al equivalente de un Hertz (una oscilación por segundo), unas vibraciones que pueden alcanzar 2,400,000 [Hertz]. Ayudándose de componentes electrónicos diversos, Adelsberger y Sheibe inventaron un divisor de frecuencias (en base a lo que hoy conocemos como Flip-Flops), que reducía estas vibraciones.

En 1934, sus trabajos desembocaron en un reloj que tenía las medidas de un armario... Era ya enormemente preciso, permitía solamente una desviación de una millonésima de segundo por día.

Posteriormente, el reloj atómico realizado en 1955 por los ingleses Esen y Parry mejoraría este resultado.

Sin embargo, quedaba todavía un largo camino por recorrer antes de llegar al reloj de cuarzo actualmente tan común. Sin el viaje a la Luna de los años sesenta y sin el programa de investigaciones espaciales de la NASA, este reloj no existiría aún. En efecto,

los imperativos de la astronáutica potenciaron la construcción de sistemas cada vez mas miniaturizados, particularmente en el campo de la electrónica, y provocaron un salto tecnológico con la puesta en marcha del famoso circuito integrado.

Hoy, en una pastilla de silicio de 2 [mm]² se pueden integrar hasta 80000 transistores elementales. Con lo que respecta al reloj de cuarzo, el circuito integrado cumple un papel muy concreto. Este debe reemplazar el armario de Adelsberg y Sheibe, transformando las oscilaciones de cuarzo en una señal de tiempo legible para cualquier individuo.

Para realizar esta operación, el circuito integrado está equipado de numerosos transistores internos interconectados, que van a proporcionar un tratamiento de los impulsos del oscilador de cuarzo, como si se tratara de una cadena de montaje. El primer circuito es el divisor de frecuencias. Se consideran frecuencias ideales las de 32768 y 16384 [Hz] ¿Porqué? Simplemente porque estas cifras son potencias de 2 y permiten ser divididas y redivididas entre 2 hasta

alcanzar una vibración por segundo. Esta única vibración por segundo será utilizada para indicar el tiempo.

Los relojes de cuarzo funcionan de forma muy diferente según sea el modelo. En los que tienen apariencia clásica, con manecillas y esfera, el impulso segundo es enviada a un amplificador, cuyo papel consiste en reforzar el impulso, demasiado débil para el trabajo que se le pide; se le incorpora un micromotor para convertir los impulsos en movimiento, a fin de que paso a paso, en cada impulso eléctrico, arrastrase el minuterero y el segundero como si se tratara de un reloj tradicional. Los relojes digitales (*digit: cifra*), no poseen ningún mecanismo, ni engranaje, ni manecillas: Son 100% electrónicos.

Los dígitos iluminados en rojo suelen ser de leds, (*light emitting diodes*), y los que representan sus dígitos en negro son los relojes con indicación por cristales líquidos, LCD (*liquid crystal display*), que consumen muy poca energía.

El contador de segundos cuenta de 1 a 60.

Cuando llega a esta última cifra se crea un nuevo impulso que llega hasta el contador de minutos. El impulso minuto es enviado al contador de las horas... Toda la información recogida en estos contadores se transmite a un decodificador que la transforma en cifras legibles sobre el módulo indicador de la esfera.

Los desplegados reservados a las cifras en el indicador tienen, cada uno, 7 segmentos susceptibles de iluminarse y el decodificador se conecta a cada uno de ellos.

3. NATURALEZA DEL HABLA.

La transferencia de mensajes a través de la conversación es la forma más comúnmente usada de comunicación entre dos seres inteligentes. Afortunadamente para nosotros, es también la forma más rápida y exacta que tenemos a nuestra disposición. En efecto, nosotros nos basamos en la comunicación hablada tanto que si se nos denegara la capacidad de hablar nos frustraríamos rápidamente. Para ilustrar este punto simplemente tratemos de hacer la rutina normal del diario con al menos una hora sin hablar; si hacemos esto entonces al final de esa hora podríamos probablemente haber desarrollado un seña burda del lenguaje como sustituto, o habríamos dispuesto de un papel y un lápiz. Imaginemos la extraña posibilidad de una vida inteligente existente en algún lejano planeta sin una atmósfera que lo rodee. Como la comunicación a través de los sonidos sería imposible sin un medio de transmisión, habrían desarrollado otras formas de comunicación. Estos seres serían tan comunicativos en sus movimientos corporales como nosotros lo somos con nuestros sonidos. Si en dicho planeta tampoco existiera la luz, la vista no habría podido existir, por lo que quizás se habrían

desarrollado distintas formas de comunicación que utilizarían el tacto o hasta telepatía mental. Sin embargo, regresando a la realidad nosotros sí tenemos el medio de transmisión para la voz y al parecer hemos ciertamente aprendido a usarlo.

¿Qué es exactamente el habla? ¿Qué ocurre físicamente cuando nosotros hacemos el acto común de conversar con un compañero? La respuesta a estas preguntas no es trivial. Las características de las comunicaciones habladas y su contenido semántico asociado son extremadamente complejos. A continuación presentamos un estudio elemental de la voz y de sus propiedades.

3.1 PRODUCCION Y ANALISIS DEL HABLA.

Cuando alguien empieza a generar un mensaje hablado, un gran número de cosas suceden casi simultáneamente. Primero el mensaje es generado dentro del cerebro como el supuesto portador de información. La figura 3-1 da un simplificado diagrama de flujo que ilustra una posible forma de describir el proceso del habla.

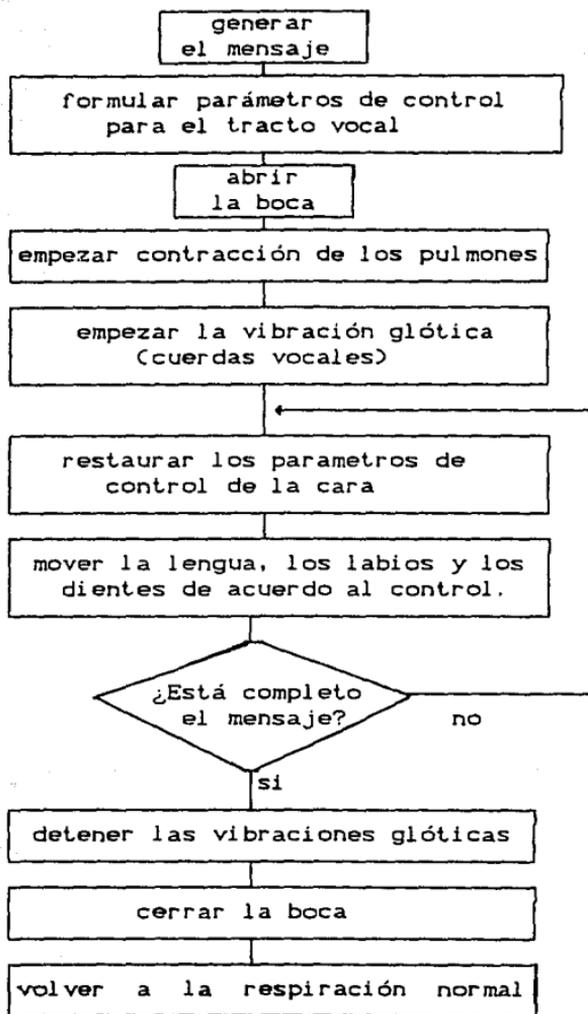


Fig. 3-1. Proceso del Habla

Después de la generación del mensaje en el cerebro, los parámetros de control para el tracto vocal son formulados desde la memoria y el habla comienza. Mientras el verdadero orden en que la secuencia de eventos ocurre puede ser abierto a debates, las funciones dadas en el diagrama de flujo generalmente ocurrirán durante un mensaje hablado. Realmente, el verdadero orden o secuencia que es seguido depende mucho de la palabra que se esté hablando. Las palabras que empiezan con la letra M, por ejemplo, tienen un comienzo nasal; entonces, la boca no se abre sino hasta después de la contracción de los pulmones y de que las vibraciones glóticas hayan empezado. El primer sonido simplemente viene a través de la nariz.

Los sonidos del habla son producidos al escapar el aire de los pulmones. El flujo estable de aire que viene de los pulmones es parcialmente convertido en energía pulsante por uno o más de los siguientes mecanismos:

- 1) La acción de las cuerdas vocales, que convierte el flujo estable de aire en pulsos cortos.

2) Turbulencia creada en constricciones delgadas en el tracto vocal.

3) Repentino libramiento de presión excesiva después de que el tracto vocal haya sido cerrado completamente en algún punto a lo largo de su longitud.

La señal acústica creada de esta manera, llamada *función de excitación* de voz (o del habla) puede tomar 3 formas diferentes, de acuerdo a los 3 modos de conversión de flujo de aire antes mencionados:

1) Pulsos cuasi-periódicos.

2) Ruido continuo.

3) Un pulso único.

La mayoría de los sonidos del habla son generados por una de estas funciones de excitación. El primer modo produce el habla *vocalizada*, dentro de la cual podemos encontrar las cinco vocales del español. El segundo de los modos mencionados es el causante de los sonidos no vocalizados llamados *fricativos*, como los creados al pronunciar las letras s, f, etc. El tercer modo de conversión de flujo es el causante de los sonidos llamados *plosivos*, tales como p,

t, y k. Existen además sonidos del habla que se crean en base a una combinación de 2 tipos de funciones de excitación. Inclusive en algunos idiomas existen sonidos creados por los 3 modos de excitación.

En total, hay probablemente cerca de cuarenta diferentes sonidos para hacer cualquier lenguaje hablado. A éstos se les llama fonemas. Ellos en total comprenden un arreglo de sonidos hablados distinguibles mutuamente exclusivos que pueden ser encontrados en casi cualquier lenguaje hablado. En la tabla 3-1 presentamos unos ejemplos de fonemas en el idioma español (de México).

Los fonemas tienen características espectrales que pueden no ser inicialmente obvias. Algunos pertenecen al grupo de fonemas llamados continuos. Estos fonemas en particular, por la falta de movimiento del tracto vocal durante su pronunciación, tienen un espectro de frecuencia estable y constante. Los sonidos continuos son sostenidos. (Cabe aclarar que no es lo mismo el flujo continuo del aire, el cual no produce voz, que los sonidos continuos).

Tipo de fonema:	Fonema:	Ejemplo:
Vocales	a	casa
	e	poner
	i,y	ir
	o	oso
	u	uno
Consonantes		
Fricativas vocalizadas	rr,r	hierro
Fricativas no vocalizadas	f	fosa
	j,g,x	jalar
	s,z,c,x	solo
Plosivas vocalizadas	g,gu	ganar
	d	dado
	b,v	bueno
Plosivas no vocalizadas	k,c,q	caer
	p	papa
	t	tirar
	ch	chango
Nasales	m	mar
	n	nuca
Nasales fricativas	ñ	niño
Glides	y,ll	llave
Semivocales (líquidas)	l	pala
	r	arar

Tabla 3-1. Fonemas del idioma español.

Incluidos en esta categoría están las vocales, los fricativos, los nasales y los semivocales. El resto de las clases de fonemas, los plosivos y los *glides*, son considerados como sonidos dinámicos no continuos, que normalmente se acoplan con los fonemas que lo rodean en una manera semejante a los diptongos.

Los diptongos existen como una clase de sonidos hablados caracterizados por un movimiento extremo del tracto vocal al acoplarse con otros fonemas. Estos son generados durante el movimiento de la boca de la posición de un fonema a la del siguiente. Si pudiéramos hablar fonemas puros concatenados, sin intentar unir estos sonidos, entonces los diptongos no existirían. Sin embargo, como el tiempo de respuesta de los músculos dentro de nuestra garganta y boca tiende a apresurar el movimiento de un fonema al siguiente, generamos un gran número de diptongos dentro de nuestra secuencia de voz.

Si una pequeña porción del espectro del habla fuera tomada durante una pronunciación continua, podría probablemente ser identificada de su espectro como un fonema

específico. Sin embargo, si un segmento corto se tomara de un diptongo, de un plosivo o de un *glide*, probablemente un fonema específico no podría ser identificado.

Los espectros de las funciones de excitación cubren un amplio rango de frecuencias, con componentes especterales significativas en gran parte del rango de frecuencias audibles. Para producir los diferentes sonidos del habla para cada modo de excitación, las señales de excitación son "filtradas" por el tracto vocal. En la figura 3-2 se muestra el espectro del "habla" vocalizada antes de ser filtrada, es decir, los llamados *pulsos glóticos* o bien *vibración glótica* que es el sonido que se produce por la vibración de las cuerdas vocales. Este espectro (el de la fig. 3-2) es el que se generaría si se colocara un micrófono para tomar la señal exactamente sobre las cuerdas vocales y por debajo de las estructuras resonantes del tracto vocal.

El espectro de los pulsos glóticos está constituido por armónicas del período básico de ciclo, el cual es el rango fundamental de frecuencias producidas por las cuerdas

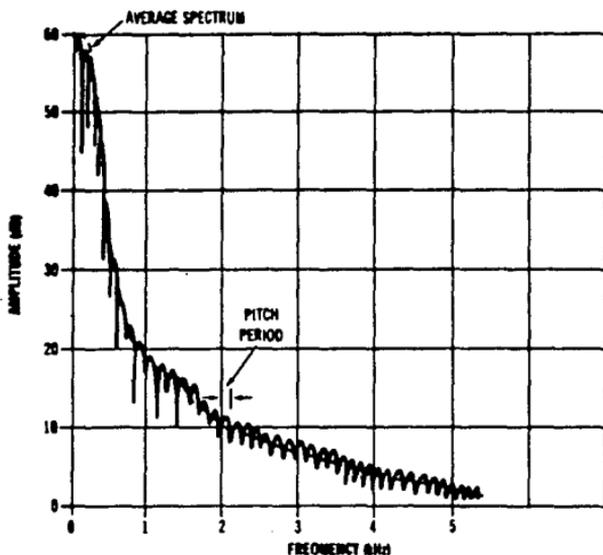


Figura 3-2. El espectro de pulsos glóticos.

vocales. A pesar de que el espectro tiene una muy fuerte componente cerca de la frecuencia fundamental (aprox. 50[Hz]), es muy rico en armónicas, por lo que tiene componentes de frecuencia incluso arriba de los 5[kHz].

En si mismo, el espectro de frecuencias de las vibraciones glóticas tiende a dar muy poca información como señal de habla. Sin embargo, si el micrófono de grabación fuese

trasladado fuera del tracto vocal cerca de los labios, entonces el espectro tomaría una apariencia diferente. Ahora tendría ya la información del habla. El efecto físico que ocurre en este proceso es la *resonancia*.

La respuesta en frecuencia del tracto vocal depende de la posición de la lengua, los labios y otros órganos de articulación. Esta respuesta del tracto vocal se caracteriza por cierto número de resonancias o *formantes* (lóbulos de frecuencia). Existen típicamente 3 lóbulos abajo de los 3000[Hz], estos dan forma a la función de excitación y causan las características de calidad y timbre de cada sonido.

Al pasar la vibración glótica a través de las cámaras del tracto vocal, ciertos volúmenes tienden a resonar con frecuencias variantes de manera análoga a los tonos producidos por un órgano tubular. Una ilustración de las cámaras de resonancia en el tracto vocal se da en la figura 3-3. Esta es una ilustración muy simplificada de los efectos de la nariz, garganta y boca sobre el espectro glótico, y está muy de acuerdo con la analogía con las cámaras de un órgano

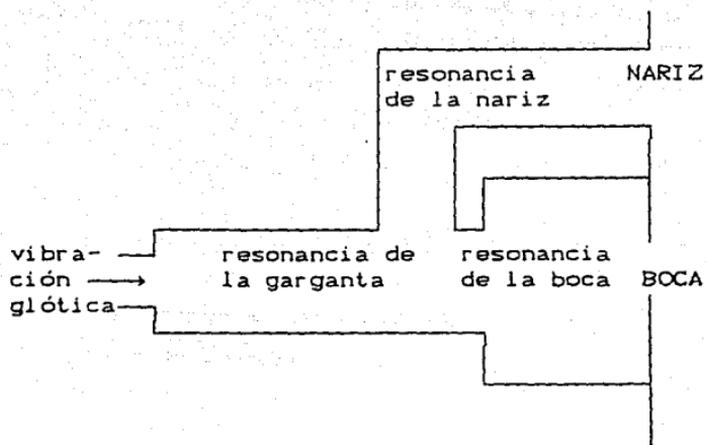


Figura 3.3. Resonancias del Tracto Vocal.

tubular. Como era de esperarse, como tenemos 3 cámaras de resonancia, entonces habrá 3 frecuencias principales de resonancia, enfatizadas en el espectro glótico de vibración. Estas principales frecuencias de resonancia son conocidas como *formants frequencies* (frecuencias de lóbulo). Aunque puede demostrarse que el habla normal humana tiene más de 3 lóbulos principales de frecuencia, aquellos lóbulos por encima del tercero, contienen relativamente poca energía, y pueden ser descartados al considerar el espectro total del habla. Esto

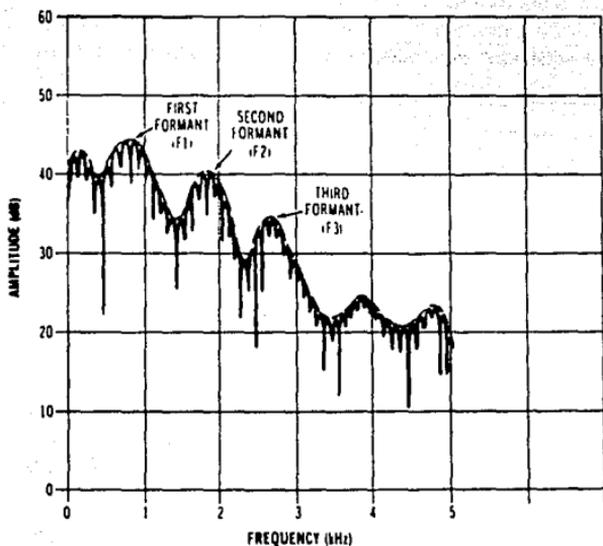


Figura 3-4. Espectro del habla con lóbulos.

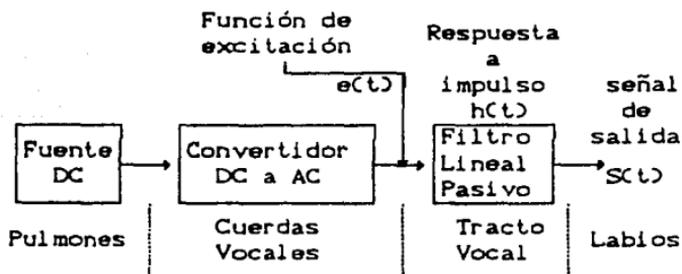
es lo que efectivamente sucede cuando uno habla con otra persona a través del teléfono. El mecanismo que causa la pérdida de los lóbulos de alto valor es un filtro paso-bajas generado en el circuito electrónico del teléfono; el filtrado únicamente deja pasar frecuencias por debajo o menores a los 3[kHz]. Una representación gráfica del espectro de frecuencias que muestra los lóbulos de frecuencia se muestra en la figura 3-4. Una relación entre este espectro y las cámaras resonantes de la figura 3-3 mostraría

que el lóbulo de más baja frecuencia (F1), está formado por la cámara de resonancia de la garganta. El siguiente lóbulo de frecuencia (F2), se relaciona con la cámara resonante nasal, y el tercero es asociado con la resonancia de la boca.

Cuando el tracto vocal presenta un área transversal uniforme a lo largo de toda su longitud entre las cuerdas vocales y los labios, sus frecuencias de lóbulo están aproximadamente en 500[Hz], 1300[Hz], y 2500[Hz]. Estas frecuencias corresponden a las resonancias $\lambda/4$, $3\lambda/4$, y $5\lambda/4$ de un tubo cuya longitud es de 17.5[cm], y cuya área es de 5[cm]² y que está abierto en un extremo y cerrado en el otro. Esta condición es aproximada durante la llamada vocal neutra, correspondiente a la letra *e* del español. Para casi todas las vocales, el área transversal dista mucho de ser uniforme, y las frecuencias de lóbulo son alteradas hacia arriba o hacia abajo de su posición neutra. Generalmente, abajo de los 5[kHz], el tracto vocal puede ser considerado como una línea de transmisión no uniforme con impedancias terminales variables: La impedancia de guía representada por las cuerdas vocales y la

impedancia de radiación de la abertura de la boca. En frecuencias mayores, existe propagación transversal del sonido, lo que hace que ya no se pueda aplicar la misma analogía.

Este es un simplificado diagrama de bloques que representa la producción de sonidos vocalizados del habla:



El flujo estable ("DC") de aire desde los pulmones es convertido en energía acústica pulsante ("AC") por las cuerdas vocales, y su espectro es modificado por las resonancias del tracto vocal. La señal de habla de salida $s(t)$ que aparece en los labios es la convolución de la función de excitación $e(t)$, que representa el flujo de aire en las cuerdas vocales, con la respuesta a impulso del filtro que representa al tracto vocal:

$$s(t) = \int_{-\infty}^t e(t) \cdot h(t-\tau) d\tau$$

Después de aplicar la transformada de Fourier en ambos lados de la ecuación, la convolución se convierte en una multiplicación,

$$S(f) = E(f) \cdot H(f)$$

De esta manera el espectro de la señal de habla es el producto del espectro de la excitación $E(f)$, y la función de transferencia del tracto vocal $H(f)$.

Suele usarse más comúnmente el espectro de amplitud de la señal de habla:

$$|S(f)| = |E(f)| \cdot |H(f)|$$

En la práctica, tanto el espectro de la excitación como la función de transferencia varían con el tiempo, por lo tanto, en lugar de utilizarse el análisis espectral a lo largo de todo el tiempo, se usa un análisis espectral en períodos cortos de tiempo, haciendo uso de las llamadas *ventanas de tiempo* que suelen tener una duración semejante a la del fonema más corto. Se ha

demostrado que para obtener una buena resolución, se requiere de ventanas de 40[ms] de duración.

Para hacer un análisis del habla, suele ésta dividirse en las ventanas de tiempo antes mencionadas, y se considera a lo largo de la duración de toda una ventana, que el filtrado realizado por el tracto vocal es estacionario, es decir, que el tracto vocal no cambia su forma durante este período. Así, puede hacerse un análisis de contenido de frecuencias de cada uno de estos períodos y obtener diferentes espectros de frecuencia, utilizando un espectrógrafo.

3.1.1 Análisis de las vocales.

Las figuras 3-5 a 3-9 muestran una tabulación del área transversal en función a la posición a lo largo del tracto vocal para cada una de las 5 vocales del idioma español. Estas figuras muestran además el espectro de frecuencias y la forma de onda producida al emitir cada una de dichas vocales. Cada una de las figuras muestra además la frecuencia central y el ancho de banda de sus primeros lóbulos (*formants*).

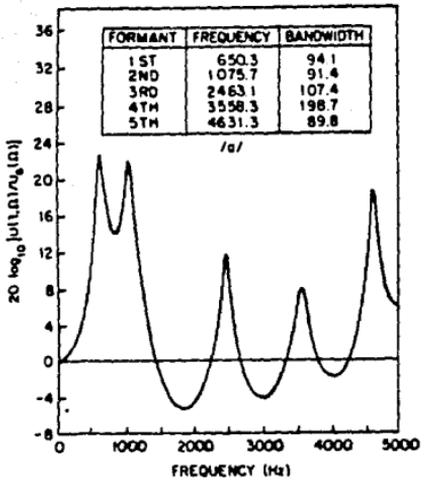
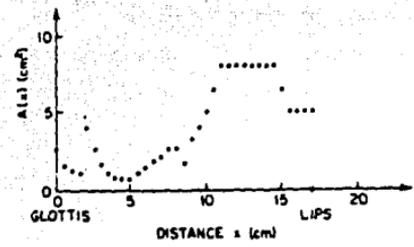


Figura 3-5. Función del área, respuesta en frecuencia y forma de onda para la vocal /a/.

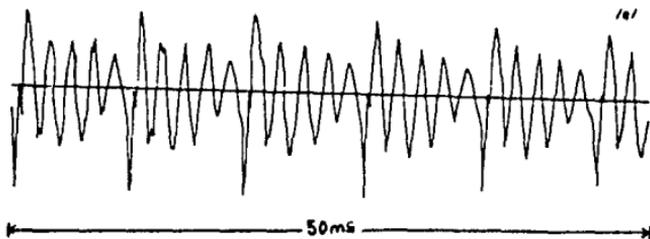
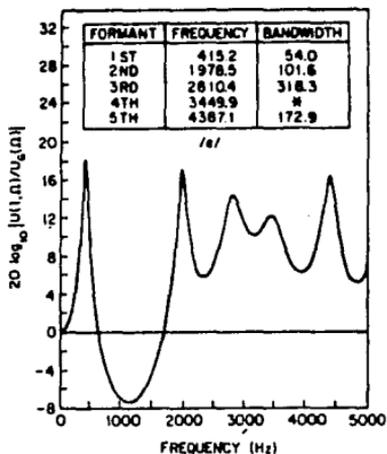
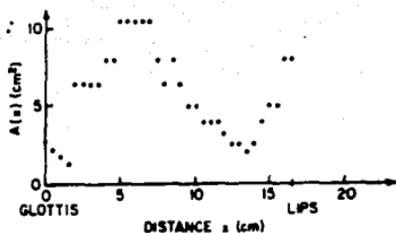


Figura 3-6. Función del área, respuesta en frecuencia y forma de onda para la vocal /e/.

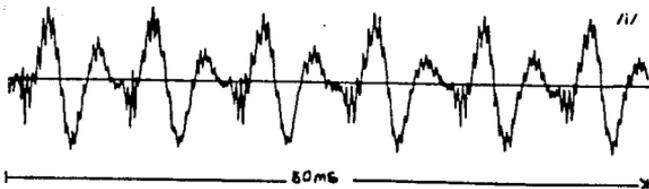
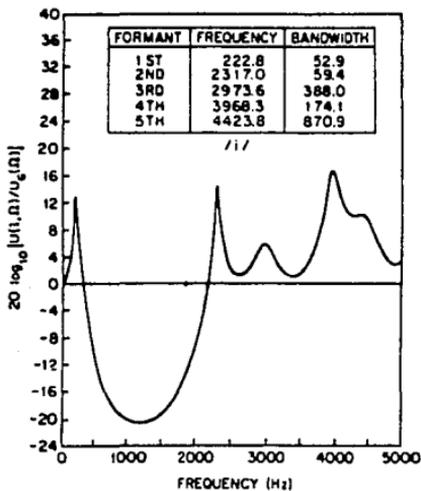
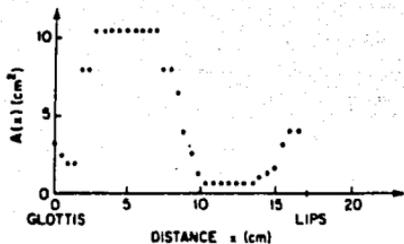


Figura 3-7. Función del área, respuesta en frecuencia y forma de onda para la vocal /i/.

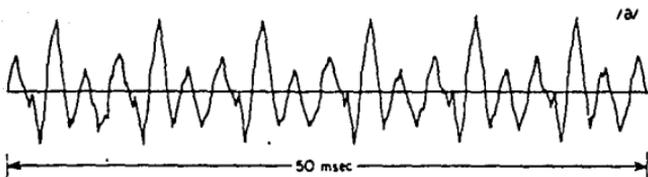
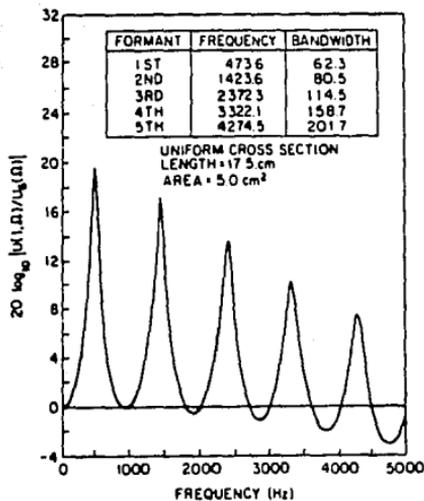


Figura 3-8. Función de respuesta en frecuencia y forma de onda para la vocal neutra, equivalente a la /a/. Para esta vocal, el área transversal a lo largo del tracto vocal es prácticamente constante.

3.1.2 Análisis de las consonantes.

Hemos agrupado en la tabla 3-1 los diferentes fonemas del idioma español. Ahora se hace un análisis particular por grupos de fonemas, considerando la manera en que son emitidos, su forma de onda y su espectrograma.

a) Semivocales y *Glides*: Estos 2 grupos suelen ser agrupados dentro de un mismo grupo por algunos autores. Tienen una naturaleza parecida a la de las vocales. En el español se pueden incluir los fonemas /l/, /r/ y /y/ ó /w/, otro fonema de este grupo, no existente en nuestro idioma es /w/. Las características acústicas de estos sonidos son sumamente influenciadas por el contexto en las que estas ocurren.

b) Nasales: Dentro de este grupo podemos incluir a los fonemas /m/ y /n/. Estos son producidos con una excitación glótica al tener el tracto vocal totalmente cerrado en algún punto del conducto oral. El vélum baja, ocasionando que el aire fluya a través del tracto nasal de manera que el sonido sea radiado por las fosas nasales. El tracto vocal, a pesar de que queda cerrado en la

/ə ma/

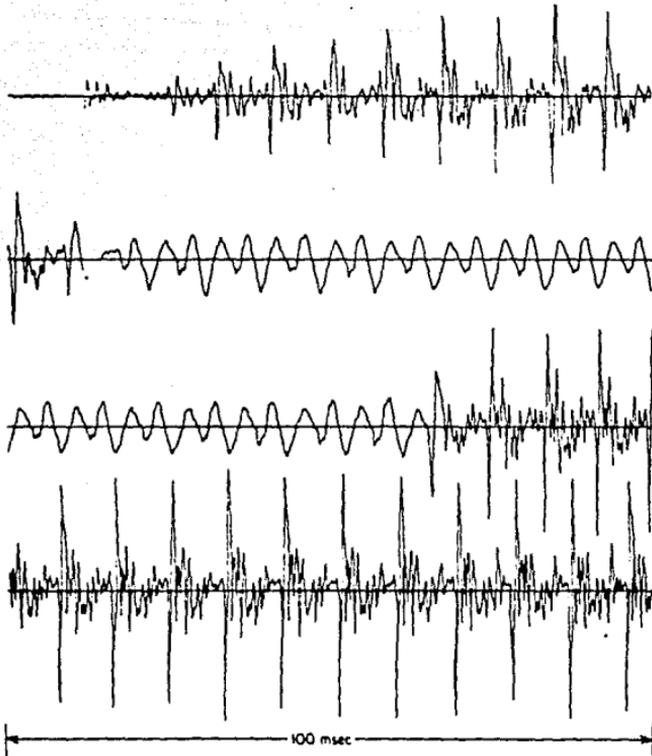


Figura 3-10. Formas de onda y espectrogramas para los vocables /əma/ y /əna/

parte anterior de la boca, se mantiene acoplado con la faringe, entonces la boca sirve como una cámara de resonancia que atrapa energía acústica de ciertas frecuencias.

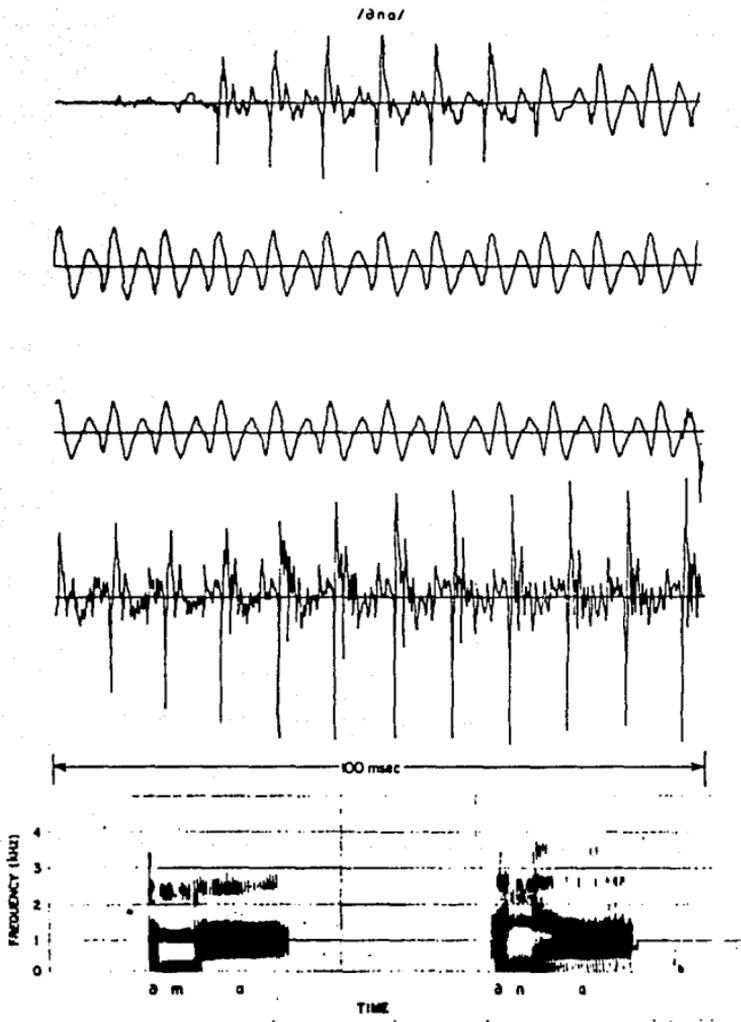


Figura 3-10. (Continuación).

