UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

CODIFICACION DE VOZ EN TIEMPO REAL A BAJA TASA DE TRANSMISION (4800 bits/seg.), UTILIZANDO L.P.C.

T E S I S

Que para obtener el Título de
Ingeniero Mecánico Electricista
p r e s e n t a n

VICTOR GARCIA GARDUÑO
MIGUEL MOCTEZUMA FLORES
SERGIO POPOCATL NAJERA

Director de Tesis
DR. LUIS ANDRES BUZO DE LA PEÑA

Ciudad Universitaria, D. F., Octubre 1986

UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

INDICE

INTRODUCCION

El gran avance tecnológico ha hecho posible la construcción de mejores computadoras capaces de procesar información rapidamente, esto ha significado avances en el área del Procesamiento Digital de Señales de Voz. Siendo posible la realización de un sistema Análisis-Síntesis de voz en tiempo real.

Diferentes modelos han sido postulados para describir y cuantificar los procesos involucrados en el proceso del habla. Una de las técnicas comunmente usadas, que ha resultado ser apta para Análisis-Síntesis de voz es la Predicción Lineal (LPC). Desde que se presento por vez primera la técnica de LPC a señales de voz, (Itakura y Saito, 1968, Atal y Hanaver, 1971), ha gozado de gran interés, siendo objeto de varios trabajos de investigación. La estructura de LPC corresponde a una codificación muy eficiente de la señal de voz, no requiere un intervalo de señal grande para su análisis, como los sistemas que trabajan en el dominio de la frecuencia.

El objetivo del presente trabajo es la implementación en el microprocesador TMS32010, de un sistema Análisis-Síntesis de voz en tiempo real a baja tasa de transmisión (4800 bits/seg.), utilizando LPC.

En el primer capítulo se hace un breve estudio del proceso de habla, en los tres siguientes de la técnica de LPC aplicada al procesamiento de voz: Análisis, Detección de tono, Síntesis. En el capítulo número 5 se describe su implementación en el TMS 32010.

1._ INTRODUCCION

1.1 ESTUDIO DEL HABLA

El sistema de transmisión de la voz está constituido por el aparato fonador como emisor, el aparato auditivo como receptor y la atmósfera como medio de transmisión. La voz es una onda acústica generada a partir del chorro de aire expulsado por los pulmones y modulado después, por los diferentes órganos que componen el aparato fonador. Estos se pueden dividir en tres grupos:

- Cavidades infraglóticas. Constituidas fundamentalmente por los pulmones.
- Cavidades laringeas ú órgano fonador.
- Cavidades supraglóticas. Constituidas por la cavidad bucal, la nasal y el velo del paladar.

Un esquema correspondiente a lo mencionado se muestra en la fig. 1.1.

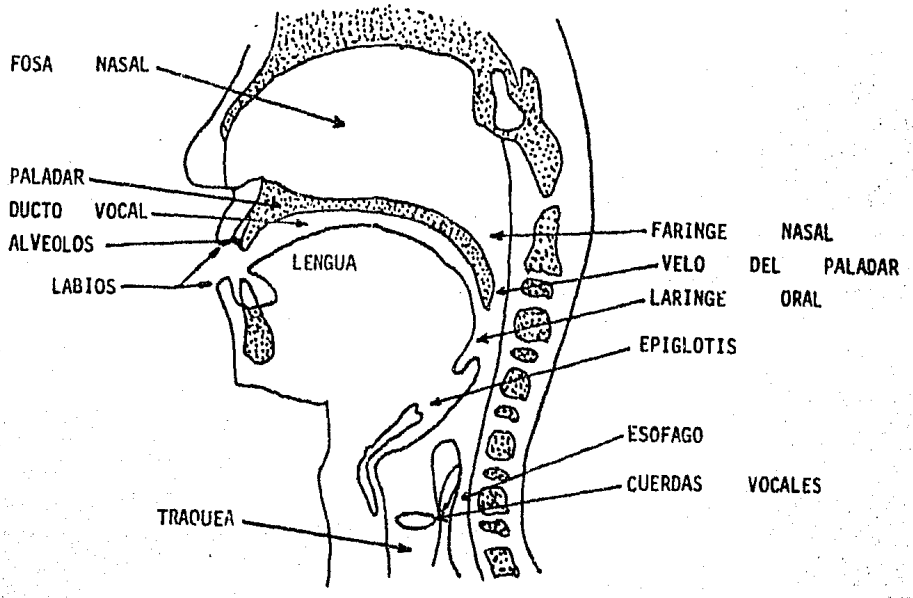


Fig. 1.1 Aparato vocal humano

El conjunto de cavidades supragloticas constituye, un resonador ó filtro acústico que conforma el espectro de la excitación aplicada que puede ser por ejemplo, un tren de impulsos generados por la vibración de las cuerdas vocales. La energía del sistema es proporconada por los pulmones al tratar de expeler el aire.

Se pueden distingir tres tipos de excitación, que además pueden aparecer simultáneamente :

A).La vibración de las cuerdas vocales. Se produce al tratar de expi-

rar y mantener la glotis cerrada, la presión subglotal llega a ser suficiente para separar las cuerdas, lo que provoca la salida de un pulso de aire al mismo tiempo que se reduce la presión y permite que los ligamentos vocales se vuelvan a cerrar repitiéndose el ciclo, la onda generada es aproximádamente triangular, de frecuencia entre 100 y 200 hz. El espectro se compone del fundamental y rayas de las armónicas con amplitud decreciente 12db/octava. Este periodo fundamental caracteriza la altura tonal de los sónidos articulados llamados sonoros, frente a los sordos en los que no se produce vibración de las cuerdas vocales y recibe en la literatura inglesa el nombre de "pitch" siendo parámetro importante en análisis síntesis de voz .

B). Sónidos fricativos. Consisten en la generación de una turbulencia de aire por estrechamiento del aparato fonador, la pronunciación de la "j" por ejemplo, produciéndoce un ruido acústico aproximádamente blanco

C) Sónidos oclusivos. Este tipo de excitación se produce cuando el aparato fonador se mantiene cerrado en algún punto, mientras se crea una presión de aire por detrás el caso de la "p" por ejemplo.

1.2 PROCESAMIENTO DE VOZ CON PREDICCION LINEAL

En esta sección se describe de manera general la técnica de L.P.C. en secciones posteriores se analizá con más detalle: análisis, detección de tono y síntesis tomando como base su implementacion en el microprocesador TMS32010.

Tomando como base el sintetizador de D.H. Klat, (de Klat 1980), y para mayor facilidad de comprensión se reproduce aquí el modelo de ingeniería fig 1.2.

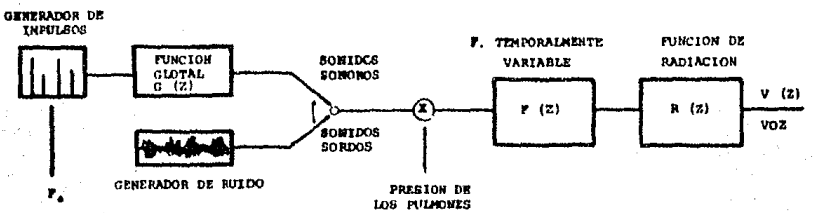


Fig 1.2 Modelo de producción de voz

Para los sónidos vocálicos, la excitación consiste en un tren de impulsos cuya frecuencia fundamental, F0, determina el tono de la voz. La funcion glotal le da la forma característica triangular del pulso glotal. Para el subconjunto de sonidos consonánticos, la excitación corresponde a una fuente de ruido la presión de aire de los pulmones determina la energía del sonido. La onda sonora resultante entra en un tubo acústico que es el tracto vocal, cuyas dimensiones varían en conformidad con la letra que se quiere articular. Estas caractarísticas permanecen fijas durante al menos 10 a 20 mseg, tiempo corto para

consonantes explosivas "p" ó "t", mientras que para vocales su tiempo de duracion es mayor. A la salida se representa el efecto de radiación por los labios que acentúa las frecuencias altas y compensa en parte el filtraje de la función glotal. De esta forma la salida en transformada z está dada por:

$$V(Z)=A\ G(Z)\ F(Z)\ R(Z)\ I(Z) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.1)$$

donde G(Z), F(Z), R(Z), representan carácteristicas de la funcion glotal, ducto vocal y la función de radiación respectivamente, I(Z) es el generador de impulsos.

En análisis con L.P.C. se procura obtener las carácteristicas de cada bloque. Sea como primer paso la función glotal G(Z) que se representa como una función de dos polos:

$$G(Z)=\frac{K_1}{(1-Z_a Z^{-1})(1-Z_b Z^{-1})} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2)$$

El filtraje paso alto está dado por: $R(Z)=K_2(1-Z^{-1})$ ..... (2.3) combinando G(Z) y R(Z).

$$G(Z)R(Z)=\frac{K_1\ K_2\ (1-Z^{-1})}{(1-Z_a Z^{-1})(1-Z_b Z^{-1})} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.4)$$

Se apróxima como:

$$\frac{K_1\ K_2}{[1\ +\ (1-Z_a)Z^{-1}](1-Z_b Z^{-1})}$$

Si Za=1 entonces:

$$G(Z)R(Z)=\frac{K}{1\ -\ \not{\alpha}\ Z^{-1}}$$

Donde $\not{\alpha}$ es el coeficiente de preacentuación (preénfasis). Entonces se pueden eliminar los efectos de la función glotal y de radiación al pasar la señal de voz por un filtro inverso de estás funciones llamado preacentuación T(Z).

$$T(Z)=1\ -\ Z^{-1}=\frac{K}{G(Z)\ R(Z)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.5)$$

El valor óptimo de es funcion de la entrada y varía de 0.5 a 1

sea: $S(Z)=V(Z)T(Z)=A\ G(Z)F(Z)R(Z)T(Z)E(Z)=A'F(Z)E(Z)$ ..... (2.7)

Para la señal resultante S(Z) se encuentra un filtro predictor P(Z), el cual es función de los últimos M valores de S[n], se genera un valor S'[n] que se aproxima a S[n], ver figura 1.3

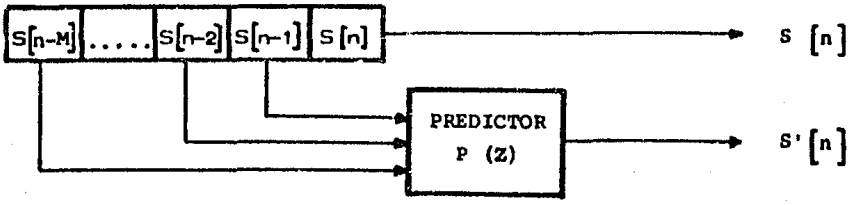


Fig 1.3 Filtro predictor P(Z)

donde: $S'[n]=a_1 S[n-1] + a_2 S[n-2]....a_m S[n-M]$ ..... (2.8)

$$=\sum_{J=1} a_j S[n-j]$$ ..... (2.9)

entonces P(Z) está dado por:

$$P(Z)=\sum_{J=1}^{M} a_j Z^{-j} \quad y \quad S'(Z)= P(Z)\, S(Z)$$

Los coeficientes aj se escogen de tal forma que la diferencia ó error d[n] entre S[n] y S'[n] sea mínima:

$$d[n]=S[n] - S'[n]=\sum_{J=0}^{M} a_j S[n-j] \text{ , donde } ao=1$$

ó $D(Z)=(1-P(Z))S(Z)$ ..... (2.10)

el criterio a seguir es minimizar $d^2[n]$ en el intervalo (no,n1) es decir:

$$\alpha =\sum_{n=no}^{n_1} d^2[n]=\sum_{n=no}^{n_1} \left[\sum_{j=0}^{M} a_j S(n-j)\right]^2$$ .....(2.11)

se minimiza si tomamos sus derivadas parciales e igualamos a cero sea:

$$\frac{\partial \alpha}{\partial a_j} = 0 = 2\sum_{i=0}^{M} a_i C_{ij}$$ .....(2.12)

donde:

$$C_{ij}=\sum_{n=no}^{ni} S[n-i]S[n-j]$$

que son los coeficientes de correlación de S[n] y como ao=1

$$\sum_{i=1}^{M} a_i C_{ij} = -C_{oj} \qquad j = 1, 2 \ldots M \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.13)$$

El orden M del predictor, determina la exactitud con que se representa S[n].

En sistemas LPC se considera cada formante de la voz como una resonancia, que puede representarse por un par de polos complejos conjugados para el tracto vocal se genera una resonancia por cada Khz en consideracion, en nuestro caso si muestreamos a 6.4 Khz tendremos un filtro de 6 polos, pero como LPC no separa los efectos de radiación de los labios ó cuerdas vocales del ducto vocal, se agregan 2 ó 4 polos más para compensar, quedando para 6.4 Khz de muestreo un filtro de M=10.

El efecto del predictor de orden M es el de restar la contribución del ducto vocal:

$$1- P(Z) = [F(Z)]^{-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.15)$$

La señal diferencia de salida es la excitación:

$$D(Z)=S(Z)-P(Z)S(Z)=S(Z)(1-P(Z))= A'E(Z) \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.16)$$

esquemáticamente representado en la fig. 1.4.

Para encontrar el filtro predictor hay que calcular los coeficientes Cij de correlación, después solucionar las M ecuaciones de la expresión 2.13. Para hallar las M incógnitas aj varias técnicas son empleadas para solucionar este problema, una de ellas es por ejemplo:

Utilizar una variante de autocorrelación y el método recursivo Levinson-Durbin.

Si se considera que la señal S[n] es igual a cero fuera de una ventana. La minimización del error nos lleva a las ecuaciones:

$$\sum_{j=1}^{M} a_j R_{i-k} = -R_i \qquad 1 < i < M \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.17)$$

donde:

$$Ri = \sum_{n=0}^{N-|i|} S[n]S[n+|i|]$$

que es la función de autocorrelación de S[n].

Estas ecuaciones pueden resolverse en forma recursiva para los coeficientes aj, como se indica a continuación:

$$Eo=Ro \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.18a)$$

$$K_i = -\left(R_i + \sum_{j=1}^{i-1} a_j^{(i-1)} R_{i-j}\right) / E_{i-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.18b)$$

$$a_i^{(i)} = K_i \qquad \text{.....(2.18c)}$$

$$a_j^{(i)} = a_j^{(i-1)} + K_i\, a_{i-j}^{(i-1)} \qquad 1 \leq j \leq i-1 \qquad \text{.....(2.18d)}$$

$$E_i = (1 - K_i^2)E_{i-1} \qquad \text{.....(2.18e)}$$

el proceso se repite para i=1,2 .... M y se obtiene la solución:

$$a_j = a_j^{(p)} \qquad 1 \leq j \leq P \qquad \text{.....(19)}$$

con éste método obtenemos los coeficientes de reflexión K_i y el valor de la energía de la diferencia residual $= E_m$, estimador de la amplitud cuadrado de la señal.
El análisis con LPC da lugar a varias transformaciones derivadas de él, ya que existen además de los coeficientes de PL, las raíces del polinomio del predictor, los coeficientes cepstrales, la respuesta a impulso del filtro, los coeficientes de autocorrelación de la señal y de los coeficientes, los coeficientes de la función logaritmica del área del tubo acústico, correspondientes al tracto vocal y los coeficientes de reflexión, que tienen la ventaja de: Dar estabilidad al filtro después de su cuantificación, de proveer un ordenamiento natural de los parámetros, que permite su manejo estadístico de valores comunes para poder elaborar códigos de transmisión eficientes. Su interpretación física es muy interesante pues corresponden al coeficiente de reflexión de la onda acústica, del modelo del tubo acústico correspondiente al tracto vocal. Estos parámetros sirven para controlar un filtro de rejilla que es un sintetizador de voz.

Para terminar con análisis nos resta codificar la exitación E(Z) de acuerdo al modelo de la fig. 1.2. La exitación ó genera pulsos ó es fuente de ruido, esta decisión se toma en base al tono de la señal, su estimación es vital en la síntesis de voz natural. En la fig. 1.4 se resume el sistema de análisis con LPC.

El filtro a la entrada asegura que la señal sea de banda limitada y además sirve para eliminar el ruido de la linea, la banda de paso de este filtro es de 50 a 3200 hz, implementado en realidad con dos filtros un paso alto y un paso bajo.