La diferencia entre el sonido causado por cada una de las dos consonantes nasales se debe a la posición en la cual se cierra el tracto oral; para la /m/, la constricción se hace en los labios, mientras que para la /n/ la constricción se hace justo detrás de los dientes. En algunos idiomas, existe un sonido nasal en el que la constricción se hace en el vélum mismo. Para el caso de la /ɲ/, se hace una constricción incompleta detrás de los dientes, que permite un ligero flujo de aire a través del tracto oral. Por tal motivo, este fonema no es considerado nasal puro. La fig. 3-10 muestra la forma de onda y el espectrograma para cada uno de los vocablos: /oma/ y /ona/.

c) Fricativas no vocalizadas: Este tipo de fonemas se producen al excitar el tracto vocal por un flujo de aire estable que se transforma en turbulento en la región de una constricción en algún punto del tracto vocal. En el alfabeto español, podemos incluir a los sonidos equivalentes a las letras /f/, /s/, /ʃ/ y (solo en España) /z/ (representado también como /s/). Otro sonido incluido en este grupo es /sʎ/, que, pese a no existir oficialmente en nuestro idioma, suele

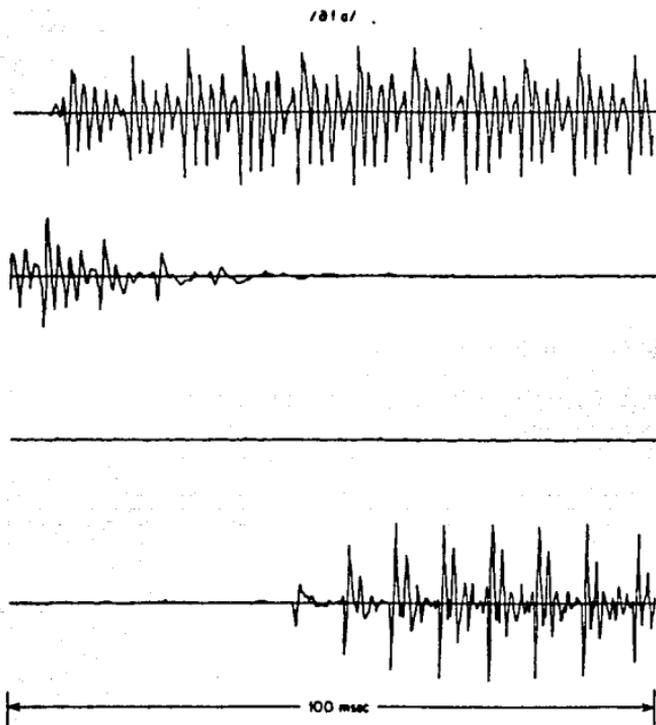


Fig. 3-11. Formas de onda y espectrogramas para los vocábulos /əfə/, /əsa/ y /əsha/.

utilizarse en palabras mexicanas con la letra /x/. El punto en el cual ocurre la constricción sirve para determinar qué sonido fricativo es producido. Para /f/, la constricción ocurre cerca de los labios; para /s/, ésta está cerca de los dientes; para /x/

/ðso/

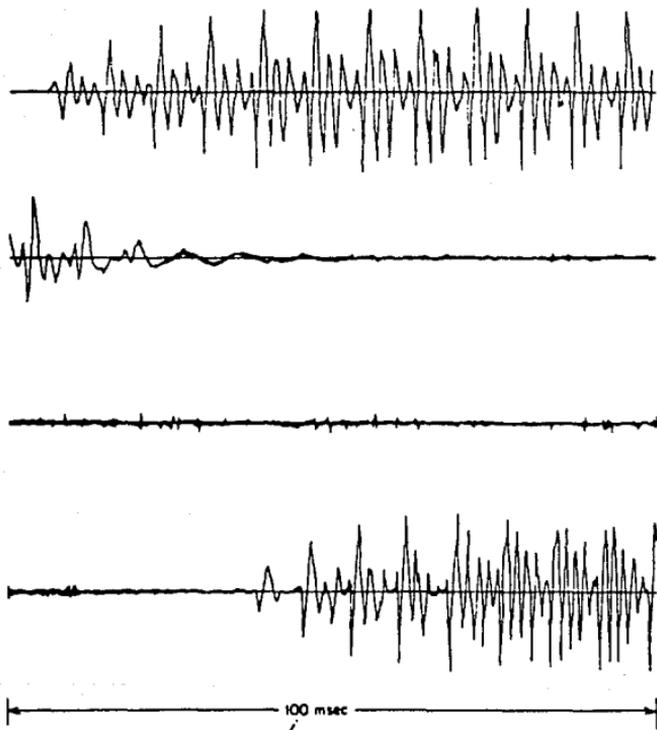


Figura 3-11. (Continuación)

la constricción está en la zona media del tracto oral; para /s/ ésta ocurre cerca de la parte posterior del tracto oral; y finalmente para /j/, la constricción ocurre justo adelante del vélum.

/ðʃa/

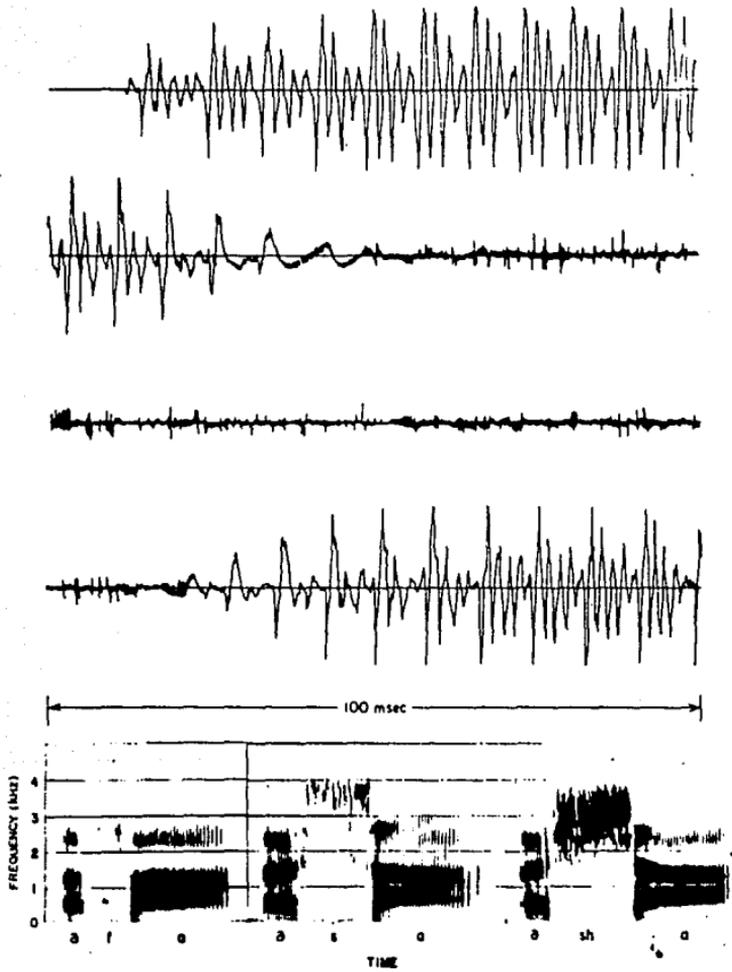


Figura 3-11. (continuación).

Es así como los sonidos fricativos no vocalizados se producen por una fuente de ruido en una constricción, misma que separa el tracto vocal en dos cavidades, lo que, como en las nasales, hace que se atrape energía en la cavidad posterior, y se produzcan los sonidos que se emiten en la parte anterior de la boca. La fig. 3-11 muestra las formas de onda y los espectrogramas para los fonemas /k/, /s/ y /sh/. En las gráficas de forma de onda se observa la no periodicidad de este tipo de sonidos.

d) Fricativas vocalizadas: En la pronunciación de este tipo de fonemas, tienen que ver dos diferentes fuentes de excitación. Aquí, las cuerdas vocales están vibrando, y por tanto, una fuente de excitación está en la glotis. Sin embargo, como el tracto vocal hay una constricción en algún punto, el flujo de aire se hace turbulento en las cercanías de la constricción. En algunos idiomas existen varios sonidos de este tipo, sin embargo en el español solo tenemos /w/. En algunos lugares, suele aún conservarse la pronunciación de la /v/ como una /k/ vocalizada, lo que la hace caer dentro de este grupo. En Argentina suele pronunciarse

/ova/

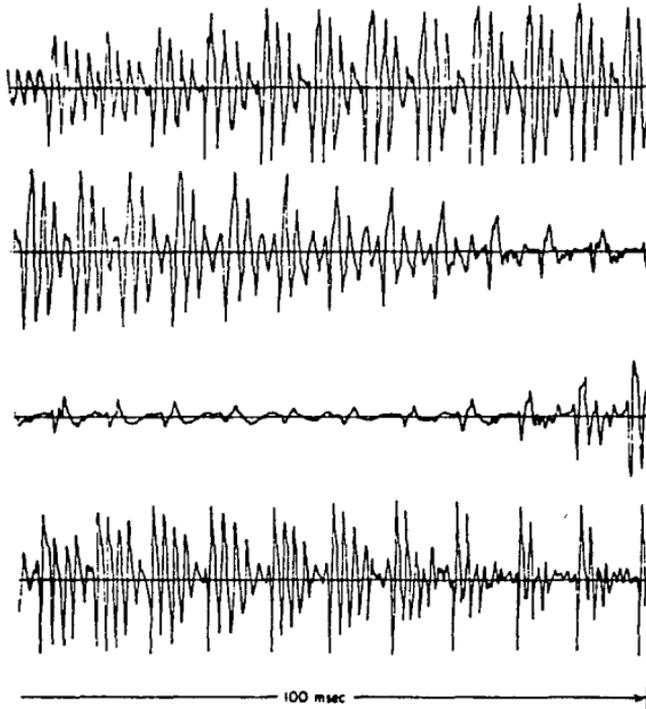


Fig. 3-12 Formas de onda y espectrogramas para los vocales /ova/ y /ozha/

el sonido /zh/, que es fricativo vocalizado, este se usa en vez del sonido /ll/ de otros lugares. La figura 3-12 muestra espectrogramas y formas de onda de los fonemas /ova/ y /ozha/

/Bzha/

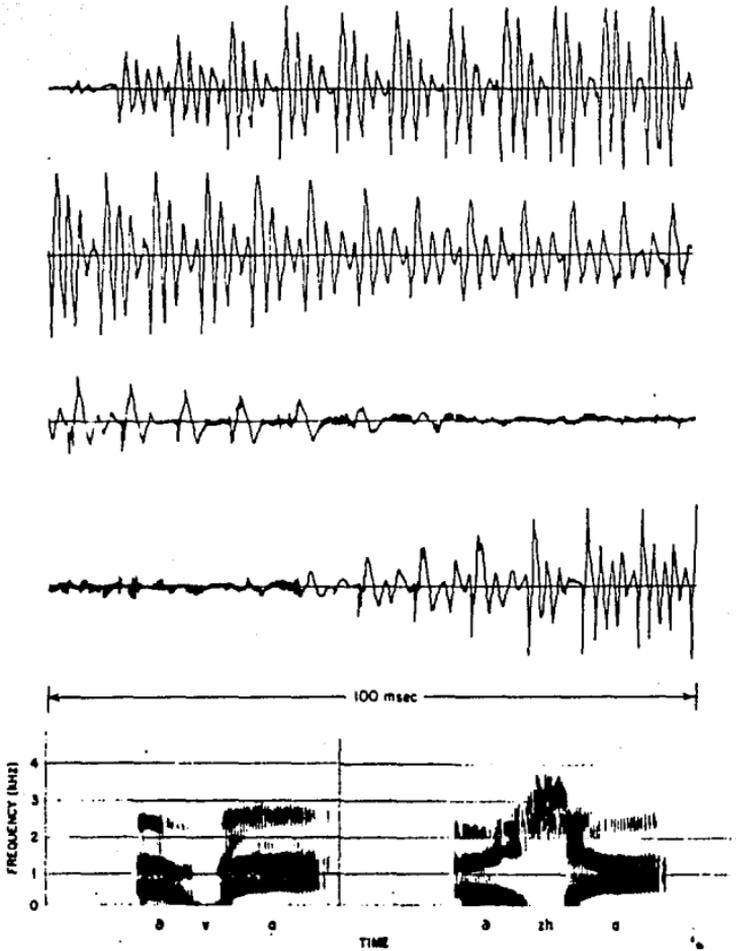


Figura 3-12. (Continuación).

e) Plosivas vocalizadas: En este grupo se encuentran los fonemas /b/, /g/ y /d/, que son transitorios y no continuos. Estos sonidos son producidos al ejercer presión atrás de una constricción completa en algún punto del tracto oral, y repentinamente dejar de ejercer esta presión. Para el caso de la /b/, la constricción ocurre en los labios; para la /d/ la constricción es detrás de los dientes; y para la /g/, ésta ocurre cerca del vélum. Durante el período de la constricción completa del tracto oral, no se radia ningún sonido a través de los labios; sin embargo, suele existir una pequeña emisión de energía de baja frecuencia a través de las paredes de la garganta. Esto pasa cuando las cuerdas vocales pueden vibrar aún cuando el tracto vocal está cerrado en algún punto.

Como los sonidos plosivos son dinámicos por naturaleza, sus propiedades están altamente influenciadas por la vocal que le sigue. La figura 3-13 muestra la forma de onda y el espectrograma del vocablo /oba/. En la forma de onda para la /b/ puede distinguirse la excitación vocalizada y falta de componente de alta frecuencia.

/oba/

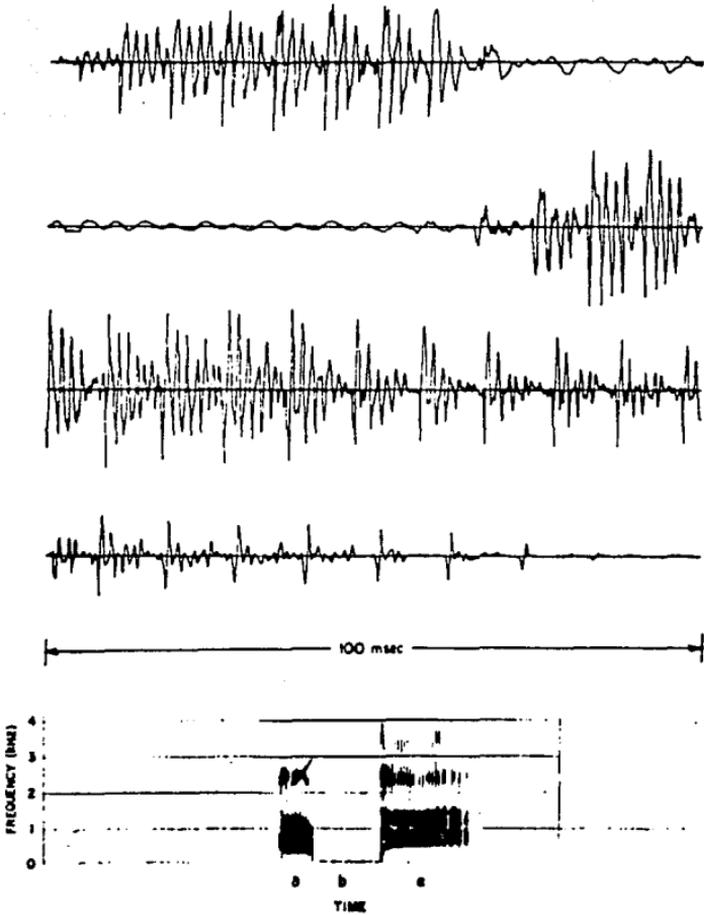


Figura 3-13. Forma de onda y espectrograma para el vocable /oba/.

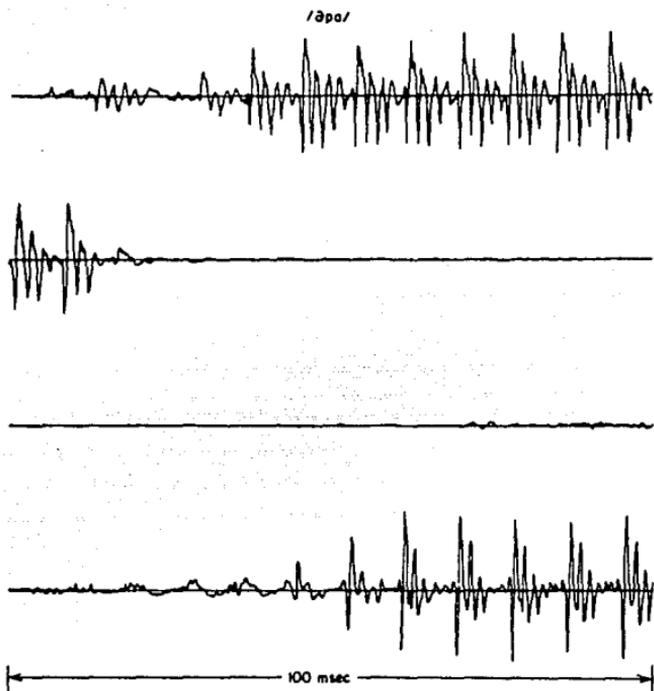


Figura 3-14. Forma de onda y espectrograma para los vocables /əpə/ y /ətə/.

f) Plosivas no vocalizadas: Aquí se pueden incluir los fonemas /p/, /t/ y /k/, que son similares a su contraparte vocalizada /b/, /d/ y /g/. La diferencia de las plosivas vocalizadas con las no vocalizadas está en que para éstas, las cuerdas vocales no

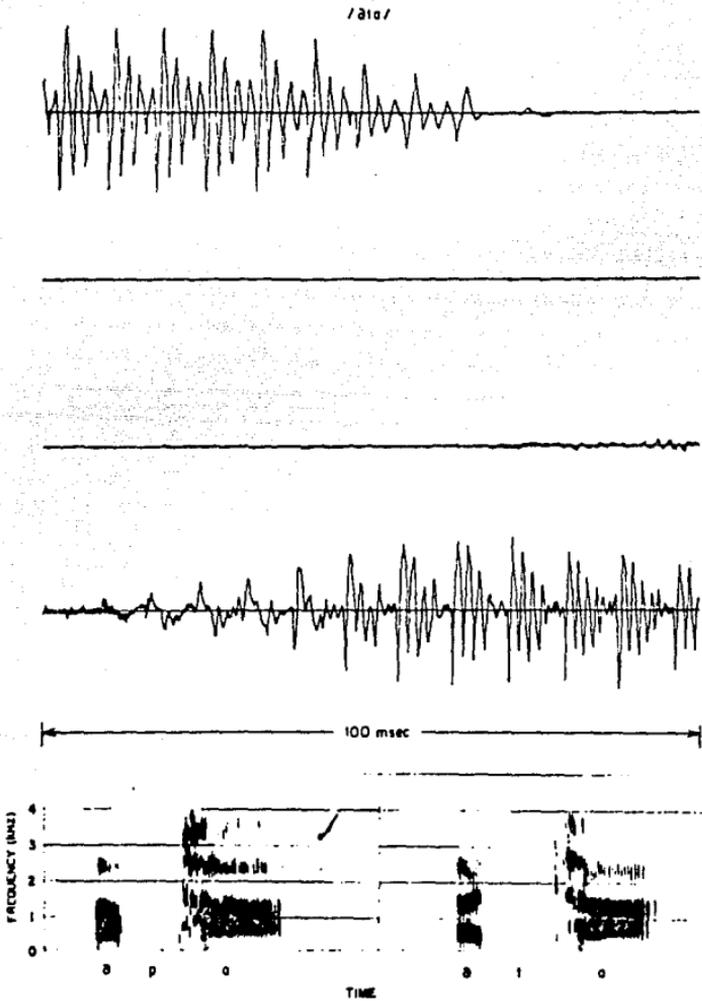


Figura 3-14. (Continuación).

vibran. De esta manera, después de la constricción, ocurre un pequeño intervalo de fricción. Dentro de este grupo podemos también incluir al fonema /ch/, que presenta un período de fricción más prolongado. Por estas características se le podría llamar también plosiva-fricativa. En la figura 3-14 se muestra la forma de onda y el espectrograma de plosivas no vocalizadas /p/ y /t/.

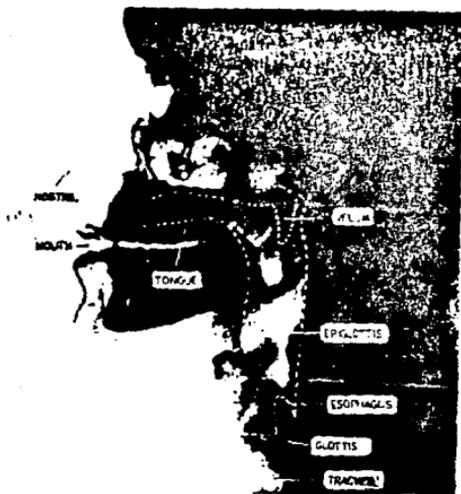


Figura 3-15. Rayos-X del plano sagital del aparato vocal humano.

3.2 ANALISIS AUDITIVO DEL HABLA.

Existen tres propiedades básicas relacionadas con la percepción humana del habla, de las cuales puede tomarse ventaja para el análisis y la sintetización del habla:

- 1) El oído realiza un análisis espectral de intervalos cortos.
- 2) Para la percepción monoaural, el oído es prácticamente insensible a la fase.
- 3) El oído es excesivamente sensible a la frecuencia de las señales del habla.

Los primeros en considerar al oído como un analizador de espectros insensible a la fase fueron los científicos Ohm y Von Helmholtz, quienes pensaron que la membrana basilar del oído interno funcionaba como un arreglo de resonadores entonados, semejante a las cuerdas de un piano. Esta idea fue después rectificada por Von Békésy, quien demostró la existencia de ondas que viajan a través de la membrana basilar. El modelo de la membrana basilar como un simple banco de resonadores tuvo que ser abandonada a cambio de una nueva descripción que consideraba al

oído como una línea de transmisión no uniforme con una resolución espectral limitada. La discriminación de altas frecuencias del oído no puede considerarse únicamente mecánica, se asume que ésta ocurre más bien en el cerebro.

Se ha podido demostrar que los cambios de fase en una señal de habla si pueden tener un mínimo efecto en la calidad de la señal percibida, sin embargo, esto no tiene ningún efecto en cuanto a la inteligibilidad de la señal escuchada. Lo que sí presenta problemas es una distorsión de fase mayor a los 50(ms), puesto que este período tiene una longitud tal que puede percibirse por el oído.

El oído presenta una altísima capacidad de discriminación de frecuencias. Por ejemplo para componentes de frecuencia cercanas a los 1000 Hz. se han reportado percepciones de variaciones tan pequeñas como de 3 Hz. Estos datos fueron tomados por los científicos E.G. Shower y R. Biddulph. Este asombroso resultado conduce a la conclusión de que no es propiamente en el oído donde se lleva a cabo este filtrado, puesto que mecánicamente no puede explicarse; por ello, se ha llegado

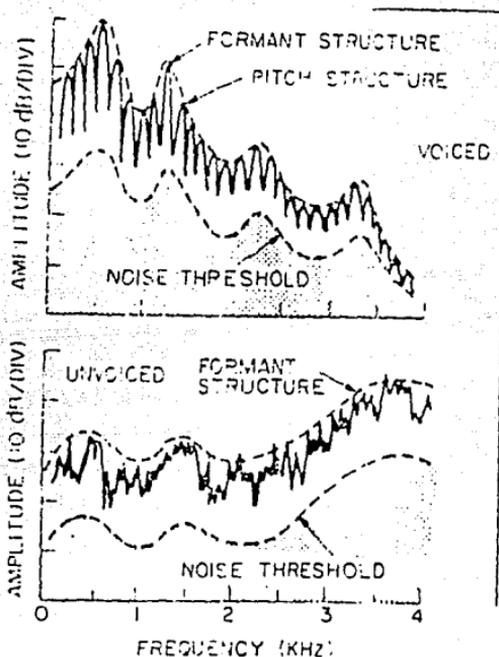


Fig. 3-16. Modelos espectrales para señales de habla vocalizada y no vocalizada, considerando el fenómeno de enmascaramiento auditivo.

a la conclusión de que esto se lleva a cabo en un proceso neuronal.

Otra particularidad que presenta el oído es que presenta diferentes controles

automáticos de ganancia a distintas frecuencias de una manera análoga a un banco de filtros (ecualizador). Además, por las propias características del oído, las señales muy fuertes de baja frecuencia interfieren a las de alta frecuencia, reduciendo su sensibilidad. Esto conduce a un fenómeno llamado enmascaramiento auditivo, como se ilustra en las áreas sombreadas de la fig. 3-16. Esencialmente, cualquier ruido o distorsión añadida a una señal no se percibe si cae por debajo del umbral (ver fig. 3-16).

El enmascaramiento auditivo se presenta en cualquier persona, inclusive cuando el oír el ruido "enmascarado" pudiera salvar su vida, como podría suceder con un peatón que atraviesa una calle escuchando su *walkman*, y no advierte la presencia de un automóvil porque el sonido del claxon está por debajo del umbral de la señal auditiva que el vá escuchando.

El enmascaramiento auditivo ofrece ventajas en la sintetización de voz, pues permite que el ruido de cuantificación de una señal se haga inaudible siempre y cuando esté por debajo del umbral.

Una particularidad psicoacústica que se ha observado es que existen bandas críticas enmascaradas de acuerdo a las frecuencias de prominencias espectrales tales como los lóbulos. Se ha determinado que si la frecuencia central de esta prominencia es menor a 600[Hz], se produce un enmascaramiento de las frecuencias de 100[Hz], y si la frecuencia central es mayor a 600[Hz], la banda crítica (o sea, la que no se escucha) se ubica en $1/6$ de la frecuencia central.

El enmascaramiento que se produce sobre sonidos de frecuencia mayor a la del sonido enmascarante sí tiene una explicación fisiológica: En el oído interno, el sonido se propaga por medio de ondas que viajan de los lugares donde se detectan las frecuencias altas a los lugares donde se detectan las frecuencias bajas; por tal motivo las frecuencias bajas interfieren (enmascaran) con las frecuencias altas, y no las altas a las bajas.

En este capítulo se ha hecho un estudio básico de la Naturaleza del Habla, los

conceptos que aquí se han presentado nos permiten ya entrar al tema de la sintetización de voz, mismo que se presenta en el siguiente capítulo.

4. SINTETIZACION DE VOZ

El objetivo de este capítulo es el de presentar a grandes rasgos las bases de la sintetización de voz, abarcando algunas de las técnicas más utilizadas para este fin. Hacemos notar que este capítulo no pretende hacer un análisis exhaustivo de cada una de estas técnicas. Sin embargo, lo que en este capítulo presentamos ayuda a comprender el funcionamiento del sintetizador de voz que nosotros ocupamos para nuestro diseño. (El desarrollo del diseño se presenta en otro capítulo). Además, dado que existe muy poco material sobre sintetización de voz en español, este capítulo puede servir como una introducción al interesado en el estudio de este tema, con la ventaja de que está hecho en nuestro idioma.

4.1 GENERALIDADES DE LA SINTETIZACION DEL HABLA.

La generación automática del habla por medio de datos digitales se ha desarrollado rápidamente en años recientes. Antes de que los chips de procesamiento de señales especializados fueran introducidos hace una década, el habla sintética era principalmente generada mediante grandes computadoras

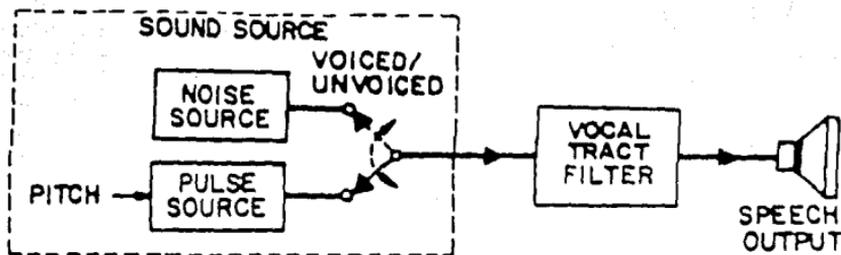


Figura 4-1. Modelo eléctrico de la producción del habla.

interconectadas con un modelo análogo al tracto vocal. Hoy en día existen dispositivos para producir voz que van desde los programas de software baratos para computadoras caseras hasta máquinas capaces de leer en ayuda a invidentes, que permiten leer cualquier tipo de texto impreso. La reciente expansión de sintetizadores comerciales se debe principalmente a los avances en la tecnología de circuitos integrados, pero también a las mejoras realizadas en la metodología de sintetización de voz.

Las representaciones del habla varían en complejidad, desde simples muestras de formas de onda de voz tomadas periódicamente en tiempo hasta estimaciones de parámetros del modelo de la figura 4-1 que equivale a la

división de sonidos en vocalizados y no vocalizados (capítulo 3). La selección de la representación digital esta gobernada por tres factores principales: Complejidad de proceso, tasa de información, y flexibilidad. Por complejidad, entendemos la cantidad de procesamiento requerido para obtener la representación escogida. En muchos casos la complejidad de procesamiento es una medida del costo de implementación del sistema en hardware. Una representación simple puede generalmente ser implementada más económicamente que una representación compleja. Entonces la complejidad es a veces una consideración difícil en algunas aplicaciones de transmisión donde el bajo costo terminal es crucial. La tasa de información o de bits es una medida de la redundancia en la señal de habla que ha sido eliminada por el procesamiento. Una baja tasa de bits significa que la representación digital de la señal de habla puede ser transmitida por un canal de baja capacidad, o guardada eficientemente en memoria digital. Finalmente, flexibilidad es una medida de cómo el habla puede ser manipulada o alterada por aplicaciones diferentes a las de transmisión, tales como respuesta de voz,

reconocimiento de habla y verificación de un orador.

En general, una mayor complejidad es el precio pagado para bajar la tasa de bits e incrementar la flexibilidad. En transmisión y en respuesta de voz, la calidad y la inteligibilidad del habla reconstruida son también consideraciones importantes.

La sintetización del habla requiere de la conversión de una secuencia de comandos o texto de entrada (palabras u oraciones) a formas de onda del habla, utilizando algoritmos y voz previamente codificada. El texto puede ser introducido mediante teclado, reconocimiento óptico de caracteres, o desde una base de datos almacenada previamente. (fig 4-2)

Los sintetizadores de voz pueden ser caracterizados tanto por el tamaño de las unidades de habla que concatenan para conseguir una emisión hablada, como por el método utilizado para codificar, guardar, y sintetizar la voz. Las unidades grandes de habla, tales como frases y oraciones, pueden



Figura 4-2. Diagrama de bloques de un sintetizador de voz.

dar una alta calidad de voz (con grandes requerimientos de memoria). Los métodos eficientes de codificación reducen el tamaño de memoria, pero usualmente degradan la calidad del habla.

La mayoría de los sintetizadores tienen un ancho de banda de entre 3[kHz] (para aplicaciones de telefonía) y 5[kHz] (para más alta calidad. Las vocales son bien percibidas a frecuencias de hasta 3[kHz], puesto que ellas pueden ser especificadas por las tres primeras resonancias (lóbulos de frecuencia) del tracto vocal. La percepción de algunas consonantes es ligeramente deteriorada si las frecuencias alrededor de 3 y 5 [kHz] son omitidas. Las frecuencias mayores a 5[kHz] son quizás útiles para mejorar la claridad

del habla y su naturalidad, pero no ayudan gran cosa en lo que respecta al entendimiento de la voz.

Cuando se tiene una calidad semejante a la de la red telefónica, se dice que se tiene una calidad *toll quality*, la cual equivale al habla que tiene una relación señal-ruido de 60(dB).

4.2 UNIDADES FONÉTICAS DE ALMACENAMIENTO DE VOZ.

Los sintetizadores más simples concatenan palabras almacenadas o frases. Este método nos entrega una alta calidad del habla (dependiendo del método utilizado por el sintetizador), pero está limitado por la necesidad de almacenar en la memoria de la computadora (ROM), todas las frases a ser sintetizadas después de que éstas hayan sido pronunciadas, ya sea aisladamente o en un conjunto de oraciones. Para obtener una mayor naturalidad en la pronunciación del sintetizador, cada palabra u oración deberá originalmente ser pronunciada con la apropiada secuencia y entonación para todas las oraciones en las cuales podrían ser utilizadas. Entonces, una palabra puede tener

diferentes estructuras de sintaxis. la pronunciación deberá ser grabada y almacenada utilizando cadenas de oraciones simulando distintos contextos. Al concatenar palabras originalmente habladas de una manera aislada, se produce una pérdida de inteligibilidad y naturalidad, degradando la calidad original presentada por este método.

Sistemas más avanzados generan habla por medio de una secuencia de fonemas, la mínima unidad de diferencia entre las palabras. Las características espectrales de estos sonidos cortos (50-200[ms]) deben ser suavizados en los límites de los fonemas, para evitar el brincoteo y discontinuidad de la voz. (El problema es que la pronunciación de un fonema en una palabra o en una frase depende mucho del contexto del fonema.) Las reglas de suavización y la necesidad de calcular una entonación apropiada pueden dar como resultado sintetizadores complejos y un lenguaje aún menos natural.

Los sintetizadores híbridos (más recientes), concatenan unidades de tamaño intermedio, de voz almacenada, tales como sílabas, demisílabas, y difonemas, usando

suavización de parámetros espectrales en las fronteras de las unidades. Una sílaba consiste comúnmente de una vocal más su consonante vecina. Las demisílabas son unidades de habla obtenidas al dividir sílabas por la mitad, haciendo el corte en la mitad de la vocal (donde la lengua y los labios se mueven poco, contrario a las transiciones rápidas que ocurren en las fronteras de los fonemas). Los difonemas se obtienen por la división de la forma de onda del habla en unidades de tamaño de fonema, haciendo el corte a la mitad de cada fonema. Cuando las demisílabas o los difonemas son concatenados en una secuencia diferente a la de su grabación original, se obtiene usualmente habla suave, porque los sonidos unidos en las fronteras son espectralmente similares.

4.3 TECNICAS DE CODIFICACION DEL HABLA.

Los métodos de codificación del habla pueden ser agrupados en 3 diferentes grupos; 7080D8R, CODIFICADORES POR FORMA DE ONDA Y METODOS PARAMETRICOS E HIBRIDOS.

4.3.1 Vocodero.

Fué en Octubre de 1928 cuando Homer Dudley, de la compañía *Bell Telephone Laboratories* bosquejó un dispositivo que después sería llamado *Vocoder* (Un término derivado de las palabras *voice-voz* y *coder-codificador*). Estos sistemas fueron los pioneros de la sintetización de voz.

Originalmente el *vocoder* fué concebido como un método de transmisiones eficientes de señales de voz, especialmente en líneas telefónicas. En telefonía, el ancho de banda ordinario es de unos 3000(Hz) (desde 200(Hz) hasta 3200(Hz)). Esta porción del espectro del habla es suficiente para asegurar calidad e inteligibilidad adecuadas. Sin embargo, Dudley demostró que las señales de voz pueden ser transmitidas en una fracción de este ancho de banda si se utilizan métodos adecuados de codificación.

Dudley demostró que gran parte de la señal de habla natural es redundante y que pueden utilizarse métodos que sacan ventaja de esta característica, tanto para transmisión como para almacenamiento de señales de voz. Desde su invención, los

vocoders han sido perfeccionados y utilizados en diferentes áreas, tomando ventaja de los desarrollos tecnológicos que han aparecido en el transcurso del tiempo.

El análisis y la sintetización de voz se basan en una descripción paramétrica del espectro de corto tiempo (una ventana espectral) o alguna representación equivalente de la señal del habla. Pueden existir diferentes esquemas de análisis y sintetización dependiendo de diferentes selecciones de parámetros.

El primer método de análisis y sintetización que empleó una descripción paramétrica en un espectro de una ventana de tiempo fué el *spectrum channel vocoder*. Dentro de éste, la envolvente espectral está representada por 10 a 20 muestras espaciadas a lo largo del eje de la frecuencia. Además, la frecuencia fundamental (equivalente a la primera espiga del espectro), queda representada por un parámetro extra, el cual es nulo para períodos no vocalizados o de silencio. En la figura 4-3 se muestra un diagrama de bloques de este tipo de *vocoder*. La señal de habla se separa en bandas

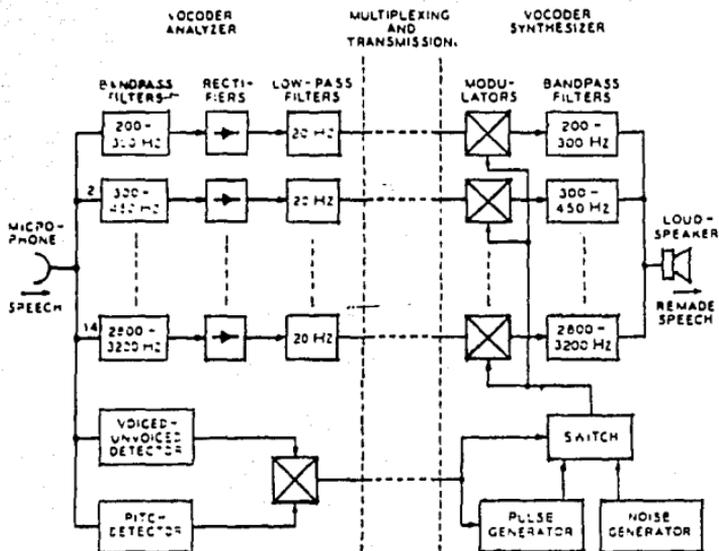


Figura 4-3. Diagrama de bloques de un spectrum channel vocoder.

espectrales adyacentes, cuyos anchos de banda son de 100-400[Hz] y que abarcan el espectro de frecuencias desde 200[Hz] hasta 3200[Hz]. Este rango de frecuencias, típico de señales telefónicas, permite una alta inteligibilidad y una buena calidad. La salida de cada filtro se conecta a un rectificador y a un filtro paso-bajas cuya salida representa la componente de DC (el promedio) de la amplitud para cada banda de frecuencia en un cierto

intervalo de tiempo. En conjunto, estas 14 señales representan la envolvente del espectro de una señal en un período pequeño de tiempo.

También se muestra en la figura 4-3 un detector de períodos vocalizados o no vocalizados, y un detector de frecuencia fundamental (presente sólo en intervalos vocalizados). Estas 15 señales se combinan para formar una sola señal cuyo ancho de banda total es de $15 \times 20[\text{Hz}] = 300[\text{Hz}]$. Por tanto, el ancho de banda total para la transmisión es de únicamente una décima parte del ancho de banda requerido para la señal original.

Para una transmisión digital, el ahorro es incluso mayor. Para una tasa de transmisión de 40 muestras por segundo, usando 3 bits por muestra para cada uno de los 14 canales y 6 bits por muestra para la frecuencia fundamental, la señal de salida del *vocoder* puede ser representada por una tasa de bits de $40 \times (14 \times 3 + 6) = 1920[\text{bits/s}]$. Esto es menos de un veinteavo de la tasa de bits requerida para representar la señal de habla original (7000 muestras/seg. por 7

bits/muestra \cong 50000(bits/s)). Sin embargo, los *vocoders* que funcionan con tales tasas de bits presentan una baja calidad.

En el sintetizador, cada una de las señales de los distintos canales son recuperadas y utilizadas para controlar la respuesta en frecuencia de un filtro variable en el tiempo (que consiste de moduladores y filtros paso-banda) para reconstruir la envolvente medida en el analizador. La entrada a este filtro variable en el tiempo es una señal espectralmente plana que contiene una estructura adecuada (pulsos cuasi-periódicos para los segmentos vocalizados o ruido blanco para los segmentos no vocalizados).

Hemos dado a conocer el *vocoder* original presentado por H. W. Dudley en la exposición *New York World's Fair* en 1939. El mismo principio que originalmente se utilizó en transmisión de voz, se ha utilizado ampliamente en almacenamiento y sintetización de voz, esto debido a la muy baja tasa de bits que este método maneja.

A través de los años se ha tratado de

perfeccionar el esquema original del *vocoder*. Para tal efecto, se ha hecho uso de diferentes técnicas tales como la utilización de muestras recientes para hacer una predicción posterior de la señal (Esta técnica se utiliza en ciertos codificadores paramétricos, y se presenta en la sección 4.3.3.). Otra técnica que se utiliza es la de manejar escalas logarítmicas para efecto de la codificación de amplitud en las diferentes bandas de frecuencia; esto permite que en partes de baja amplitud se tenga una mayor resolución que en las de alta amplitud, tomando ventaja de que el oído responde en amplitud (además de en frecuencia) en una escala logarítmica, y de que las partes de alta amplitud son poco frecuentes en una señal de habla.

Discriminación de Bandas: Una modificación que se le ha hecho al *vocoder* original es la de seleccionar únicamente las bandas de frecuencia cuya amplitud sea de las mayores; por ejemplo, pueden sólo seleccionarse las 5 bandas con más alta amplitud y discriminarse las demás, esto puede llevarse a cabo considerando que el oído, al captar una señal de habla, tiende a enmascarar el

contenido de poca amplitud con las de mayor amplitud. Con este método se puede reducir mucho la señal a transmitirse o sintetizarse, sin embargo, se requieren algunos bits extra que permitan indicar de que bandas se han tomado las señales. La reducción de la tasa de bits es de casi el 50%.

Vocoders por lóbulos: Los llamados *Formant Vocoders* pretenden sintetizar voz al determinar la frecuencia en la que se localizan sus lóbulos más importantes. Las señales correspondientes a las frecuencias de los lóbulos se transmiten y se utilizan para controlar las resonancias de un sintetizador que consiste de 3 o más circuitos resonadores entonados en una frecuencia.

Medición de las frecuencias de lóbulos: La manera más simple de medir la frecuencia de un lóbulo es separando por medio de filtros un lóbulo de otros, y midiendo la tasa de cruces por el eje de la señal resultante de un solo lóbulo.

Un método de medición de frecuencias de lóbulo se basa en los momentos espectrales, tales como el centroide del espectro de la

señal en una banda de un lóbulo de frecuencia. El centroide o frecuencia media puede ser medido subdividiendo la banda total de frecuencia del habla en varias sub-bandas, y midiendo la amplitud de la señal en cada sub-banda, la frecuencia media se puede aproximar por medio de:

$$f_{\text{media}} = \frac{\sum_n f_n / n}{\sum_n 1/n}$$

donde f_n es la frecuencia central de la n -ésima sub-banda.

M. R. Schroeder demostró en 1956 que los momentos espectrales pueden medirse operando la señal en el dominio del tiempo sin ayuda de filtros, por medio de la siguiente fórmula:

$$F_{\text{media}} \cong \frac{1}{2\pi} \left| \frac{d}{dt} s(t) \right| \Big/ |s(t)|$$

donde las barras horizontales indican promedios en el tiempo y las barras verticales denotan valores absolutos

Sin embargo, los métodos de medición de frecuencias de lóbulos mencionados aquí no resuelven el problema más difícil: La

separación de la señal del habla en sub-bandas que contengan un solo lóbulo cada una. Este problema se debe a que los lóbulos se traslapan.

Una manera de resolver este problema es el llamado *analysis by synthesis*, en el cual se genera un espectro artificial en una computadora y mediante iteraciones, se vá ajustando al espectro de la señal de habla real. Los lóbulos generados por la computadora pueden tomarse como los de la señal original.

Un método ingenioso que también evita el problema de traslape de lóbulos fue descrito por C. H. Coker en 1963. En este método, el pico más grande de la envolvente del espectro logarítmico de la señal de habla es localizado, y después un lóbulo centrado en el pico mayor es restado del espectro de habla. El proceso se repite unas 4 o 5 veces tomando el máximo pico de la señal restante. Experimentalmente se demuestra que los tres valores de menor frecuencia de entre los picos obtenidos equivalen a los primeros tres lóbulos de la señal.

Vocoders por ajuste de formas: Este método consiste en seleccionar distintos espectros preestablecidos que sean lo más aproximado posible a los diferentes espectros de ventana tomados durante la señal de habla original. C. P. Smith estimó que con 1024 patrones de espectros equiprobables (10 bits), se obtenía una aceptable aproximación de la señal original.

Los *vocoders* fueron los primeros dispositivos capaces de sintetizar voz, teniendo una eficiente tasa de bits. Sin embargo, el habla generada a través de estas unidades no es de alta calidad. En la actualidad, con los avances tecnológicos que existen, pueden hacerse sintetizadores con una calidad superior a la de los *vocoders*, pero con una tasa de bits más elevada y una complejidad frecuentemente alta. Por estas razones en la actualidad se siguen utilizando los *vocoders* en procesos de sintetización de voz que no requieran una calidad muy elevada.

4.3.2 Codificadores por Forma de Onda.

El objetivo principal en el procesamiento de una señal de habla, es el de obtener una representación de la información que contiene dicha señal. Los codificadores por forma de

onda pretenden generar precisamente la forma de onda de la señal del habla al sintetizarla. Para esto, se debe muestrear y almacenar la señal de habla que después será sintetizada.

Los codificadores por forma de onda (que trabajan en el dominio del tiempo), muestrean periódicamente la forma de onda de las unidades de habla a la tasa de Nyquist, y almacenan las muestras utilizando técnicas digitales tales como PCM.

La mayoría de los codificadores que trabajan en el dominio del tiempo tratan de minimizar la diferencia ("error") entre la señal de habla original y su versión regenerada. Este error se deriva del ruido de cuantificación, debido a la representación del habla analógica usando un número finito de niveles de amplitud. El resultado es una degradación perceptible en el habla reproducida, la cual puede ser reducida mediante el incremento del número de niveles de cuantificación (y consecuentemente la tasa de almacenamiento de bits) o decrementando el rango dinámico de la cuantificación de entrada.

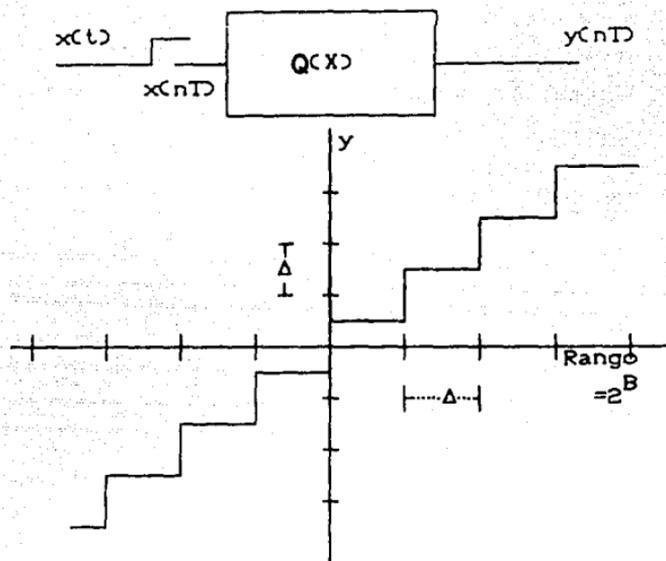


Figura 4-4. Muestreo y Cuantificación de una señal analógica.

PCM: PCM significa *codificación modulada por pulsos*, y es la más simple técnica de codificación de forma de onda.

Al aplicar el teorema de muestreo a una representación digital de habla, nos encontramos con 2 factores principales; la tasa de muestreo y la cuantificación. Estos son mostrados en la figura 4-4. Si el ancho de banda de la señal es de ν hertz, entonces

el período de muestreo debe ser $T < 1/(2W)$ (Nyquist). Puesto que las muestras $x(nT)$ de la señal generalmente se toman de un rango continuo de valores, estos deben ser cuantificados para la transmisión o para el almacenamiento digital. Al representar las muestras como palabras binarias de bits, la tasa de bits es $2BW$ [bits/s] (B =número de bits de cuantificación). El valor de W requerido para la señal de habla, depende del uso que se les dé a las muestras. Se sabe de estudios teóricos que los sonidos de habla como los fricativos tienen grandes anchos de banda (de hasta 10[kHz]). Por otro lado, mucha de la información requerida para que el habla sea inteligible, está contenida en la variación de los primeros tres *formants* (lóbulos de frecuencia), que forman el habla vocal, los que están en frecuencias inferiores a 3[kHz]. Por lo cual, una tasa de muestreo de entre 6 y 20 [kHz] es usada generalmente.

Sin importar que tasa de muestreo se utilice, la señal de habla debe ser filtrada por un filtro paso-bajas antes del proceso de muestreo para eliminar las altas frecuencias no deseadas y parte del ruido (de alta frecuencia).

La selección del número de bits por muestra B, está relacionada con el uso que se le vaya a dar a las muestras. Si el propósito es la transmisión o almacenamiento en sistemas de cómputo seguido por una conversión a una señal analógica, lo único que importa es que la señal analógica sea aceptable para el oído. Sin embargo, el proceso de muestreo descrito también suele ser el primer paso en técnicas de análisis digital de voz.

Como los errores de cuantificación se propagan, si se pretenden representaciones digitales más adecuadas, debe considerarse un amplio margen de error al seleccionar la tasa de muestreo y el número de bits por muestra.

Sin embargo, se puede afirmar que la cantidad de procesamiento requerido para implementar la mayoría de los sistemas es proporcional a la tasa de muestreo. Por lo tanto, se debe tratar de mantener la tasa de transmisión tan baja como sea posible, tanto como los objetivos lo permitan.

Como no se puede bajar la tasa de

muestreo, para reducir la tasa de bits, se tiene que reducir el número de bits por muestra, la clave para tales reducciones radica en la consideración de una de las propiedades básicas de la señal del habla: el amplio rango dinámico que ésta posee. Podemos ver en la figura 4-4, que si B es fijo, entonces el tamaño del intervalo Δ debe ser seleccionada tal que $\Delta * 2^B$ se adecúe al máximo valor pico-a-pico de la señal. Por tanto, el tamaño de la unidad mínima de cuantificación se determina por la amplitud de los segmentos vocales del habla, mientras que una buena representación de segmentos no vocales requeriría de un tamaño mucho menor de dicha unidad.

El PCM lineal simplemente cuantifica las muestras del habla sin tomar ventaja de las redundancias en la señal del habla. Otros métodos usan propiedades estadísticas de la señal de habla para reducir la tasa de bits.

PCM Logarítmico: Dos variaciones de PCM muy utilizadas son PCM μ -law y PCM A -law, éstos se basan en la cuantificación no uniforme de la amplitud de la señal de acuerdo a una escala logarítmica en lugar de una escala

lineal. Estos codificadores toman ventaja de las características dinámicas de la amplitud en el habla y de la percepción logarítmica de amplitud en el oído.

Codificadores Adaptables: La mayoría de las muestras de voz son pequeñas en comparación con los poco frecuentes picos de amplitud. Un histograma de muestras de habla tomadas durante muchos segundos semeja una distribución de probabilidad Gaussiana. La compresión de amplitudes en una escala logarítmica antes de la codificación, aplanar la distribución y reduce el ruido producido por la cuantificación. El habla es estacionaria durante intervalos cortos de tiempo (10-30[ms]), puesto que el tracto vocal se mueve a una velocidad relativamente baja. Por tanto, los codificadores adaptables cambian las características de cuantificación con el tiempo, usando tiempos de muestreo proporcionales a la energía de la señal hablada. El tamaño del período de muestreo es directamente proporcional a la cuantificación del error, esto permite que haya menos ruido durante las secciones con menor variación de la señal de habla, reduciendo así el ruido que habría con un tamaño fijo del período de

muestreo, dando como resultado una mejor calidad de voz.

Codificadores Diferenciales: La mayoría de los sonidos hablados tienen concentración de energía a bajas frecuencias, que causa un alto grado de correlación en el tiempo entre muestras. Los codificadores diferenciales cuantifican la señal después de que la componente predecible ha sido eliminada, lo que permite el uso de menos bits por muestra con el mismo ruido de salida. Por ejemplo, la modulación delta reduce la diferencia entre muestras mediante el muestreo de la señal analógica de habla a una tasa de varias veces la de Nyquist, permitiéndonos el uso de muestreo con un solo bit (2 niveles) de cuantificación. El concepto de diferencial puede ser extendido mediante el estudio de la invariancia del espectro en tiempos cortos.

Para ilustrar el funcionamiento de la modulación delta, mostramos un diagrama (Figura 4-5). Como se vé, el nuevo valor de la muestra sera "1" si la amplitud es mayor que en la anterior muestra, y será "-1" si la amplitud es menor. En la segunda parte de la figura(4-5-b), se muestra la modulación delta

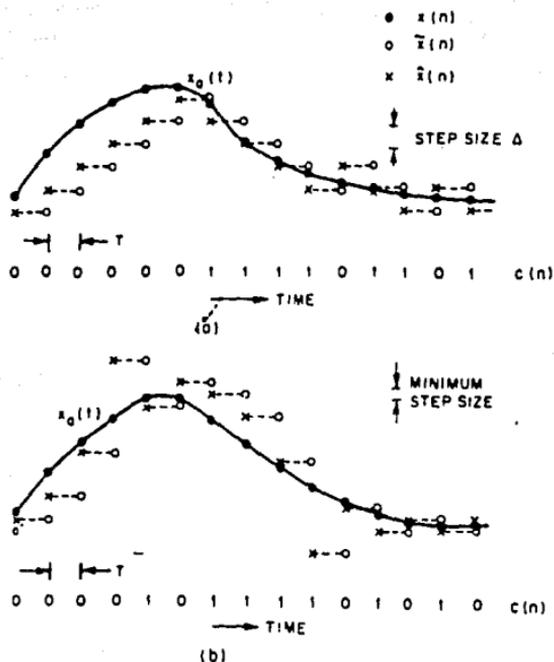


Figura 4-5. Modulación Delta. (a) Tamaño de unidad de cuantificación fija. (b) adaptable.

con el tamaño de unidad de cuantificación adaptable (ADPCM).

APC: En el análisis APC (Adaptive Predictive Coding), la cuantificación codifica la diferencia entre el espectro original y la versión retardada por un período fundamental.

Puesto que la mayor parte del habla es casi periódica durante cortos períodos de tiempo, el error de cuantificación resultante es muy reducido.

En todos estos codificadores adaptables y diferenciales, las estaciones sintetizadoras utilizan un filtro para compensar cualquier filtrado hecho en el análisis antes de la cuantificación. El ruido en el habla de salida usualmente es igual al error de cuantificación. Así, estos codificadores intentan minimizar el error de cuantificación para ganar calidad, a la vez que reducen la tasa de bits (la tasa de muestreo veces el número de bits de cuantificación por muestra) para lograr eficiencia. En algunos casos, un tipo de información extra debe también ser almacenada, elevando la tasa de bits ligeramente, dado que el sintetizador debe conocer un tamaño adaptable de intervalo y/o parámetros espectrales que describan el filtrado de salida. Estos datos extra son usualmente ajustados a una tasa mucho menor que la tasa de muestreo original (por ejemplo, 20-100/s vs. 8000-10000/s).

Codificadores de sub-banda: Este tipo de

codificadores dividen el espectro del habla en 4 a 8 bandas de frecuencias, y codifican cada banda usando PCM adaptable. Después de bajar las tasas de muestreo en cada banda, una tasa de bits menor puede ser obtenida, manteniendo la calidad del habla, por medio de la reducción de bits por muestra en bandas de alta frecuencia que son menos perceptibles y por tanto, menos importantes. Las bandas con poca energía usan intervalos de muestreo pequeños. Inclusive, el ruido de una banda no afecta otras bandas de frecuencia.

Codificación Mozer: Esta técnica toma ventaja de la insensibilidad del oído para detectar la fase así como de la tendencia de la mayoría de la voz a ser casi periódica a intervalos cortos de tiempo, con tal de reducir el espacio requerido de almacenamiento. La codificación Mozer localiza pequeños períodos en el habla de entrada y después codifica estos como sigue: En la transformada discreta de Fourier, la fase del espectro es puesta a cero. De esta forma, se procede a obtener una señal en el dominio del tiempo cuyo espectro de amplitud es el mismo que el de la señal original y cuyo espectro de fase es igual a cero. La

señal resultante es simétrica en el tiempo, lo cual permite descartar la mitad de la forma de onda. Además, la mayoría de las muestras de gran amplitud se concentran en la mitad del período, permitiendo que las secciones de baja amplitud se descarten, siendo este efecto casi imperceptible.

4.3.3 Métodos Paramétricos e Híbridos:

Estos métodos son aquellos que de alguna manera combinan las características de los *vocoders* y de los codificadores por forma de onda.

Vocoder de Alta Fidelidad: En 1957, J. R. Pierce sugirió el uso de "*semivocoders*" para transmitir habla de alta calidad con un ancho de banda de 10[kHz] sobre líneas telefónicas convencionales, cuyo ancho de banda es de 3[kHz]. En un *semivocoder*, una porción del espectro del habla, comúnmente una banda de baja frecuencia; es transmitida sin procesar (por el analizador *vocoder*) mientras que la porción restante del espectro sí se "vocoderifica". El hecho de poder obtener la señal de habla con un ancho de banda de 10[kHz] indica que su calidad es mucho mejor. El primer problema que se presentó fué el de

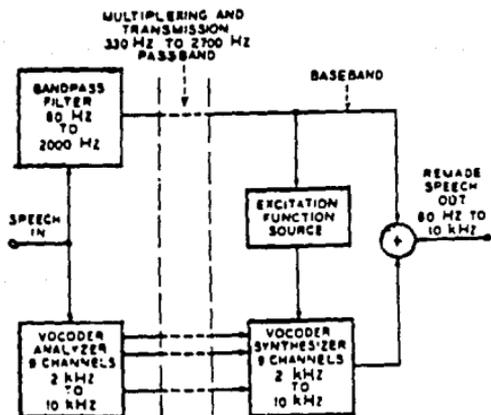


Figura 4-6. Vocóder excitado por voz ideado para transmitir señales de habla de 10 [kHz] sobre líneas telefónicas de 3 [kHz].

la excitación, puesto que el período fundamental de la porción "vocodificada" debía ser igual que el de la señal original de habla.

La solución para el problema de excitación en el vocóder de alta fidelidad provocó el desarrollo de una nueva manera de excitación llamada *excitación por voz*. Un vocóder excitado por voz (VEV) se ilustra en la figura 4-6. La señal de excitación para el sintetizador del vocóder se obtiene de la

"banda-base" no codificada por un proceso llamado *aplanamiento de espectro*. Un aplanador de espectro tiene la función de esparcir por medio de una distorsión no lineal el espectro de la banda-base, después, los componentes de frecuencia son ecualizados para formar una señal de excitación de espectro plano.

Los sistemas digitales que presentan el principio de los VEVs tanto en transmisión como en almacenamiento de señales de habla, utilizan PCM para la codificación de la banda-base, es decir, de la banda no "vocodificada". La tasa de bits requerida para codificar esta porción es baja en relación con sistemas PCM que codifican todo el espectro del habla (La frecuencia máxima a codificar vendría siendo de solo 2[kHz]).

Con los esquemas que operan con el principio antes mencionado se ha realmente obtenido una mejor calidad, se ha perdido el "sumbido" que presentaban los *vocoders* originales. El período fundamental y los intervalos vocalizados y no vocalizados de la señal de habla original se reproducen con una buena fidelidad. Con este método se puede

identificar de quién era la voz original antes de ser procesada, cosa que era difícil con los *vocoders* originales.

LPC: (*Linear Predictive Coding*). La codificación lineal predecible surgió del hecho de que en el habla vocalizada cada muestra está altamente correlacionada con la muestra correspondiente que ocurre un período fundamental antes. Además, cada muestra está también correlacionada con las muestras inmediatamente anteriores a ella; se ha demostrado que para una tasa de muestreo de 8[kHz], bloques de 80 muestras presentan una apreciable correlación entre ellas.

Basándonos primero en las correlaciones de corto tiempo, se pretende aproximar una muestra por una combinación lineal de las muestras s_{n-k} inmediatamente precedentes a ella. Manfred R. Schroeder demostró basándose en conservación de las cuatro primeras resonancias del espectro del habla, que con 12 muestras precedentes se obtiene una buena aproximación de la muestra predicha con la real.

La representación matemática de esta

predicción es:

$$s_n = -a_1 s_{n-1} - a_2 s_{n-2} - \dots - a_p s_{n-p} + e_n \quad (1)$$

donde p es el orden de la predicción y e_n es el residuo de la predicción, o sea la parte de s_n que no puede ser representada por la combinación lineal de las p muestras previas. Los coeficientes de la ecuación (1) son llamados *coeficientes de predicción*. El signo negativo se pone por conveniencia notacional.

Ahora el problema está en la manera de obtener los coeficientes. El método más simple es el de minimizar el cuadrado del residuo de predicción promediado en N muestras:

$$E = \sum_{n=1}^N e_n^2 = \sum_{n=1}^N \left(\sum_{k=0}^p \alpha_k s_{n-k} \right)^2, \quad \alpha_0=1 \quad (2)$$

Un valor típico para N es 80, que corresponde a una ventana de tiempo de 10[ms]. Valores mayores de N implicarían sumatorias sobre intervalos de tiempo mayores, que interferirían con las transiciones perceptualmente importantes entre sonidos sucesivos de habla. Si N fuera mucho menor,

el análisis sería más propenso al ruido.

La minimización de E con respecto a a_k requiere de la igualación a cero de las derivadas parciales $\delta E / \delta a_m$:

$$\frac{\delta E}{\delta a_m} = \sum_{n=1}^N 2s_{n-m} \sum_{k=0}^P a_k s_{n-k} = 0 \quad (3)$$

Invirtiendo el orden de las 2 sumatorias tenemos:

$$\sum_{k=0}^P r_{mk} a_k = 0 \quad (4)$$

donde se han introducido los coeficientes de correlación:

$$r_{mk} = \sum_{n=1}^N (s_{n-m})(s_{n-k}) \quad (5)$$

Una vez determinados los coeficientes de correlación r_{km} de la señal de habla, puede utilizarse la ecuación (4) para calcular los coeficientes de predicción a_k ($\omega=1$).

Tenemos:

$$\sum_{k=1}^P r_{mk} a_k = -r_{m0} \quad (6)$$

O en notación matricial:

$$R a = -r_0 \quad (7)$$

donde R es la matriz ($P \times P$) cuyos coeficientes son r_{mk} y a y r_0 son los vectores columna

correspondientes a α_k y r_{m0} en la ecuación (8) respectivamente. Despejando de la ecuación (7) obtenemos finalmente los coeficientes de predicción:

$$\alpha = -R^{-1}r_0$$

La figura 4-7 muestra una señal de habla (el trazo superior) y el residuo de predicción (o señal residual) después del proceso (es el trazo central, y está amplificado 10[dB]). Como se esperaba, la correlación a corto tiempo ha sido casi completamente eliminada. Pero la correlación a largo tiempo, entre periodos fundamentales, aún persiste, tal como se observa en los grandes picos entre intervalos de período fundamental. Para eliminar éste tipo de correlación, se tiene que encontrar para qué período fundamental existe un máximo en la señal residual y cuánto vale. Después, se resta el residuo de predicción atrasado de sí mismo. El resultado es el trazo inferior de la figura 4-7, que muestra la nueva señal amplificada por otros 10[dB]. La señal residual resultante, después de que se le aplicó predicción de corto tiempo (de *lóbulo*) y de largo tiempo (de *período*

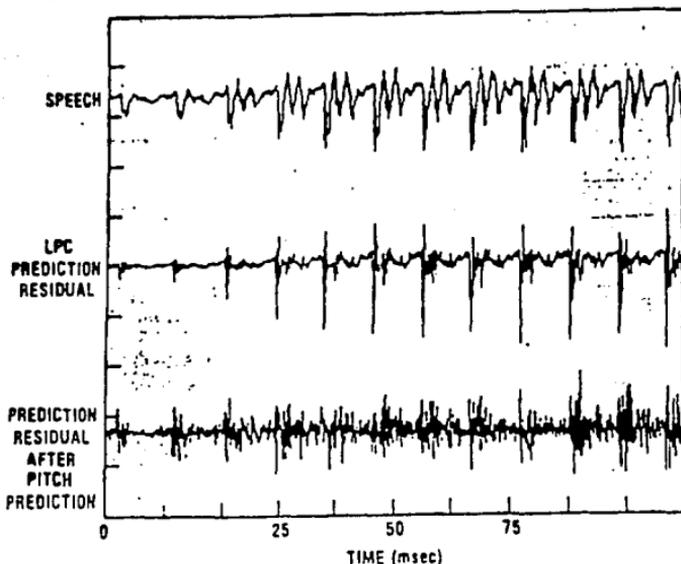


Fig. 4-7. Comparación de la señal de habla con las señales de error de predicción.

fundamental) es impredecible y de poca potencia (unos 20[dB] inferior a la señal original)

Una vez que se ha obtenido el residuo de predicción en, es posible obtener la señal original a partir de este residuo; para ello, se representa la ecuación (1) en el dominio Z de la siguiente manera:

$$s(z)A(z) = e(z) \quad (1a)$$

donde:

$$A(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p} \quad (10)$$

Ahora bien, la señal de habla puede ser obtenida así:

$$s(z) = e(z) \frac{1}{A(z)} \quad (11)$$

en donde se dice que a $e(z)$ se le aplica un *filtrado* $1/A(z)$ para obtener $s(z)$.