El análisis de la señal es por tramos, cada 20 mseg para poder seguir los cambios bruscos de la señal de voz, se agrega al sistema además la multiplicación por una ventana de Hamming con traslape, con lo cual se pueden obtener buenos resultados.

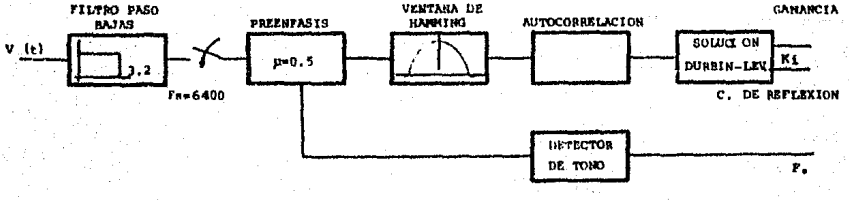


Fig. 1.4 Sistema de análisis de predicción lineal

SINTESIS:

Para sintetizar la señal se invierte la ecuación 2.10

$$S[n]=\sum_{j=1}^{M} a_j\, S[n-j] + d[n]$$

la cual se representa en la fig. 1.5.

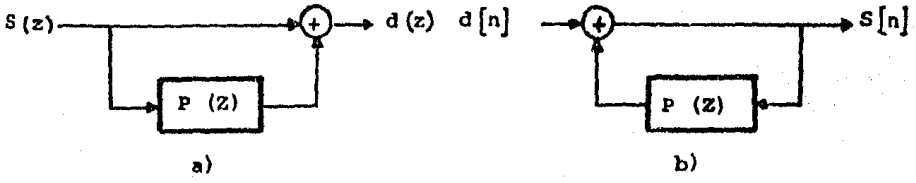


Fig 1.5 Modelo de LPC a) Análisis b) Síntesis.

Donde la funcion de transferencia del sistema anterior esta dada por:

$$F(Z)=\frac{1}{1-\sum_{j=1}^{M} a_j\, Z^{-j}}$$

La exitación para sónidos sonoros es producida por un generador de pulsos, su amplitud depende de α y su priodo de fo. La exitacion tipo ruido es producida por un generador de ruido blanco, el efecto de preacentuación se compensa con su inverso, el sistema de síntesis con LPC queda entonces, como el mostrado en la fig. 1.2 anterior.

2. ANALISIS

INTRODUCCION

En este capitulo, se describe brevemente el desarrollo del Modelo de Análisis encontrado en el capitulo 1 y su solución a través del Método de Autocorrelación, el cual permite encontrar los coeficientes necesarios que representan el Modelo del Tubo Acústisco del Sistema Vocal.

Se presenta un breve desarrollo de las operaciones necesarias para obtener dichos coeficientes, además de la ganancia que es requerida en el Modelo de Síntesis.

Cabe hacer notar que el Metodo de Autocorrelación se utilizo por considerarlo bastante flexible en el aspecto del desarrollo matemático y se mencionará posteriormente en el transcurso del capítulo, el porque, se uso este método.

El desarrollo del Modelo de Análisis, está basado en el estudio de los algoritmos y su solución, en las fórmulas del producto interno y las relaciones ortogonales ya conocidas en el Area de Ingeniería.

2.1 DESARROLLO Y SOLUCIONES

El problema principal que hay que resolver, es el de minimizar una expresion de la forma:

$$\alpha = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i c_{ij} a_j$$

donde $a_0=1$. Resolviendo las ecuaciones simultáneas.

$$\sum_{i=0}^{M} a_i c_{ij} = -c_{0j} \quad \text{para} \quad j=1,2,\ldots,M$$

los coeficientes c_{ij} son obtenidos a través de una correlación de secuencia de datos de entrada.

$$c_{ij}=c_{ji}= \sum_{n=n_0}^{n_1} x[n-i]x[n-j] \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

A partir del Método de Autocorrelación, las entradas son truncadas de modo que $x[n]=0$ para $n<0$ y para $n>N-1$. Tomando la suma en la ec.2.1 sobre los valores de n desde $n_0=-\infty$ a $n_1=\infty$, se obtiene:

$$c_{ij}=r[i-j] \quad \text{[Método de Autocorrelación]}$$

donde r[i-j] es el término de los coeficientes de autocorrelación.

Las razones primordiales por las cuales se estudian estas ecuaciones son:

- El número de propiedades que se obtienen son importantes.
- Se obtienen un grupo de parámetros k_m, que en el modelo de autocorrelación simulan un tubo acustico de un sistema vocal, en donde la voz es pre-enfatizada y los paramétros k_m serán referidos como coeficientes de reflexión.

El desarrollo del Método de Correlación (autocorrelación), está basado en el producto interno, por lo que:

$$\alpha = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i c_{ij} a_j$$

y $$\sum_{i=1}^{M} a_i c_{ij} = -c_{0j} \quad \text{para} \quad j=1,2,\ldots,M$$

por lo que se expresa:

$$\alpha = a^t Ca + 2a^t C + C^t C \quad y \quad Ca = -C$$

donde t indica transpuesta, C es la matriz de orden [MxM]. Como la matriz C es simétrica $C_{i,j} = C_{j,i}$, su solución se realiza por medio del procedimiento de Ortogonalizacion de Gram-Schmidt.

A traves del producto interno, se representa el resultado de dos filtros o polinomios F(Z) y G(Z) dados por $\langle F(Z), G(Z) \rangle$.

$$\text{donde} \quad F(Z) = \sum_{i=0}^{\infty} f_i z^{-i}$$

$$y \quad G(Z) = \sum_{i=0}^{\infty} g_i z^{-i}$$

$$\text{entonces} \quad \langle F(Z), G(Z) \rangle = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_i C_{i,j} g_j$$

$$\text{con} \quad C_{i,j} = \langle Z^{-i} Z^{-j} \rangle = \sum_{n=n_0}^{n_1} x[n-i]x[n-j]$$

Se observa que las sumas establecidas para su solución, son notablemente reducidas por el producto interno, por lo que la ec.2.1 será expresada:

$$\alpha = \langle A(Z), A(Z) \rangle$$

$$\text{donde} \quad A(Z) = \sum_{i=0}^{M} a_i Z^{-i} \qquad \text{con } a_0 = 1$$

Los coeficientes de Predicción Lineal son encontrados por medio del filtro inverso A(Z). Por otra parte, se define la norma cuadrada como el producto interno de un polinomio con respecto a si mismo.

$$\|F(Z)\|^2 = \langle F(Z), F(Z) \rangle$$

En el Método de Autocorrelación, la secuencia de entrada es truncada y los límites sobre la suma son desde $n_0 = -\infty$ hasta $n_1 = \infty$.

Así F(Z) tiene una norma cero, si y solo si F(Z) es cero, por lo que:

$$\|F(Z)\|^2 > 0 \quad \text{para} \quad F(Z) \neq 0$$

En Espacios Euclidianos, la magnitud del producto de dos vectores es siempre menor o igual al producto de sus magnitudes; por lo que se establece la desigualdad de Cauchy-Schwartz.

$$|\langle F(Z), G(Z) \rangle| \leq \|F(Z)\| \|G(Z)\|$$

el lado izquierdo de esta ecuación representa la magnitud de un número y el lado derecho, representa el producto de las normas.

El problema básico de Predicción Lineal, es encontrar el polinomio A(Z) de la forma:

$$A(Z) = \sum_{i=0}^{M} a_i Z^{-i} \qquad \text{con } a_0 = 1$$

$$y \text{ minimizar} \qquad \alpha = \|A(Z)\|^2$$

Para encontrar A(Z), se generan polinomios recursivos de la forma:

$$A_m(Z)=\sum_{i=0}^{M} a_{mi}Z^{-i} \quad \text{con } a_{mo}=1$$

y $$B_m(Z)=\sum_{i=1}^{M+1} b_{mi}Z^{-i} \quad \text{con} \quad b_{m,m+1}=1$$

que satisfaga las relaciones ortoganales:

$$\langle A_m(Z),Z^{-i}\rangle=\langle B_m(Z),Z^{-i}\rangle=0 \qquad i=1,2,\ldots,M$$

El grupo de polinomios $B_m(Z)$ se encuentran por el método de ortogonalización Gram-Schmidt de las potencias $Z^{-1},Z^{-2},\ldots,Z^{-M}$.

Como $A_{m-1}(Z)$ y $B_{m-1}(Z)$ son ortogonales a $Z^{-1},Z^{-2},\ldots,Z^{-(m-1)}$, $A_m(Z)$ será definido como:

$$A_m(Z)=A_{m-1}(Z)+k_mB_{m-1}(Z) \quad \ldots\ldots\ldots\ldots\ldots\ldots(2.2)$$

Si $A_m(Z)$ y $B_m(Z)$ son ortogonales a las potencias $Z^{-1},Z^{-2},\ldots,Z^{-M}$ entonces:

$$\langle A_m(Z),B_m(Z)\rangle=\langle 1,B_m(Z)\rangle=\langle A_m(Z),Z^{-(m+1)}\rangle$$

y el error cuadrático en el paso m, será:

$$\alpha_m=\|A_m(Z)\|^2=\langle A_m(Z),A_m(Z)\rangle=\langle 1,Am(Z)\rangle$$

$$\beta_m=\|B_m(Z)\|^2=\langle B_m(Z),B_m(Z)\rangle=\langle Z^{-(m+1)},B_m(Z)\rangle$$

Para calcular α_m. Se realiza con $A_0(Z)=1$ y usando la ec.2.2 se tiene:

$$A_m(Z)=1+\sum_{i=1}^{m} K_iB_{i-1}(Z) \qquad \text{para } m>0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(2.3)$$

Como $B_m(Z)$ es ortogonal; la norma cuadrada de la ec.2.3 es:

$$\|A_m(Z)-1\|^2=\sum_{i=1}^{m} K_i^2\,\beta_{i-1}$$

y realizando aplicaciones lineales, se obtiene:

$$\alpha_m=\|1\|^2=-\sum_{i=1}^{m} K_i^2\,\beta_{i-1}$$

reemplazando m=m+1 y sustrayendo α_m de α_{m+1}, se tiene:

$$\alpha_{m+1}=\alpha_m-K_{m+1}^2\,\beta_m$$

El primer paso para obtener los coeficientes C_{ij} están dados por:

$$C_{ij}=C_{ji}=\sum_{n=0}^{n} x[n-i]x[n-j]$$

Los coeficientes necesarios M+1 en el metodo de autocorrelación son evaluados desde:

$$C_{ok}=r(k)=\sum_{n=-\infty}^{\infty} x[n]x[n-k]=\sum_{n=0}^{N-1-k} x[n]x[n+k]$$

donde $k=0,1,2,\ldots,M$

Las condiciones iniciales son:

$$A_0(Z)=1 \quad y \quad B_0(Z)=Z^{-1}$$

o $\quad a_{00}=1 \quad y \quad b_{01}=1$

La evaluación del producto interno:

$$\langle A_m(Z),B_m(Z)\rangle=\langle 1,B_m(Z)\rangle=\langle A_m(Z),Z^{-(m+1)}\rangle$$

y $\quad \alpha_m=\|A_m(Z)\|^2=\langle A_m(Z),A_m(Z)\rangle=\langle 1,A_m(Z)\rangle$

nos da:

$$\alpha_0=\langle A_0(Z),A_0(Z)\rangle=C_{00}$$

$$\beta_0=\langle B_0(Z),B_0(Z)\rangle=C_{11}$$

el cual, en el método de autocorrelación $C_{00}=C_{11}=r(0)$.
Si $K_m=-\langle A_{m-1}(Z),Z^{-m}\rangle / \beta_{m-1}$ con $m=1$, se tiene:

$$K_1=-C_{10}/\beta_0=-C_{10}/C_{11}$$

y $\quad A_1(Z)=A_0(Z)+K_1B_0(Z)=1+K_1Z^{-1}$

con $a_{10}=1$, $a_{11}=K_1$ y $\alpha_{m+1}=\alpha_m-K_{m+1}^2\beta_m$ com $m=0$ se obtiene:

$$\alpha_1=\alpha_0-K_1^2\beta_0$$

Lo que complementa el proceso de inicialización. El filtro inverso y el error cuadrático están dados por:

$$A(Z)=A_m(Z) \quad y \quad \alpha=\alpha_m$$

Para obtener $B_m(Z)$ se utiliza el Método de Ortogonalización de Gram-Schmidt y se utiliza recursivamente.

$$\beta_m=\langle Z^{-(m+1)},B_m(Z)\rangle=\sum_{j=1}^{m+1} C_{m+1,j}b_{mj}$$

por lo que $B_m(Z)$ y β_m en el paso m es ahora completado y la nueva $A_{m+1}(Z)$ y α_{m+1} son evaluadas en el siguiente paso.
Los coeficientes C_{ij}, dependen solamente de las diferencias descritas, por lo que $B_m(Z)$ y $A_m(Z)$ se relacionan por:

$$B_m(Z)=Z^{-(m+1)}A_m(1/Z) \qquad \text{.....(2.4)}$$

y $\quad b_{mi}=a_{m,m+1-i} \quad$ para $\quad i=1,2,\ldots,m+1$

la aplicación de la ec.2.4, reduce el número de operaciones necesarias para la solución de las ecuaciones de autocorrelación del orden M^2.
Se observa que:

$$\beta_m=\alpha_m$$

incrementando en una unidad para completar el paso m.

$$A_{(m+1)}=A_m(Z)+K_{m+1}B_m(Z)$$

y de la ecuación $0=\langle A_m(Z),Z^{-m}\rangle=\langle A_{m-1}(Z)Z^{-m}\rangle+K_m\langle B_{m-1}(Z),Z^{-m}\rangle$ se obtiene:

$$0=\langle A_m(Z),Z^{-(m+1)}\rangle+K_{m+1}\,\beta_m$$

$$\text{si}\quad \beta_m\neq 0 \qquad K_{m+1}=-1/\beta_m\sum_{i=0}^{m}C_{m+1,i}\,a_{mi}$$

$$\text{si}\quad \beta_m=\alpha_m \qquad \text{y} \qquad C_{m+1}=r[m+1-i]$$

$$\text{entonces}\quad \alpha_{m+1}=\alpha_m-K_{m+1}^2\,\beta_m$$

por lo que el proceso en el paso m es completado y los parámetros K_m serán referidos como los coeficientes de reflexión que definen el Modelo del Tubo Acústico del Sistema Vocal.

En la implementación de técnicas de Predicción Lineal que requieren computación digital, se necesitan una serie de consideraciones como son la eficiencia de programación, la velocidad de cálculo y la implementación en el arreglo del punto aritmético.

Se ha dicho que en el Método de Autocorrelación, β_m es igual a la norma cuadrada del polinomio $A_m(Z)$, por lo que se relaciona con el determinante de la Matriz C, por medio del producto:

$$|C|=\prod_{m=0}^{M-1}\beta_m$$

Como los coeficientes β_m se encuentran en el denominador, un pequeño valor de β_m, aumenta errores numéricos y tales coeficientes son grandes, por lo que es necesario checar estos resultados en cada paso recursivo.

La medida más común de cubrir todas las condiciones de las matrices, es el radio del máximo y mínimo eingevalor de la matriz. En el Método de Autocorrelación, se utilizan todas las posiciones de los eingevalores junto con el rango dinámico del Espectro.

Una medida espectral que se aplica al Método de Autocorrelación es:

$$\rho_m=\frac{\left[\prod_{i=1}^{M}\lambda_i\right]^{1/M}}{1/M\sum_{i=1}^{M}\lambda_i}$$

ρ_m es una medida no deseada en el desarrollo de Análisis de orden M. Esta medida estará siempre entre uno y cero; en dicho método ρ_m se relaciona como una medida espectral llana ó insípida.

Si $\alpha_i=\beta_i$

$$\rho_m=\left[\prod_{i=0}^{M-1}(\alpha_i/\alpha_0)\right]^{1/M}$$

aprovechando el límite e incrementando en M a ρ_m:

$$\lim_{M\to\infty}\rho_m=\rho_\infty=\alpha_\infty/\alpha_0=E(x)$$

donde E(x) es la medida espectral de la entrada de una secuencia de datos. Por lo que se considera lo siguiente:

-Se toma mas precision analizando sonidos de voz, que sonidos mudos.