La ecuación (10) es un polinomio que, de acuerdo al teorema fundamental del álgebra, puede representarse por un producto:

$$A(z) = (1-z_1/z)(1-z_2/z)\dots \quad (12)$$

es decir, $A(z)$ es un polinomio que se compone de p ceros y ningún polo, por tanto, $1/A(z)$ tiene p polos y ningún cero. Por esta razón, a $1/A(z)$ se le llama *filtro de puros polos*, de esta manera, se ha logrado representar a una señal de habla por medio de polos.

A continuación se indica cómo puede representarse en sí el habla utilizando las ventajas que presenta la codificación lineal predecible. Puede observarse en la figura 4-7 que la señal residual (parte media) puede ser aproximada por pulsos periódicos sumados a una componente semejante al ruido blanco.

Originalmente se trató de codificar muestras de habla introduciendo a un filtro $1/A(z)$ una señal de excitación compuesta de ruido blanco y de una frecuencia fundamental en caso de que la señal sea vocalizada, es decir, basados en el *vocoder* concebido por Homer Dudley (utilizando el filtro $1/A(z)$ en lugar del banco de canales de frecuencia). Por ejemplo, para transmisión (análogo al almacenamiento), en la señal original se tiene un detector que identifica si la señal es vocalizada o no vocalizada, y un codificador de frecuencia fundamental (en caso de ser vocalizada), además, se envían códigos que representen a los coeficientes de predicción cada período equivalente a una ventana de tiempo. Utilizando este esquema, se ha podido sintetizar voz con una calidad superior a la de un *vocoder* y con una tasa de bits inferior a la de PCM.

Durante varios años, LPC mantuvo el esquema mencionado anteriormente, sin que se tuvieran grandes avances en este campo. Fue hasta más recientemente que surgió un método que logró obtener mayores beneficios de LPC en el procesamiento de voz. Puede verse en la fórmula (11) que si a un "filtro" $1/A(z)$ se

le introduce la señal residual, se obtiene la señal original, se ideó entonces codificar la propia señal residual para que ésta sirviera como excitación en la parte de sintetización. Con esto se mejora la calidad en una gran medida. Se ha logrado, cuantificando cada muestra de señal residual con un solo bit, una calidad mejor a la de PCM de 5 bits por muestra. Es obvio que al utilizar más bits de cuantificación en la señal residual, se obtiene una calidad mejor. De esta manera, se ha obtenido un método de sintetización de voz de alta calidad a una tasa de bits muy baja.

4.4 SINTETIZACION DEL HABLA ALMACENADA.

4.4.1 Entonación.

Una importante dificultad con el habla sintetizada es la de determinar una entonación natural correspondiente a un texto o un código de entrada. Cuando las unidades almacenadas son grandes, como palabras y frases, la frecuencia fundamental (F_0) y la intensidad son almacenadas directamente junto con los parámetros espectrales en sintetizadores LPC o en *vocoders*, o indirectamente en las formas de onda de los sintetizadores por forma de onda. Cuando se

concatenan unidades del tamaño de una palabra o menores, se produce un habla no natural a menos que la entonación guardada para esa unidad sea modificada.

Es inútil almacenar la entonación para unidades más pequeñas que una palabra, puesto que el comportamiento de la entonación está determinado principalmente en base a frases, habiendo mayores cambios en las sílabas acentuadas de las palabras enfatizadas, con una entonación subiendo y bajando, marcando los principios y los finales de frases, respectivamente. Incluso al nivel de palabra, ocurren entonaciones substancialmente diferentes, dependiendo de la posición de la palabra en la oración. Así, los sintetizadores de vocabulario limitado deberán grabar cada palabra o frase en su posición apropiada dentro de la oración. La intensidad es usualmente almacenada como un parámetro al usar unidades tan pequeñas como difonemas, puesto que las principales variaciones en la intensidad son dependientes del fonema (Hay, sin embargo, tendencias de las sílabas acentuadas a ser más intensas y de las sílabas finales de oración a ser menos intensas).

La duración es usualmente codificada implícitamente en los datos del habla. En los sintetizadores paramétricos, los parámetros son evaluados una vez cada intervalo de 10 a 30[ms]. Los sistemas de tasas variables de intervalos toman ventaja de la tendencia de muchos sonidos del habla a mantenerse en un estado estable durante varios intervalos, ajustando el tamaño del intervalo usando alguna medida de la continuidad espectral entre intervalos consecutivos, o guardando un bit de repetición única para cualquier intervalo que sea virtualmente idéntico al precedente. Sin embargo, la duración debe ser determinada explícitamente para sintetizadores de difonemas o de fonemas, puesto que las unidades por sí mismas no tienen duración implícita.

4.4.2 Modo de Operación de los Sintetizadores.

Los sintetizadores de voz comerciales pueden producir la voz ya sea por medio de códigos que el usuario utiliza para la concatenación de las unidades fonéticas deseadas; o bien por medio de texto que se introduce utilizando un teclado. Los sintetizadores del último tipo requieren de algoritmos sofisticados para que se logre una

pronunciación adecuada (por ejemplo en el idioma Inglés).

Como existen miles de sílabas diferentes (estudios indican que en idioma inglés hay más de 20,000), cada una con una duración media de 250[ms], guardar todas las sílabas requeriría de unos 50[Mbits] para la técnica más sofisticada de forma de onda o unos 10 [Mbits] utilizando métodos paramétricos. Debido a estos requerimientos de mucha memoria, el método de forma de onda ha sólo sido usado con vocabularios pequeños de hasta unos pocos minutos de voz. Los avances rápidos en VLSI han incrementado el vocabulario de los sintetizadores de forma de onda, pero para flexibilidad completa, los métodos paramétricos son necesarios.

El uso de métodos paramétricos normalmente degrada la calidad del habla sintetizada. Los sintetizadores paramétricos utilizan unidades de habla menores como difonemas, puesto que la degradación de calidad en éstos no es lo suficientemente grande como para justificar la mayor cantidad de memoria necesaria en unidades de tamaño silábico. Se ha estudiado que en el idioma

inglés, hay 1400 difonemas, lo que equivale a un almacenamiento de unos 200[kbits]. Los sintetizadores de fonemas son más populares, estos normalmente almacenan parámetros de amplitud y espectro para cada uno de los casi 40 fonemas (en el idioma inglés). Estos datos pueden ser guardados en únicamente 2[kbits].

Los sintetizadores llamados de *allophone* son modificaciones simples que utilizan hasta 230 versiones de fonemas, que varían de acuerdo al contexto fonético (por ejemplo, el fonema K podría ser representado por 2 *allophones*, uno que corresponda a la pronunciación anterior a una vocal frontal como I, y la otra anterior a una vocal trasera como U). Este tipo de sintetizadores pretenden resolver un problema importante de los sintetizadores de fonemas, que es cómo concatenar unidades cortas sin producir un habla con brincoteos. Todos los sistemas paramétricos que utilizan unidades más pequeñas que una sílaba, suavizan los cambios fonéticos en las fronteras. Puesto que las fronteras de difonemas o de demisílabas unen segmentos iguales, sus reglas de suavización son simples. Los sistemas de fonemas o de *allophones* usan por el contrario reglas de

suavización que simulan la coarticulación del tracto vocal.

No es sencillo establecer un conjunto de reglas que describan de una manera precisa como son modificados los parámetros espectrales de cada fonema por sus vecinos. Los sintetizadores de difonemas tratan de resolver este problema almacenando las transiciones de parámetros de un fonema al siguiente, puesto que los efectos primarios de la coarticulación influyen solo a los fonemas inmediatamente adyacentes. Desafortunadamente, hay muchos casos en donde la coarticulación se extiende a varios fonemas, requiriéndose así el almacenamiento de multifonemas de mayor duración. Un ejemplo podría quedar en la palabra *construir*, en donde la redondez de los labios necesaria para la vocal U causa redondeamiento anticipado durante las anteriores 3 consonantes: S, T y R. Muchos efectos de coarticulación pueden ser aproximados por reglas simples, como bajar todas las frecuencias resonantes con el redondeo de los labios; otros, como predisparo de posiciones fonéticas (siempre a excepción de en el habla lenta) son mucho más difíciles de modelar

adecuadamente.

En los sintetizadores que operan en base a texto de entrada, se requiere de un procesador lingüístico para convertir el texto en una forma adecuada para el acceso a unidades de voz almacenadas. Esto involucra la traducción de las oraciones de entrada a una secuencia de códigos para capturar las unidades almacenadas apropiadamente, y la determinación de parámetros de entonación del texto así como la variación de la F_0 y de la duración. Lo primero suele manipularse por medio de un conjunto de reglas que dependen del lenguaje para convertir una serie de letras en una secuencia de fonemas. Estas reglas examinan el contexto de cada letra, para determinar como es ésta pronunciada (En el idioma español esto es relativamente fácil, mas no para otros idiomas). Por ejemplo la letra C es pronunciada como K antes de una A, O o U, pero como S antes de una E o una I. El idioma más universal, el inglés, presenta a este respecto grandes dificultades, puesto que a pesar de ser éste un idioma germánico, tiene mucha influencia romance (tiene una doble ascendencia); y por tanto, necesita cientos de reglas para poder

convertir texto a fonemas a un nivel del 90%. Aquí, por ejemplo, muchas palabras muy utilizadas como *of* o *has*, violan reglas muy básicas de pronunciación, por lo que se requieren listas de excepciones para ser examinadas antes de que las reglas básicas sean aplicadas. Los idiomas en los que la escritura vá más de acuerdo a los fonemas, como en el nuestro (español), tienen muy pocas reglas, pues como sabemos, cada letra tiene normalmente una pronunciación.

Las reglas de pronunciación utilizadas en el sintetizador prototipo *MSTala* (Que lleva este nombre por haberse desarrollado en la universidad MIT) son precedidas por un algoritmo de descomposición de palabras, que desliga prefijos y sufijos de cada palabra. Como son pocos prefijos y sufijos, y como estos pueden afectar a la pronunciación (Recordemos que estos estudios se hacen básicamente para el idioma inglés), esto incrementa la potencia del sistema a cambio de un poco más de cómputo. El *MSTala* tiene también un diccionario de morfemas (Las unidades básicas de léxico que constituyen palabras), que contienen no solo pronunciasiones de fonemas, sino también contenido de

sintaxis del habla. Esta información es muy útil en la asignación de entonación puesto que con el uso de esto se puede saber cuando ocurren frecuencias fundamentales cruciales y efectos de duración para una entonación natural. La mayoría de los sintetizadores en realidad se basan en reglas simples pero fundamentales de entonación, o dejan libre F_0 para el usuario, dándole la posibilidad a éste de cambiarla directamente (por ejemplo, con símbolos especiales añadidos al texto de entrada). La razón principal por la que la mayoría de los sintetizadores por texto de entrada sintetizan sonidos no naturales es la pobre manipulación de la entonación.

Una forma muy utilizada en la sintetización del habla, consiste en codificar el habla del usuario con un método eficiente de almacenamiento, tal como LPC, y reproducir la voz con concatenación simple (como por ejemplo el sistema desarrollado por la Texas Instruments *Speak and Spell*). Estos sistemas son por tanto limitados en la combinación de las unidades originales del habla, usualmente con sus entonaciones originales. Esto es adecuado para muchas aplicaciones como

juguetes parlantes, sistemas de seguridad en máquinas, y asistencia automática telefónica. Sin embargo, un usuario tiene que aceptar una voz y un vocabulario estándar o enviar un audiocaset a la compañía que fabrica los sintetizadores para establecer su propio vocabulario. Una razón por la que se tiene que enviar el caset a la compañía productora se debe a que mientras ciertas partes del proceso de análisis es automática (como el algoritmo básico LPC), grandes ahorros de memoria pueden hacerse únicamente con ayuda humana, escuchando secciones de habla que pueden ser codificadas más eficientemente que otras secciones sin degradación perceptible, y modificando parámetros del habla en lugares donde el procesamiento automático imperfecto produce una voz sintética pobre.

4.4.3 Tecnologías de Hardware para la Sintetización del Habla.

Existen en la actualidad varios circuitos integrados hechos específicamente para aplicaciones de sintetización de voz que pueden ser adquiridos por usuarios para aplicarlos a diseños particulares. Existen por ejemplo sistemas completos de transmisión de voz de una manera digital para telefonía,

en base a PCM μ -law con conversiones A/D y D/A y filtros incluidos. Se pueden adquirir, también para telefonía, sistemas del tipo ADM (*Adaptive Delta Modulation*) que funcionan a 32(kbits/s). En el área de los *Vocoders*, existen chips que realizan la función descrita en la figura 4-1.

En los últimos años, la tecnología VLSI ha tenido un gran impacto en la sintetización de voz. Se han desarrollado microprocesadores de alta velocidad y circuitos integrados especializados en procesamiento de señales. La figura 4-8 muestra el diagrama de bloques de un circuito especializado en procesamiento digital de señales. Este contiene memoria para almacenar programas y una memoria para almacenar datos, un calculador de dirección, unidades seriales de entrada/salida, y una unidad aritmética. Una frecuencia de funcionamiento de varios megahertz y un alto grado de paralelismo hacen de estos chips poderosas herramientas en la sintetización de voz. Algoritmos tales como ADPCM y de Codificaciones de Sub-Banda pueden fácilmente ser desarrollados de esta manera utilizando un DSP (*Digital Signal Processor*) como codificador, uno más como decodificador

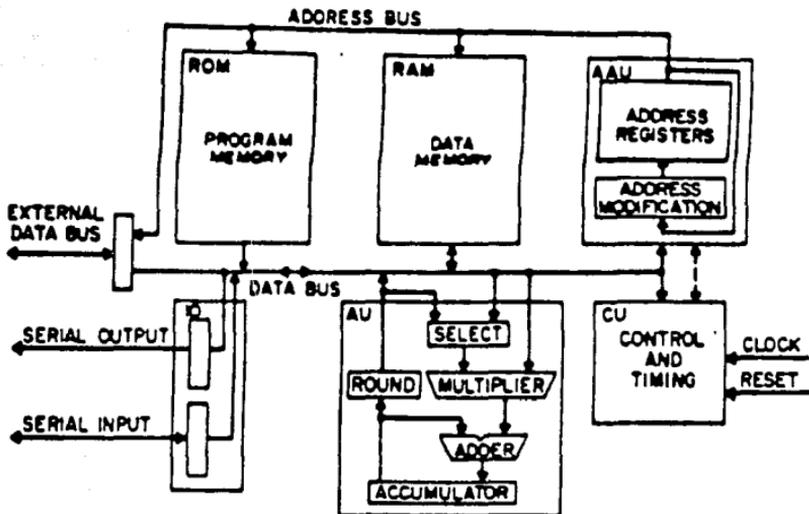


Figura 4-8. Diagrama de Bloques de un DSP.

y un chip μ -law PCM para la conversión A/D.

Para algoritmos de mayor complejidad, la utilización de un microprocesador no es suficiente. Para llevar a cabo la realización de sistemas de algoritmos sofisticados se utilizan arreglos de procesadores que funcionan a tiempo real.

Se deben tomar en cuenta 3 consideraciones en el diseño de codificadores-sintetizadores de voz:

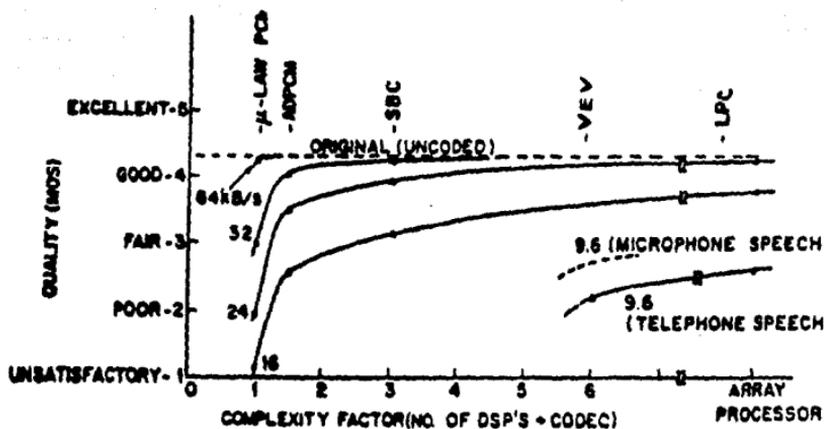


Figura 4-9. Calidad vs. Tasa de Bits y Complejidad para algunos algoritmos.

1) Calidad del habla sintética, 2) Costo o Complejidad, y 3) Tasa de bits requerida. R. E. Crochiere, R. B. Cox y J. D. Johnson han propuesto una gráfica que dá una relación aproximada de los 3 parámetros anteriores para algunas técnicas de sintetización. Esto se muestra en la figura 4-9.

5. DISEÑO DEL RELOJ PARLANTE

5.1 CARACTERISTICAS DEL RELOJ.

Presentamos a continuación algunas de las características básicas del Reloj Parlante que hemos diseñado, es decir, damos un panorama general del mismo con el fin de que el lector se familiarize con éste.

Nuestro sistema es un reloj, de tamaño medio (como de escritorio) alimentado con corriente directa, es básicamente digital, y cuenta con un desplegado en el que se da la hora en el formato clásico de 12 horas, este es del tipo 7 segmentos, que cuenta con 4 dígitos, los que corresponden a: decenas de hora, unidades de hora, decenas de minuto, y unidades de minuto. La base de tiempo de nuestro sistema es un cristal de cuarzo, el cual nos permite tener una frecuencia muy estable.

La característica principal de nuestro sistema consiste en la capacidad de poder indicar la hora de una manera parlante. La hora será hablada en el momento en que el usuario lo desee, al momento de oprimir una tecla. La hora dicha vocalmente será precisamente la hora que aparezca en el

desplegado de dígitos.

La hora y los minutos pueden ser ajustados por el usuario para poner a tiempo el reloj, esto se hace por medio de teclas.

El consumo de potencia es de 3.4[w]. Dando un punto de referencia, si comparamos este consumo de potencia con el consumo mínimo para un foco de iluminación doméstico, vemos que este último gasta unas 8 veces más potencia que nuestro sistema.

El reloj está diseñado a partir de circuitos digitales, contando con un sistema de control hecho totalmente en base a la utilización de lógica secuencial, partiendo de una carta ASM.

5.2 EL SISTEMA.

En primer lugar, antes de decidir el tipo de lógica a utilizar en nuestro sistema, hubo que tomar en cuenta ciertos factores importantes en cuanto a la manera general de llevar a cabo el sistema. Para empezar, antes de poder pensar en hacer hablar a un reloj, teníamos que idear la forma en la cual

pudiéramos llevar la cuenta del tiempo, es decir, el reloj en sí. Para esto, tuvimos que tomar en cuenta que el reloj fuera lo suficientemente preciso como para que éste nos diera la hora correcta con un índice de error bajo, a lo más tan bajo como un reloj doméstico común y corriente.

Otra característica que requeríamos de nuestra base de tiempo era la posibilidad de adquirir información de una manera sencilla y clara, con el fin de evitarnos el uso de transductores para pasar de un tipo de energía a otra, esto nos permitiría procesar la información de una manera directa, sin necesidad de introducir parámetros de escala en las entradas y salidas de los transductores. Por esta razón, decidimos la utilización de una circuitería digital que nos permitiera tener las características antes mencionadas.

Por la compatibilidad que presenta con los circuitos digitales y por la estabilidad que es capaz de mantener, decidimos partir de un cristal de cuarzo como la base de tiempo de nuestro sistema; pues como ya antes mencionamos, este tipo de cristales tienen la

característica de oscilar con una regularidad notable en su frecuencia.

Una vez tomada esta decisión, ideamos originalmente partir de la frecuencia libre de oscilación del cristal de cuarzo para la cuantificación del tiempo, dividiendo esta frecuencia hasta llegar a obtener las pulsaciones periódicas necesarias para obtener minutos y horas. Cabe mencionar que como el fin de este proyecto es el proporcionar la hora de una manera parlante, sería superfluo el indicar los segundos, pues estos varían tan rápido que para cuando se hubiera terminado de decir la hora, ya habrían transcurrido un cierto número de éstos.

Gracias al avance en la especialización de circuitos integrados, pudimos encontrar un par de circuitos capaces de cuantificar las vibraciones de un cristal de cuarzo y proporcionar, utilizando división de frecuencia, horas y minutos, de una manera digital. Los circuitos integrados que utilizamos fueron el MM5311 y el MM5369 de National, cuyas características son mostradas en los apéndices B y C.

Partiendo de esta información binaria, el objetivo era la construcción de un circuito electrónico capaz de tomar estos datos y emitirlos de una manera parlante. En este punto tuvimos que hacer algunos análisis para poder llevar a cabo el diseño de dicho circuito.

Lo primero que teníamos que resolver era la forma en que conseguiríamos la emisión de voz, en lo que inicialmente se pensó fue en el muestreo de señales de voz y el almacenamiento de dichas muestras, con el fin de poder emitir las posteriormente. El problema que aquí se tiene es el requerimiento de grandes cantidades de memoria: Si partimos del hecho de que para obtener una voz inteligible es necesario considerar un ancho de banda de unos 3[kHz], requeriríamos muestrear la voz a una velocidad mínima de 6[kHz], según el criterio de Nyquist. Así, si cuantificamos cada muestra con un byte, requeriríamos de 6 kilobytes por cada segundo de palabra almacenada. Suponiendo que la lógica digital de nuestro sistema pudiese utilizar la más eficiente concatenación de palabras, requeriríamos almacenar las siguientes palabras:

La	Nueve	Treinta
Las	Diez	Cuarenta
Una	Once	Cincuenta
Dos	Doce	En
Tres	Trece	Punto
Cuatro	Catorce	
Cinco	Quince	
Seis	Y	
Siete	Veinte	
Ocho	Veinti	

Tabla 5-1. El mínimo de palabras requeridas.

Hicimos un cálculo experimental, en el cual obtuvimos que se requieren unos 30 segundos de almacenamiento para las palabras antes mencionadas. De esta manera, para poder almacenar el mínimo de palabras requeridas, y con una calidad de voz pobre (Gran parte de los sonidos fricativos sobrepasan las frecuencias de 3[kHz], inclusive llegando a alcanzar los 10[kHz], por tal motivo, estos sonidos quedarían muy distorsionados), requeriríamos de al menos 180[kBytes] de memoria, lo que realmente es demasiado para nuestros fines.

Un método más adecuado es la sintetización de voz utilizando unidades fonéticas. Tomando en cuenta la existencia de este tipo de circuitos en el mercado, optamos por la utilización de uno de estos circuitos integrados de sintetización de voz; el SP0256 que utiliza las unidades fonéticas llamadas *allophones*, que es fabricado por la Radio-Shack, y cuyas características se muestran en un apéndice.

5.3 FUNCIONAMIENTO DEL SP0256.

Básicamente, el funcionamiento del SP0256 es como sigue:

La técnica de sintetización de habla por medio de *allophones* proporciona al usuario la capacidad de sintetizar un vocabulario ilimitado a una tasa de bits muy baja. Cincuenta y nueve sonidos discretos de habla (los llamados *allophones*) y cinco pausas son almacenadas en diferentes direcciones de una ROM interna del SP0256. Cada sonido de habla fue tomado de una palabra y analizado utilizando codificación lineal predecible (LPC). Cualquier palabra o frase puede ser creada al direccionar apropiadamente una

combinación de *allophones* y pausas. Como hay un total de 64 localidades de direccionamiento, cada una requiere 6 bits de dirección. Suponiendo que el habla contiene unos 10 sonidos por segundo, la sintetización mediante *allophones* requiere un direccionamiento de menos de 100 bits por segundo.

5.3.1 Cómo utilizar el conjunto de Allophones.

El conjunto de *allophones* (Ver tabla 5 del apéndice A) contiene 2 o 3 versiones de algunos fonemas. Puede ser necesario usar un *allophone* de un fonema particular para la posición final de una palabra u oración. Por ejemplo, DD2 suena bien en una posición inicial y DD1 suena bien en una posición final, como en *doble* y *salud*. Una de las diferencias entre las versiones inicial y final de una consonante es que la versión inicial puede ser más larga que la versión final. Por ejemplo, se puede usar para producir una S inicial 2 SSs en lugar de la SS sencilla de final de palabra, como en *solo*. La posibilidad de unir dos *allophones* iguales sin que se escuche una pausa intermedia es solo permitida en 3 consonantes: FF, SS y TH; además puede

hacerse en las vocales cortas, que son: IH, EH, AE, UH, AO, AX y AA. Para algunos sonidos (PP, BB, TT, DD, KK, GG, CH y JH), una pequeña pausa es requerida. Por esto y por las pausas inherentes del lenguaje, hay cierto número de pausas incluídas en el conjunto de *allophones*, las que varían de 10 a 200[ms]. Este conjunto incluye además algunos diptongos que permiten una mejor calidad del habla producida.

Para que el SPO256 pueda producir un *allophone*, se tiene que poner en el canal de direcciones de éste, el código del *allophone* deseado, y posteriormente enviarle una señal para que las direcciones sean cargadas. El SPO256 produce entonces una onda discretizada correspondiente a dicho sonido, después de lo cual habilita una señal que indica que el SPO256 ha terminado su ejecución, y está listo para recibir el código de un nuevo *allophone*. Debido a que la señal de habla emitida por el SPO256 está constituida por una serie de pulsos ponderados, es necesario obtener la envolvente de dicha señal para que ésta sea audible. Esto se hace haciendo pasar la señal a través de un filtro paso-bajas, con una frecuencia de corte de 5kHz.

Información complementaria sobre el SP0256 se incluye en las hojas de datos de éste, mostradas en el apéndice A.

Una vez conocidas las características del sintetizador de voz, comenzamos a plantear las posibles maneras de establecer el funcionamiento de nuestro proyecto.

5.4 CONTROL DEL SISTEMA.

El funcionamiento general del módulo de control de nuestro sistema consiste en tomar los datos digitales (horas y minutos) del MM5311 y de acuerdo con éstos, generar los códigos correspondientes a los *allophones* del SP0256, concatenándolos de tal manera que al ser emitidos por el sintetizador de voz, se obtengan de manera hablada las horas y los minutos. Este módulo debe también abarcar las señales de protocolo para establecer la comunicación con el SP0256.

Debido a que el sistema funcionará siempre bajo un algoritmo preestablecido, éste puede ser implementado en base a una lógica secuencial independiente.

El siguiente paso que se dió fué el de establecer la relación entre las señales que se iban a recibir y las decisiones que tenía que tomar el controlador. Para darle solución a esto, empezamos a elaborar un posible diagrama de flujo. En principio, se pensó en la posibilidad de que el diagrama de flujo incluyera el envío de códigos de *allophones* al SP0256, de acuerdo a los datos provenientes del MM5311. Sin embargo, esto dió como resultado un complicado diagrama de flujo, a pesar de que simplemente tendríamos que almacenar las 25 palabras mostradas en la tabla 5.1 en memoria ROM. Esta manera de desarrollar el diagrama de flujo implicaba tener complejos algoritmos que relacionaran las 25 palabras de forma tal que pudiéramos conseguir la emisión de las 720 distintas combinaciones existentes. (80 minutos para cada una de las 12 horas).

Otra posibilidad de desarrollar el diagrama de flujo era que éste considerara únicamente las señales de protocolo, que harían que las horas y los minutos provenientes del MM5311 direccionaran directamente un área de localidades en una ROM, en donde estarían almacenadas cada una

de las 720 combinaciones (como ejemplo, una combinación podría ser: *Las tres cincuenta y siete*). La parte más significativa de los bits de dirección de la ROM estaría dada directamente por los datos binarios provenientes del MM5311. Si asignamos 64 localidades de memoria para cada combinación (este sería el número mínimo, debido a que para algunas combinaciones, se requerirían más de 2^5 fonemas) requeriríamos de 76,080 bytes de memoria ROM, lo que hace que este esquema sea también eliminado.

Finalmente, tomando como base los esquemas anteriores, decidimos elaborar un diagrama de flujo intermedio entre ambos extremos, tomando ventajas de ambos esquemas, y procurando eliminar las desventajas que ellos presentaban.

Fué así como llegamos a la creación de un diagrama de flujo que nos permitiría ahorrar tanto localidades de almacenamiento en memoria de los códigos de los *allophones*, como complejidad en la lógica de control. Llegamos con esto a un algoritmo óptimo. Este consiste de lo siguiente: Se utilizan 72 diferentes áreas en memoria, éstas contienen

los codigos respectivos para 12 horas y 60 minutos.

Antes de profundizar en el desarrollo de este diagrama de flujo, presentamos la organizacion general de nuestro sistema, esta se muestra en la figura 5-1.

Como puede verse en la figura 5-1, el MM5311 proporciona los datos de: Decenas de hora, unidades de hora, decenas de minuto, unidades de minuto, decenas de segundo y unidades de segundo. Para nuestros propositos, no requerimos de los 2 últimos parametros. Los datos los dá en forma multiplexada, por lo cual, requerimos de circuiteria externa que nos permita capturar y almacenar dichos datos en flip-flops. Cada uno de estos datos es presentado en BCD. El sistema debe tener la capacidad de escoger entre la toma de las horas y la toma de los minutos. Como vemos, entre las decenas y las unidades ya sea de minutos o de horas se tienen un total de 8 bits. Estos 8 bits serán los encargados de direccionar cualquiera de las áreas de la memoria ROM, esto lo hacemos conectando cualquiera de estos 2 grupos de 8 bits a los 8 bits más significativos del bus

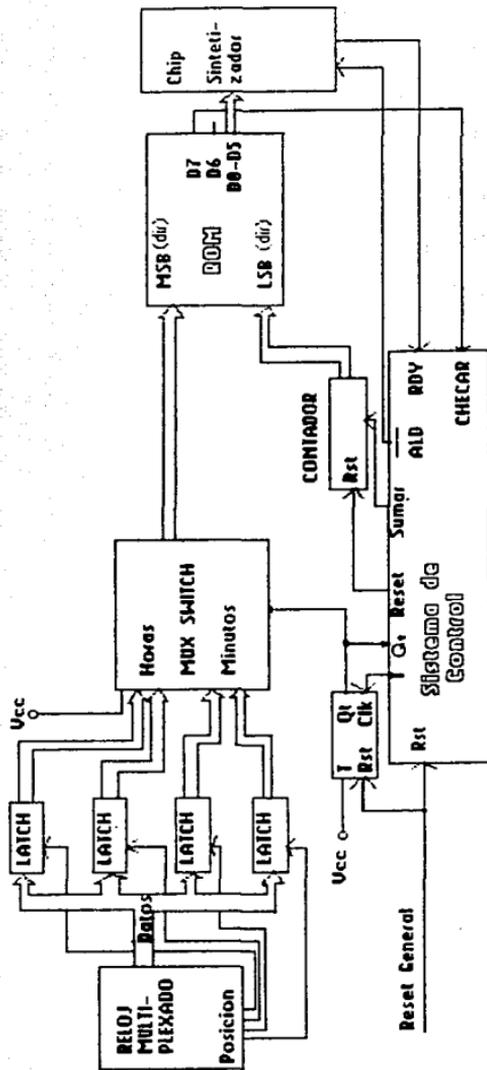


Figura 5-1. Organización General del Sistema.

de direcciones de la memoria. Para seleccionar uno de los 2 grupos de 8 bits, necesitamos una conmutación para poder conectar cualquiera de ambos grupos, para ello, utilizamos multiplexores digitales. Debe existir una señal del sistema de control capaz de seleccionar entre uno o el otro grupo. Ahora, ¿Dónde quedará guardado cada grupo de palabras? Bueno, supongamos que el MUX de conmutación está seleccionando a los minutos, y que tenemos en los 4 bits de decenas un 3 y en los 4 bits de unidades un 7, (el minuto 37). De esta manera, los 8 bits más significativos del bus de direcciones de la ROM serán: 00110111, en esta área, están almacenados los códigos correspondientes a los *allophones* que corresponden a decir: *treinta y siete*, lo mismo hacemos para cada uno de los demás minutos. Ahora, supongamos que estamos en el minuto 1, entonces recibimos 00000001, por lo que éstos serán los 8 bits más significativos del bus de direcciones de la ROM. En esta área está codificada la palabra *Uno*. Pero ¿Qué sucedería si el MUX estuviera seleccionando las horas?. Aquí se tendría que decir *La Una*, y entonces se presenta un problema, pues en la misma área de memoria se tendrían que

guardar 2 grupos de códigos diferentes.. Para solucionar esto, sabiendo que ni las decenas de hora ni las de minuto ocupan el bit más significativo, decidimos fijar a un valor de 1 lógico el bit más significativo de las decenas de hora, permitiéndonos diferenciar la dirección de almacenamiento de horas y la de minutos.

Además de la señal de selección de horas o minutos, nuestro sistema controlador requiere de una señal que permita ajustar la parte menos significativa del bus de direcciones de la ROM, para poder empezar del primer *allophone* de la secuencia de voz requerida (Puesto que en la memoria, en cada localidad almacenamos el código correspondiente a un *allophone*). El controlador debe mandar una señal al SP0256 para que éste inicie la emisión del *allophone* correspondiente al código marcado por la dirección de la ROM, debe también monitorear a una señal proporcionada por el SP0256 que indica la terminación de la emisión del *allophone*. En el momento en que el controlador recibe esta señal, este incrementa en uno el valor de la parte menos significativa del bus de direcciones, con el

fin de apuntar a la localidad en que se encuentra el código del siguiente *allophone*. Para lograr estos incrementos, utilizamos un contador binario, el cual realiza su conteo por medio de una señal del controlador.

Como la secuencia de *allophones* más larga requiere menos de 32 localidades de memoria para el almacenamiento de sus códigos, pero más de 10, decidimos que el contador binario utilizado fuera de 5 bits. Estos bits son los que conectamos a la parte menos significativa del bus de direcciones de la ROM. De esta manera, la ROM que se requiere necesita tener $8+5=13$ bits de direcciones.

El controlador necesita conocer en que momento ocurre la terminación de la secuencia de *allophones*, para esto, aprovechando que los códigos de los *allophones* requieren de 6 bits, podemos utilizar cualquiera de los 2 bits sobrantes del byte para establecer una señal de monitoreo del controlador. Decidimos utilizar el bit más significativo para este fin. Así, en la localidad siguiente a la del código del último *allophone* de cada secuencia, ponemos el código 10000000.

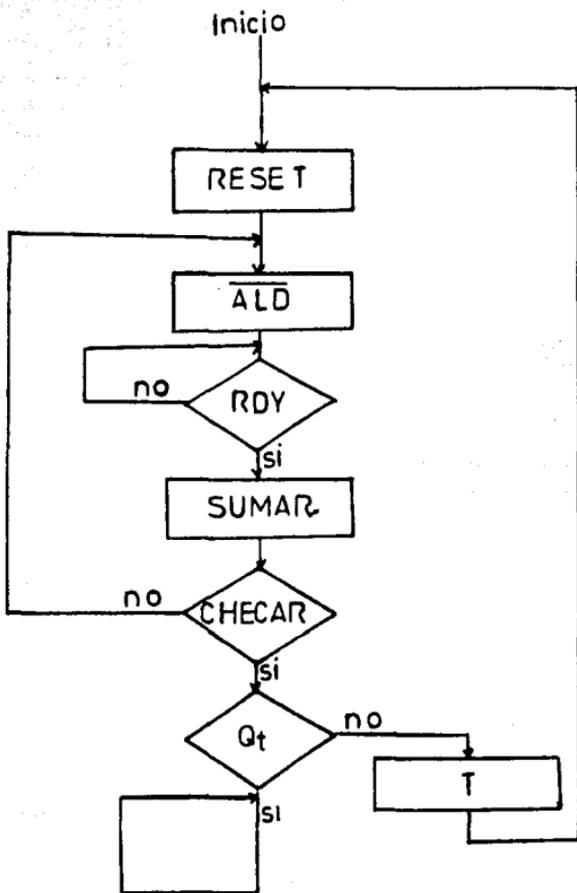


Figura 5-2. Diagrama de Flujo del Control del Sistema.

Una vez que el controlador capta la señal de terminación de secuencia, éste checa por medio de una señal de entrada (Qt) si esta secuencia fue de horas o de minutos, en caso de que ésta indique que fueron minutos, el controlador termina su operación, por el contrario, si la secuencia terminada fué de horas, el controlador manda la señal que efectúa la conmutación entre la captación de horas o de minutos. La señal en sí está conectada a un flip-flop, que cambia su valor de salida al recibirla. La salida del flip-flop se conecta a la selección del switch de conmutación y también funciona como la señal indicadora (Qt). Después de hacer la conmutación de horas a minutos, el controlador repite los pasos para la emisión de la secuencia de *allophones* correspondiente a los minutos.

5.5 DIAGRAMA DE FLUJO.

En la figura 5-2 presentamos el diagrama de flujo del controlador.

La lista de señales utilizadas en el diagrama de flujo es la siguiente:

SEÑAL:	FUNCIÓN:
→ Reset	Ajusta a ceros el contador.
→ <u>ALD</u>	Le indica al SP0256 que cargue el código y emita el <i>allophone</i> .
← RDY	Indica el SP0256 al controlador que ha terminado la emisión.
→ Sumar	Incrementa al contador en 1.
← Checar	Indica al controlador el fin de una secuencia de <i>allophones</i> .
← Qt	Indica al controlador qué se acaba de emitir (horas o minutos)
→ T	Ejecuta la conmutación en la lectura, de horas a minutos.

Nota: Se indica el sentido de la señal de acuerdo a los siguientes símbolos:

- ← Señal de entrada al controlador.
- Señal de salida del controlador.

5.5.1 Carta ASM.

A partir del diagrama de flujo del control del sistema, y con el fin de implementar el circuito de control, se desarrolló la carta ASM de la figura 5-3.

5.5.2 Reducción de la carta ASM.

Con el fin de construir el circuito que

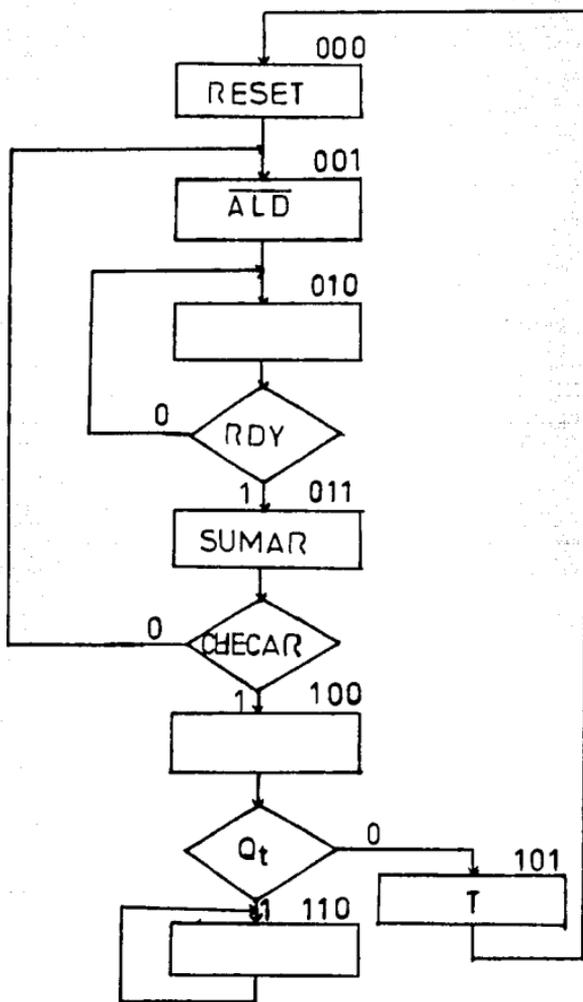


Figura 5-3. Carta ASM del Control del Sistema.

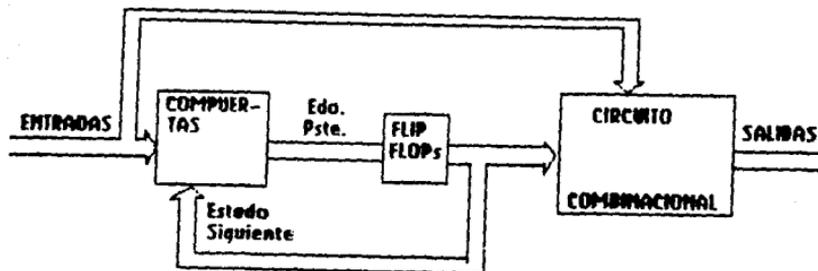


Figura 5-4 Reducción con compuertas.

represente a la carta ASM antes mostrada, e implementarla de la mejor manera posible, estudiamos una serie de métodos que permiten la reducción e implementación de cartas ASM.

A continuación mencionamos brevemente algunos métodos utilizados para este fin:

I. Reducción con compuertas: En este método, se obtienen ecuaciones para cada uno de los estados siguientes y de las salidas, partiendo de los mapas de Karnaugh. Pueden utilizarse cualquier tipo de flip-flops. Las ecuaciones de estado siguiente y las de salida se implementan mediante lógica combinacional (utilizando compuertas lógicas).

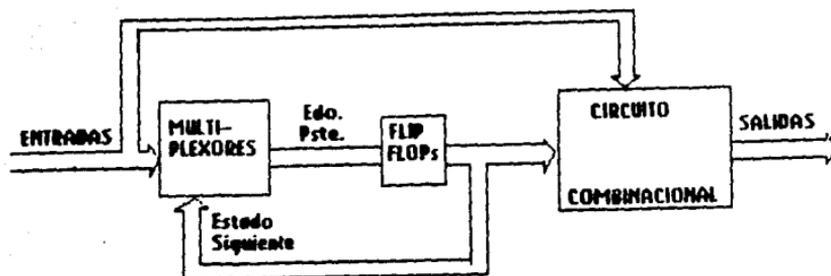


Figura 5-5. Reducción con multiplexores.

II. Reducción con multiplexores: Este método es básicamente igual al anterior. Difiere únicamente en que aquí, las ecuaciones de estado siguiente se implementan con el uso de multiplexores, dejando la implementación de las variables de salida como en el método anterior.

III. Reducción de máquinas de estado utilizando ROM: En este tipo de reducción, la memoria ROM se utiliza para almacenar la función de siguiente estado, las señales de salida, y, a veces, el manejo de las entradas. Las máquinas lógicas en base a una ROM, se distinguen entre sí por la estructura de su palabra, teniendo cuatro diferentes tipos de direccionamiento, de acuerdo a las

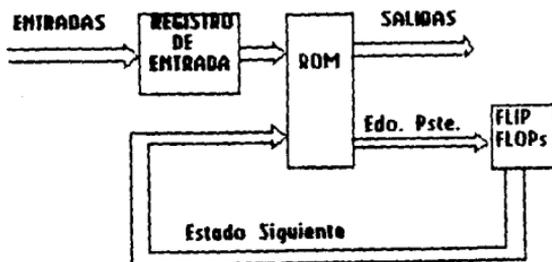


Figura 5-6. Direccionamiento por Trayectoria.

características de sus componentes y de su palabra.

a) Direccionamiento por trayectoria: Este tipo de direccionamiento almacena en una palabra de memoria el estado siguiente y las salidas, para cada estado de la carta ASM. A la parte de memoria que almacena el estado siguiente se le llama *link*, y a la que genera la salida, se le llama *instrucción*. Cada dirección de memoria se genera en función del estado presente y de las entradas, y se le llama trayectoria de la dirección (*link path address*). La carta ASM debe modificarse ligeramente cuando existen salidas condicionadas, agrupando las salidas en bloques de salidas condicionales por cada trayectoria.

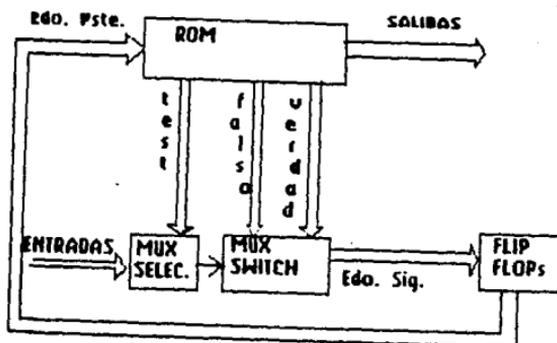


Figura 5-7. Direccionamiento Entrada-Estado.

b) Direccionamiento Entrada Estado: Este método de direccionamiento tiene un campo adicional en el campo de memoria, llamado de prueba o *test*, que sirve para seleccionar la entrada al sistema que se debe analizar. Para procesar las entradas y determinar el siguiente estado, se adicionan a la ROM dos MUX, llamados *selector* y *switch*. El campo *test* se dedica a direccionar la entrada correspondiente a la entrada presente. El MUX selecciona uno de los posibles estados siguientes que le proporciona la memoria ROM. La salida del multiplexor *switch* representa el estado siguiente de acuerdo a la entrada seleccionada por el multiplexor *selector*.

c) Direccionamiento Implícito: Una variante

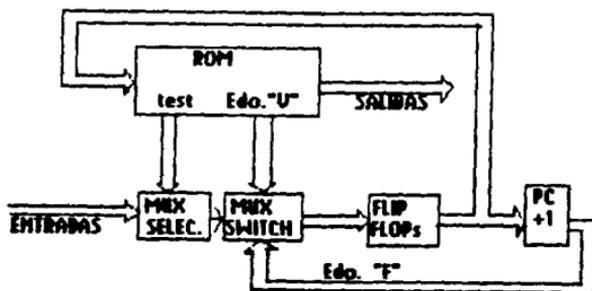


Figura 5-8. Direccionamiento implícito.

del direccionamiento Estado-Entrada, es el direccionamiento implícito, que utiliza un sólo campo de siguiente estado, que es donde se almacena la dirección cuando la entrada es verdadera, y la dirección directa o falsa se llama dirección implícita, y se calcula sumándole un 1 a la dirección presente. El incrementar en 1 el estado presente, crea transiciones, que crean trayectorias y obligan a la creación de estados de conexión que dirigen el salto hacia el estado original de la trayectoria. Para asignar código a los estados de la carta ASM, se siguen las siguientes reglas:

1. Se asigna código a cada uno de los estados que siguen trayectorias cerradas en las que

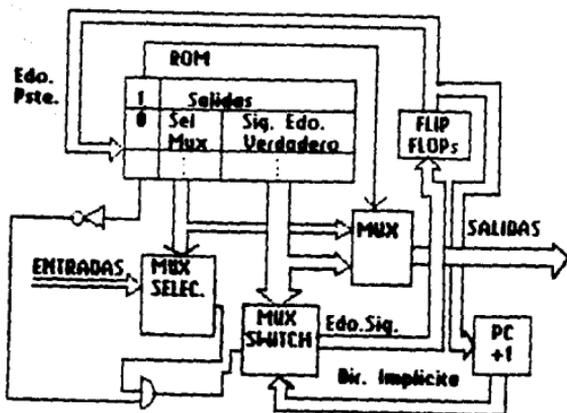


Figura 5-9. Direccionamiento de Formato Variable.

la entrada del estado sea falsa o cero, esta secuencia corresponde a la implementación de un contador de programa.

2. Para cerrar las trayectorias cuando no existe continuidad para entradas falsas, se introducen saltos hacia el estado deseado, utilizando una entrada Q_x con un valor predeterminado, esto obliga a la modificación de la carta ASM cuando existe coincidencia de entradas falsas y de estados sin entradas.

d) Direccionamiento de formato variable: En el direccionamiento de formato variable, la información que comprende microcomandos o salidas, funciones de siguiente estado y selección de entrada, se almacenan en 2 o más

palabras de memoria. El formato variable más simple, se almacena en 2 palabras, y la distribución de salidas, siguiente estado; y entradas, se determinan por el valor de un nuevo campo de n bits, llamado indicador de formato, colocado al inicio de cada palabra. El primer formato o palabra, es la parte de microcomandos que corresponde a las salidas del sistema. La segunda palabra comprende la palabra de siguiente estado, y selecciona la entrada de acuerdo a un tipo de direccionamiento implícito. Esta manera de almacenar la información, limita el ancho de la palabra, permitiendo empacarla en un menor número de ROMs. Las secciones de selección de entrada y selector de siguiente estado, que son externas a la ROM; permanecen igual que en el modo de direccionamiento implícito. La carta ASM se modifica asignando un estado a los microcomandos, y otro diferente a la evaluación de la entrada, ya sea dirección implícita o dirección de salto.

Después de analizar estos esquemas, y observando que nuestra carta ASM no requiere de la complejidad de las reducciones de estado con ROMs, decidimos llevar a cabo la reducción con multiplexores, puesto que esta

técnica cubre nuestros requerimientos.

5.5.3 Tabla de estado.

Estado Presente X Y Z	Entradas	Estado Siguiete X Y Z	Salidas			
			Res	ALD	Sum	T
0 0 0	-----	0 0 1	1	1	0	0
0 0 1	-----	0 1 0	0	0	0	0
0 1 0	RDY=0	0 1 0	0	1	0	0
	RDY=1	0 1 1	0	1	0	0
0 1 1	Checar=0	0 0 1	0	1	1	0
	Checar=1	1 0 0	0	1	1	0
1 0 0	Qt=0	1 0 1	0	1	0	0
	Qt=1	1 1 0	0	1	0	0
1 0 1	-----	0 0 0	0	1	0	1
1 1 0	-----	1 1 0	0	1	0	0
1 1 1	-----	* * *	*	*	*	*

Las ecuaciones de salida están dadas por:

$$\text{Res} = \overline{X}Y\overline{Z}$$

$$\overline{\text{ALD}} = X+Y+\overline{Z} = \overline{\overline{X}Y\overline{Z}}$$

$$\text{Sum} = \overline{X}YZ$$

$$T = X\overline{Y}Z$$

5.6 CAPTURA DE DATOS PROVENIENTES DEL MM5311.

Puesto que el MM5311 entrega sus datos de una manera multiplexada, era necesario capturarlos cíclicamente con el fin de poder

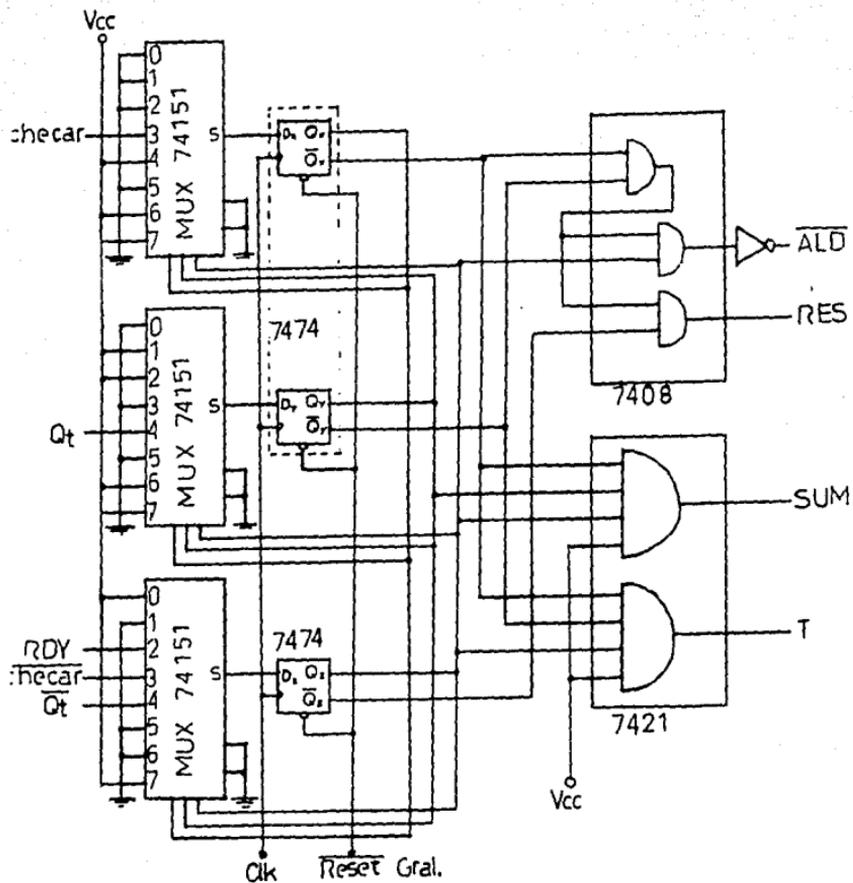


Figura 5-10. Diagrama esquemático del control del sistema.

tener acceso a los 4 datos (decenas de hora, unidades de hora, decenas de minuto y unidades de minuto) en cualquier instante.

El MM311 presenta 4 terminales que representan un dígito BCD, y 4 terminales que sirven para indicar lo que ese dígito representa (unidades de hora, decenas de minuto etc..). Por ejemplo, si en las 4 terminales de dígito BCD aparece 0101 y de las 4 terminales de selección de dígito se habilita el de decenas de minuto, sabemos que un 5 corresponderá a las decenas de minuto. En otro instante, el MM311 habilitará la terminal de unidades de minuto, y si para este momento el dígito BCD es 0010, entonces sabremos que en las unidades de minuto tendremos un 2, de esta manera, obtendremos un 52 en los minutos. De esta manera, el MM311 indica las horas y los minutos.

El dígito BCD y la selección de unidad se habilitan al mismo tiempo. Con el fin de utilizar la señal de selección de unidad para poder capturar el dígito BCD en latches, requerimos atrasar ésta (Cada una de las 4 señales atrasadas de selección está conectada

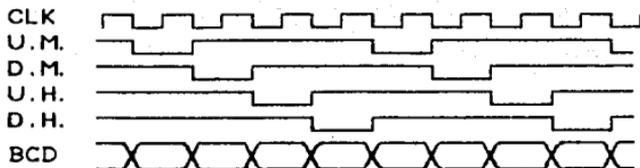


Figura 5-11. Ciclo de Multiplexaje del MM5311

a la terminal clk de un conjunto de latches, que funcionan por flanco) para que no existan transiciones en las terminales del dígito BCD al momento de la captura.

La frecuencia con la que el MM5311 ajusta el valor de sus terminales está gobernada por un oscilador externo (un C. I. 555). La fig. 5-11 muestra un diagrama de tiempo del ciclo de selección de unidad del MM5311.

Como podemos ver, la habilitación de las señales de selección ocurre en el flanco negativo del CLK de entrada, entonces, para evitar la captura de datos en el momento de transición, retrasamos las señales de selección exactamente medio ciclo, utilizando compuertas.

5.6.1 Acoplamiento de señales.

Dado que el MM5311 funciona con niveles de voltaje de 10 volts y que el resto del sistema funciona con niveles TTL, fué necesario acoplar las señales. Este acoplamiento se hizo por medio de divisores de voltaje. Experimentalmente obtuvimos que esta división debía ser de $1/3.2$ para lograr el reconocimiento de niveles de los dispositivos TTL.

La señal BCD del MM5311 fué dividida de la manera anterior, mientras que para obtener los niveles adecuados de voltaje en las señales de selección de unidad, fué necesario hacer uso de transistores. Esto se debe a que los niveles de voltaje en estas salidas son: alto 9.8[v] y bajo 9.2[v]. Por medio de transistores pudimos obtener valores de 10 y 0 volts.

Para conectar el display, se utiliza el grupo de señales de selección de unidad junto con un grupo de señales del tipo 7 segmentos con que cuenta el MM5311, y que corresponde con el dato en BCD multiplexado. Siendo que estas últimas señales tienen los mismos

niveles de voltaje que las de selección de unidad, se requirió de otro grupo de transistores, también funcionando en corte y saturación. Entre los 2 grupos de transistores se opera el display.

6. CODIFICACION DE LOS VOCABLOS Y SU DIRECCIONAMIENTO EN MEMORIA.

B.1 CODIFICACION DE LOS VOCABLOS.

Para tener la mejor codificación de *allophones* posible, conectamos el SP0256 con una computadora PC, y de esta manera pudimos enviar diferentes secuencias de códigos con el fin de escucharlos y determinar la mejor manera de codificar los *allophones*. El uso de una PC nos permitió un gran ahorro en trabajo, puesto que si hubiéramos tratado de hacer estas pruebas directamente en memoria ROM, el proceso hubiera sido mucho más tardado. La conexión entre el SP0256 y la computadora se hizo por medio del puerto paralelo que esta última presenta.

Finalmente, llegamos a la codificación que mejor representaba los vocablos requeridos. Esta codificación se muestra a continuación:

Palabra(s):	Códigos en hexadecimal:
<i>La una</i>	2D,18,18,03,1E,1E,0B,18,04
<i>Las dos</i>	2D,18,18,2B,03,21,17,17,2B 04
<i>Las tres</i>	2D,18,18,2B,03,11,01,0E,07 07,2B,04

Palabra(s):	Códigos en hexadecimal:
<i>Las cuatro</i>	2D,18,18,2B,03,08,01,1E,1E 18,18,11,01,0E,01,17,04
<i>Las cinco</i>	2D,18,18,2B,03,2B,13,0B,08 17,04
<i>Las seis</i>	2D,18,18,2B,03,2B,07,07,13 2B,04
<i>Las siete</i>	2D,18,18,2B,03,2B,13,07,07 11,01,07,04
<i>Las ocho</i>	2D,18,18,2B,03,17,17,32,02 17,04
<i>Las nueve</i>	2D,18,18,2B,03,0B,1E,1E,07 07,23,07,04
<i>Las diez</i>	2D,18,18,2B,03,21,19,07,07 2B,04
<i>Las once</i>	2D,18,18,2B,03,17,17,0B,2B 07,04
<i>Las doce</i>	2D,18,18,2B,03,21,17,17,2B 07,04
<i>en punto</i>	07,07,0B,03,09,1E,1E,0B,11 01,17,04
<i>uno</i>	1E,1E,0B,17,04
<i>dos</i>	21,17,17,2B,04
<i>tres</i>	11,01,0E,01,07,07,2B,04
<i>cuatro</i>	0B,01,1E,1E,18,11,01,0E 01,17,04
<i>cinco</i>	2B,13,0B,0B,01,17,04
<i>seis</i>	2B,07,07,13,2B,04

Palabra(s):	Códigos en hexadecimal:
<i>siete</i>	2B,13,07,07,11,01,07,04
<i>ocho</i>	17,17,32,02,17,04
<i>nueve</i>	0B,1E,07,07,23,07,04
<i>diez</i>	21,19,07,07,2B,04
<i>once</i>	17,17,0B,2B,07,04
<i>doce</i>	21,17,17,2B,07,04
<i>trece</i>	11,01,0E,07,07,2B,07,04
<i>catorce</i>	0B,01,1B,11,17,0E,2B,07,04
<i>quince</i>	0B,01,13,0B,2B,07,04
<i>dieciséis</i>	21,19,07,2B,13,2B,07,07,13 2B,04
<i>diecisiete</i>	21,19,07,2B,0C,2B,13,07,07 11,07,04
<i>dieciocho</i>	21,19,07,2B,13,17,17,32,02 17,04
<i>diecinueve</i>	21,19,07,2B,0C,0B,1E,1E,07 07,23,23,07,04
<i>veinte</i>	23,07,07,13,0B,11,01,07,07 04
<i>veintuno</i>	23,07,13,0B,11,01,13,1E,1E 0B,01,17,04
<i>veintidos</i>	23,07,13,0B,11,01,13,21,01 17,17,2B,04
<i>veintitres</i>	23,07,13,0B,11,01,13,11,01 0E,07,07,2B,04
<i>veinticuatro</i>	23,07,13,0B,11,01,13,0B,01 1E,2E,1B,1B,11,01,0E,17,04

Palabra(s):	Códigos en hexadecimal:
<i>veinticinco</i>	23,07,13,0B,11,01,13,2B,13 0B,08,17,04
<i>veintiseis</i>	23,07,13,0B,11,01,13,2B,07 07,13,2B,04
<i>veintisiete</i>	23,07,13,0B,11,01,13,2B,13 07,07,11,01,07,04
<i>veintiocho</i>	23,07,13,0B,11,01,13,17,17 32,02,17,04
<i>veintinueve</i>	23,07,13,0B,11,01,13,0B,1E 1E,07,07,23,07,04
<i>treinta</i>	11,01,0E,01,07,07,13,0B,11 01,18,04
<i>treinta y uno</i>	11,02,0E,01,07,13,0B,11,01 0B,13,1E,1E,0B,17,04
<i>treinta y dos</i>	11,02,0E,01,07,13,0B,11,01 0B,13,21,17,17,2B,04
<i>treinta y tres</i>	11,01,0E,01,07,13,0B,11,01 0B,13,11,01,0E,01,07,07,2B 04
<i>treinta y cuatro</i>	11,01,0E,01,07,13,0B,11,01 0B,13,0B,01,1E,1E,18,18,11 01,0E,01,17,04
<i>treinta y cinco</i>	11,01,0E,01,07,13,0B,11,01 0B,13,2B,13,0B,08,01,17,04
<i>treinta y seis</i>	11,01,0E,01,07,13,0B,11,01 01,0B,3B,13,2B,07,07,13,2B 04

Palabra(s):	Códigos en hexadecimal:
<i>treinta y siete</i>	11,01,0E,01,07,13,0B,11,01 06,13,2B,13,07,07,11,01,07 04
<i>treinta y ocho</i>	11,01,0E,01,07,13,0B,11,01 06,13,17,17,32,02,17,04
<i>treinta y nueve</i>	11,01,0E,01,07,13,0B,11,01 06,13,0B,1E,07,07,23,07,04
<i>cuarenta</i>	08,01,1E,18,27,07,07,0B,11 01,18,04
<i>cuarenta y uno</i>	08,01,1E,18,27,07,0B,11,01 06,13,1E,1E,0B,17,04
<i>cuarenta y dos</i>	08,01,1E,18,27,07,0B,11,01 06,13,17,17,2B,04
<i>cuarenta y tres</i>	08,01,1E,18,27,07,0B,11,01 06,13,11,01,0E,01,07,07,2B 04
<i>cuarenta y cuatro</i>	08,01,1E,18,27,07,0B,11,01 06,13,0B,01,1E,1E,18,18,11 01,0E,01,17,04
<i>cuarenta y cinco</i>	08,01,1E,18,27,07,0B,11,01 06,13,2B,13,0B,08,01,17,04
<i>cuarenta y seis</i>	08,01,1E,18,27,07,0B,11,01 06,13,2B,07,07,13,2B,04
<i>cuarenta y siete</i>	08,01,1E,18,27,07,0B,11,01 06,13,2B,13,07,07,11,01,07 04

Palabra(s):	Códigos en hexadecimal:
<i>cuarenta y ocho</i>	08,01,1E,18,27,07,0B,11,01 08,13,17,17,32,02,17,04
<i>cuarenta y nueve</i>	08,01,1E,18,27,07,0B,11,01 08,13,0B,1E,07,07,23,07,04
<i>cinquenta</i>	2B,13,0B,08,1E,07,07,0B,11 01,18,04
<i>cinquenta y uno</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,1E,1E,0B,17,04
<i>cinquenta y dos</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,21,17,17,2B,04
<i>cinquenta y tres</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,11,01,0E,01,07,07,2B 04
<i>cinquenta y cuatro</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,08,01,1E,1E,18,11,01 0E,01,17,04
<i>cinquenta y cinco</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,2B,13,0B,18,01,17,04
<i>cinquenta y seis</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,2B,07,07,13,2B,04
<i>cinquenta y siete</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,2B,13,07,07,11,01,07 04
<i>cinquenta y ocho</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,17,17,32,02,17,04
<i>cinquenta y nueve</i>	2B,13,0B,08,1E,07,0B,11,01 08,13,0B,1E,07,07,23,07,04

6.2 DIRECCIONAMIENTO DE LOS VOCABLOS EN MEMORIA.

El siguiente paso fué localizar la dirección de comienzo de cada área de codificación de secuencia de *allophones*. Esto se hizo de acuerdo al concepto de división de la memoria en 72 áreas como se describió en el capítulo 5. A continuación se muestran estas áreas de dirección.

NOTA: Tenemos que considerar que cuando en las decenas de hora no se tiene ningun valor, en las terminales BCD del MM5311 se tiene el código binario 1111.