-Incrementando el rango de muestreo, se incrementa la precisión computacional.
-Un pre-enfasis en la secuencia de entrada, incrementa la cantidad de precisión computacional.

Para realizar el pre-enfásis de datos de voz, se efectua por medio de un filtro de la forma $1-\mu Z^{-1}$, con el propósito de reducir las condiciones no deseadas. Este filtro será de primer orden.

2.2 MODELADO DEL TUBO ACUSTICO

Atal[1970], demostró que las frecuencias de formato y ancho de banda son suficientes para determinar las áreas de un tubo acústico con un número específico de secciones. Demostró que una función de transferencia de M polos, es siempre realizable como una función de transferencia de un tubo acústico de M secciones cilíndricas de igual longitud. Wakita demostró que el mismo modelo del tubo acústico, es equivalente a representarlo como un filtro inverso A(Z), obtenido por Predicción Lineal de las formas de onda de voz. Demostró que si la voz es adecuadamente pre-enfatizada y los límites de las condiciones del tubo acústico son escogidos; se obtienen formas razonables del Sistema Vocal que son estimadas por medio del Método de Autocorrelación de Predicción lineal.

Del desarrollo del Tubo Acústico se establece lo siguiente:
-Los resultados matemáticos son aceptables.
-Los parámetros del modelo son obtenidos a partir de Predicción Lineal de la forma de onda de voz.
-El Modelo del Tubo Acústico resulta muy aproximado a la forma del Sistema Vocal Humano.

Las bases usadas en la derivación del modelo es como sigue:
-El Sistema Vocal es representado como un grupo consistente de M secciones interconectadas de igual longitud. Cada sección individual es una área uniforme.
-Las dimensiones transversales de cada sección es pequeña comparada con una longitud de onda, de modo que la propagación de sonido a través de una seccion, permita tratarla como una onda plana.
-Las secciones son rígidas, por lo que las pérdidas a la vibración de las paredes, viscocidad y conducción de calor son nulas.
-Las ecuaciones elementales de propagación de onda, son válidas.
-El modelo es lineal.
-Los efectos del sistema nasal son ignorados.

Un modelo esquemático del Sistema Vocal Humano mostrado en la fig.1.1, se encuentra en la fig.2.1, como una serie de 8 secciones uniformes cilíndricas. Cada sección consta de una área A_m, el cual el valor estimado, es el promedio del área de las secciones no uniformes del sistema vocal.

Otras representaciones de este modelo son mostradas en las fig.2.1b y 2.1c., éstas figuras representan el área A_m en dos dimensiones, desde la glotis a los labios. La forma de la fig.2.1c. se usa en mas ejemplos sugeridos por Wakita y los resultados gráficos son representados como una función de área discreta ó una función de área.

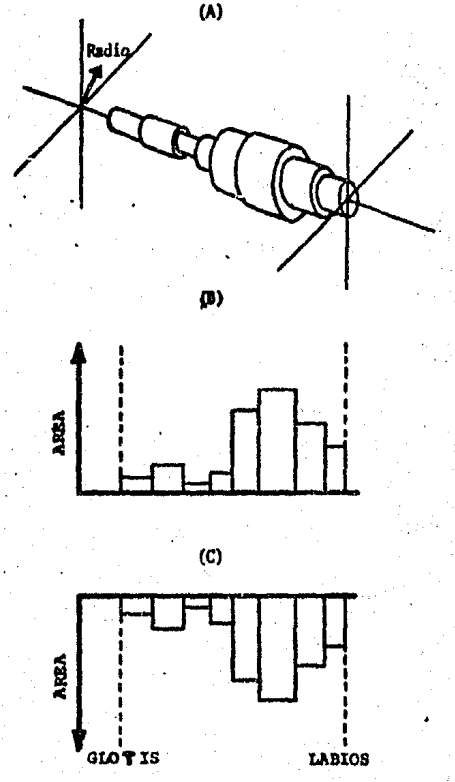


Fig.2.1 a) Serie de secciones cilíndricas uniformes.
b) Función discreta de area.
c) Función de área con eje invertido.

2.2.1 Derivación de secciones

Las ecuaciones que gobiernan la presión y velocidad de volumen, las cuales satisfacen la ecuación del momento son:

$$\frac{\partial P_m(x,t)}{\partial x} = -\frac{\rho \partial U_m(x,t)}{A_m(x)\,\partial t}$$

y la ecuación de continuidad de masa:

$$\frac{\partial U_m(x,t)}{\partial x} = -\frac{A_m(x)\,\partial P_m(x,t)}{\rho c^2\,\partial t}$$

Las variables $P_m(x,t)$ y $U_m(x,t)$, definen la presion y velocidad de volumen respectivamente, junto con la sección M del tubo acústico como una función del tiempo y la distancia x. El termino ρ define la densidad de aire y c es la velocidad del sonido. Estas ecuaciones combinadas, dan las ecuaciones de Webster.

$$\frac{\partial P_m(x,t)}{\partial x} = -\frac{\rho}{A_m(x)} \frac{\partial U_m(x,t)}{\partial t^2}$$

Su solucion es una combinanción lineal de ondas viajando hacia atrás y hacia adelante. (+)=adelante, (-)=atrás.

$$U_m(x,t) = U_m^+(t-x/c) - U_m^-(t+x/c)$$
$$P_m(x,t) = P_m^+(t-x/c) - U_m^-(t+x/c)$$

donde x/c tiene unidades de tiempo.

Las ondas viajando hacia adelante se mueven en dirección de la glotis a los labios y las ondas viajando hacia atrás en sentido contrario.

La representación de la velocidad de volumen son mostradas en la fig.2.2.

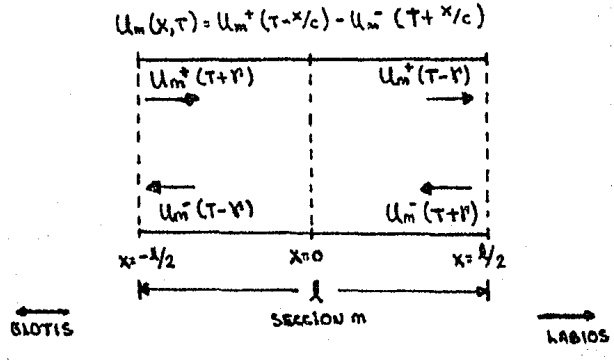


Fig.2.2 Trayectoria de la forma de onda hacia atrás y hacia adelante definidas para la velocidad de volumen en la seccion m.

La velocidad de volumen $U_m(x,t)$ en cualquier lugar y tiempo (x,t) junto con la sección M, es la diferencia entre la onda viajando hacia adelante $U_m^+(t-x/c)$ y la onda viajando hacia atrás $U_m^-(t+x/c)$. El centro de la sección es definido como x=0. La longitud de la sección es definida como L, de modo que hacia la izquierda y hacia la derecha son respectivamente -L/2 y +L/2.

Las ecuaciones que gobiernan la presión y velocidad de volumen pueden ser usadas para describir $P_m(x,t)$ en terminos de los componentes de la velocidad de volumen, obteniéndo:

$$P_m(x,t) = \frac{\rho_a}{A_m} [U_m^+(t-x/c) + U_m(t+x/c)]$$

los términos constantes son omitidos debido a la variación de presión.

2.2.2 Condiciones de Continuidad

La velocidad de volumen en la seccion de la derecha m(x=+L/2) es:

$$U_m(L/2,t)=U_m^+(t-\gamma)-U_m^-(t+\gamma)$$

de manera similar, la velocidad de volumen en la sección de la izq.es:

$$U_{m-1}(-L/2,t)=U_{m-1}^+(t+\gamma)-U_{m-1}^-(t-\gamma)$$

estas ecuaciones de continuidad son ilustradas en la fig.2.3.

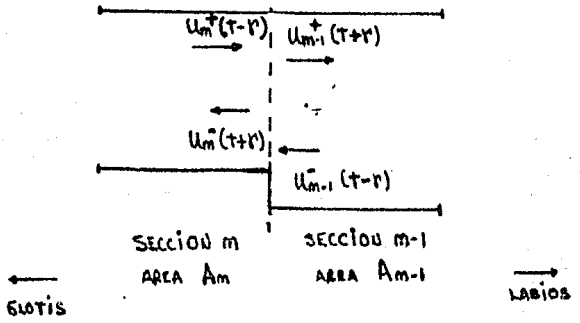


Fig.2.3 Condiciones de continuidad entre la sección m-1 y la sección m para la velocidad de volumen.

En un sistema físico, la presión y velocidad de volumen entre dos secciones será continua.
Por lo que se obtiene el coeficiente de reflexión μ_m:

$$\mu_m = \frac{A_{m-1}-A_m}{A_{m-1}+A_m}$$

$$\frac{A_m}{A_{m-1}} = \frac{1-\mu_m}{1+\mu_m}$$

A_m y A_{m-1} tienen áreas iguales y no hay reflexión ($\mu_m=0$).

2.2.3 Relaciones entre el Tubo Acústico y Predicción Lineal.

Las relaciones entre la velocidad de volumen en la sección m y en la seccion m-1, son desarrolladas a lo largo de la glotis y los labios con un grupo específico de condiciones.
Los resultados obtenidos, muestran que las relaciones recursivas son equivalentes al diseño del filtro inverso. Si el número de coeficientes de reflexión M en el filtro inverso, es igual al número

secciones M en el modelo de: tubo acústico, entonces los coeficientes de reflexion μ_m, el cual definen el radio del area del Modelo del Tubo Acustico; son obtenidos a traves del Analisis de Prediccion Lineal de la forma de onda de voz.

La relacion que existe entre la frecuencia de muestreo $f_m=1/t$, el numero de secciones M, la longitud del tubo acustico L=Ml y la velocidad del sonido c, está dada por:

$T=2l/c=2Ml/Mc$ o $f_m=Mc/2L$ L=longitud del sistema vocal.

La forma de onda acustica de voz s(t) a los labios, es considerada como una primera aproximacion de la derivada de la velocidad de volumen a los labios. Por lo que la transformada Z de la medida de voz a la distancia (o) desde los labios, se da por:

$$L(Z)=1-\mu Z^{-1} \qquad \text{con} \qquad 0\leq\mu\leq 1$$

S(Z) es desarrollado en términos de la transferencia vocal V(Z) por:

$$S(Z)=L(Z)V(Z)G(Z)E(Z) \simeq \frac{V(Z)E(Z)}{L(Z)}$$

los terminos de la forma glotal G(Z) es aproximado por $G(Z)=1/L^2(Z)$.

El filtro inverso, es descrito como una version pre-enfatizada de voz: $X(Z)=P(Z)S(Z)$, por lo que la funcion de transferencia del filtro inverso es:

$$\frac{E(Z)}{V(Z)} \simeq \frac{L(Z)}{V(Z)P(Z)} = A(Z)$$

y se obtiene el factor de radiación de los labios: $P(Z)=L(Z)\simeq 1-\mu z^{-1}$ junto con el resultado $V(Z)\simeq 1/A(Z)$.

En Análisis de Predicción Lineal con la señal de voz pre-enfatizada, un filtro inverso se obtiene, su reciproco 1/A(Z), es una función de transferencia estimada del Sistema Vocal.

En la fig.2.4 se muestran la comparación de funciones de área con las ventajas del pre-enfasis de una señal de voz desde la glotis a los labios.

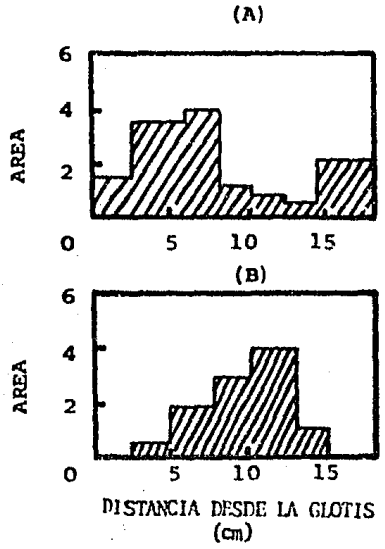


Fig.2.4 Comparaciones de funciones de área.
a) Pre-enfasis.
b) No Pre-énfasis.

3. PITCH.

INTRODUCCION

Un detector de frecuencia fundamental (pitch), es un componente escencial en los sistemas de procesamiento de voz, que dá valiosas guías dentro de la fuente de excitación natural para la producción de voz. El contorno del tono de una expresión es usada para reconocer voces (recognizing speakers), y es requerido en casi todos los sistemas análisis - síntesis (vocoders).

De ahi que una gran variedad de algoritmos hayan sido propuestos en la literatura de procesamiento de voz.

En el presente capítulo se analizan las características generales de 7 algoritmos, haciendo mayor énfasis en 2 de ellos, que fuerón implementados en el microprocesador TMS 32010.

Clásificación espectral de los sonidos

La descripsión espectral esta basada en el contenido de frecuencias de los sonidos. Al excitar la caja resonante del tracto vocal con una fuente sonora, existen frecuencias a las cuales la transmisión de sonidos es más fuerte y estas representan las resonancias naturales, llamadas formantes. Los formantes son denominados: el primero (F1), el segundo (F2), el tercero (F3), en orden ascendente desde las más baja frecuencia como se muestra en la fig. 3.1 .

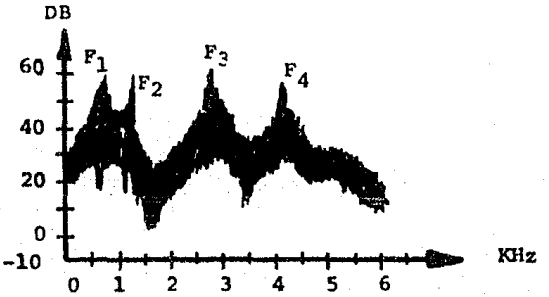


Fig. 3.1 Espectro de sonido de la "o".

En el espectro de frecuencias tambien se aprecia si el sonido es sordo ó sonoro. En el caso de sonidos sonoros sonoros como la "o" la energía esta concentrada en valores discretos de frecuencia. Para sonidos sordos como la "s", en donde la excitación corresponde a una fuente de ruido y desaparece la estructura armónica característica de los sonidos sonoros fig. 3.2.

La frecuencia fundamental Fo ó alternativamente , la diferencia de frecuencia entre picos fija el tono de la voz, su valor varía de una persona a otra, siendo relativamente alta 250 hz para niños, media 200 hz para mujeres y baja 125 hz para hombres.

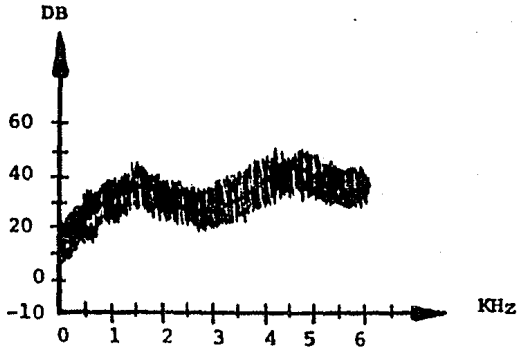


Fig. 3.2 Espectro de sonido de la "s".

3.2 Problemas en la deteccion de Fo.

La exactitud en la medida del periodo del tono, de una señal de voz desde la forma de onda de la presión acústica por sí misma, es excesivamente difícil por muchas razones, una de ellas es que la forma de onda de la excitación glotica , no es un tren perfecto de pulsos periodicos. Encontrar el periodo de una señal periodica puede ser cierta, no asi la medición en una señal de voz, ya que hay variaciones tanto en el periodo como en la estructura de la señal dentro de un mismo periodd, lo que dificulta su medición. Una segunda dificultad es la interacción entre el ducto vocal y la excitación glotica, en determinada instancia los formantes del ducto vocal pueden alterar significativamente, la estructura de la forma de onda de la glotis y por consiguiente el periodo del tono actual es dificilmente detectado, tales interacciones son más nocivas durante movimientos rápidos de las articulaciones, cuando los formantes estan cambiando tambien rápidamente. Un tercer problema en la medición fidedigna de Fo, es la dificultad inherente en definir el exacto principio y fin de cada periodo de tono , esta eleccion es frecuentemente arbitraria. Por ejemplo basado en la forma acústica por sí misma, algunos candidatos para definir el principio y fin del periodo incluye: el máximo valor durante el periodo, el cruce por cero antes del máximo etc. la fig. 3.3 muestra dos posibles estimaciones para definir una marca en el tono , basada directamente en la medida de la forma de onda.

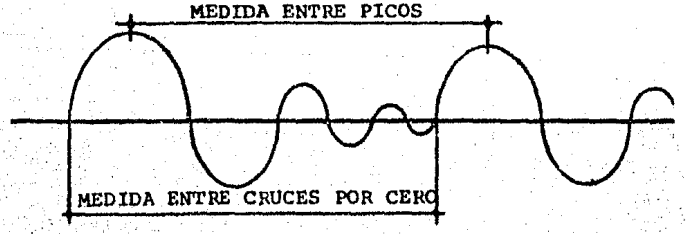


Fig. 3.3 Posibles marcas de tono

Las dos medidas de la forma de onda de la figura anterior dan diferentes valores para el tono, las discrepancias no son debidas solamente a la forma cuasiperiodica de la señal, por ejemplo: la medida entre picos es sensible a la estructura de los formantes, durante el periodo de tono, el cruce por ceros es sensible a formantes, ruido y a un determinado nivel de d.c. de la señal.

La cuarta dificultad es la distinción entre no señal de voz y un bajo nivel de señal. En muchos casos la transision entre segmentos de no-voz y segmentos de bajo nivel son muy agudos lo que dificulta hacer una buena decisión.

Se pueden además mencionar problemas adicionales como el de una señal de voz que ha sido transmitida telefónicamente, tales efectos en sistemas telefónicos incluyen: filtrado lineal, atenuación de Fo, procesos no lineales, distorsión de fase ruido etc . De modo que estos objetos obscurecen la estructura periodica de la señal y por consiguiente la detección de Fo.

3.3 Tipos de detectores de Fo

Basicamente un detector de tono es un dispositivo que realiza la decisión de voz- no voz y durante un periodo de voz provee, la medida del periodo del tono. Algunos detectores realizan la cuantización del tono y utilizan otra técnica diferente para la decisión voz - no voz .

Los algoritmos para la detección de Fo se pueden dividir en:

A) Utilizan principios en el dominio del tiempo.
B) Utilizan principios en el domonio de la frecuencia.
C) Utilizan una combinacion de ambos.

Los detectores en el dominio de del tiempo operan directamente sobre la forma de onda para estimar Fo , la medida es sobre picos y valles cruces por cero medidas de autocorrelación. La asumción básica es que si, una señal cuasiperiodica ha sido convenientemente procesada, para minimizar los efectos de la estructura de los formantes, entonces la simple medida en el dominio del tiempo proven una buena estimación.

La clase en el dominio de la frecuencia utiliza la propiedad de que si, la señal es periodica en el tiempo entonces el espectro de frecuencia de la señal consistirá, de una serie de impulsos en la frecuencia fundamental y sus armonicas, con una simple medida de ese espectro se puede obtener Fo, ó por medio de una versión de transformada no lineal como usa el detector cepstral [1].

Para los detectores híbridos se pueden usar por ejemplo: técnicas en el dominio de la frecuencia, para dar un espectro en tiempo de la forma de onda y entonces usar medidas de autocorrelacion para estimar Fo.

3.4 Criterios de evaluación de detectores de Fo.

Uno de los problemas más difíciles es encontrar un criterio de evaluación que nos permita escoger el mejor detector de Fo, ya que determinado criterio puede ser bueno para ciertas aplicaciones y para otras no. Hay muchas características en los algoritmos las cuales influyen en su elección, algunos factores son:

a) Exactitud en la estimación.
b) Exactitud en la decisión voz - no voz.
c) Robustez en la medida. Ellos deben ser modificables para diferentes

condiciones de transmision, de voces.
d) Velocidad de operación.
e) Complejidad del algoritmo.
f) Fácil implementación en hardware.
g) Bajo costo de implementación.

Para nuestro caso el realizar un sistema análisis - síntesis de voz en tiempo real, nos interesa implementar un algoritmo de poca complejidad y gran velocidad de operación, además de una buena exactitud en la decisión voz- no voz y en la detección de Fo. Estas características son deseables ya que su implementación se hará en el TMS 32010 de aritmética entera con limitaciones de memoria pero de gran rapidez.

Otro criterio no mencionado y que puede tomarse en cuenta: es la exactitud perceptual del detector de tono por ejemplo, la pregunta de que tan fiel el contorno del tono medido por el detector, iguala el contorno del tono de exitación natural en término de la calidad sintetica de la voz.

En las siguientes secciones de este capítulo se analizan 7 algoritmos para la detección de Fo, dos toman principios en el dominio del tiempo : Método de reducción de datos y procesamiento en paralelo , dos que utilizan autocorrelación: Correlación modificada, función de diferencias de magnitud promedio, uno en frecuencia: Cepstrum y dos hibridos: Tecnica simplificada de filtro inverso , ecualizacion espectral LPC.

En el presente trabajo se implementarón los dos primeros algoritmos: Método de procesamiento en paralelo [2], el de función de diferencias de magnitud promedio [1], que utiliza correlación. Estos algoritmos se especifican con más detalle que los posteriores. En su implementación se logró una mejor calidad no-sintetica con el segundo algoritmo.

3.5 Procesamiento en paralelo

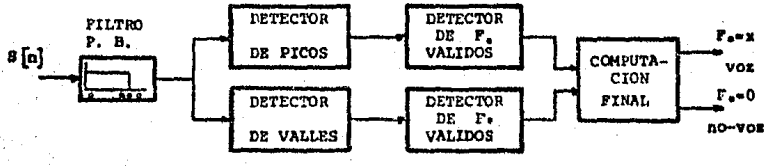


Fig. 3.4 Procesamiento en paralelo

Este algoritmo fué propuesto por Gold y Rabiner [2], puede ser dividido en cuatro partes:

a) Filtrado de la señal de voz.

El próposito del filtrado es seleccionar aproximádamente la primera región de formantes, otra información no es necesaria, picos causados por altos formantes tienden a reducir la exactitud del tono subsequente es importante la elección de este filtro, ya que de ello dependerá que

este presente la frecuencia fundamental, por ejemplo en un filtro paso bajas a 600 hz esta presente Fo, en un paso banda de 300 a 900 hz al menos las dos armónicas más altas están presentes. En nuestro caso utilizamos un filtro elíptico paso bajas de sexto orden.

b) Detector de picos y valles

En esta sección sólo se detactan los picos valles, tomando en cuenta sus cruces por cero se tomaron bloques de análisis de 20 mseg, para una frecuencia de muestreo de 6.4 khz equivale a 128 muestras, se les asigna magnitud y posición dentro del bloque.

c) Detector de Fo válidos

Guardados ya en memoria un determinado número de vectores con magnitud y posición dentro del bloque, uno para picos y otro para valles la detección es como sigue :

Se empieza con la primera magnitud en el bloque, se procede a buscar el siguiente valor note segun la figura 3.5 que hay un intervalo de tiempo, en la que no se puede detectar el siguiente vector hasta que no termine un determinado tiempo lamado muerto, después del cual viene un decaimiento de la amplitud original hasta encontrar un valor mayor a esta en ese instante en que ocurre la segunda magnitud pico ó valle , si se encuentra ese vector se resetea el sistema y comienza una nueva busqueda si no, seguirá el decaimiento exponencial hasta encontrar un vector que cumpla con esta condición, la constante de tiempo de decaimiento exponencial debe ser tal que un vector A1 cualquiera alcanze la mitad de su valor original en 5 mseg. por lo tanto se tiene:

$$A1\, e^{(-0.005\, t)} = A1/2 \qquad \text{por lo tanto} \quad t = -200 \ln 0.5$$

Si ese tiempo de decaimiento es mayor que 16mseg. se resetea el sistema y detecta no-voz es decir Fo=0. La duracion del tiempo muerto fué de 2mseg. con la cual se obtubieron buenos resultados.

Hay que hacer notar que el decaimiento exponencial no inicia hasta que no termina el tiempo muerto.

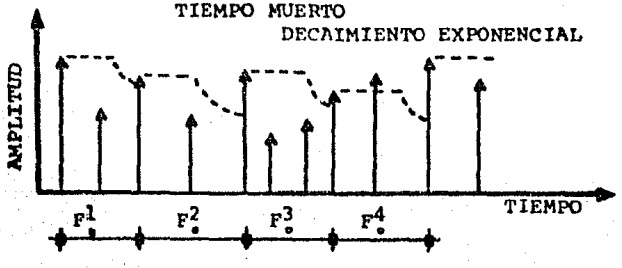


Fig. 3.5 Detección de valores válidos

d) Estimación final de Fo

Después de detectar las Fo válidas de la sección anterior, tanto de picos como de valles se procede a sacar simplemente una media de estos

valores o sea:

$$Fo = \frac{1}{N} \sum_{i=1}^{N} Fo^{i}$$

La detección no tiene historia y al analizar el siguiente bloque de 20 mseg se resetea todo, los valores posibles de Fo que se pueden detectar son:

$$100 \text{ hz} \leq Fo \leq 250 \text{ hz}.$$

3.6 Funcion de diferencias de magnitud promedio [1].

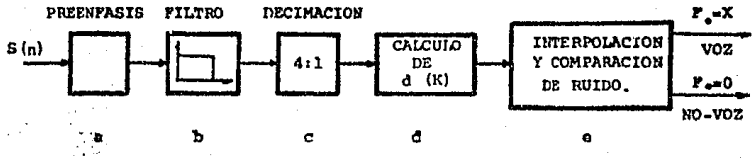


Fig. 3.6 Funcion de diferencias de magnitud promedio

Este algoritmo puede ser dividido en cinco partes:

3.6 a) Preénfasis

El preénfasis es importante porque acentua los formantes altos y se obtienen funciones razonables del área del ducto vocal, la funcion usada en Z es:

$$S(Z) = 1 - 0.5\ Z^{-1}$$

3.6 b) Filtrado

Se utilizó un filtro elíptico de sexto orden con frecuencia de corte a 800 hz.

3.6 c) Decimación

Se tomó un bloque de 40 mseg. que para 6400 hz de muestreo equivalen a 256 muestras, obteniendose después de la decimación 4:1 un bloque de de 64 muestras.

3.6 d) Cálculo de las f.d.m.p.

Se procede a calcular las funciones:

$$d[k] = \frac{\sum_{n=k+1}^{N=64} |S[n-k] - S[n]|}{64 - k} \qquad \text{para } k=1,2 \ldots 26 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6.1)$$

donde $d[k]^N$ significa el calculo de la función d[k] hasta la muestra N = 64. Esta función definida en la ecuacion 3.6.1 puede ser cálculada en el momento que llega la siguiente muestra N +1 , sin necesidad de esperar el bloque completo, de la ecuación anterior se puede facilmente verificar que:

$$d[k]^{N+1} = d[k]^N - |S[1] - S[1+K]| + |S[n_o - k] - S[n_o]| \quad \text{para } k = 1,2 \ldots 26 \qquad \ldots\ldots (3.6.2)$$

Por ejemplo si nos llega la muestra 65, nn = 65 ya tenemos cálculada la función d[k]N en N=64 por consiguiente para k=5,6...26 de la ecuación 3.6.2.

$$d[5]^{65} = d[5]^{64} - |S[1] - S[6]| + |S[60] - S[65]|$$

$$d[6]^{65} = d[6]^{64} - |S[1] - S[7]| + |S[59] - S[65]|$$

.

.

.

$$d[26]^{65} = d[26]^{64} - |S[1] - S[27]| + |S[39] - S[65]| \quad \ldots\ldots\ldots\ldots (3.6.3)$$

El problema en esta sección es calcular $d[k]^{N+1}$ siempre que aparezca una nueva muestra decimada anteriormente. Ya que obtuvimos las funciones de diferencia d[k], obtenemos la mínima función para esto hay que darles el mismo peso, pues las sumatorias son de magnitudes distintas, se normaliza dividiendo entre 64-K .

Obtención del posible valor de tono Fo.

Una vez obtenidas las funciones d[k] con el mismo peso, se procede a escoger la mínima función d[k], él por qué se escoge la mínima función se puede entender si observamos las ecuaciones 3.6.3, esas funciones estan definidas como una diferencia de valores absolutos con k variable esa sumatoria será igual a cero, si se hace la diferencia de la misma señal defasada 360º .

Ya obtenida la posible frecuencia fundamental Fo la real será igual a k+4 por la decimación. Si se tomara esa decisión se estaría perdiendo precisión en la elección de Fo por qué se escogió a la k que hace mínima a d[k] sobre 64 muestras decimadas y no sobre las 256 muestras originales, por lo que es necesario hacer una interpolación. La parábolica fué usada y esta dada por:

AA = d[k] - d[k]

BB = (AA + d[k+1] - d[k]) / 2

CC = (d[k+1] - d[k-1]) / 4

Fo =(k + 4) - BB / CC

En donde d[k] es la función mínima obtenida de las ecuaciones 3.6.3.

Decisión voz no-voz

Una vez obtenida la función mínima d[k], se procede a cumpara ese valor con uno fijado anteriormente para el ruido del sistema, si esa d[k] es mayor que el nivel fijado para el ruido hay voz y su Fo será como el cálculado anteriormente, si es mayor el ruido se hace Fo = 0 y

hay decisión de no-voz.

3.7 Método modificado de autocorrelación

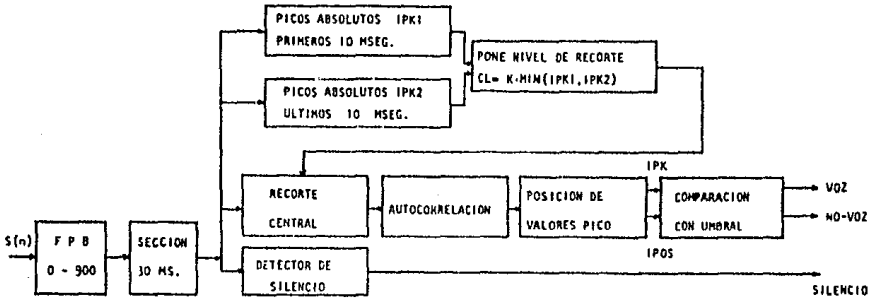


Fig 3.7 Método modificado de autocorrelación

Esta basado en el método de recorte en el centro de Sondhi, la frecuencia de muestreo es de 10 khz, la detección del pitch se hace cada 10 mseg., el primer paso es la computación del nivel de recorte Cl de los 30 primeros mseg, el valor de recorte es puesto a 64% del pico más pequeño de las secciones anteriores, en el recorte central resulta una señal que asume tres posibles valores: +1, -1 ó 0. Después se realizan las fuinciones de autocorrelación sobre un rango de retraso desde 20 a 200 muestras, es computada además la correlación de 0 retraso para propósitos de normalización, después es buscado el máximo valor normalizado de la autocorrelación, si ese valor excede un determinado umbral (0.3) es clasificado como voz. El detector de silencio sólo clasifica a la señal como silencio ó no de acuerdo a un determinado nivel de la señal de entrada.

3.8 Método Cepstral

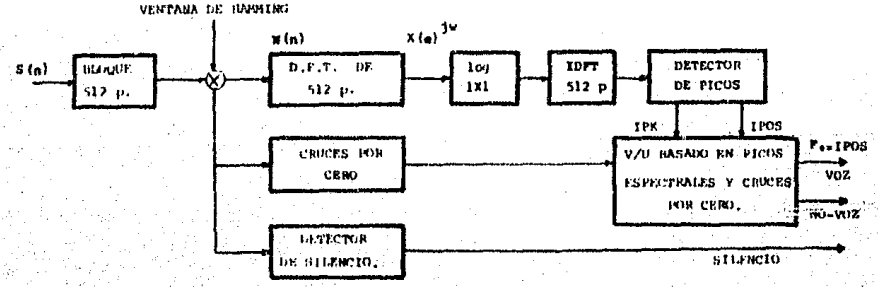


Fig. 3.8 Método Cepstral

Los valores picos espectrales y su localización son determinados y si ese valor pico excede el umbral fijo es voz y la frecuencia fundamental Fo es la localización de ese pico, si el pico no excede el umbral el cruce por cero es hecho, si el conteo de cruces por cero excede el umbral dado detecta no-voz, de otro modo hay voz y Fo es la localización del pico máximo. El detector de silencio se basa en el nivel de la señal.