Secuencia:	Dirección de inicio:	
	Binaria	Hexadecimal
<i>La una</i>	1111000100000	1E20
<i>Las dos</i>	1111001000000	1E40
<i>Las tres</i>	1111001100000	1E80
<i>Las cuatro</i>	1111010000000	1E80
<i>Las cinco</i>	1111010100000	1EA0
<i>Las seis</i>	1111011000000	1EC0
<i>Las siete</i>	1111011100000	1EE0
<i>Las ocho</i>	1111100000000	1F00
<i>Las nueve</i>	1111100100000	1F20
<i>Las diez</i>	1001000000000	1200

Dirección de inicio:

Secuencia:	Binaria	Hexadecimal
<i>Las once</i>	1001000100000	1220
<i>Las doce</i>	1001001000000	1240
<i>en punto</i>	0000000000000	0000
<i>uno</i>	0000000100000	0020
<i>dos</i>	0000001000000	0040
<i>tres</i>	0000001100000	0060
<i>cuatro</i>	0000010000000	0080
<i>cinco</i>	0000010100000	00A0
<i>seis</i>	0000011000000	00C0
<i>siete</i>	0000011100000	00E0
<i>ocho</i>	0000100000000	0100
<i>nueve</i>	0000100100000	0120
<i>diez</i>	0001000000000	0200
<i>once</i>	0001000100000	0220
<i>doce</i>	0001001000000	0240
<i>trece</i>	0001001100000	0260
<i>catorce</i>	0001010000000	0280
<i>quince</i>	0001010100000	12A0
<i>dieciseis</i>	0001011000000	02C0
<i>diecisiete</i>	0001011100000	02E0
<i>dieciocho</i>	0001100000000	0300
<i>diecinueve</i>	0001100100000	0320
<i>veinte</i>	0010000000000	0400
<i>veintiuno</i>	0010000100000	0420
<i>veintidos</i>	0010001000000	0440
<i>veintitres</i>	0010001100000	0460

Dirección de inicio:

Secuencia:	Binaria	Hexadecimal
<i>veinticuatro</i>	0010010000000	0480
<i>veinticinco</i>	0010010100000	04A0
<i>veintiseis</i>	0010011000000	04C0
<i>veintisiete</i>	0010011100000	04E0
<i>veintiocho</i>	0010100000000	0500
<i>veintinueve</i>	0010100100000	0520
<i>treinta</i>	0011000000000	0600
<i>treinta y uno</i>	0011000100000	0620
<i>treinta y dos</i>	0011001000000	0640
<i>treinta y tres</i>	0011001100000	0660
<i>treinta y cuatro</i>	0011010000000	0680
<i>treinta y cinco</i>	0011010100000	06A0
<i>treinta y seis</i>	0011011000000	06C0
<i>treinta y siete</i>	0011011100000	06E0
<i>treinta y ocho</i>	0011100000000	0700
<i>treinta y nueve</i>	0011100100000	0720
<i>cuarenta</i>	0100000000000	0800
<i>cuarenta y uno</i>	0100000100000	0820
<i>cuarenta y dos</i>	0100001000000	0840
<i>cuarenta y tres</i>	0100001100000	0860
<i>cuarenta y cuatro</i>	0100010000000	0880
<i>cuarenta y cinco</i>	0100010100000	08A0
<i>cuarenta y seis</i>	0100011000000	08C0
<i>cuarenta y siete</i>	0100011100000	08E0
<i>cuarenta y ocho</i>	0100100000000	0900
<i>cuarenta y nueve</i>	0100100100000	0920

Dirección de inicio:

Secuencia:	Binaria	Hexadecimal
<i>cincuenta</i>	010100000000	0A00
<i>cincuenta y uno</i>	010100010000	0A20
<i>cincuenta y dos</i>	010100100000	0A40
<i>cincuenta y tres</i>	010100110000	0A80
<i>cincuenta y cuatro</i>	010101000000	0A80
<i>cincuenta y cinco</i>	010101010000	0AA0
<i>cincuenta y seis</i>	010101100000	0AC0
<i>cincuenta y siete</i>	010101110000	0AE0
<i>cincuenta y ocho</i>	010110000000	0B00
<i>cincuenta y nueve</i>	010110010000	0B20

B.3 MAPA DE MEMORIA.

Una vez que se obtuvieron los códigos para cada secuencia de *allophones* y que se obtuvieron las direcciones de memoria en las que estos códigos debían quedar, se procedió a realizar la grabación del circuito integrado de memoria. Este chip es un UVEPROM (*Ultra Violet Erasable Programmable Read Only Memory*). La grabación de la memoria se llevó a cabo utilizando un equipo de cómputo especial para esto. El mapa de memoria se muestra a continuación. Se omiten grandes áreas de memoria no ocupadas. A la izquierda se muestra la primera de cada 16 direcciones.

0000:07 07 08 03 09 1E 1E 0B 11 01 17 04 80 55 55 55UUU
0010:55 55 55 55 55 55 55 55 55 55 55 55 55 55	UUUUUUUUUUUUUUUU
0020:1E 1E 0B 17 04 80 55 55 55 55 55 55 55 55 55UUUUUUUUUU
0030:55 55 55 55 55 55 55 55 55 55 55 55 55 55	UUUUUUUUUUUUUUUU
0040:21 17 17 2B 04 80 55 55 55 55 55 55 55 55 55	!.,.,UUUUUUUUUU
0050:55 55 55 55 55 55 55 55 55 55 55 55 55 55	UUUUUUUUUUUUUUUU
0060:11 01 0E 07 07 2B 04 80 FF FF FF FF FF FF FF FF+.....
0070:FF FF
0080:9B 01 1E 1E 1B 1B 11 01 0E 17 04 80 FF FF FF FF
0090:FF FF
00A0:2B 13 0B 0B 01 17 04 80 FF FF FF FF FF FF FF FF	+.....
00B0:FF FF
00C0:2B 07 07 13 2B 04 80 FF AA AA AA AA AA AA AA	+..+.....
00D0:AA AA
00E0:2B 13 07 07 11 07 04 80 AA AA AA AA AA AA AA	+.....
00F0:AA AA
0100:17 17 32 02 17 04 80 AA AA AA AA AA AA AA AA	..2.....
0110:AA AA
0120:0B 1E 1E 07 07 23 23 07 04 80 AA AA 00 00 00 0000.....
0130:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0140:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0150:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0160:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0170:00 00 00 00 00 00 00 00 00 00 00 00 00 00

0180:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0190:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01A0:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01B0:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01C0:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01D0:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01E0:38 38 38 38 38 38 38 38 38 38 38 38 38 38	8888888888888888
01F0:38 38 38 38 00 00 00 00 00 00 00 00 00 00	8888.....
0200:21 19 07 07 28 04 80 00 00 00 00 00 00 00	!...+.....
0210:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0220:17 17 08 28 07 04 80 00 00 00 00 00 00 00	...+.....
0230:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0240:21 17 17 28 07 04 80 00 00 00 00 00 00 00	!...+.....
0250:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0260:11 01 0E 07 07 28 07 04 80 00 00 00 00 00+.....
0270:00 00 00 00 00 00 00 00 00 00 00 00 00 00
0280:08 01 18 11 17 17 03 28 07 04 80 00 00 00 00+.....
0290:00 00 00 00 00 00 00 00 00 00 00 00 00 00
02A0:08 01 13 08 28 07 04 80 00 00 00 00 00 00+.....
02B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00
02C0:21 19 07 28 13 28 07 07 13 28 04 80 00 00 00	!...+...+.....
02D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00
02E0:21 19 07 28 0C 28 13 07 07 11 07 04 80 00 00	!...+...+.....
02F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00

0300:21 19 07 28 13 17 17 32 02 17 04 80 00 00 00 00	!..+...2.....
0310:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0320:21 19 07 28 0C 0B 1E 1E 07 07 23 23 07 04 80 00	!..+.....#.....
0330:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0340:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0350:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0360:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0370:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0380:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0390:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
03F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0400:23 07 07 13 0B 11 01 07 07 04 80 00 00 00 00 00	#.....
0410:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0420:23 07 13 0B 11 01 13 1E 1E 0B 01 17 04 80 00 00	#.....
0430:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0440:23 07 13 0B 11 01 13 21 01 17 17 2B 04 80 00 00	#.....!...+....
0450:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0460:23 07 13 0B 11 01 13 11 01 0E 07 07 2B 04 80 00	#.....+....
0470:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

0480:23 07 13 0B 11 01 13 0B 01 1E 2E 1B 1B 11 01 0E	#.....
0490:17 14 80 00 00 00 00 00 00 00 00 00 00 00 00
04A0:23 07 13 0B 11 01 13 2B 13 0B 0B 17 04 80 00 00	#.....+
04B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
04C0:23 07 13 0B 11 01 13 2B 07 07 13 2B 04 80 00 00	#.....+...+
04D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
04E0:23 07 13 0B 11 01 13 2B 13 07 07 11 01 07 04 80	#.....+
04F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0500:23 07 13 0B 11 01 13 17 17 32 02 17 04 80 00 00	#.....2.....
0510:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0520:23 07 13 0B 11 01 13 0B 1E 1E 07 07 23 07 04 80	#.....#...
0530:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0540:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0550:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0560:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0570:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0580:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0590:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
05F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

0600:11 01 0E 01 07 07 13 0B 11 01 1B 04 80 00 00 00
0610:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0620:11 02 0E 01 07 13 0B 11 01 06 13 1E 1E 0B 17 04
0630:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0640:11 02 0E 01 07 13 0B 11 01 06 13 21 17 17 2B 04!..+
0650:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0660:11 01 0E 01 07 13 0B 11 01 06 13 11 01 0E 01 07
0670:07 2B 04 80 00 00 00 00 00 00 00 00 00 00 00	..+.....
0680:11 01 0E 01 07 13 0B 11 01 06 13 0B 01 1E 1E 1B
0690:1B 11 01 1E 01 17 04 80 00 00 00 00 00 00 00 00
06A0:11 01 0E 01 07 13 0B 11 01 06 13 2B 13 0B 0B 01+....
06B0:17 04 80 00 00 00 00 00 00 00 00 00 00 00 00
06C0:11 01 0E 01 07 13 0B 11 01 06 13 2B 07 07 13 2B+...+
06D0:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
06E0:11 01 0E 01 07 13 0B 11 01 06 13 2B 13 07 07 11+....
06F0:01 07 04 80 00 00 00 00 00 00 00 00 00 00 00
0700:11 01 0E 01 07 13 0B 11 01 06 13 17 17 32 02 172..
0710:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
0720:11 01 0E 01 07 13 0B 11 01 06 13 0B 1E 07 07 23#
0730:07 04 80 00 00 00 00 00 00 00 00 00 00 00 00
0740:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0750:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0760:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0770:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

0780:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0790:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
07F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0800:08 01 1E 18 27 07 0B 11 01 18 04 80 00 00 00 00
0810:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0820:08 01 1E 18 27 07 0B 11 01 06 13 1E 1E 0B 17 04
0830:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0840:08 01 1E 18 27 07 0B 11 01 06 13 17 17 2B 04 80+
0850:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0860:08 01 1E 18 27 07 0B 11 01 06 13 11 01 0E 01 07
0870:07 2B 04 80 00 00 00 00 00 00 00 00 00 00 00 00+
0880:08 01 1E 18 27 07 0B 11 01 06 13 0B 01 1E 1E 18
0890:18 11 01 0E 01 17 04 80 00 00 00 00 00 00 00 00
08A0:08 01 1E 18 27 07 0B 11 01 06 13 2B 13 0B 0B 01+
08B0:17 04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
08C0:08 01 1E 18 27 07 0B 11 01 06 13 2B 07 07 13 2B+...+
08D0:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
08E0:08 01 1E 18 27 07 0B 11 01 06 13 2B 13 07 07 11+...+
08F0:01 07 04 80 00 00 00 00 00 00 00 00 00 00 00 00

0880:08 01 1E 18 27 07 08 11 01 06 13 08 01 1E 1E 18'
0890:1E 11 01 0E 01 17 04 80 00 00 00 00 00 00 00 00
08A0:08 01 1E 18 27 07 08 11 01 06 13 2B 13 08 08 01'.....t....
08B0:17 04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
08C0:08 01 1E 18 27 07 08 11 01 06 13 2B 07 07 13 2B'.....t...t
08D0:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
08E0:08 01 1E 18 27 07 08 11 01 06 13 2B 13 07 07 11'.....t....
08F0:01 07 04 80 00 00 00 00 00 00 00 00 00 00 00 00
0900:08 01 1E 18 27 07 08 11 01 06 13 17 17 32 02 17'.....2..
0910:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0920:08 01 1E 18 27 07 08 11 01 06 13 0B 1E 07 07 23'.....B
0930:07 04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
0940:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0950:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0960:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0970:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0980:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0990:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
09F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

0A00:2B 13 0B 0B 1E 07 07 0B 11 01 1B 04 80 80 00 00	+.....
0A10:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0A20:2B 13 0B 0B 1E 07 0B 11 01 06 13 1E 1E 0B 17 04	+.....
0A30:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0A40:2B 13 0B 0B 1E 07 0B 11 01 06 13 21 17 17 2B 04	+.....!..+
0A50:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0A60:2B 13 0B 0B 1E 07 0B 11 01 06 13 11 01 0E 01 07	+.....
0A70:07 2B 04 80 00 00 00 00 00 00 00 00 00 00 00	.+.....
0A80:2B 13 0B 0B 1E 07 0B 11 01 06 13 0B 01 1E 1E 1B	+.....
0A90:11 01 0E 01 17 04 80 00 00 00 00 00 00 00 00
0AA0:2B 13 0B 0B 1E 07 0B 11 01 06 13 2B 13 0B 0B 01	+.....+...+
0AB0:17 04 80 00 00 00 00 00 00 00 00 00 00 00 00
0AC0:2B 13 0B 0B 1E 07 0B 11 01 06 13 2B 07 07 13 2B	+.....+...+
0AD0:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
0AE0:2B 13 0B 0B 1E 07 0B 11 01 06 13 2B 13 07 07 11	+.....+...+
0AF0:01 07 04 80 00 00 00 00 00 00 00 00 00 00 00
0B00:2B 13 0B 0B 1E 07 0B 11 01 06 13 17 17 32 02 17	+.....2..
0B10:04 80 00 00 00 00 00 00 00 00 00 00 00 00 00
0B20:2B 13 0B 0B 1E 07 0B 11 01 06 13 0B 1E 07 07 23	+.....8
0B30:07 04 80 00 00 00 00 00 00 00 00 00 00 00 00
0B40:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0B50:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0B60:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0B70:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

1180:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1190:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
11F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1200:2D 18 18 2B 03 21 19 07 07 2B 04 80 00 00 00 00	-..+!...+.....
1210:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1220:2D 18 18 2B 03 17 17 0B 2B 07 04 80 00 00 00 00	-..+!...+.....
1230:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1240:2D 18 18 2B 03 21 17 17 2B 07 04 80 00 00 00 00	-..+!...+.....
1250:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1260:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1270:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1280:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1290:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12A0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12B0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12C0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12D0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12E0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
12F0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

1DB0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1D90:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DA0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DB0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DC0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DD0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DE0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1DF0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E00:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E10:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E20:2D 18 18 03 1E 1E 0B 18 04 80 00 00 00 00 00 00	~.....
1E30:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E40:2D 18 18 2B 03 21 17 17 2B 04 80 00 00 00 00 00	~..+!..f.....
1E50:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E60:2D 18 18 2B 03 11 01 0E 07 07 2B 04 80 00 00 00	~..+.....+.....
1E70:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1E80:2D 18 18 2B 03 0B 01 1E 1E 18 18 11 01 0E 17 04	~..+.....
1E90:80 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1EA0:2D 18 18 2B 03 2B 13 0B 0B 01 17 04 80 00 00 00	~..+..+.....
1EB0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1EC0:2D 18 18 2B 03 2B 07 07 13 2B 04 80 00 00 00 00	~..+..+..+.....
1ED0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1EE0:2D 18 18 2B 03 2B 13 07 07 11 07 04 80 00 00 00	~..+..f.....
1EF0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

1F00:2D 18 18 2B 03 17 17 32 02 17 04 80 00 00 00 00	-...2.....
1F10:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F20:2D 18 18 2B 03 0B 1E 1E 07 07 23 23 07 04 80 00	-...#.....
1F30:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F40:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F50:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F60:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F70:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F80:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1F90:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FA0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FB0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FC0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FD0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FE0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
1FF0:00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

7. CONSTRUCCION Y PRUEBAS DEL RELOJ PARLANTE

7.1 UTILIZACION DE TABLETAS EXPERIMENTALES.

Todo el diseño que se ha descrito en capítulos anteriores, fué armado en tabletas experimentales (*project-board*), probando primero el reloj independiente, es decir sin considerar la parte que lo hace parlante. Después armamos el sistema controlador, probando que la secuencia de estados fuera la correcta, de acuerdo con las variables de entrada, que primeramente fueron asignadas manualmente, también fueron checadas las habilitaciones de señales de salida correspondientes a los distintos estados.

Después de algunas correcciones en cuanto al alambrado, logramos hacer funcionar el sistema controlador en sí. Fué entonces cuando empezamos a armar la parte de captura de datos digitales del reloj. Aquí tuvimos un problema que no estaba contemplado; la captura no se estaba realizando de la manera correcta. Después de investigar detenidamente el error, logramos apreciar que éste se debía a que los latches, que deberían adquirir el dato únicamente en el flanco positivo de su CLK, lo hacían también en el negativo. Dedujimos, después de estudiar el

funcionamiento del sistema con detenimiento, que este error era causado por rebotes que se presentaban en las señales de entrada al CLK de los latches. Para solucionar este problema, colocamos capacitores, que al amortiguar las señales, eliminaban los rebotes.

Cabe mencionar dentro de las pruebas, las que se realizaron para obtener la relación 1/3.2 de los divisores de voltaje que se ocuparon en el acoplamiento de las señales entre el MM311 y la parte de captura de datos del sistema de control.

Una vez armado todo el sistema, habiendo interconectado sus módulos, encontramos que se producían más rebotes en distintas señales del controlador. El funcionamiento del sistema era ya apreciable, pero se percibía una especie de *tartamudeo* y de fraccionamiento de los vocablos. En base a capacitores y circuitos schmitt trigger, fue posible eliminar los rebotes de dichas señales, con lo cual se podía apreciar de una manera clara la emisión de los fonemas.

Las pruebas de la parte de sintetización

de la voz se realizaron en una computadora PC, la que mediante su puerto paralelo fué conectada al SP0256. Después de conseguir la vocalización adecuada para las frases requeridas, obtuvimos los códigos que más se apegaron a lo que se pretendía. Hay que hacer notar que el uso de la computadora nos ahorró una gran cantidad de trabajo en cuanto a la prueba de los diferentes códigos para los distintos *allophones*, cosa que no habríamos logrado si éstos hubieran sido generados mediante memorias ROM.

Una vez que quedó concluido el proyecto hecho en *project-board*, realizamos una prueba que demostraría la funcionalidad de nuestro sistema: Elegimos a un individuo ajeno al proyecto y le pedimos que no observara el *display* y tratara de escuchar qué hora emitía el reloj parlante. Entonces hicimos funcionar varias veces el sistema en horas y minutos puestos al azar y todas fueron escuchadas precisamente como deberían de haber sido, es decir, se reconocieron cada una de las horas y los minutos que marcaba el reloj.

La figura 7-1 muestra un diagrama esquemático de toda la circuitería utilizada



Fotografía 7-1. El Reloj Parlante construido sobre tableros experimentales.

para el reloj parlante. En éste se muestran los circuitos integrados y los dispositivos electrónicos utilizados, incluidas las interconexiones entre cada una de las terminales de los diferentes elementos.

En la fotografía 7-1 se muestra el reloj parlante construido sobre *project-board*, tal y como quedó al momento de funcionar adecuadamente.

7.2 EL CIRCUITO IMPRESO.

Una vez que concluimos la construcción del sistema sobre *project-board*, procedimos a construir el circuito impreso del reloj parlante.

Primeramente, hicimos modelos de papel de cada uno de los circuitos integrados y demás dispositivos utilizados en el proyecto, todos ellos proporcionales a su tamaño real. El papel con el que fueron hechos era cuadriculado, en donde la longitud del lado de un cuadro correspondía a la distancia entre 2 terminales adyacentes de un circuito integrado común (o sea, 100 milésimas de pulgada, que es igual a 2.54[mm]).

Una vez que se tenían todos los modelos, se distribuyeron cada uno de ellos de la manera más apropiada para el futuro circuito impreso. Para esto, se tomó como base un pliego de papel cuadriculado, sobre el cual se colocaron los modelos de papel. La distribución fué hecha de tal manera que las líneas de conexión entre los diferentes elementos tuvieran el espaciamiento óptimo para el circuito impreso.

Después de obtener la distribución más adecuada, hicimos uso del programa de cómputo llamado *Smart Work* e hicimos las mascarillas del que sería el circuito impreso. Una de las mascarillas correspondía al lado anverso del circuito impreso (al lado de *componentes*), y la otra correspondía al lado reverso (o sea el lado de *soldaduras*). Tratamos básicamente de seguir la regla de dejar todas las líneas verticales de un lado y todas las líneas horizontales del otro, con algunas excepciones donde no convenía. Las uniones entre líneas de lados diferentes se marcaron mediante círculos que después indicarían la existencia de una perforación (llamada *through-hole*).

Una vez que en la computadora se concluyó el diseño de las mascarillas, el siguiente paso fué el graficar las mascarillas con ayuda de un *plotter*. Las gráficas estaban hechas a una escala 2-a-1 con respecto al tamaño real del circuito impreso. El motivo por el cual se utilizó un *plotter* (en lugar de una impresora común de matriz de puntos) y una escala doble fué el

de obtener una resolución muy buena en el trazado de las pistas.

Después de obtener las gráficas de las mascarillas, se les tomó una fotografía a cada una de ellas y se obtuvieron negativos y positivos de las gráficas en tamaño real, impresas en papel fotográfico transparente. En la figuras 7-2 se muestran los positivos de las mascarillas (Incluyendo una figura-guía de distribución). Se adquirió una tableta fenólica virgen de doble cara, a la cual se le bañó con un químico fotorresistente en ambas caras. Cada uno de los positivos de las mascarillas fueron colocados en las superficies de la tableta fenólica fijándolas de tal manera que las terminales de los chips y demás elementos correspondiera en ambas caras.

Se procedió entonces a exponer a la luz a ambas caras de la tableta fenólica con el fin de que las áreas de cobre iluminadas pudieran reaccionar con otro químico. De esta manera, en estas zonas se eliminó el cobre; con lo que pudimos obtener el circuito impreso del reloj parlante.

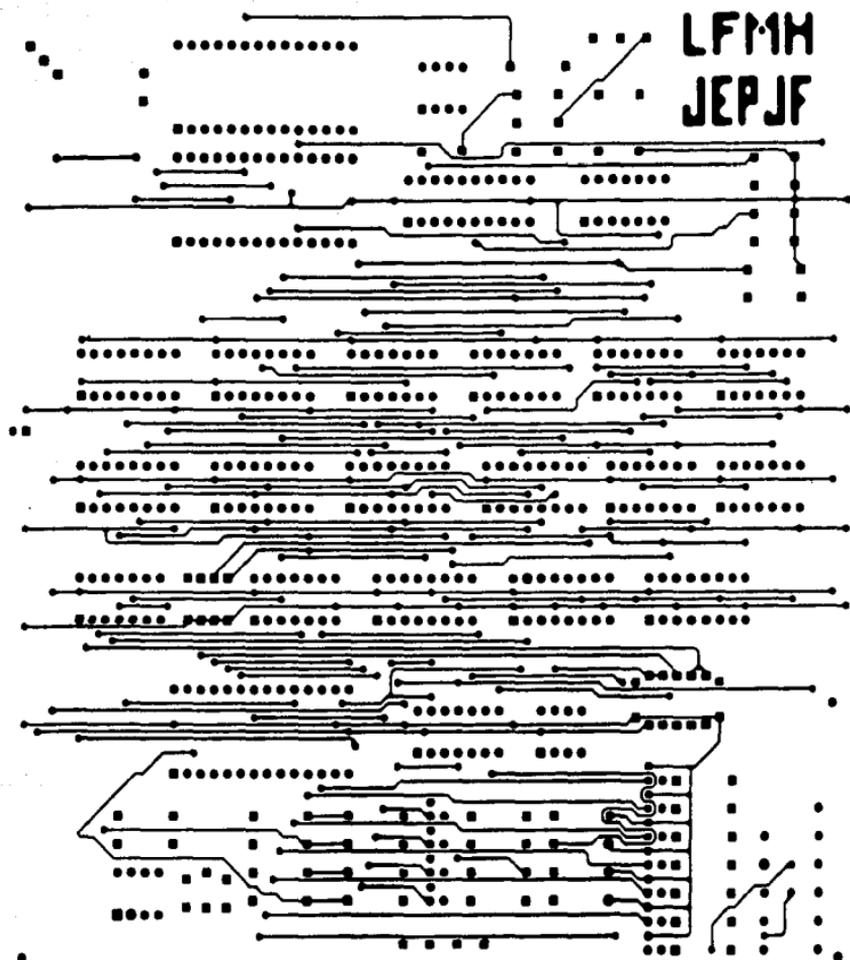


Figura 7-2-a. Mascarilla del lado anverso del circuito impreso.

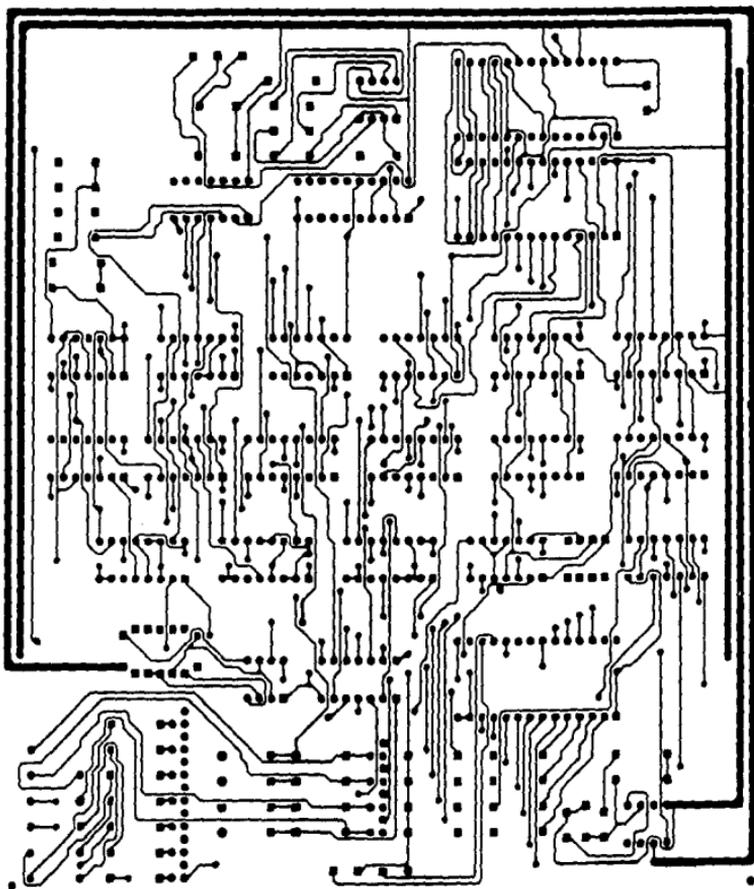


Figura 7-2-8. Mascarilla del lado reverso del
circuitos impresos.

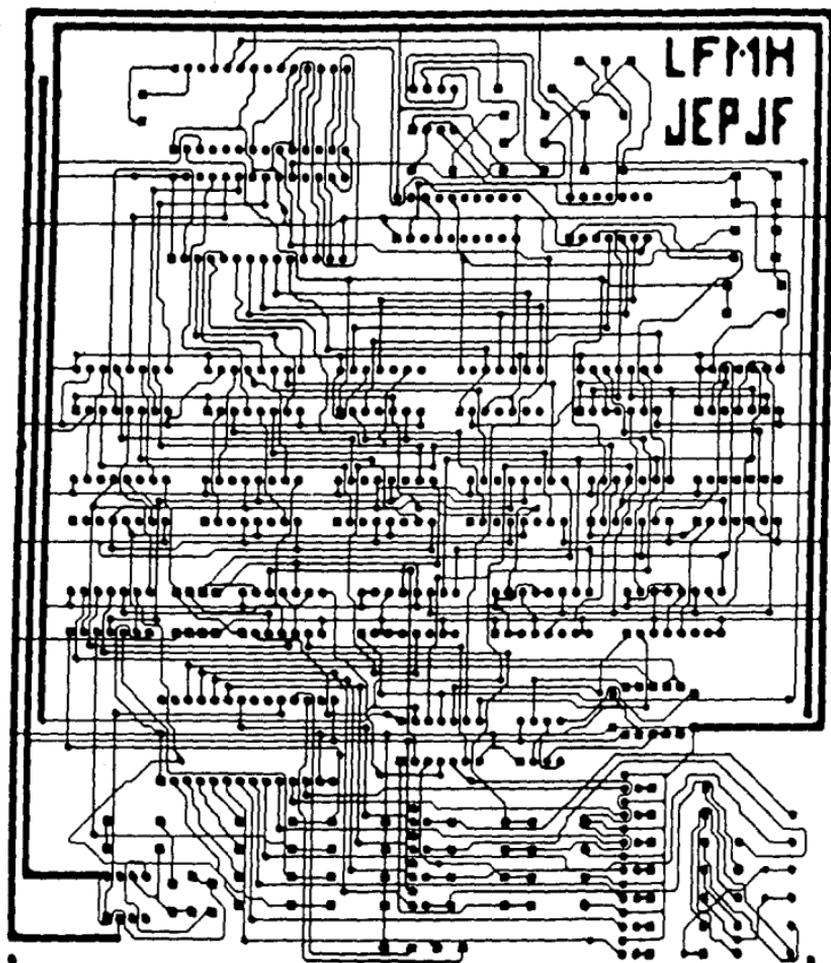


Figura 7-2-c. Superposición de las mascarillas del circuito impreso.

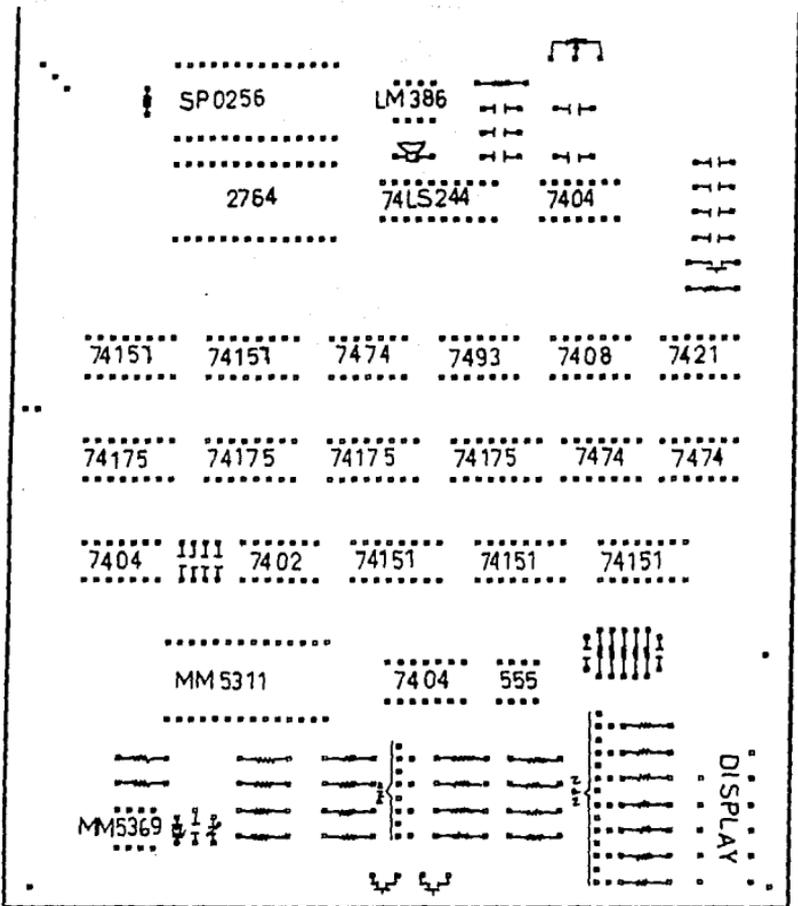


Figura 7-2-d. Distribución de componentes en el Circuito Impreso.

Al circuito impreso se le hicieron las perforaciones correspondientes a todas las terminales de los diferentes elementos, así como las correspondientes a los *through-holes*.

Se hicieron entonces las interconexiones de los *through-holes*, soldando conductores que unían ambas caras. También fueron soldados todos los chips (algunos de ellos requirieron de bases para aislarlos del calor de la soldadura), dispositivos y terminales de alimentación.

El circuito impreso estaba ya en condiciones de ser probado. Se procedió a esto y se encontraron un gran número de fallas. La mayoría de estos errores se debieron a fallas de conductividad en las pistas, debido principalmente a que para realizar la construcción del C.I. no se contó con un equipo especializado. Este tipo de errores es difícil de localizar, por lo que en esa parte del trabajo se consumió más tiempo del que teníamos contemplado. Para corregir estos errores, unimos con soldadura, pero principalmente con alambres independientes, las pistas afectadas.

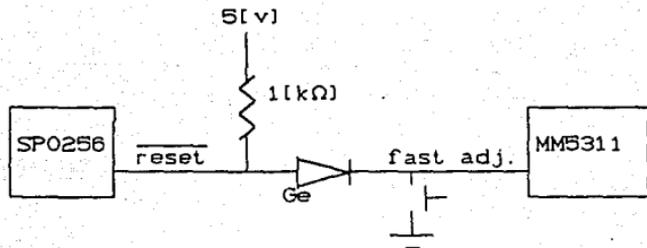


Figura 7-3 Reset del SP0256.

Hubo también algunos problemas de otra índole, tales como el ruido que se producía en el oscilador del SP0256, el cual se corrigió utilizando capacitores para filtrado. O como el que se presentaba al oprimir el botón de emisión de hora, el cual producía un ligero subido antes de hablar la hora. Este fué corregido cambiando la entrada del reset del SP0256 a uno de los botones de ajuste de hora en lugar de mantenerlo en el de emisión de hora; para lo cual se diseñó lo mostrado en la figura 7-3.

También se hizo el siguiente cambio: La señal de tiempo del control del sistema, en lugar de obtenerse de un divisor de tensión se obtuvo de la caída de voltaje que

La fotografía 7-2 muestra el reloj
parlante hecho con circuito impreso.

8. CONCLUSIONES

Con la presente investigación pudimos darnos cuenta de que el hombre, desde tiempos remotos, ha tenido la necesidad de medir el tiempo. Los diferentes instrumentos de medición del tiempo han ido evolucionando a través de los años. El desarrollo de los relojes ha tendido a mejorar la precisión y la presentación de los mismos. Sin embargo, esta presentación ha estado básicamente dirigida hacia una percepción visual, por lo que aquellos individuos que no pueden hacer uso de la vista, han estado limitados en este aspecto.

Al decidir diseñar un Reloj Parlante, hicimos un estudio del área de la sintetización de voz, en el cual pudimos darnos cuenta de que existe una gran variedad de técnicas relacionadas con el tema, cada una con características particulares. Pudimos ver que los codificadores *vocoders* presentan una muy baja tasa de bits, sin embargo la calidad de voz sintética no es buena; analizamos por otro lado los codificadores por forma de onda, éstos presentan una calidad de habla bastante buena, no obstante, la tasa de bits es muy elevada. En el grupo de codificadores paramétricos e híbridos

podimos encontrar un equilibrio en que se presentaba una buena calidad y una baja tasa de bits. En este punto es preciso decir que actualmente, la tasa de bits está relacionada con el costo del sistema, por lo que el sacrificio en la calidad a cambio de una disminución de la tasa de bits se ha hecho por ventajas económicas. En base a este estudio, seleccionamos un sintetizador de voz de alta versatilidad, que nos permitió llevar a cabo de una manera satisfactoria nuestro proyecto. Ahora bien, con la rápida evolución en la tecnología de producción de circuitos integrados (especialmente chips de almacenamiento), la tasa de bits se está convirtiendo en un factor menos crucial en el costo de los sistemas, por lo que existe una tendencia a futuro de utilizar codificadores de forma de onda, y así aprovechar las ventajas de calidad que estos métodos presentan (Como ejemplo de la buena calidad que se alcanza con PCM, basta mencionar que los Compact Disks utilizan esta técnica, y tienen calidad de sonido incluso mejor que sistemas analógicos de grabación, además, poco a poco se están haciendo más accesibles en cuanto a su costo).

En el diseño del Reloj Parlante, logramos cumplir con nuestras pretensiones, construyendo un prototipo no muy complejo, que permite comprender la hora que proporciona. Comprobamos que un individuo ajeno al proyecto puede entender la hora hablada, siempre y cuando el sujeto esté conciente de que se trata de un reloj. Sin embargo, hay que aceptar que la voz sintética no puede igualar la calidad de una voz natural. Consideramos importante el hecho de haber llevado el proyecto hasta la producción del circuito impreso, pues ello lo hace más práctico, es decir, lo convierte en un diseño que ha dejado de ser *de laboratorio*, presenta ya características que le permitirían ser producido en serie. El costo del material en el diseño hecho en circuito impreso fué de \$150,000 aproximadamente (en tabletas de experimentación el costo se eleva hasta unos \$250,000). Es oportuno mencionar que la fuente de voltaje elevaría el costo del sistema en unos \$30,000.

En un futuro, a este proyecto se le pueden adicionar distintas características como el de contar con una alarma, un control

remoto etc. O inclusive, se pueden utilizar esquemas semejantes a los de este diseño para hacer instrumentos de medición, control, seguridad etc., haciendo uso de sintetización de voz.

Además, este trabajo cumplió con el objetivo primario de toda tesis; pues nos hizo adquirir experiencia en cuanto a la investigación y diseño de un sistema utilizando algunos conocimientos adquiridos durante nuestros estudios universitarios. Esto nos alienta a seguir en el campo de la investigación y el diseño, tan necesarios en un país en desarrollo como el nuestro.

BIBLIOGRAFIA.

- 1 ENCICLOPEDIA BARSA, TOMO XIII, 1970.
- 2 BRITANNICA MACROPAEDIA, TOMO 28, 1986.
- 3 BRITANNICA MACROPAEDIA, TOMO 11, 1986.
- 4 REVISTA MUY INTERESANTE, AÑO 1, NUMERO 1.
- 5 REVISTA MUY INTERESANTE, AÑO 4, NUMERO 2.
- 6 REVISTA MUY INTERESANTE, AÑO 6, NUMERO 2.
- 7 ELECTRONICALLY HEARING: COMPUTER SPEECH RECOGNITION. JOHN P. CATER, 1984.
- 8 IEEE; VOCODERS: ANALYSIS-SYNTHESIS OF SPEECH. M.R. SCHROEDER, IEEE 54, 1966.
- 9 IEEE; DIGITAL REPRESENTATIONS OF SPEECH SIGNALS. P.W. SCHAFFER Y L.R. RABINER, ABRIL, 1975.
- 10 IEEE; CURRENT PERSPECTIVES IN DIGITAL SPEECH. P.W. GROCHIERE Y J.L. FLANAGAN, ENERO, 1983.
- 11 IEEE; AUTOMATIC SPEECH SYNTHESIS. D.O'SHAUGHNESSY, DICIEMBRE, 1983.
- 12 DIGITAL PROCESSING OF SPEECH SIGNALS. LAWRENCE R. RABINER Y RONALD W. SCHAFFER, 1978.
- 13 APUNTES DE LA CATEDRA DE DISEÑO DE SISTEMAS DIGITALES, FAC. DE INGENIERIA.
- 14 IEEE LINEAR PREDICTIVE CODING OF SPEECH: REVIEW AND CURRENT DIRECTIONS. MANFRED R. SCHROEDER, Agosto, 1985.
- 15 NATIONAL SEMICONDUCTOR CORPORATION-LINEAR

- DATA BOOK 1, 1988.
- 16 NATIONAL SEMICONDUCTOR CORPORATION-LINEAR
DATA BOOK 2, 1988.
- 17 NATIONAL SEMICONDUCTOR CORPORATION-
LS/S/TTL LOGIC DATABOOK, 1987.
- 18 CIRCUITOS INTEGRADOS LINEALES Y
AMPLIFICADORES OPERACIONALES. R.F.
COUGHLIN Y F.F. DRISCOLL, 1982.

APENDICE A.
HOJAS DE DATOS DEL C.I. SP0256



TECHNICAL DATA

AN EXCLUSIVE RADIO SHACK SERVICE TO THE EXPERIMENTER

SP0256 NARRATOR™ SPEECH PROCESSOR

Features

- Natural Speech
- Stand Alone Operation with Inexpensive Support Components
- Wide Operating Voltage
- Word, Phrase, or Sentence Library, ROM Expandable
- Expandable to 491K of ROM Directly
- Simple Interface to Most Microcomputers or Microprocessors
- Supports L.P.C. Synthesis: Formant Synthesis; Allophone Synthesis

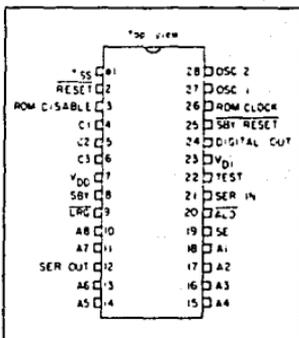
General Description

The SP0256 (Speech Processor) is a single chip N-Channel MOS LSI device that is able, using its stored program, to synthesize speech or complex sounds.

The achievable output is equivalent to a flat frequency response ranging from 0 to 5 kHz, a dynamic range of 42dB, and a signal to noise ratio of approximately 35dB.

The SP0256 incorporates four basic functions:

- A software programmable digital filter that can be made to model a VOCAL TRACT.
- A 16K ROM which stores both data and instructions (THE PROGRAM).
- A MICROCONTROLLER which controls the data flow from the ROM to the digital filter, the assembly of the "word strings" necessary for linking speech elements together, and the amplitude and pitch information to excite the digital filter.
- A PULSE WIDTH MODULATOR that creates a digital output which is converted to an analog signal when filtered by an external low pass filter.



PIN CONFIGURATION

verted to an analog signal when filtered by an external low pass filter.

Allophone Based Speech Processor — SP0256-AL2

One example of a preprogrammed SP0256 is the AL2 pattern.

Allophone Usage with a Microprocessor

The SP0256-AL2 requires the use of a processor to concatenate the speech sounds to form words.

The SP0256 is controlled using the address pins (A1-A8), ALD (Address Load), and SE (Strobe Enable). The object for controlling the chip is to load an address into it which contains the desired allophone. The speech data for the allophone set is contained within the internal 16K ROM of the SP0256-AL2.

CUSTOM PACKAGED IN U.S.A. BY RADIO SHACK A DIVISION OF TANDY CORPORATION

This particular application (Allophone Set) requires only six address pins (A1-A6) to address all the 59 allophones plus five pauses, a total of 64 locations. For simplicity, since only six address pins are needed to address the 64 locations, pins A7 and A8 can be tied low (to ground) and now any further references to the address bus will include A1-A6 and A7=AB-0.

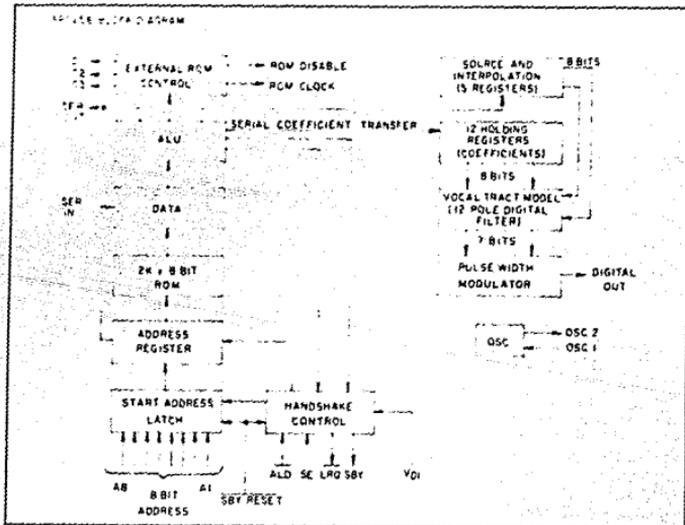
There are two modes available for loading an address into the chip. SE (Strobe Enable) controls the mode that will be used.

Mode 0 (SE=0) will latch in an address when any one or more of the address pins makes a low to high transition. For example, to load the address one (1), A2 to A6=0 and A1 is pulsed high. To load the address twelve (12 octal), A1 A3=A5=A6=0, A2 and A4 are pulsed high simultaneously. (Note that an address of zero cannot be loaded using this mode).

Mode 1 (SE=1) will latch in an address using the ALD pin. First, setup the desired address on the address bus (A1-A6)

and then pulse ALD low. Any address can be loaded using this mode, but certain setup and hold times are required (refer to the attached timing diagram for the specific times).

Two microprocessor interface pins are available for quick loading of addresses. They are LRQ and SBY. LRQ (Load Request) tells the processor when the input buffer is full. SBY (Stand By) tells the processor that the chip has stopped talking and no new address has been loaded. Either interface pin can be used when concatenating allophones. LRQ is an active low signal, when LRQ goes low it is time to load a new address to the chip. If LRQ is high, then simply wait for it to go low before loading the address. SBY will stay high until an address is loaded, then it will go low and stay low until all the internal instructions (Speech Code) from that one address are completed. Once this signal goes high, it is time to load a new address. Since speech does not require very fast address loading, it would be acceptable to use SBY to interface to the processor.



To end a word using allophones it is necessary to load a pause to complete the word. For example, the word "TWO"

can be implemented using the following allophones, TT2-VW2-PA1. PA1 is actually not an allophone but a pause which is needed to end the word.

ELECTRICAL CHARACTERISTICS

Maximum Ratings*

Allophone with reset: 0.3 to 8.0 V

Storage Temperature: -25 C to 125 C

Standard Conditions

Clock - Crystal Frequency: 3.120 MHz

Operating Temperature: 0 C to 70 C

DC CHARACTERISTICS/SP0 256

*Exceeding these ratings could cause permanent damage to the device. This is a stress rating only and functional operation of this device at these conditions is not implied. Operating ranges are specified in Standard Conditions. Exposure to absolute maximum rating conditions for extended periods may affect device reliability.

Data labeled "typical" is presented for design guidance only and is not guaranteed.

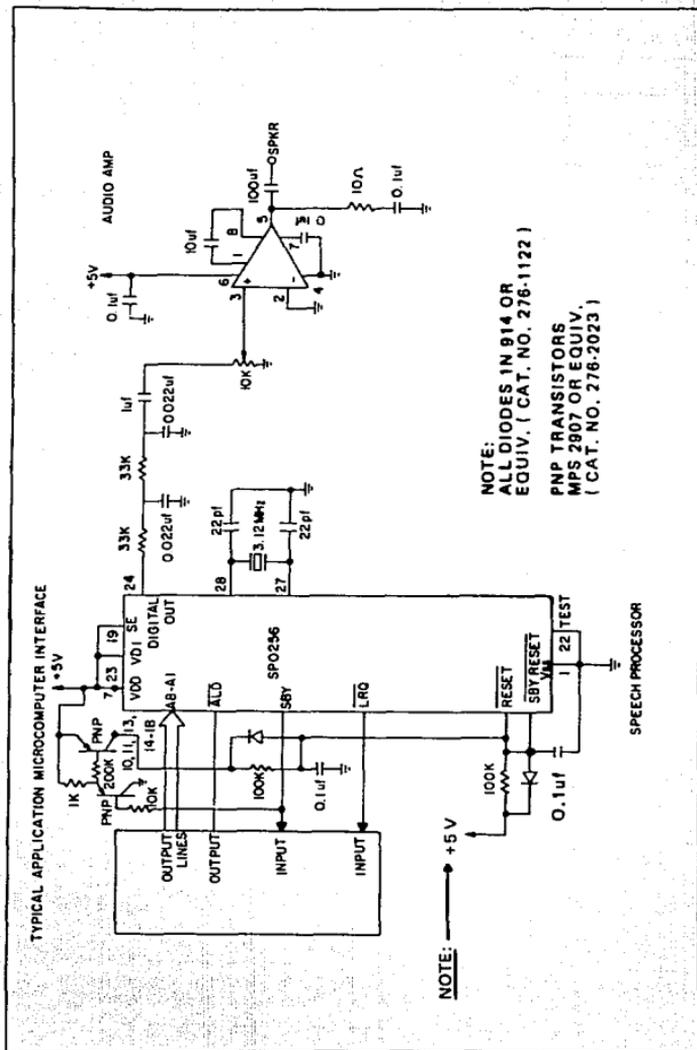
Characteristic	Sym	Min	Typ	Max	Units	Conditions
Supply Voltage	V_{DD}	4.6	—	7.0	V	
	V_{SS}	4.6	—	7.0	V	
Supply Current	I_{CC}	—	—	90	mA	$T_A = 25 C$, $V_{DD} = 7.0 V$ Reset & SBY Reset high All outputs floating
	I_{DD}	—	—	21	mA	Same as above
INPUTS						
A1 A5 \overline{ALD} SERIN TEST SE						
LOGIC 0	V_{IL}	0.0	—	0.6	V	
LOGIC 1	V_{IH}	2.4	—	V_{DD}	V	
CAPACITANCE	C_{in}	—	—	10	pf	0 Volts bias, $f = 3.12 MHz$
LEAKAGE	I_{L}	—	—	-10	μA	$V_{DD} = 7.0 V$ Other Pins = 0.0 V
RESET SBY RESET						
LOGIC 0	V_{IL}	0.0	—	0.6	V	
LOGIC 1	V_{IH}	3.6	—	V_{DD}	V	
OUTPUTS						
SBY Dig 13 Out C1 C2 C3						
\overline{LRQ} ROM DIS ROM CLK						
SEROUT						
LOGIC 0	V_{OL}	0.0	—	0.6	V	$I_{OL} = 0.72ma$ (2LS TTL Loads)
LOGIC 1	V_{OH}	2.5	—	V_{DD}	V	$I_{OH} = -50 \mu A$ (2LS TTL Loads)
OSCILLATOR						
OSC 2 (Output)						When driven from external source
LOGIC 0	V_{OL}	0.0	—	0.6	V	OSC 1 (Input) = 3.90 V MIN
LOGIC 1	V_{OH}	2.5	—	V_{DD}	V	OSC 1 (Input) = 0.60 V MAX

AC CHARACTERISTICS	SYM	MIN	TYP	MAX	UNITS	CONDITIONS
$\overline{\text{ALD}}$	t_{pw1}	200	—	1100	ns	$200 \leq \overline{\text{ALD}} \leq 1100$ ns
A1-A8 Set Up	t_{s1}	0	—	—	ns	
Hold	t_{h1}	160	—	—	ns	
$\overline{\text{LRQ}}$	t_{pd1}	—	—	300	ns	
SBY	t_{pd2}	—	—	300	ns	

AC CHARACTERISTICS	SYM	MIN	TYP	MAX	UNITS	CONDITIONS
$\overline{\text{ALD}}$	t_{pw2}	1100	—	—	ns	$\overline{\text{ALD}} > 1100$ ns
A1-A8 Set Up	t_{s2}	0	—	—	ns	
Hold	t_{h2}	1120	—	—	ns	
$\overline{\text{LRQ}}$	t_{pd1}	—	—	300	ns	
SBY	t_{pd2}	—	—	300	ns	

AC CHARACTERISTICS	SYM	MIN	TYP	MAX	UNITS	CONDITIONS
Clock Frequency	F	—	3.120	—	MHz	Crystal Oscillator
Clock Duty Cycle	—	48	—	52	%	driven from
Reset	t_{pw3}	100	—	—	us	external.
SBY Reset	t_{pw4}	100	—	—	us	

ADDENDUM



PIN FUNCTIONS

PIN NUMBER	NAME	FUNCTION
1	V _{SS}	Ground
2	RESET	A logic 0 resets that portion of the SP powered by V _{DD} . Must be returned to a logic 1 for normal operation.
3	ROM DISABLE	For use with an external serial speech ROM, a logic 1 disables the external ROM.
4, 5, 6	C1, C2, C3	Output control lines for use with an external serial speech ROM. Refer to the SPRO16 Data Sheet for details.
7	V _{DD}	Power supply for all portions of the SP except the microprocessor interface logic.
8	SBY	STANDBY. A logic 1 output indicates that the SP is inactive and V _{DD} can be powered down externally to conserve power. When the SP is reactivated by an address being loaded, SBY will go to a logic 0.
9	LRQ	LOAD REQUEST. LRQ is a logic 1 output whenever the input buffer is full. When LRQ goes to a logic 0, the input port may be loaded by placing the 8 address bits on A1-A8 and pulsing the ALD output.
10, 11, 13, 14, 15, 16, 17, 18	A8, A7, A6, A5, A4, A3, A2, A1	8 bit address which defines any one of 256 speech entry points.
12	SER OUT	SERIAL ADDRESS OUT. This output transfers a 16-bit address serially to an external speech ROM.
19	SE	STROBE ENABLE. Normally held in a logic 1 state. When tied to ground, ALD is disabled and the SP will automatically latch in the address on the input bus approximately 1 μ s after detecting a logic 1 on any address line.
20	ALD	ADDRESS LOAD. A negative pulse on this input loads the 8 address bits into the input port. The negative edge of this pulse causes LRQ to go high.
21	SER IN	SERIAL IN. This is an 8-bit serial data input from an external speech ROM.

Pin Functions Continued

PIN NUMBER	NAME	FUNCTION
22	TEST	This pin should be grounded for normal operation.
23	VD1	Power supply for the microprocessor interface logic and controller.
24	DIGITAL OUT	Pulse width modulated digital speech output which, when filtered by a 5KHz low pass filter and amplified, will drive a loudspeaker.
25	SBY RESET	STANDBY RESET. A logic 0 resets the microprocessor interface logic and the address latches. Must be returned to a logic 1 for normal operation.
26	ROM CLOCK	This is a 1.56MHz clock output used to drive an external serial speech ROM.
27	OSC1	XTAL IN. Input connection for a 3.12MHz crystal.
28	OSC2	XTAL OUT. Output connection for a 3.12MHz crystal.

ALLOPHONE SPEECH SYNTHESIS

Introduction

The General Instrument allophone speech synthesis technique provides the user with the ability to synthesize an unlimited vocabulary at a very low bit rate. Fifty-nine discrete speech sounds (called allophones) and five pauses are stored at different addresses in the SP0256 internal ROM. Each speech sound was excised from a word and analyzed using linear predictive coding (LPC). Any English word or phrase can be created by addressing the appropriate combination of allophones and pauses. Since there is a total of 64 address locations each requires a 6 bit address. Assuming that speech contains 10 to 12 sounds per second, allophone synthesis requires addressing less than 100 bits per second.

Linguistics

A few basic linguistic concepts will help you start your own library of "allophone words". (See Table 1 for the General Instrument Allophone Dictionary). First, there is no one-to-one correspondence between written letters and speech sounds; secondly, speech sounds are acoustically different depending upon their position within a word; and lastly, the human ear may perceive the same acoustic signal differently in the context of different sounds.

The first point compares to the problem that a child encounters when learning to read. Each sound in a language may be represented by more than one letter and, conversely each letter may represent more than one sound. (See the examples in Table 2.) Because of these spelling irregularities, it is necessary to think in terms of sounds, not letters, when using allophones.

The second, and equally important, point to understand, is that the acoustic signal of a speech sound may differ depending upon its position within a word. For example, the initial K sound in *coop* will be acoustically different from the K's in *keep* and *speak*. The K's in *coop* and *keep* differ due to the influence of the vowels which follow them, and the final K in *speak* is usually not as loud as initial K's.

Finally, a listener may identify the same acoustic signal differently depending on the context in which it is perceived. Don't be surprised, therefore, if an allophone word sounds slightly different when used in various phrases.

Phonemes Of English

The sounds of a language are called phonemes, and each language has a set which is slightly different from that of other languages. Table 3 contains a chart of all the consonant phonemes of English, Table 4 all the vowel phonemes.

Consonants are produced by creating an occlusion or constriction in the vocal tract which produces an aperiodic sound source. If the vocal cords are vibrating at the same time, as in the case of the voiced fricatives VV, DH, ZZ, and ZH, (See Table 5) there are two sound sources: one which is aperiodic and one which is periodic.

Vowels are usually produced with a relatively open vocal tract and a periodic sound source provided by the vibrating vocal cords. They are classified according to whether the front or back of the tongue is high or low (See Table 4 whether they are long or short, and whether the lips are rounded or unrounded). In English all rounded vowels are produced in or near the back of the mouth (UW, UH, OW, AO, OR, AW).

Speech sounds which have features in common behave in similar ways. For example, the voiceless stop consonants PP, TT, and KK (See Table 3) should be preceded by 50-80 msec of silence, and the voiced stop consonants BB, DD, and GG by 10-30 msec of silence.

Allophones

Phoneme is the name given to a group of similar sounds in a language. Recall that a phoneme is acoustically different depending upon its position within a word. Each of these positional variants is an allophone of the same phoneme. An allophone, therefore, is the manifestation of a phoneme in true speech signal. It is for this reason that our inventory of English speech sounds is called an allophone set.

How To Use The Allophone Set

(See Table 1 for instructions on how to create all the sample words mentioned in this section.) The allophone set (Refer to Table 5) contains two or three versions of some phonemes. It may be necessary to use one allophone of a particular phoneme for word-or-syllable-final position. A detailed set of guidelines for using the allophones is given in Table 5. Note that these are suggestions, not rules.

For example, DD2 sounds good in initial position and DD1 sounds good in final position, as in "daughter" and "collide". One of the differences between the initial and final versions of a consonant is that an initial version may be longer than the final version. Therefore, to create an initial SS, you can use two SSs instead of the usual single SS at the end of a word or syllable, as in "sister". Note that this can be done with TH, and FF, and the inherently short vowels (to be discussed below), but with no other consonants. You will want to experiment with some consonants such as str, cl) to discover which version works best in the cluster. For example, KK1 sounds good before LL as in "clown", and KK2 sounds good before WW as in "square". One allophone of a particular phoneme may sound better before or after back vowels and another before or after front vowels. KK3 sounds good before UH and KK1 sounds good before IY, as in "cookie". Some sounds (PP, BB, TT, DD, KK, GG, CH, and JH) require a brief duration of silence before them. For most of these, the silence has already been added but you may decide you want to add more. Therefore there are several pauses included in the

allophone set varying from 10-200 msec. To create the final sounds in the words "letter" and "little" use the allophones ER and EL.

Remember that you must always think about how a word sounds, not how it is spelled. For example, the NG sound is represented by the letter N in "uncle". And remember that some sounds may not even be represented in words by any letters, as the YY in "computer".

As mentioned earlier there are some vowels which can be doubled to make longer versions for stressed syllables. These are the inherently short vowels IH, EH, AE, AX, AA, and UH. For example, in the word "extent" use one EH in the first syllable, which is unstressed and two EHs in the second syllable which is stressed. Of the inherently long vowels there is one, UW, which has a long and

short version. The short one, UW1, sounds good after YY in computer. The long version, UW2, sounds good in monosyllabic words like "two". Included in the vowel set is a group called R-colored vowels. These are vowel + R combinations. For example, the AR in "alarm" and the OR in "score". Of the R-colored vowels there is one, ER, which has a long and short version. The short version is good for polysyllabic words with final ER sounds like "letter", and the long version is good for monosyllabic words like "fir". One final suggestion is that you may want to add a pause of 30-50 msec between words, when creating sentences, and a pause of 100-200 msec between clauses.

Note: Every utterance must be followed by a pause in order to make the chip stop talking the last allophone.

Table 1:

NUMBERS:			
zero	ZZ YR OW	seventeen	SS SS EH VV TH NN1 PA2 PA3 TT2 IY NN1
one, won	WW SX AX NN1	eighteen	EY PA2 PA3 TT2 IY NN1
two, to, too	TT2 UW2	nineteen	NN1 AY NN1 PA2 PA3 TT2 IY NN1
three	TH RR1 IY	twenty	TT2 WH EH EH NN1 PA2 PA3 TT2 IY
four, for, fore	FF FF OR	thirty	TH ER2 PA2 PA3 TT2 IY
five	FF FF AY VV	forty	FF OR PA3 TT2 IY
six	SS SS IH IH PA3 KK2 SS	fifty	FF FF IH FF FF PA2 PA3 TT2 IY
seven	SS SS EH EH VV IH NN1	sixty	SS SS IH PA3 KK2 SS PA2 PA3 TT2 IY
eight, ate	EY PA3 TT2	seventy	SS SS EH VV IH NN1 PA2 PA3 TT2 IY
nine	NN1 AA AY NN1	eighty	EY PA3 TT2 IY
ten	TT2 EH EH NN1	ninety	NN1 AY NN1 PA3 TT2 IY
eleven	IH LL EH EH VV IH NN1	hundred	HH2 AX AX NN1 PA2 DD2 RR2 IH IH PA1 DD1
twelve	TT2 WH EH EH LL VV	thousand	TH AA AW ZZ TH PA1 PA1 NN1 DD1
thirteen	TH ER1 PA2 PA3 TT2 IY NN1	million	MM IH IH LL YY1 AX NN1
fourteen	FF OR PA2 PA3 TT2 IY NN1		
fifteen	FF IH FF PA2 PA3 TT2 IY NN1		
sixteen	SS SS IH PA3 KK2 SS PA2 PA3 TT2 IY NN1		

Table 1 Continued

DAY OF THE WEEK:

Sunday	SS SS AX AX NN1 PA2 DD2 EY	K L M N O P Q R S T U V W	KK1 EH EY EH EH EL EH EH MM EH EH NN1 OW PP IY KK1 YY1 UW2 AR EH EH SS SS TT2 IY YY1 UW2 VV IY DD2 AX PA2 BB2 EL YY1 UW2
Monday	MM AX AX NN1 PA2 DD2 EY	X Y Z	EH EH PA3 KK2 SS SS WW AY ZZ IY
Tuesday	TT2 UW2 ZZ PA2 DD2 EY		
Wednesday	WW EH EH NN1 ZZ PA2 DD2 EY		
Thursday	TH ER2 ZZ PA2 DD2 EY		
Friday	FF RR2 AY PA2 DD2 EY		
Saturday	SS SS AE PA3 TT2 PA2 DD2 EY		

MONTHS:

January	JH AE AE NN1 YY2 XR IY
February	FF EH EH PA1 BR RR2 UW2 XR IY
March	MM AR PA3 CH
April	EY PA3 PP RR2 IH IH LL
May	MM EY
June	JH UW2 NN1
July	JH UW1 LL AY
August	AO AO PA2 GG2 AX SS PA3 TT1
September	SS SS EH PA3 PP PA3 TT2 EH EH PA1 BB2 ER1
October	AA PA2 KK2 PA3 TT2 OW PA1 BB2 ER1
November	NN2 OW VV EH EH MM PA1 BB2 ER1
December	DD2 IY SS SS EH EH MM PA1 BB2 ER1

LETTERS:

A	EY
B	BB2 IY
C	SS SS IY
D	DD2 IY
E	IY
F	EH EH FF FF
G	JH IY
H	EY PA2 PA3 CH
I	AA AY
J	JH EH EY

DICTIONARY:

alarm	AX LL AR MM
bathe	BB2 EY DH2
bather	BB2 EY DH2 ER1
bathing	BB2 EY DH2 IH NG
beer	BB2 YR
bread	BB1 RR2 EH EH PA1 DD1
by	BB2 AA AY
calendar	KK1 AE AE LL EH NN1 PA2 DD2 ER1
clock	KK1 LL AA AA PA3 KK2
clown	KK1 LL AW NN1
check	CH EH EH PA3 KK2
checked	CH EH EH PA3 KK2 PA2 TT2
checker	CH EH EH PA3 KK1 ER1
checkers	CH EH EH PA3 KK1 ER1 ZZ
checking	CH EH EH PA3 KK1 IH NG
checks	CH EH EH PA3 KK1 SS
cognitive	KK3 AA AA GG3 NN1 IH PA3 TT2 IH VV
collide	KK3 AX LL AY DD1
computer	KK1 AX MM PP1 YY1 UW1 TT2 ER
cookie	KK3 UH KK1 IY

coop	KK3 UW2 PA3 PP	fir	FF ER2
correct	KK1 ER2 EH EH PA2 KK2 PA2 TT1	freeze	FF FF RR1 IY ZZ
corrected	KK1 ER2 EH EH PA2 KK2 PA2 TT2 IH PA2 DD1	freezer	FF FF RR1 IY ZZ ER1
correcting	KK1 ER2 EH EH PA2 KK2 PA2 TT2 IH NG	freezers	FF FF RR1 IY ZZ ER1 ZZ
corrects	KK1 ER2 EH EH PA2 KK2 PA2 TT1 SS	freezing	FF FF RR1 IY ZZ IH NG
crown	KK1 RR2 AW NN1	frozen	FF FF RR1 OW ZZ EH NN1
date	DD2 EY PA3 TT2	gauge	GG1 EY PA2 JH
daughter	DD2 AO TT2 ER1	gauged	GG1 EY PA2 JH PA2 DD1
day	DD2 EH EY	gauges	GG1 EY PA2 JH IH ZZ
divided	DD2 IH VV AY PA2 DD2 IH PA2 DD1	gauging	GG1 EY PA2 JH IH NG
emotional	IY MM OW SH AX NN1 AX EL	hello	HH EH LL AX OW
engage	EH EH PA1 NN1 GG1 EY PA2 JH	hour	AW ER1
engagement	EH EH PA1 NN1 GG1 EY PA2 JH MM EH EH NN1 PA2 PA3 TT2	infinitive	IH NN1 FF FF IH IH NN1 IH PA2 PA3 TT2 IH VV
engages	EH EH PA1 NN1 GG1 EY PA2 JH IH ZZ	intrigue	IH NN1 PA3 TT2 RR2 IY PA1 GG3
engaging	EH EH PA1 NN1 GG1 EY PA2 JH IH NG	intrigued	IH NN1 PA3 TT2 RR2 IY PA1 GG3 PA2 DD1
enrage	EH NN1 RR1 EY PA2 JH	intrigues	IH NN1 PA3 TT2 RR2 IY PA1 GG3 ZZ
enraged	EH NN1 RR1 EY PA2 JH PA2 DD1	intriguing	IH NN1 PA3 TT2 RR2 IY PA1 GG3 IH NG
enrages	EH NN1 RR1 EY PA2 JH IH ZZ	investigate	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT2
enraging	EH NN1 RR1 EY PA2 JH IH NG	investigated	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT2 IH PA2 DD1
escape	EH SS SS PA3 KK1 PA2 PA3 PP	investigator	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT2 ER1
escaped	EH SS SS PA3 KK1 PA2 PA3 PP PA2 TT2	investigators	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT2 ER1 ZZ
escapes	EH SS SS PA3 KK1 PA2 PA3 PP SS	investigates	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT1 SS
escaping	EH SS SS PA3 KK1 PA2 PA3 PP IH NG		
equal	IY PA2 PA3 KK3 WH AX EL		
equals	IY PA2 PA3 KK3 WH AX EL ZZ		
error	EH XR OR		
extent	EH KK1 SS TT2 EH EH NN1 TT2		