3.9 Técnica simplificada de filtro inverso.

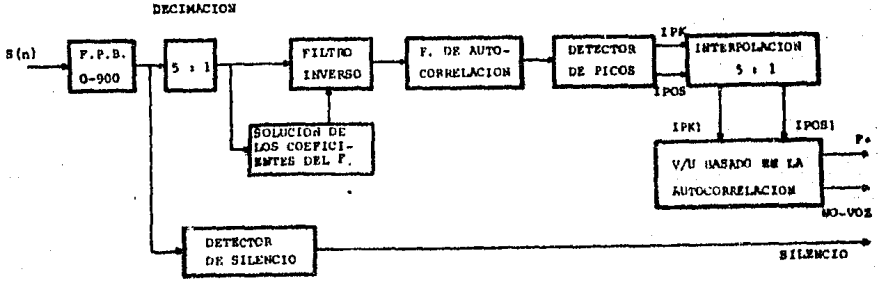


Fig. 3.9 Técnica simplificada de filtro inverso.

La frecuencia de muestreo es de 10 khz. se procesan 400 muestras los coeficientes de un filtro inverso de cuarto orden son obtenidos utilizando el método de autocorrelación LPC, la decisión de voz no -voz es hecha en base a la amplitud de los picos de la función de autocorrelación, el umbral en la prueba se situa a 0.4 de los picos de autocorrrelación.

3.10 Método de reducción de datos

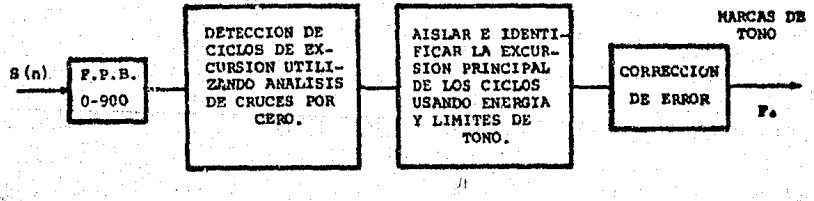


Fig. 3.10 Método de reducción de datos.

Este detector pone marcas directamente del filtro, detector de Fo síncronos. Para obtener las marcas detecta la excursión de los ciclos de la forma de onda, tomando solo las excursiones válidas, de acuerdo al valor esperado de Fo, no hay un cálculo inherente de voz no-voz.

3.11 Ecualización espectral LPC utilizando el método de Newton.

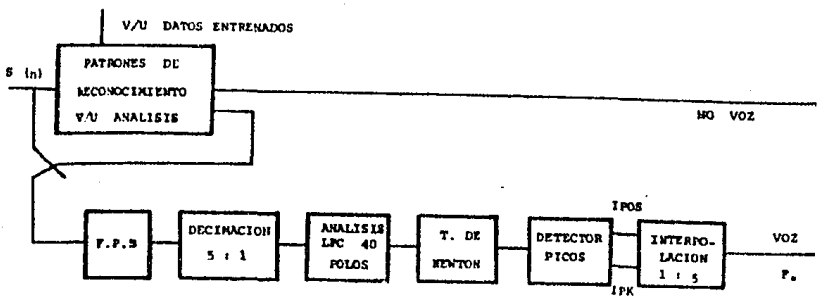


Fig. 3.11 Ecualizacion espectral LPC.

Esquema propuesto por Atal. La primer sección se compone de un detector, el cual usa técnicas de reconocimiento de patrones para clasificar cada intervalo de 10 mseg. como voz ó no-voz, ya clasificada la señal como voz, un análisis LPC de 41 polos es realizado en bloques de 40mseg la transformación de Newton es uzada para hacer un aplanado espectral de la señal y así transformar la señal en una que tenga los picos más pronunciados en los impulsos de Fo y es aproximadamente cero en cualquier otro caso. El éxito de este método depende primordialmente de que tan válidos sean los datos de entrada, que caracterizan las diferentes voces. Rabiner [1] con un cuidadiso tren de datos obtuvo el 99% de éxito en tal decisión.

3.12 Conclusiones.

Tomando como base el estudio relizado por Rabiner [1], en el cual se hace un análisis extensivo de errores en: La detección de Fo, en la decisión voz no-voz, así como en el tiempo de computación requerido para cada algoritmo que fué simulado en fortran. Siendo el tiempo más bajo para el método de Reducción de Datos y Procesamiento en Paralelo con artitmética entera que llevó sólo 5 y 7.5 seg. respectivamente y los más lentos los métodos Cepstral y LPC , con aritmética de tipo flotante que se llevarón 300 y 400 segundos de computación.

En el presente trabajo se implementarón en el TMS 32010 los métodos: Procesamiento en Paralelo (P.P.), Función de Diferencias de Magnitud Promedio (F.D.M.P.), tomando en cuenta su tiempo total de computación se obtuvieron los siguientes resultados:

Para P.P. el tiempo fué de 106.8 microsegundos y para F.D.M.P. 400.4 microsegundos, obteniendose mejor calidad no sintética para síntesis utilizando el segundo método.

4. SINTESIS

La técnica empleada para Síntesis cae dentro de los codificadores de voz que suponen un conocimiento de como la señal se ha generado en la fuente, denominandosele al sintetizador paramétrico Codificador de Fuente o comunmente Vocoders (contracción de Voice Coders).

Como se presentó en la Sección 1, el proceso de Síntesis esta basado en la utilización de los coeficientes de reflexión del modelo del Tubo Acústico visto en el capítulo anterior. La expresión representativa del Proceso Síntesis es $G(Z)=P(Z)/A(Z)$, donde los polinomios P(Z) y A(Z) son de la forma

$$P(Z)=P_M(Z)=\sum_{m=0}^{M} P_{Mm}Z^{-m}$$

$$A(Z)=A_M(Z)=\sum_{m=0}^{M} a_{Mm}Z^{-m}$$

4.1 Estabilidad

La estabilidad de cualquier filtro de la forma P(Z)/A(Z) o 1/A(Z) está determidada por las raices del polinomio A(Z). Puesto que estos coeficientes son modificados cada pocos milisegundos, no es facil determinar la estabilidad del filtro. Por lo que en el estudio de la estabilidad se les suele considerar como constantes y si alguna de las raices se encuentra sobre o fuera del circulo unitario $|Z|=1$ se considera el filtro como inestable.

Existen estudios que demuestran que si $A_m(Z)$ tiene sus raices dentro del círculo unitario, entonces $A_{m-1}(Z)$ tiene tambien sus raices dentro del círculo unitario. Puede afirmarse que

- -Si $1/A_m(Z)$ es estable, $1/A_{m-1}(Z)$ es estable.
- -La estabilidad de $1/A_{m-1}(Z)$ es necesaria para la estabilidad de $1/A_m(Z)$.

Expresado por los coeficientes de reflexión, la estabilidad de 1/A(Z) se cumple si

$$|Km|<1 \quad \text{para } m=M,\ M-1,\ \ldots\ 1$$

4.2 Estructura General del Filtro

En el capítulo de Análisis fueron definidos los polinomios $A_m(Z)$ y $ZB_m(Z)$, ademas, se vio que el filtro inverso A(Z) podia ser obtenido recursivamente utilizando

$$A_m(Z)=A_{m-1}(Z)+Km\ B_{m-1}(Z) \qquad (4.21a)$$

$$ZB_m(Z)=Km\ A_{m-1}(Z)+B_{m-1}(Z) \qquad (4.21b)$$

con $m=1,2,\ldots\ m$; $A_0(Z)=1$; $ZB_0=1$

P(Z) expresado mediante el polinomio $ZB_m(Z)$ es

$$P(Z)-P_M(Z)=\sum_{m=0}^{M}\nu_m ZB_m(Z)$$

donde ν_m es constante de multiplicación. Sustituyendo P(Z) en la expresión del filtro Síntesis

$$G(Z)=P(Z)/A(Z)$$

$$=\sum_{m=0}^{M}\nu_m[ZB_m(Z)/A(Z)]$$

Definiendo E(Z) como la entrada y X(Z) como salida, la salida del filtro es $X(Z)=G(Z)E(Z)$

$$=\sum_{m=0}^{M}\nu_m[ZB_m(Z)E(Z)/A(Z)]$$

Contando con las anteriores expresiones, es posible obtener una estructura general del Filtro Síntesis, como se desea, a partir de los coeficientes de reflexión y dos tipos de parámetros mas: Parámetros de Derivación(Tap) ν_m y Parámetros-pi(π_m).

Las expresiones (4.21) al ser multiplicadas por $E(Z)/A_M(Z)=E(Z)/A(Z)$ resultan en

$$\frac{A_m(Z)E(Z)}{A(Z)}=\frac{A_{m-1}(Z)E(Z)}{A(Z)}+\frac{K_mB_{m-1}E(Z)}{A(Z)}$$

$$\frac{ZB_m(Z)E(Z)}{A(Z)}=\frac{K_mA_{m-1}(Z)E(Z)}{A(Z)}+\frac{B_{m-1}(Z)E(Z)}{A(Z)}$$

$$\frac{ZBo(Z)E(Z)}{A(Z)}=\frac{Ao(Z)E(Z)}{A(Z)}=\frac{E(Z)}{A(Z)}=X(Z) \qquad (4.22)$$

En la última expresión, desde luego, $Ao(Z)=1$.

Para $m=M$ $E(Z)=A_M(Z)E(Z)/A(Z)$. Se introducen a continuación los parámetros pi

$$\hat{A}_m(Z)=A_m(Z)\pi_m \qquad (4.23a)$$

$$\hat{B}_m(Z)=B_m(Z)\pi_m \qquad (4.23b)$$

$$\hat{\nu}_m=\nu_m/\pi_m \qquad (4.23c)$$

Dado $\pi_M=1$, $\hat{A}_M(Z)=A_M(Z)=A(Z)$

Definiendo ahora

$$E_m^{+}(Z)=\hat{A}_m(Z)E(Z)/A(Z) \qquad (4.24a)$$

$$E_m^{-}(Z)=\hat{B}_m(Z)E(Z)/A(Z) \qquad (4.24b)$$

Despues de afectar a (4.21) mediante las expresiones (4.23) y (4.24) se obtiene:

$$E_{m-1}^{+}(Z)=\frac{\pi_{m-1}}{\pi_m}E_m^{+}(Z)-K_mE_{m-1}^{-}(Z) \qquad (4.25a)$$

$$ZE_m^{-}(Z)\frac{\pi_{m-1}}{\pi_m}=K_mE_{m-1}^{+}(Z)+E_{m-1}^{-}(Z) \qquad (4.25b)$$

Obteniendose ahora la respuesta X(Z) como

$$X(Z)=G(Z)E(Z)$$

$$L(Z)=E_M^{+}(Z)=\hat{A}_M(Z)\,E(Z)/A_M(Z)$$

$$X(Z)=\sum_{m=0}^{M}\hat{\nu}_m Z\hat{B}_m(Z)E(Z)/A(Z)$$

y finalmente

$$X(Z)=\sum_{m=0}^{M} \hat{v}_m E_m^-(Z) \qquad (4.26)$$

En la figura 4.1 se muestra la extructura de las expresiones (4.25) y (4.26).

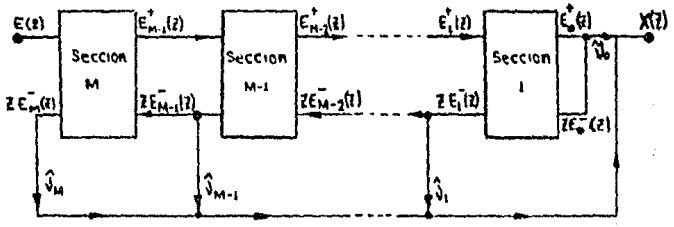


Fig 4.1 Estructura General de P(Z)/A(Z).

4.3 Filtro tipo Rejilla

Durante las dos últimas décadas se han desarrollado diversas estructuras que eficientemente implementan el filtro P(Z)/A(Z). Una de las mas conocidas es el Tipo Rejilla Todo-Polos Doble-Multiplicación, el cual es el utilizado en el presente trabajo.

Los parámetros pi se definen como $\pi_m=1$, para m=0,1,...,M y sustituidos en las expresiones (4.25) se obtiene:

$$E_{m-1}^+(Z)= E_m^+(Z) - KmE_{m-1}^-(Z) \qquad (4.31a)$$
$$zE_m^-(Z)= KmE_{m-1}^+(Z) + E_{m-1}^-(Z) \qquad (4.31b)$$

Para m=M, M-1,...,1

Los parámetros de Derivación se reducen a

$$\hat{v}_m= v_m/\pi_m= v_m$$

La estructura obtenida se muestra en la figura 4.2

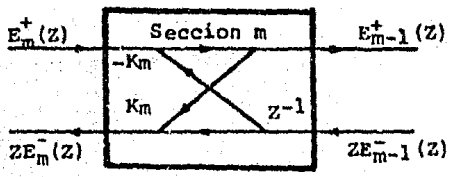


Fig. 4.2 Estructura del Filtro Rejilla Doble Multiplicación.

Durante la generación de sonidos nasales y fricativos es necesaria la presencia de polos y ceros en la función de transferencia del ducto vocal. Mediante un modelo Todo-Polos no se podrian generar estos sonidos, sin embargo, si el orden del filtro es suficientemente grande, es posible mediante un modelo Todo-Polos generar casi todos los sonidos de la voz. Se optó por la implementación de un modelo de este tipo por reducirse a una estructura sencilla.

En este caso, de la expresión (4.26) se considera:

$$ZE_o^-(Z)/\Pi_o = E_o^+(Z)/\Pi_o$$
$$X(Z) = \hat{V}_o ZE_m^-(Z) = ZE_m^-(Z)/\Pi_o$$

y finalmente $X(Z) = E_o^+(Z)/\Pi_o$

Se observa que los parámetros $\hat{V}_1 = \hat{V}_2 = \ldots = \hat{V}_M = 0$, influyendo solamente $\hat{V}_o$ en la respuesta del filtro, asi

$$x(n) = e_o^+(n)\hat{V}_o = e_o^+(n)/\Pi_o$$

4.4 Exitacion y Programa Sintesis

En el trabajo desarrollado cada muestra sintetizada presenta dos componentes:

- -Valores de una exponencial compleja decayente q(n) del anterior período de frecuencia fundamental (pitch).
- -Una salida u(n) que representa a una exitación sin considerar los efectos del período de frec. fundamental anterior.

En la figura 1.2 se muestra un esquema general de Síntesis.

La exitación e(n) es una serie de pulsos periódicos para señal tipo voz, o un generador de valores semi-aleatorios para señal tipo No-Voz.

Mediante un factor de ganancia g se puede representar la salida $\hat{s}(n)$ para la nueva porción de síntesis

$$\hat{s}(n) = q(n) + g\,u(n) \qquad (4.41)$$

Se usa una barra "——" indicando la suma de N muestras. Si se considera ahora que u(n) contiene tambien a q(n), la energia de (4.41) se expresa como

$$\overline{s^2(n)} = \overline{\hat{s}^2(n)} = g^2\,\overline{u^2(n)}$$

Se encuentra ahora la exitación e(n) apartir del parámetro de ganancia calculado por análisis($G = \alpha\,\sigma^{1/2}$). Para N datos muestra por cada franja de análisis el error cuadrático obtenido es $\sigma^2 = \alpha G$

Si el generador aleatorio de muestras g(n) posee variancia σ_g^2, para igualar la energia sobre las N muestras, e(n) debe expresarse como

$$e(n) = g(n)G/(\sigma_g N^{1/2}) \qquad \text{para No-Voz}$$

Algunos autores [5] consideran que no es factor muy importante la función de distribución de probabilidad de las muestras aleatorias.

Para el caso de señal tipo Voz, e(n) debe ser un tren de pulsos unitarios separados por un número entero I de muéstras, representando el período de frec. fund. (pitch). Esta exitación debe tener, ademas, un

valor medio de cero:

$$e(n)=\begin{cases} \sqrt{(I/N)^{1/2}} & \text{para } n=0,1,2I,\ldots \\ -\sqrt{(I/N)^{1/2}/(I-1)} & \text{para } n\neq 0,1,2I,\ldots \end{cases}$$

El generador de pulsos emplado se muestra en la figura 4.3. Se expresa como

$$e(n)=\begin{cases} a & \text{para } n=1 \\ b & \text{para } 1<n\leq P \end{cases}$$

Donde **a** y **b** estan determinados por

$$a + (P-1)b=0$$
$$[\ a^2 + (P-1)b^2\ ]/P=1$$

Con valor medio cero y ademas energia unitaria promedio para un periodo dado de frec. fundamental P(Pitch).

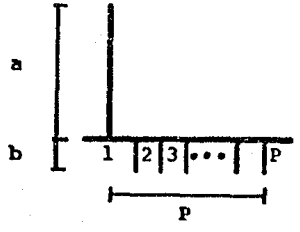


Fig. 4.3 Forma del Generador de pulsos para señal Tipo-Voz.

El diagrama de flujo del programa se muestra en la Sección 5,algunas consideraciones acerca de éste son:

- -Se emplea el filtro tipo rejilla de Doble-Multiplicación de la sección 4.3.
- -Se efectuan interpolaciones segun valor de frec. fundamental para calcular la ganancia y nuevo valor de frec. fundamental empleados por sintesis.
- -Como se mencionó antes, la ganancia se calcula mediante la energía de la señal de error calculada por análisis.

Se utilizan para la interpolación dos franjas de datos de análisis, referidas como franja izquierda y derecha, por lo que el proceso de sintesis proporciona resultados con un atrazo de al menos una franja de datos con respectos a la señal que entra a análisis. Las dos primeras franjas de muestras sintetizadas son solo para cargado de condiciones iniciales, por tanto, sus resultados no son válidos.

Si el contador IPC > IPITCH, IPC se convierte en 1 comenzando un nuevo período de frec. fundamental. Cuando IPC es menor a la constante longitud de franja(IFLGTH), se prueba si ambas franjas son señal tipo Voz, de ser así, se interpola para obtener el nuevo período de frec.

fundamental y el factor de ganancia(GAIN). Después de interpolar se continua sintetizando muestras hasta que el contador IPC exceda a IPITCH.

Cuando el contador de franja IPC exceda a la constante IFLGTH, se actualizan las franjas derecha e izquierda, transfiriendo los datos de la derecha a la izquierda. Los nuevos datos leídos son guardados en la franja derecha. Estos nuevos datos son el resultado del proceso de Análisis efectuado por el microprocesador correspondiente, los datos son M=10 coeficientes de reflexión, el valor de ganancia y el de frecuencia fundamental (pitch).

5. Implementacion

Se mencionan a continuación las principales consideraciones que se tuvieron en cuenta para la determinación y aplicación de ciertos parámetros importantes, asimismo el arreglo de las tarjetas con los microprocesadores y Diagramas de Flujo de los programas implementados.

5.1 Consideraciones

5.1.1 Frecuencia de Muestreo Prefiltrado

Considerando 17 cm. como valor promedio de longitud del ducto vocal, las tres primeras frecuencias del formante caerán aproximadamente en el rango de frecuencias de 250-280 Hz. Para ductos vocales mas cortos, por ejemplo los de mujeres y niños, la frecuencia de los formantes estará entre 300-3500 Hz (puede notarse que la frecuencia formante es inversamente proporcional a la longitud del ducto vocal).

La razón de muestreo se desea sea baja, pero sin destruir las características significativas de la señal analizada, por lo que se debe muestrear según el Criterio de Nyquist.

En una estimación precisa de la voz, la frecuencia de muestreo debe ser mayor a 6 KHz para tener un ancho de banda de al menos 3 KHz.

La percepción de algunas consonantes disminuye ligeramente si se omiten frecuencias entre 3 y 5 KHz. Frecuencias sobre 5KHz no son de mucha ayuda en el mejoramiento de la claridad y naturaleza de la voz.

Para el caso de sonidos fricativos, por ser éstos de baja intensidad es necesario tener una razón Señal-ruido muy grande, además de una frecuencia de muestreo mas alta debido a que su frecuencia puede ser entre 8 y 10 KHz.

Antes de muestrear es necesario prefiltrar la señal, pasándola primero por un filtro paso-alto de frecuencia de corte en 100 Hz para eliminar el ruido de la red eléctrica y los comprendidos bajo este rango, sin suprimir la frecuencia fundamental de vibración de las cuerdas vocales. Después se pasa la señal por un filtro paso-bajo con frecuencia de corte en aproximadamente $f_m/2$ para limitar en banda la señal.

La frecuencia de muestreo usada fué de 6.4 KHz.

5.1.2 Orden del Filtro

Para representar el ducto vocal en circunstancias ideales, la memoria del filtro A(Z) debe ser dos veces el tiempo necesario para que las ondas sonoras recorran el ducto vocal, esto es $2l/c$, siendo L la longitud del ducto vocal y c la velocidad del sonido. Por lo que si $c \approx 34$ cm/ms y L=17 cm es necesaria una memoria de 1 ms.

La relación entre frecuencia de muestreo f_m, orden del filtro M, L y c es

$$f_m = 1/T$$

$$T = 2l/c = 2Ml/Mc; \qquad f_m = Mc/2L$$

Para $f_m = 6.4$ KHz, el orden del filtro M debe ser al menos de 7 redondeado al entero mas alto. Pero como el modelo glotal y la radiación caracteristica de los labios no esta tomada en cuenta, M=7 debe ser tomado como límite inferior.

Considerando además que la señal muestreada ha sido ya prefiltrada, es necesario modificar el orden del filtro, J. Markel sugiere una corrección basada en resultados experimentales y que considera la frecuencia de muestreo:

$$M=MN + f_m$$

Con f_m en KHz, siendo MN igual a 4 o 5. Otros autores sugieren que MN tome valores en un rango mas amplio, esto es, entre 2 y 5.

A los filtros implementados, A(Z) y 1/A(Z) se les asignó un orden de M=10.

3.1.3 Preenfasis

Si las caracteristicas espectráles del ducto vocal sin los efectos glotales y de radiación labial son estimados, la señal de voz debe ser preenfatizada antes de su análisis. Como fue presentado en el Capítulo 1, la forma de hacerlo es aplicando un filtro de primer orden de la forma $1-\mu Z^{-1}$.

Puede optarse tambien por un preenfasis adaptivo dado por $\mu = r_m(1)/r_m(0)$, donde μ es usualmente pequeño para sonidos No-Voz y aproximado a la unidad para sonidos Voz.

3.1.4 Ventana Hamming

El procesamiento directo de la forma de onda de la voz cae dentro de los métodos de procesamiento en el dominio del tiempo. Por medio de éstos se evalua la energía y las funciones de autocorrelación.

De la observación directa de las variaciones de la voz, se concluye que ésta varia muy lentamente para porciones pequeñas de tiempo.

Para su análisis se le considera como estacionaria y puede dividirse en franjas de entre 10-40 ms de duración.

La elección de la duración de la franja esta determinada principalmente, y ya discutido en la Sección 3, por el método empleado para la detección de la frecuencia fundamental.

El dividir la señal de voz en franjas equivale a la multiplicación por una Ventana Rectangular de magnitud constante (equivalente a su vez a un filtro paso-bajo). En este trabajo se prefirio el empleo de Ventanas Hamming, ya que son muchas ventajas que éstas presentan sobre la ventana rectangular, como lo son:

- -Para una misma longitud, el ancho de banda de una ventana Hamming es casi el doble del ancho de banda de una ventana rectangular.
- -La ventana Hamming proporciona mayor atenuación en la banda de rechazo que una ventana rectangular.
- -El espectro de una señal, pasada por una ventana Hamming tiene una forma mas plana que el equivalente de una ventana rectangular.

Las ventanas empleadas son de 256 puntos, presentando traslape en 128 de ellos.

5.2 Arreglo de las Tarjetas

Las operaciones necesarias para implementar el Análisis y Síntesis del presente trabajo fueron desarrolladas en dos juegos idénticos de tarjetas de computación. En la figura 5.1 se muestra el arreglo.
Cada juego de tarjetas esta compuesto por:

-Un modulo de Evaluación para procesamiento digital de señales(EVM).
-Una tarjeta de interface Analógica (AIB).

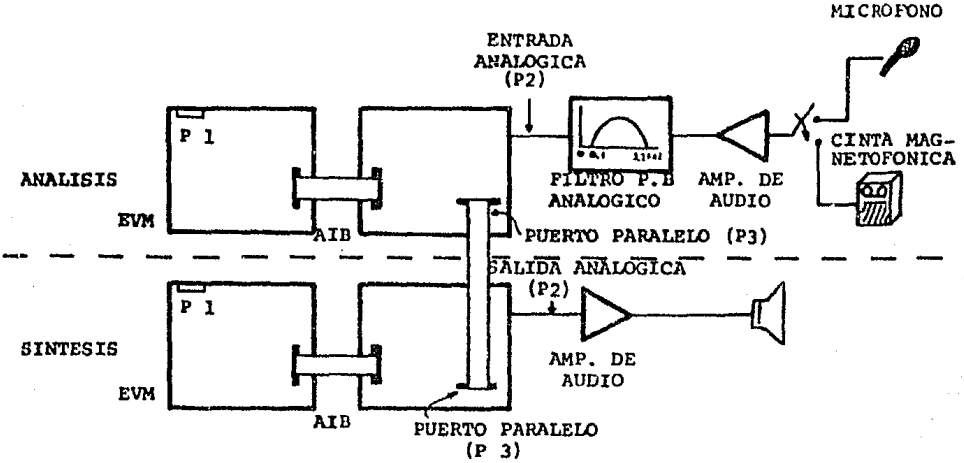


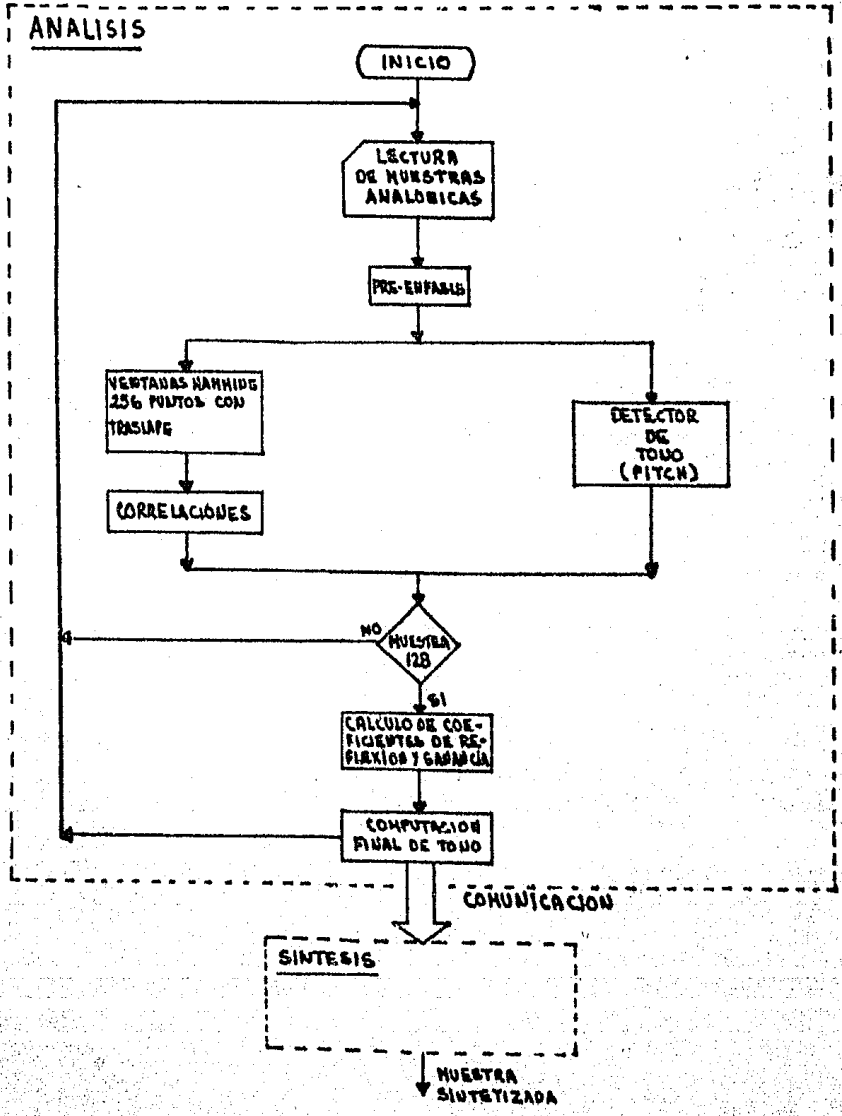
Fig. 5.1 Arreglo de las Tarjetas usadas

El Modulo de Evaluación es una tarjeta de desarrollo para el microprocesador TMS32010 con funciones de edición, ejecución y depurado de programa. La tarjeta AIB posee convertidores D/A y A/D de 12 bits a ±10 volts. Mayor información sobre las tarjetas y el microprocesador TMS32010 se encuentra en el Apéndice de este trabajo.

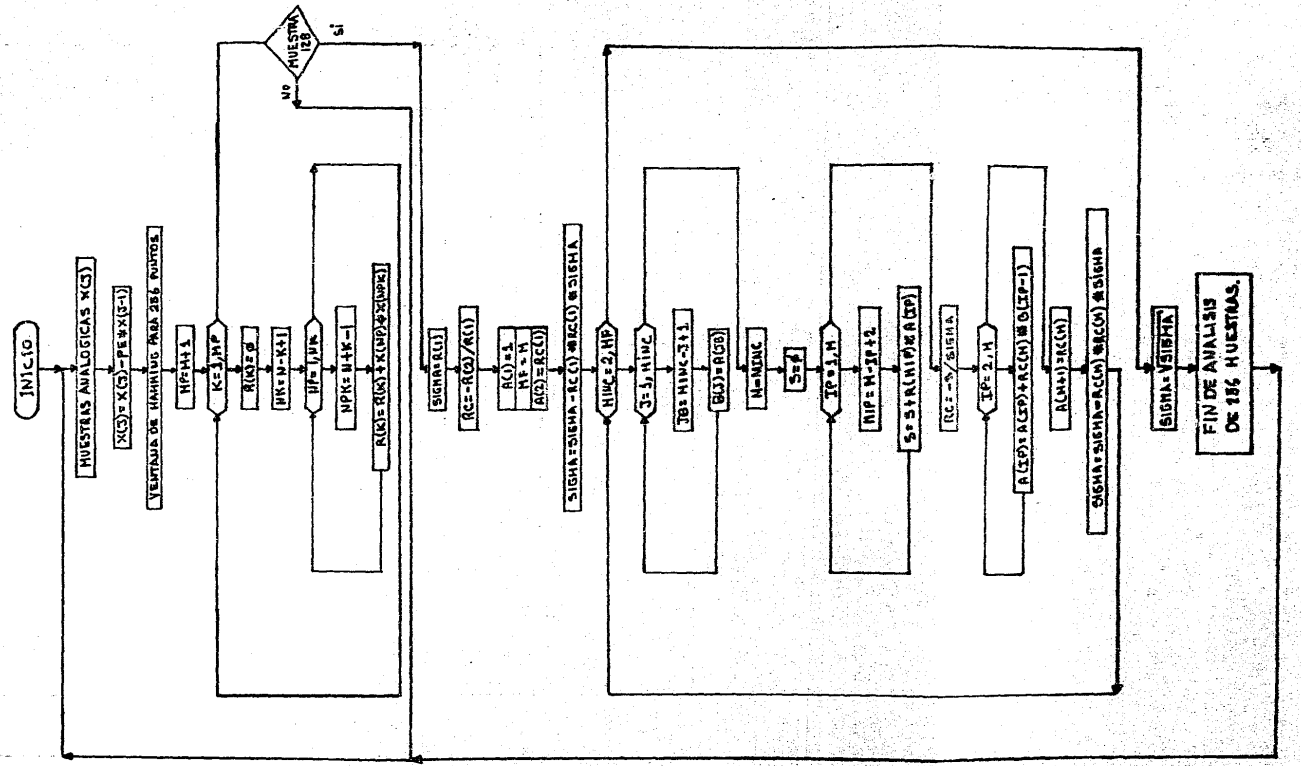
Durante el Análisis la tarjeta de interface se encarga de muestrear la señal analogica, poniendola a disposición del EVM en forma digital. Se efectuan entonces las correlaciones de Analisis y principia la detección de la Frecuencia Fundamental (pitch), hasta que se alcanzan 128 muestras. Se procede a terminar entonces el Análisis y la detección de la Frecuencia Fundamental.
Cuando se tienen calculados los parámetros de todo análisis (10 coeficientes de reflexión, 1 valor de ganancia y 1 valor de frec. fundamental expresado en muestras), se envian al otro juego de tarjetas, via Puerto 3 (paralelo) de la tarjeta de interface para su síntesis.
En la tarjeta de Síntesis, las muestras sintetizadas son sacadas al exterior via Puerto 2 de la tarjeta de Interface correspondiente, a la misma frecuencia de muestreo que en Análisis (8.4 KHz).
Las tarjetas de Análisis y Síntesis trabajan en forma totalmente independiente necesitando solamente sincronia entre ellas durante la transferencia de los resultados del Análisis hacia Síntesis.