Table 1 Continued

investigating	IH IH NN1 VV EH EH SS PA2 PA3 TT2 IH PA1 GG1 EY PA2 TT2 IH NG KK1 IY	pledging	PP LL EH EH PA3 JH IH NG
key	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 TT2	plus	PP LL AX AX SS SS
legislate	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 TT2	ray	RR1 EH EY
legislated	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 TT2 IH DD1	rays	RR1 EH EY ZZ
legislates	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 TT1 SS	ready	RR1 EH EH PA1 DD2 IY
legislating	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 TT2 IH NG	red	RR1 EH FH PA1 DD1
legislature	LL EH EH PA2 JH JH SS SS LL EY PA2 PA3 CH ER1	robot	RR1 OW PA2 BB2 AA PA3 TT2
letter	LL EH EH PA3 TT2 ER1	robots	RR1 OW PA2 BB2 AA PA3 TT1 SS
litter	LL IH IH PA3 TT2 ER1	score	SS SS PA3 KK3 OR
little	LL IH IH PA3 TT2 EL	second	SS SS EH PA3 KK1 IH NN1 PA2 DD1
memory	MM EH EH MM ER2 IY	sensitive	SS SS EH EH NN1 SS SS IH PA2 PA3 TT2 IH VV
memories	MM EH EH MM ER2 IY ZZ	sensitivity	SS SS EH EH NN1 SS SS IH PA2 PA3 TT2 IH VV IH PA2 PA3 TT2 IY
minute	MM IH NN1 IH PA3 TT2	sincere	SS SS IH IH NN1 SS SS YR
month	MM AX NN1 TH	sincerely	SS SS IH IH NN1 SS SS YR LL IY
nip	NN1 IH IH PA2 PA3 PP	sincerity	SS SS IH IH NN1 SS SS EH EH RR1 IH PA2 PA3 TT2 IY
nipped	NN1 IH IH PA2 PA3 PP PA3 TT2	sister	SS SS IH IH SS PA3 TT2 ER1
nipping	NN1 IH IH PA2 PA3 PP IH NG	speak	SS SS PA3 IY PA3 KK2
nips	NN1 IH IH PA2 PA3 PP SS	spell	SS SS PA3 PP EH EH EL
no	NN2 AX OW	spelled	SS SS PA3 PP EH EH EL PA3 DD1
physical	FF FF IH ZZ IH PA3 KK1 AX EL	speller	SS SS PA3 PP EH EH EL ER2
pin	PP IH IH NN1	spellers	SS SS PA3 PP EH EH EL ER2 ZZ
pinned	PP IH IH NN1 PA2 DD1	spelling	SS SS PA3 PP EH EH EL IH NG
pinning	PP IH IH NN1 IH NG1	spells	SS SS PA3 PP EH EH EL ZZ
pins	PP IH IH NN1 ZZ	start	SS SS PA3 TT2 AR PA3 TT2
pledge	PP LL EH FH PA3 JH	started	SS SS PA3 TT2 AR PA3 TT2 IH PA1 DD2
pledged	PP LL EH EH PA3 JH PA2 DD1	starter	SS SS PA3 TT2 AR PA3 TT2 ER1
pledges	PP LL EH EH PA3 JH IH ZZ		

starting SS SS PP3 TT2 AR
 PA3 TT2 IH NG
starts SS SS PP3 TT2 AR
 PA3 TT1 SS
stop SS SS PA3 TT1 AA
 AA PA3 PP
stopped SS SS PA3 TT1 AA
 AA PA3 PP PA3 TT2
stopper SS SS PA3 TT1 AA
 AA PA3 PP ER1
stopping SS SS PA3 TT1 AA
 AA PA3 PP IH NG
stops SS SS PA3 TT1 AA
 AA PA3 PP SS
subject (noun) SS SS AX AX PA2
 BB1 PA2 JH EH PA3
 KK2 PA3 TT2
subject (verb) SS SS AX PA2 BB1
 PA2 JH EH EH PA3
 KK2 PA3 TT2
sweat SS SS WW EH EH
 PA3 TT2
sweated SS SS WW EH EH
 PA3 TT2 IH PA3
 DD1
sweater SS SS WW EH EH
 PA3 TT2 ER1
sweaters SS SS WW EH EH
 PA3 TT2 ER1 ZZ
sweating SS SS WW EH EH
 PA3 TT2 IH NG
sweats SS SS WW EH EH
 PA3 TT2 SS
switch SS SS WH IH IH
 PA3 CH
switched SS SS WH IH IH
 PA3 CH PA3 TT2
switches SS SS WH IH IH
 PA3 CH IH ZZ2
switching SS SS WH IH IH
 PA3 CH IH NG2
system SS SS IH IH SS SS
 PA3 TT2 EH MM
systems SS SS IH IH SS SS
 PA3 TT2 EH MM ZZ
talk TT2 AO AO PA2
 KK2
talked TT2 AO AO PA3
 KK2 PA3 TT2
talker TT2 AO AO PA3
 KK1 ER1
talkers TT2 AO AO PA3
 KK1 ER1 ZZ
talking TT2 AO AO PA3
 KK1 IH NG
talks TT2 AO AO PA2
 KK2 SS

thread TH RR1 EH EH
 PA2 DD1
threaded TH RR1 EH EH
 PA2 DD2 IH PA2
 DD1
threader TH RR1 EH EH
 PA2 DD2 ER1
threaders TH RR1 EH EH
 PA2 DD2 ER1 ZZ
threading TH RR1 EH EH
 PA2 DD2 IH NG
threads TH RR1 EH EH
 PA2 DD2 ZZ
then DH1 EH EH NN1
time TT2 AA AY MM
times TT2 AA AY MM ZZ
uncle AX NG PA3 KK3 EL
whale WW EY EL
whaler WW EY LL ER1
whalers WW EY LL ER1 ZZ
whales WW EY EL ZZ
whaling WW EY LL TH NG
year YY2 YR
yes YY2 EH EH SS SS

TABLE 2 – EXAMPLES OF SPELLING IRREGULARITIES

	Same sound represented by different letters	Different sounds represented by the same letters
Vowels	mEAt	vEIn
	fEEt	forElgn
	pEte	dElsm
	pEOple	dElcer
	pennY	gElsha
Consonants	SHip	althouGH
	tenSton	GHastly
	preClous	couGH
	naTion	hiccouGH

TABLE 3 — CONSONANT PHONEMES OF ENGLISH**

		LABIAL	LABIO-DENTAL	INTER-DENTAL	ALVEO-DALAR	PALATAL	VELAR	GLOTTAL
Stops:	Voiceless	PP			TT		KK	
	Voiced	BB			DD		GG	
Fricatives:	Voiceless	WH	FF	TH	SS	SH		HH
	Voiced		VV	DH	ZZ	ZH*		
Affricates:	Voiceless					CH		
	Voiced					JH		
Nasals:	Voiced	MM			NN		NG*	
Resonants:	Voiced	WW			RR,LL	YY		

*These do not occur in word-initial position in English.

Labial: Upper and Lower Lips
Touch or Approximate

Labio-Dental: Upper Teeth and Lower Lip Touch

Inter-Dental: Tongue Between Teeth

Alveolar: Tip of Tongue Touches or Approximates Alveolar Ridge (just behind upper teeth)

Palatal: Body of Tongue Approximates Palate (roof of mouth)

Velar: Body of Tongue Touches Velum (posterior portion of roof of mouth)

Glottal: Glottis (opening between vocal cords)

TABLE 4 — VOWEL PHONEMES OF ENGLISH

	FRONT	CENTRAL	BACK
High	YR IY IH*		UW# UH* #
Mid	EY EH* XR	ER AX*	OW# OY#
Low	AE*	AW# AY AR AA*	AO* # OR#

* Short Vowels

Rounded Vowels

TABLE 5 — GUIDELINES FOR USING THE ALLOPHONES

Silence

- PA1 (10 ms) — before BB, DD, GG, and JH
- PA2 (30 ms) — before BB, DD, GG, and JH
- PA3 (50 ms) — before PP, TT, KK, and CH, and between words
- PA4 (100 ms) — between clauses and sentences
- PA5 (200 ms) — between clauses and sentences

Short Vowels

- */IH/ — sitting, stranded
- */EH/ — extent, gentlemen
- */AE/ — extract, acting
- */UH/ — cookie, full
- */AO/ — talking, song
- */AX/ — lapel, instruct
- */AA/ — pottery, cotton

Long Vowels

- /IY/ — treat, people, penny
- /EY/ — great, statement, tray
- /AY/ — kite, sky, mighty
- /OY/ — noise, toy, voice
- /UW1/ — after clusters with YY: computer
- /UW2/ — in monosyllabic words: two, food
- /OW/ — zone, close, snow
- /AW/ — sound, mouse, down
- /EL/ — little, angle, gentlemen

R-Colored Vowels

- /ER1/ — letter, furniture, interrupt
- /ER2/ — monosyllables: bird, fern, burn
- /OR/ — fortune, adorn, store
- /AR/ — farm, alarm, garment
- /YR/ — hear, earring, irresponsible
- /XR/ — hair, declare, stare

Resonants

- /WW/ — we, warrant, linguist
- /RR1/ — initial position: read, write, x-ray
- /RR2/ — initial clusters: brown, crane, grease
- /LL/ — like, hello, steel
- /YY1/ — clusters: cute, beauty, computer
- /YY2/ — initial position: yes, yarn, yo-yo

Voiced Fricatives

- /VV/ — vest, prove, even
- /DH1/ — word-initial position: this, then, they
- /DH2/ — word-final and between vowels: bathe, bathing
- /ZZ/ — zoo, phase
- /ZH/ — beige, pleasure

Voiceless Fricatives

- */FF/ —) These may be doubled for initial position and used singly in final position
- */TH/ —)
- */SS/ —)
- /SH/ — shirt, leash, nation
- /HH1/ — before front vowels: YR, IY, IH, EY, EH, XR, AE
- /HH2/ — before back vowels: UW, UH, OW, OY, AO, OR, AR
- /WH/ — white, whim, twenty

Voiced Stops

- /BB1/ — final position: rib; between vowels: fiber, in clusters: bleed, brown
- /BB2/ — initial position before a vowel: beast
- /DD1/ — final position: played, end
- /DD2/ — initial position: down; clusters: drain
- /GG1/ — before high front vowels: YR, IY, IH, EY, EH, XR
- /GG2/ — before high back vowels: UW, UH, OW, OY, AX; and clusters: green, glue
- /GG3/ — before low vowels: AE, AW, AY, AR, AA, AO, OR, ER; and medial clusters: anger; and final position: peg

Voiceless Stops

- /PP/** – pleasure, ample, trip
/TT1/ – final clusters before SS: tests
its
/TT2/ – all other positions: test, street
/KK1/ – before front vowels: YR, IY,
IH, EY, EH, XR, AY, AE,
ER, AX; initial clusters: cute,
clown, scream
/KK2/ – final position: speak; final
clusters: task
/KK3/ – before back vowels: UW, UH,
OW, OY, OR, AR, AD; initial
clusters: crane, quick, clown,
scream

Affricates

- /CH/** – church, feature
/JH/ – judge, injure

Nasal

- /MM/** – milk, alarm, ample
/NN1/ – before front and central vowels:
YR, IY, IH, EY, EH,
XR, AE, ER, AX, AW, AY,
UW; final clusters: earn
/NN2/ – before back vowels: UH, OW,
OY, OR, AR, AA
/NG/ – string, anger

* These allophones can be doubled.

TABLE 6 - ALLOPHONE ADDRESS TABLE

DECIMAL ADDRESS	OCIAL ADDRESS	ALLOPHONE	SAMPLE WORD	DURATION	DECIMAL ADDRESS	OCIAL ADDRESS	ALLOPHONE	SAMPLE WORD	DURATION
0	000	PA1	PAUSE	10MS	32	040	/AW/	Out	370MS
1	001	PA2	PAUSE	30MS	33	041	/DD2/	Do	160MS
2	002	PA3	PAUSE	50MS	34	042	/GG3/	Wig	140MS
3	003	PA4	PAUSE	100MS	35	043	/VV/	Vest	190MS
4	004	PA5	PAUSE	200MS	36	044	/GG1/	Got	80MS
5	005	/OY/	Boy	420MS	37	045	/SH/	Ship	160MS
6	006	/AY/	Sky	260MS	38	046	/ZH/	Azure	190MS
7	007	/EH/	End	70MS	39	047	/RR2/	Brain	120MS
8	010	/KK3/	Comb	120MS	40	050	/FF/	Food	150MS
9	011	/PP/	Pow	210MS	41	051	/KK2/	Sky	190MS
10	012	/JH/	Dodge	140MS	42	052	/KK1/	Can't	160MS
11	013	/NN1/	Thin	140MS	43	053	/ZZ/	Zoo	210MS
12	014	/IH/	Sit	70MS	44	054	/NG/	Anchor	220MS
13	015	/TT2/	To	140MS	45	055	/LL/	Lake	110MS
14	016	/RR1/	Rural	170MS	46	056	/WW/	Wool	180MS
15	017	/AX/	Succeed	70MS	47	057	/XR/	Repair	360MS
16	020	/MM/	Milk	180MS	48	060	/WH/	Whig	200MS
17	021	/TT1/	Part	100MS	49	061	/YY1/	Yes	130MS
18	022	/DH1/	They	290MS	50	062	/CH/	Church	190MS
19	023	/IY/	See	250MS	51	063	/ER1/	Fir	160MS
20	024	/EY/	Beige	280MS	52	064	/ER2/	Fir	300MS
21	025	/DD1/	Could	70MS	53	065	/OW/	Beau	240MS
22	026	/UW1/	To	100MS	54	066	/DH2/	They	240MS
23	027	/AO/	Aught	100MS	55	067	/SS/	Vest	90MS
24	030	/AA/	Hot	100MS	56	070	/NN2/	No	190MS
25	031	/YY2/	Yes	180MS	57	071	/HH2/	Hoe	180MS
26	032	/AE/	Hat	120MS	58	072	/OR/	Store	330MS
27	033	/HH1/	He	130MS	59	073	/AR/	Alarm	290MS
28	034	/BB1/	Business	80MS	60	074	/YR/	Clear	350MS
29	035	/TH/	Thin	180MS	61	075	/GG2/	Guest	40MS
30	036	/UH/	Book	100MS	62	076	/EL/	Saddle	190MS
31	037	/UW2/	Food	260MS	63	077	/BB2/	Business	50MS

NOTES

RADIO SHACK, A DIVISION OF TANDY CORPORATION

U.S.A.: FORT WORTH, TEXAS 76102
CANADA: BARRIE, ONTARIO L4M 4W5

TANDY CORPORATION

AUSTRALIA

91 KURRAJONG ROAD
MOUNT DRIHTT N S W 2770

BELGIUM

PARC INDUSTRIEL DE NANINNE
5140 NANINNE

U K

BILSTON ROAD WEDNESBURY
WEST MIDLANDS WS10 7JN

484

Printed in USA

223

APENDICE B
HOJAS DE DATOS DEL C.I. MM53II



MM5309, MM5311, MM5312, MM5313, MM5314, MM5315 digital clocks

general description

These digital clocks are monolithic MOS integrated circuits utilizing P-channel low-threshold, enhancement mode and ion implanted depletion mode devices. The devices provide all the logic required to build several types of clocks. Two display modes (4 or 6-digits) facilitate end-product designs of varied sophistication. The circuits interface to LED and gas discharge displays with minimal additional components, and require only a single power supply. The timekeeping function operates from either a 50 or 60 Hz input, and the display format may be either 12 hours (with leading-zero blanking) or 24 hours. Outputs consist of multiplexed display drives (BCD and 7-segments) and digit enables. The devices operate over a power supply range of 11V to 19V and do not require a regulated supply. These clocks are packaged in dual-in-line packages.

features

- 50 or 60 Hz operation
- 12 or 24-hour display format

- Leading zero blanking (12 hour format)
- 7-segment outputs
- Single power supply
- Fast and slow set controls
- Internal multiplex oscillator
- For features of individual clocks, see Table 1

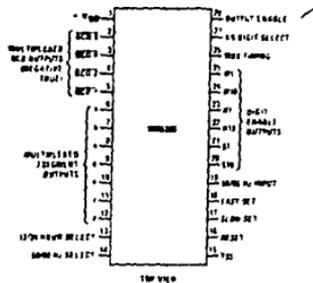
applications

- Desk clocks
- Automobile clocks
- Industrial clocks
- Interval Timers

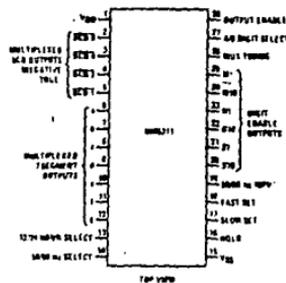
TABLE 1

FEATURES	MM5309	MM5311	MM5312	MM5313	MM5314	MM5315
BCD Outputs	X	X	X	X		X
4 or 6 Digit Display Mode	X	X		X	X	X
MM53 Count Control		X		X	X	X
1 Hz Output	X		X	X		
Output Enable Control	X	X			X	
Reset	X					X

connection diagrams (Dual In-Line Packages)



Order Number MM5309N
See Package 23



Order Number MM5311N
See Package 23

absolute maximum ratings

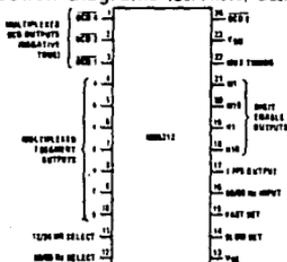
Voltage at Any Pin
 Operating Temperature
 Storage Temperature
 Lead Temperature (Soldering, 10 seconds)

$V_{SS} + 0.3$ to $V_{SS} - 20V$
 $-25^{\circ}C$ to $+70^{\circ}C$
 $-65^{\circ}C$ to $+150^{\circ}C$
 $300^{\circ}C$

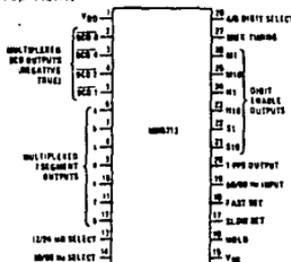
electrical characteristics T_A within operating range, $V_{SS} = 11V$ to $19V$, $V_{DD} = 0V$, unless otherwise specified.

PARAMETER	CONDITIONS	MIN	TYP	MAX	UNITS
Power Supply Voltage	$V_{SS} (V_{DD} = 0V)$	11		19	V
Power Supply Current	$V_{SS} = 14V$, (No Output Loads)			10	mA
50/60 Hz Input Frequency		dc	50 or 60	60k	Hz
50/60 Hz Input Voltage					
Logical High Level		$V_{SS}-1$	V_{SS}	V_{SS}	V
Logical Low Level		V_{DD}	V_{DD}	$V_{SS}-10$	V
Multiplex Frequency	Determined by External R & C	0.100	1.0	60	kHz
All Logic Inputs	Driven by External Timebase	dc		60	kHz
Logical High Level	Internal Depletion Device to V_{SS}	$V_{SS}-1$	V_{SS}	V_{SS}	V
Logical Low Level		V_{DD}	V_{DD}	$V_{SS}-10$	V
\overline{BCD} and 7 Segment Outputs					
Logical High Level	Loaded 2 k Ω to V_{DD}	2.0		20	mA source
Logical Low Level				0.01	mA source
Digital Enable Outputs					
Logical High Level				0.3	mA source
Logical Low Level	Loaded 100 Ω to V_{SS}	5.0		25	mA sink

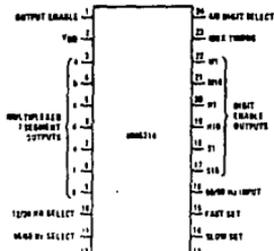
connection diagrams (Continued) Dual-In-Line Packages (Top Views)



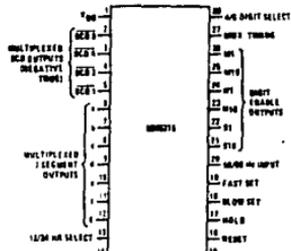
Order Number MMS312N
 See Package 22



Order Number MMS313N
 See Package 23



Order Number MMS314N
 See Package 22



Order Number MMS315N
 See Package 23

functional description

A block diagram of the MM5309 digital clock is shown in Figure 1. MM5311, MM5312, MM5313, MM5314 and MM5315 clocks are bonding options of MM5309 clock. Table 1 shows the pinouts for these clocks.

50 or 60 Hz Input: This input is applied to a Schmitt Trigger shaping circuit which provides approximately 5V of hysteresis and allows using a filtered sine-wave input. A simple RC filter such as shown in Figure 10 should be used to remove possible line voltage transients that could either cause the clock to gain time or damage the device. The shaper output drives a counter chain which performs the timekeeping function.

50 or 60 Hz Select Input: This input programs the prescale counter to divide by either 50 or 60 to obtain a 1 Hz timebase. The counter is programmed for 60 Hz operation by connecting this input to VDD. An internal depletion device is common to this pin, simply leaving this input unconnected programs the clock for 50 Hz operation. As shown in Figure 1, the prescale counter provides both 1 Hz and 10 Hz signals, which can be brought out as bonding options.

Time Setting Inputs: Both fast and slow setting inputs, as well as a hold input, are provided. Internal depletion devices provide the normal timekeeping function. Switching any of these inputs (one at a time) to VDD results in the desired time setting function.

The three gates in the counter chain (Figure 1) are used for setting time. During normal operation, gate A connects the shaper output to a prescale counter (÷50 or ÷60), gates B and C cascade the remaining counters. Gate A is used to inhibit the input to the counters for the duration of slow, fast or hold time-setting input activity. Gate B is used to connect the shaper output directly to a seconds counter (÷60), the condition for slow advance. Likewise, gate C connects the shaper output directly to a minutes counter (÷60) for fast advance.

Fast set then, advances hours information at one hour per second and slow set advances minutes information at one minute per second.

12 or 24-Hour Select Input: This input is used to program the hours counter to divide by either 12 or 24, thereby providing the desired display format. The 12-hour display format is selected by connecting this input to VDD; leaving the input unconnected (internal depletion device) selects the 24-hour format.

Output Multiplexer Operation: The seconds, minutes, and hours counters continuously reflect the time of day. Outputs from each counter (indicative of both units and tens of seconds, minutes, and hours), are time-division multiplexed to provide digit-sequential access to the time data. Thus, instead of requiring 42 leads to interconnect a 6-digit clock and its display (7 segments per digit), only 13 output leads are required. The multiplexer is addressed by a multiplex divider decoder,

which is driven by a multiplex oscillator. The oscillator and external timing components set the frequency of the multiplexing function and, as controlled by the 4 or 6-digit select input, the divider determines whether data will be output for 4 or 6 digits. A zero-blanking circuit suppresses the zero that would otherwise sometimes appear in the tens-of-hours display; blanking is effective only in the 12-hour format. The multiplexer addresses also become the display digit-enable outputs. The multiplexer outputs are applied to a decoder which is used to address a programmable (code converting) ROM. This ROM generates the final output codes, i.e., BCD and 7-segment. The sequential output order is from digit 6 (unit seconds) through digit 1 (tens of hours).

Multiplex Timing Input: The multiplex oscillator is shown in Figure 2. Adding an external resistor and capacitor to this circuit via the multiplex timing input (as shown in Figure 4a) produces a relaxation oscillator. The waveform at this input is a quasi-sawtooth that is squared by the shaping action of the Schmitt Trigger in Figure 2. Figure 3 provides guidelines for selecting the external components relative to desired multiplex frequency.

Figure 4 also illustrates two methods of synchronizing the multiplex oscillator to an external timebase. The external RC timing components may be omitted and this input may be driven by an external timebase; the required logic levels are the same as 50 or 60 Hz input.

Reset: Applying VDD to this input resets the counters to 0:00:00.00 in 12-hour format and 00:00:00.00 in 24 hour formats leaving the input unconnected (internal depletion pull-up) selects normal operation.

4 or 6-Digit Select Input: Like the other control inputs, this input is provided with an internal depletion pull-up device. With no input connection the clock outputs data for a 4-digit display. Applying VDD to this input provides a 6-digit display.

Output Enable Input: With this pin unconnected the BCD and 7-segment outputs are enabled (via an internal depletion pull-up). Switching VDD to this input inhibits these outputs. (Not applicable to MM5312, MM5313, and MM5315 clocks.)

Output Circuits: Figure 5a illustrates the circuit used for the BCD and 7-segment outputs. Figure 5b shows the digit enable output circuit. Figure 6 illustrates interfacing these outputs to standard and low power TTL. Figures 7 and 8 illustrate methods of interfacing these outputs to common anode and common cathode LED displays, respectively. A method of interfacing these clocks to gas discharge display tubes is shown in Figure 9. When driving gas discharge displays which enclose more than one digit in a common gas envelope, it is necessary to inhibit the segment drive voltage(s) during inter-digit transitions. Figure 9 also illustrates a method of generating a voltage for application to the output enable input to accomplish the required inter-digit blanking.

functional description (Continued)

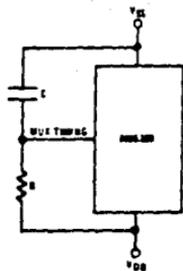


FIGURE 4a. Relaxation Oscillator

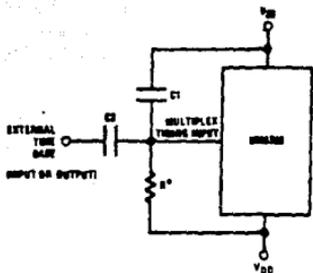


FIGURE 4b. External Time Base

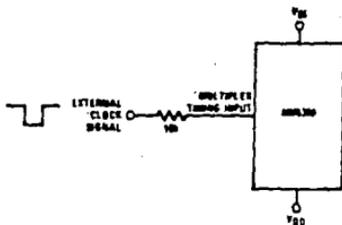


FIGURE 4c. External Clock

Note: Free running frequency should be set to run slightly lower than system frequency over temperature. External time base may be input or output.
* R=100k.

FIGURE 4. Synchronizing or Triggering Multiplex Oscillators

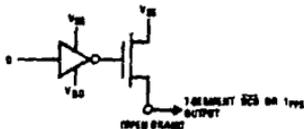


FIGURE 5a

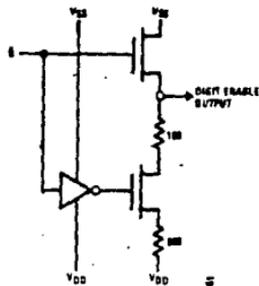


FIGURE 5b

FIGURE 5. Output Circuits

functional description (Continued)

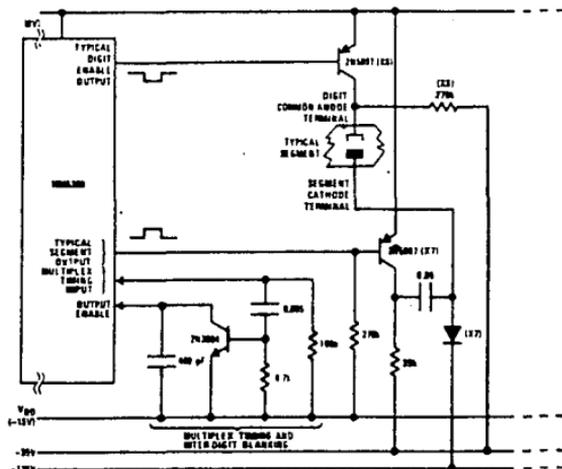


FIGURE 9. Interface Penplex II* Neon Display Tube

*TM of Burroughs Corp.

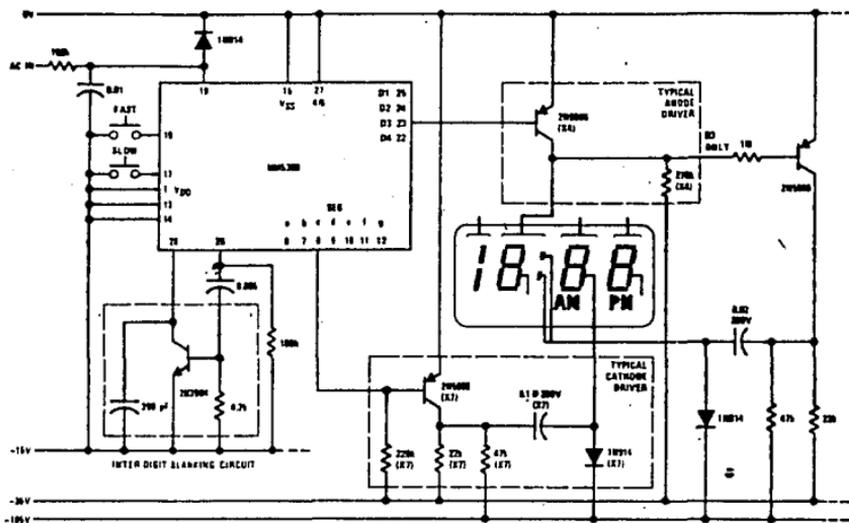


FIGURE 10. MM5309 Driving Gas Discharge Display, Typical Applications

81

APENDICE C.
HOJAS DE DATOS DEL MM5369



MM5369 17-stage programmable oscillator/divider

general description

The MM5369 is a CMOS integrated circuit with 17 binary divider stages that can be used to generate a precise 60 Hz reference from commonly available high frequency quartz crystals. An internal pulse is generated by mask programming the combinations of stages 1 through 4, 16 and 17 to set or reset the individual stages. The programmable number the circuit will divide by can vary from 10000 to 98000. The MM5369 is advanced one count on the positive transition of each clock pulse. Two buffered outputs are available: the crystal frequency for tuning purposes and the 17th stage 60 Hz output. Mask options are available for use with commonly available, low cost, high frequency crystals. Therefore, this design can be "customized" by special order to design specific programmable divider limits whereby the maximum divide by can be 98,000 and the minimum divide by can be 10,000. The MM5369 is available in an 8-lead dual-in-line epoxy package.

features

- Crystal Oscillator
 - Two buffered outputs
 - Output 1 crystal frequency
 - Output 2 full division
 - High speed (4 MHz at $V_{DD} = 10$)
 - Wide supply range 3-15V
 - Low Power
 - Fully static operation
 - 8 lead dual-in line package
 - Low current
- Standard MM5369N Only
- 3.58 MHz (color TV oscillator) input frequency
 - 60 Hz output frequency

connection diagram

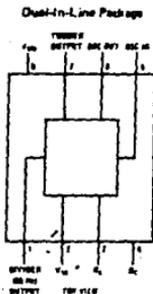


FIGURE 1.

Order Number MM5369N
See Package 17

block diagram

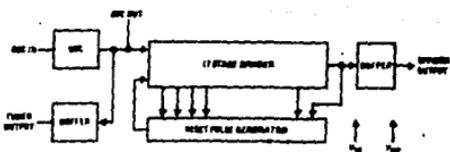


FIGURE 2.

absolute maximum ratings

Voltage at Any Pin	-0.3V to $V_{CC} + 0.3V$
Operating Temperature	0°C to +70°C
Storage Temperature	-65°C to +150°C
Package Dissipation	500 mW
Maximum V_{CC} Voltage	16V
Operating V_{CC} Range	3V to 15V
Lead Temperature (Soldering 10 seconds)	300°C

electrical characteristics

T_A within operating temperature range. $V_{GS} = GND$, $3V \leq V_{DD} \leq 15V$ unless otherwise specified.

PARAMETER	CONDITIONS	MIN	TYP	MAX	UNITS
Quiescent Current Drain	$V_{DD} = 15V$			10	μA
Operating Current Drain	$V_{DD} = 10V$, $f_{IN} = 4.19 MHz$		1.2	2.5	mA
Frequency of Oscillation	$V_{DD} = 10V$	DC		4.5	MHz
	$V_{DD} = 6V$	DC		2	MHz
Output Current Levels	$V_{DD} = 10V$				
	$V_{OUT} = 5V$				
Logical "1" Source		500			μA
Logical "0" Sink		500			μA
Output Voltage Levels	$V_{DD} = 10V$				
	$I_O = 10 \mu A$				
Logical "1"		9.0			V
Logical "0"				1.0	V

functional description

A connection diagram for the MM5369 is shown in Figure 1 and a block diagram is shown in Figure 2.

TIME BASE

A precision time base is provided by the interconnection of a 3,579,545 Hz quartz crystal and the RC network shown in Figure 3 together with the CMOS inverter/amplifier provided between the OSC IN and the OSC OUT terminals. Resistor R1 is necessary to bias the inverter for class A amplifier operation. Capacitors C1 and C2 in series provide the parallel load capacitance required for precise tuning of the quartz crystal.

The network shown provides > 100 ppm tuning range when used with standard crystals trimmed for $C_L = 12 pF$. Tuning to better than $\pm 2 ppm$ is easily obtainable.

DIVIDER

A pulse is generated when divider stages 1 through 4, 16 and 17 are in the correct state. By mask options, this pulse is used to set or reset individual stages of the counter, thus varying the modulus of the counter from 10000 to 98000. Figure 4 shows the relationship between the duty cycle and the programmed modulus.

OUTPUTS

The Tuner Output is a buffered output at the crystal oscillator frequency. This output is provided so that the crystal frequency can be obtained without disturbing the crystal oscillator. The Divide Output is the input frequency divided by the mask programmed number. Both outputs are push-pull outputs. A typical application of the MM5369 is shown in Figure 5.