5.3 Diagramas de Flujo

5.3.1 Diagrama General

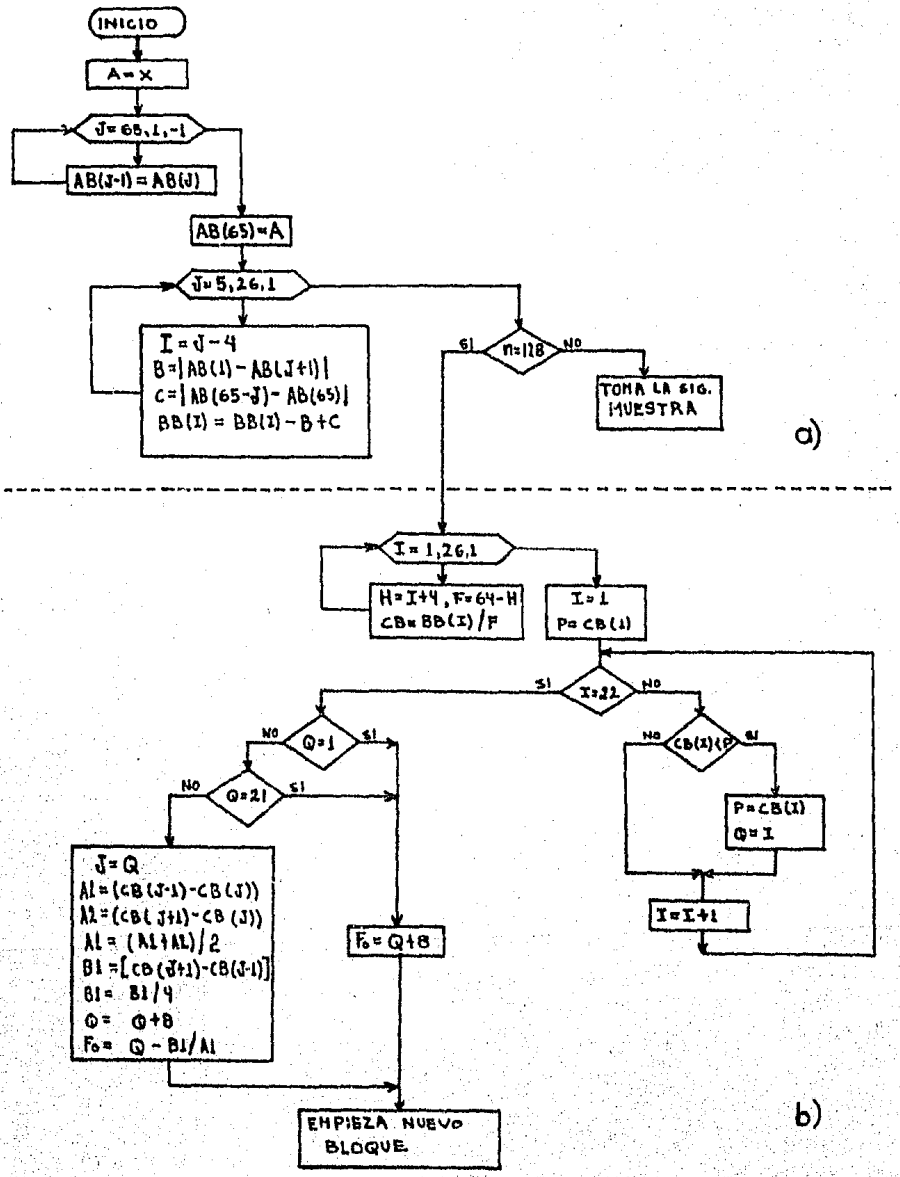


5.3.2 Análisis.

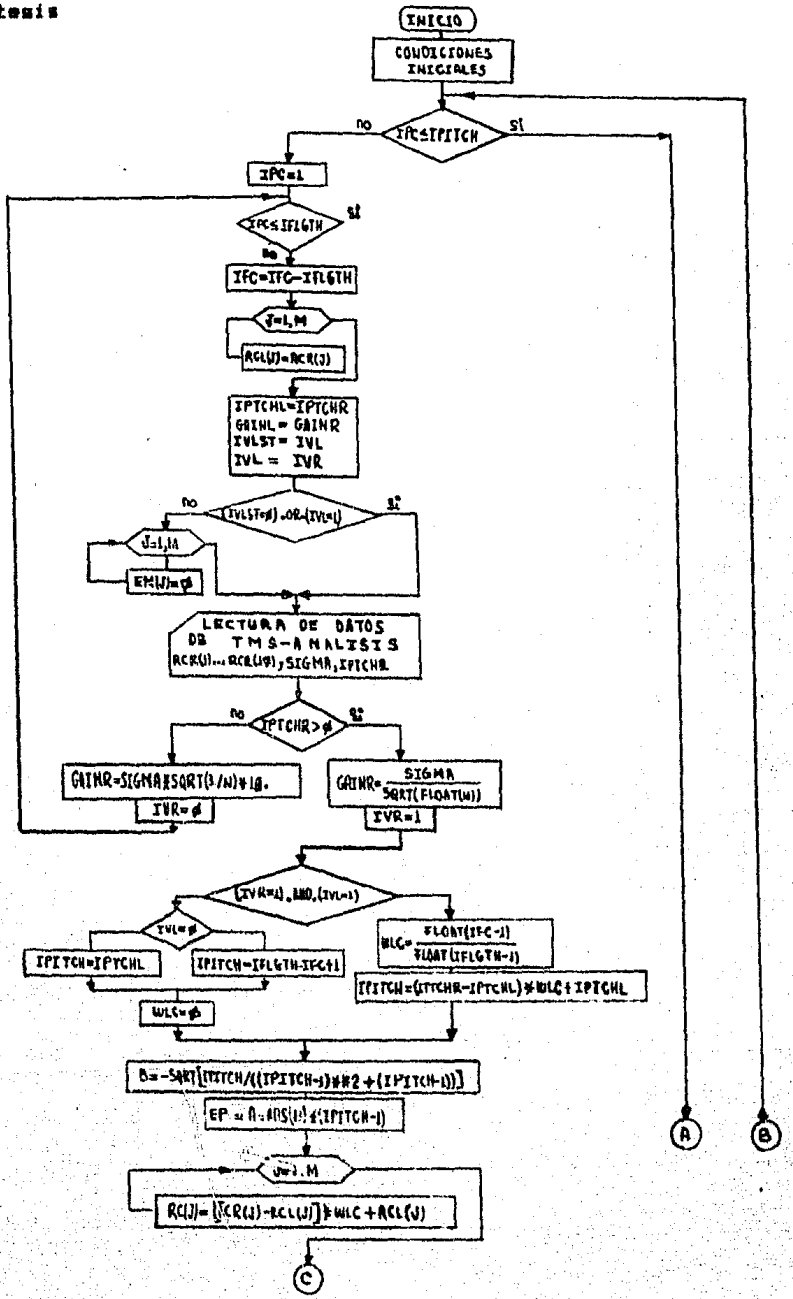


5.3.3 Detección de Tono

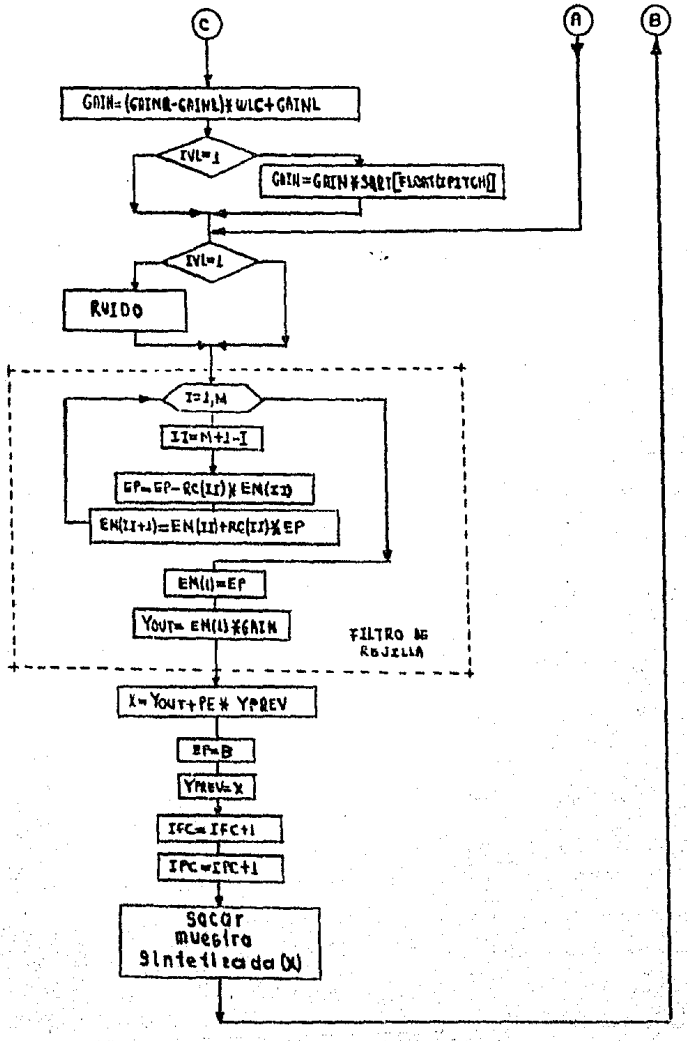
a) Detector de Tono
b) Computación Final de Tono



5.3.4 Síntesis



5.3.4 Continuacion...



CONCLUSIONES

La codificación de voz, dependiendo de la calidad deseada, presenta un alto grado de complejidad en sus algoritmos y en consecuencia, su tiempo de ejecución aumenta en una computadora.

El sistema de codificación de voz implementado, fué escogido de tal manera que fuera lo menos complejo posible, para poder ser implementado en tiempo real. Los resultados obtenidos se presentan en base a:

-Robustez.

El sistema implementado, no es completamente robusto; ya que hay ciertos tipos de voces que no pueden ser codificadas correctamente; este problema es inherente de LPC.

-Calidad Perceptual.

La calidad de voz obtenida es mejor que la sintética, sin llegar por completo a voz telefónica. Es inteligible, se puede distinguir lo que se está diciendo.

-Taza de Transmisión.

La taza de transmisión para codificadores de fuente, como LPC; anda sobre los 4800 bit/seg como máximo, pero con calidad sintética. En este renglón, la calidad se mejoró.

-Tiempo de Ejecución.

Los algoritmos implementados en el microprocesasdor TMS32010 aumentaron en complejidad, dadas sus limitaciones en instrucciones y a la capacidad de memoria, no así en rapidez. Para algoritmos similares a los implementados; una computadora VAX-730, tarda aproximadamente 1.171 seg. para procesar un bloque de 256 muestras, esto sin tomar en cuenta el tiempo de lectura y escritura en disco. Mientras que los programas implementados en el TMS32010 para un mismo bloque, tarda aproximadamente 1.093 mili-segundos.

Perspectivas

Las perspectivas que se presentan son muy amplias, por un lado se pretende mejorar la calidad de la voz y por otro, disminuir la taza de transmisión, sin perder la condición de trabajo en tiempo real.

Consideramos que estas posibles metas, se pueden alcanzar si se construye un sistema hibrido que se base en LPC y otra codificación de forma de onda, como la modulación delta (DM). Seguramente esta alternativa aumentará la complejidad del sistema de codificación de voz, pero esto no será limitante, ya que en la actualidad, se cuenta en el mercado con microprocesadores más veloces y con mayor capacidad de memoria, como el microprocesador TMS32020.

ARQUITECTURA DEL TMS32010

A.1 Microprocesador TMS32010

El microprocesador TMS32010, es el primer miembro de la nueva familia TMS320 para procesamiento de señales, diseñado para soportar un gran numero de operaciones numéricas a gran velocidad, microcomputador de 16/32 bits.

La familia TMS320, contiene el primer microcomputador HDS capaz de ejecutar 5 millones de instrucciones por segundo; resultado de una serie de instrucciones fáciles, comprensivas y eficientes; gracias a su tipo de arquitectura.

Contiene una serie de instrucciones especiales para poder ejecutar algoritmos usados en el Procesamiento de Señales. Como aplicaciones tipicas podemos mencionar:

-Procesamiento de Señales: Filtrado Digital, Correlacion, Ventaneo, FFT, Filtros Adapativos, Generadores de Señales, etc.
-Procesamiento de Voz: Análisis y Síntesis, Reconocimiento de Patrones Vocoders, etc.
-Telecomunicaciones: Ecualizadores Adaptivos, Modems, etc.

A.1.2 Arquitectura

El TMS32010 utiliza una arquitectura modificada Harvard; (Su estructura interna es mostrada en la fig.A.1.) en la cual, la memoria de programa y memoria de datos están separados.

Esto permite una mayor eficiencia en la búsqueda y ejecución de instrucciones.

La Memoria de Programa; consiste de 4K palabras de 16 bits, externas al microprocesador; este modo de operación es controlado por la pata MC/MP, en el modelo que se tiene sólo podemos trabajar en modo microprocesador. Las instrucciones en Memoria de Programa son ejecutadas a gran velocidad, por lo que se requieren memorias con tiempo de acceso de 100 nano-segundos.

La Memoria de Datos; se encuentra dentro del TMS32010 y consiste de 144x16 bits de memoria RAM. Los operandos de las instrucciones a ejecutar son buscados desde esta RAM, no se permite operandos en memoria de programa. Sin embargo, datos en memoria de programa pueden ser escritos en memoria RAM, con ayuda de la instrucción TBLR, lo cual implica mayor tiempo de ejecución, pues primero hay que escribir el dato de Memoria de programa (PM) a Memoria de Datos (DM) y despues ejecutar la instruccion; se puede hacer el proceso inverso utilizando la instrucción TBLW. La fig.A.1.2 muestra los entre-cruces de la pre-búsqueda y ejecución de instrucciones.

En la fig.A.1.2, podemos observar que hay instantes en los cuales se están ejecutando dos instrucciones a la vez, esto es posible por su operación interna "Pipeline".

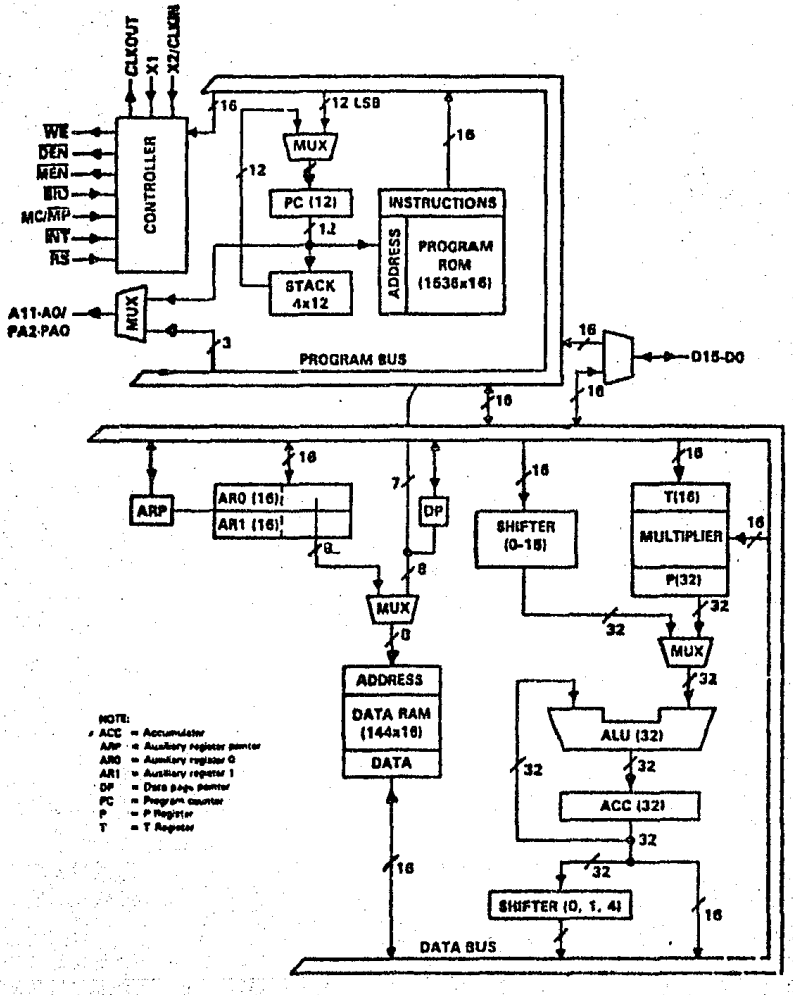


Fig.A.1 Estructura Interna del TMS32010.

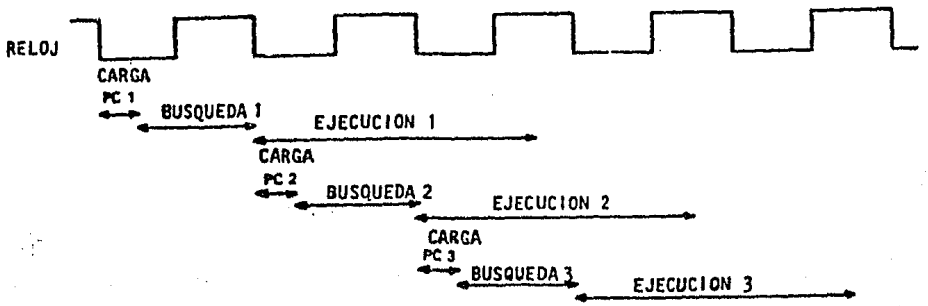


Fig.A.1.2 Arquitectura Harvard.

A.1.3 Instrucciones

El conjunto de instrucciones del TMS32010 se muestran el tabla A-1. Consisten de instrucciones que se realizan en un ciclo de reloj, permitiendo así la ejecución de 5 millones de instrucciones por segundo, sólo saltos e instrucciones de entrada y salida, llamados a subrutinas; son de dos ciclos. Se pueden realizar operaciones Booleanas que combinadas con corrimientos, permiten la manipulación de bits. Operaciones de doble precisión son ejecutadas; por ejemplo, la la instrucción ADDS que manipula un número de 32 bits. La arquitectura del TMS32010, permite que por medio de la instrucción MPY, una multiplicación sea ejecutada en un ciclo. La instrucción SUBC, realiza corrimientos y saltos condicionados necesarios para implementar una división.

BRANCH INSTRUCTIONS

MNEMONIC	DESCRIPTION	NO. OF CYCLES	NO. OF WORDS
BANZ	Branch on auxiliary register not zero	2	2
BV	Branch on overflow	2	2
BIOZ	Branch on BIO = 0	2	2
B	Branch unconditionally	2	2
BLZ	Branch if accumulator < 0	2	2
BLEZ	Branch if accumulator ≤ 0	2	2
BGZ	Branch if accumulator > 0	2	2
BGEZ	Branch if accumulator ≥ 0	2	2
BNZ	Branch if accumulator ≠ 0	2	2
BZ	Branch if accumulator = 0	2	2
CALL	Call subroutine immediate	2	2
CALA	Call subroutine from accumulator	2	1
RET	Return from subroutine	2	1

I/O AND DATA MEMORY OPERATIONS

MNEMONIC	DESCRIPTION	NO. OF CYCLES	NO. OF WORDS
IN	Input data from port	2	2
OUT	Output data to port	2	2
TBLR	Table read from program memory to data RAM	3	1
TBLW	Table write from data RAM to program memory	3	1
DMOV	Shift contents of data memory forward one address	1	1

Tabla A-1 Instrucciones del TMS32010.

ACCUMULATOR INSTRUCTIONS			
MNEMONIC	DESCRIPTION	NO OF CYCLES	NO OF WORDS
ADD	Add to accumulator with shift	1	1
SUB	Subtract from accumulator with shift	1	1
LAC	Load accumulator with shift	1	1
SACL	Store low-order accumulator bits	1	1
SACH	Store high-order accumulator bits with shift	1	1
ADDH	Add to high-order accumulator bits	1	1
ADDS	Add to accumulator with no sign extension	1	1
SUBH	Subtract from high-order accumulator bits	1	1
SUBS	Subtract from accumulator with no sign extension	1	1
SUBC	Conditional subtract (for divide)	1	1
ZALH	Zero accumulator and load high-order bits	1	1
ZALS	Zero accumulator and load low-order bits	1	1
LACK	Load accumulator immediate	1	1
ABS	Absolute value of accumulator	1	1
ZAC	Zero accumulator	1	1
XOR	Exclusive OR with accumulator	1	1
AND	AND with accumulator	1	1
OR	OR with accumulator	1	1

AUXILIARY REGISTERS AND DATA PAGE POINTER INSTRUCTIONS			
MNEMONIC	DESCRIPTION	NO. OF CYCLES	NO OF WORDS
SAR	Store auxiliary register	1	1
LAR	Load auxiliary register	1	1
MAR	Modify auxiliary register and pointer	1	1
LDPK	Load data memory page pointer immediate	1	1
LDP	Load data memory page pointer	1	1
LARK	Load auxiliary register immediate	1	1
LARP	Load auxiliary register pointer immediate	1	1

T REGISTER, P REGISTER, AND MULTIPLY INSTRUCTIONS			
MNEMONIC	DESCRIPTION	NO OF CYCLES	NO OF WORDS
LT	Load T Register	1	1
LTA	Load T Register and accumulate product	1	1
LTD	Load T Register, accumulate product, and move data in memory forward one address	1	1
MPY	Multiply with T Register, store product in P Register	1	1
PAC	Load accumulator from P Register	1	1
APAC	Add P Register to accumulator	1	1
SPAC	Subtract P Register from accumulator	1	1
MPYK	Multiply T Register with immediate operand, store product in P Register	1	1

CONTROL INSTRUCTIONS			
MNEMONIC	DESCRIPTION	NO. OF CYCLES	NO. OF WORDS
LST	Load status register	1	1
SST	Store status register	1	1
NOP	No operation	1	1
DINT	Disable interrupt	1	1
EINT	Enable interrupt	1	1
ROVM	Reset overflow mode	1	1
SOVM	Set overflow mode	1	1
POP	Pop stack to accumulator	2	1
PUSH	Push stack from accumulator	2	1

Tabla A-1 Instrucciones del TMS32010.

A.2 Módulo de Evaluación

El Modulo de Evaluacion (RTC/EVM320A), es una tarjeta con un sistema de desarrollo para el Procesador Digital de Señales TMS32010; con funciones de edición, depurado y ejecucion de programas.

La tarjeta contiene dos microprocesadores en configuracion maestro-esclavo. El procesador TMS9995 funciona como maestro, mientras que el TMS32010 (esclavo), es el encargado de ejecutar el código del usuario en tiempo real.

El EVM proporciona tres puertos para entrada/salida de programa (texto y codigo objeto), para almacenado y/o desplegado. Dos de los puertos son serie tipo RS232C, denominados Puerto 1 y Puerto 2. El EVM acepta archivos desde un CPU huésped a traves de cualquiera de los dos puertos serie. El tercer puerto está asignado a una interface para cinta de audio.

Las funciones de los puertos son:

-Puerto 1: Terminal de Usuario.
-Puerto 2: CPU huésped o conexión de impresora.
-Puerto 3: Cinta de Audio.

En la fig.A.2 se muestra la configuracion general de la tarjeta.

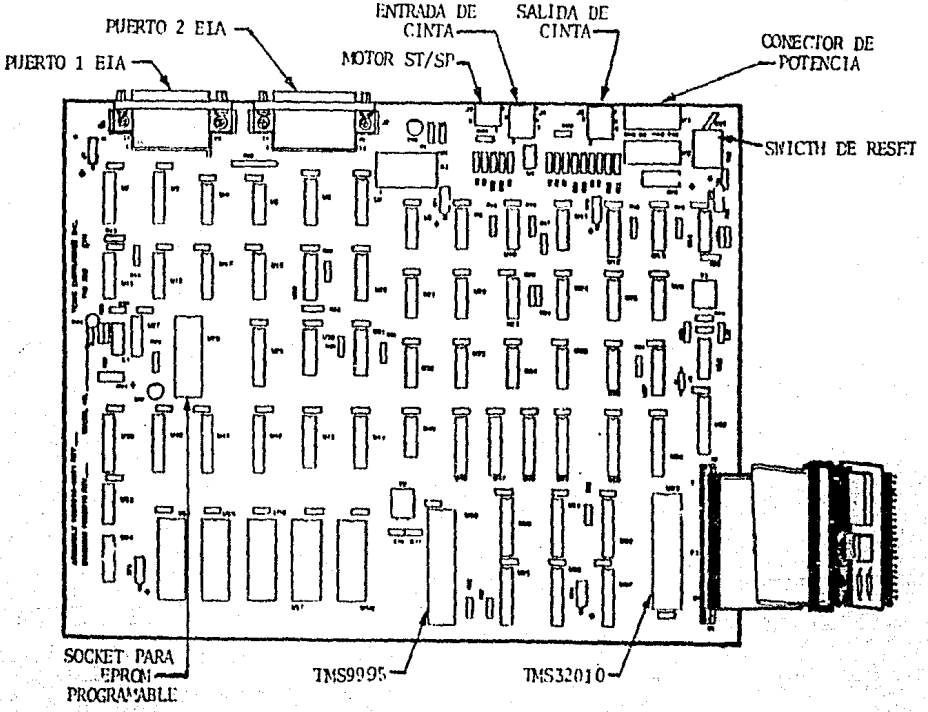


Fig.A.2 Configuracion General del EVM.

El Sistema Operativo (firmware) del EVM, reside en memorias EPROM dividido en las siguientes partes:

-Programa de Monitor: Conteniendo aproximadamente 80 comandos y subcomandos para depurado de programas y manejo del monitor.
-Ensamblado y Des-ensamblado de programas.
-Editor de Texto: para crear programas en lenguaje fuente.
-Utilerías en memoria PROM TMS2764.

A.3 Tarjeta de Interface Analógica [AIB].

Como su nombre lo indica, esta tarjeta sirve de interface a emuladores como: Módulo de Evaluación (EVM), TMS32010-XDS y tiene flexibilidad para poderse utilizar con cualquier otro emulador.

Está provista de convertidores: analógico/digital (A/D) y digital/analógico (D/A), además de dos puertos de expansión P1 y P2. Donde P1 es un puerto de expansión para posibles convertidores adicionales, ó en otro caso, como canal de comunicación en paralelo; P2 es un puerto de expansión de memoria: ver fig.A.3.

El reloj de la frecuencia de muestreo de la AIB, es derivada de la salida CLKOUT del TMS32010, puede ser programada para dar salidas/entradas analógicas periódicas. Hay dos filtros paso bajas en la tarjeta.

Un filtro a la entrada del convertidor A/D que sirve para limitar el ancho de banda y minimizar los efectos de aplanando de la señal analógica. El otro filtro a la salida D/A, suaviza la señal analógica.

La respuesta en frecuencia de los filtros es controlada por componentes externos, que pueden ser fácilmente cambiados para variar esta respuesta. La frecuencia de corte actual, esta puesta a 4.7 Khz.

La tarjeta cuenta también con un amplificador de audio que puede alimentar una bocina de 8 ohms si es necesario.

Especificaciones Generales.

-Convertidor analógico/digital de 12 bits con retén y muestreador, entrada analógica de -10v. a +10v.; tiempo de conversión máximo de 25 micro-segundos.
-Convertidor digital/analógico de 12 bits y salida analógica; tiene un tiempo de conversión de 250 nano-segundos.
-Frecuencia de Muestreo de 76.29 Hz a 5Mhz.
-Extensión de memoria. Sockets para 8192x16 bits con dos memorias RAM HM6264P-12 (no disponibles).
-Cuenta con un decodificador para poder direccionar hasta 64K palabras.

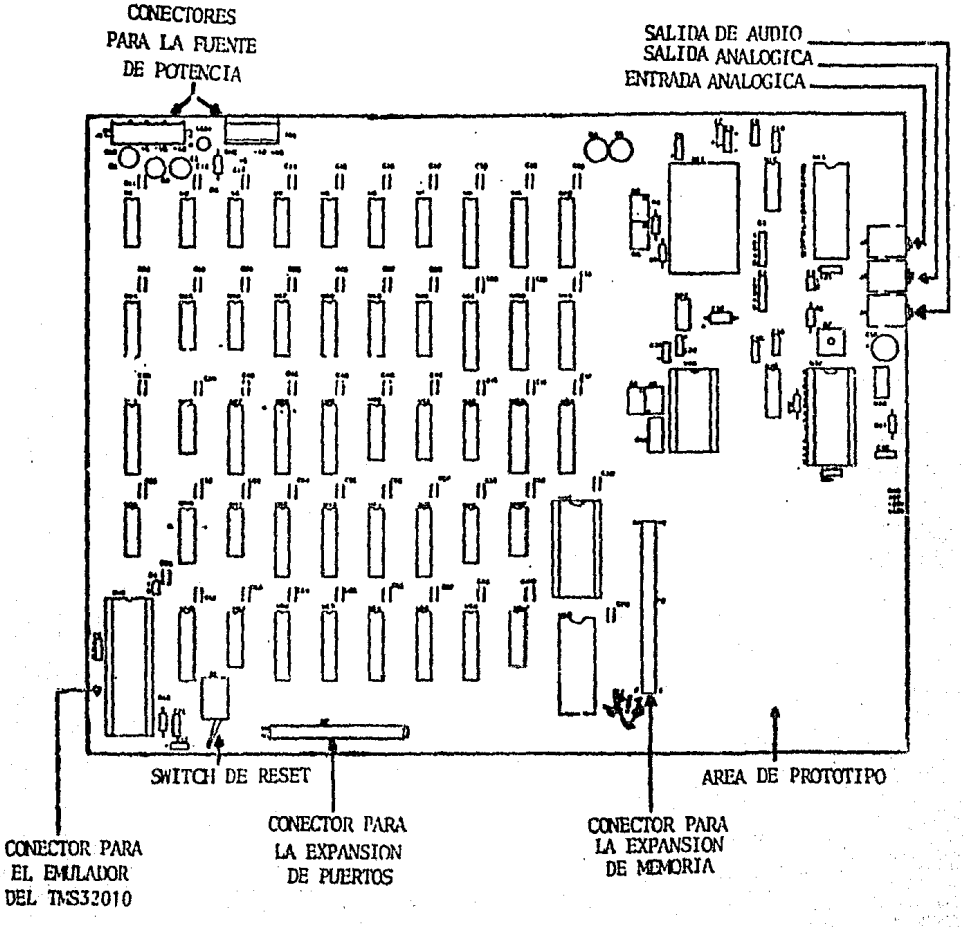


Fig.A.3 Tarjeta de Interface Analógica.

Bibliografía

1.- Lawrence R. Rabiner, M.J. Cheng, A.E. Rosenberg, "A comparative performance study of several Pitch Detection Algorithms", IEEE Trans. on Acustics, Vol. Assp-24, No. 5, October 1976.

2.- B. Gold and L.R. Rabiner, "Parallel processing techniques for estimating Pitch Periods of Speech in the time domain", J. Acust. Soc. Am., Vol. 46. No. 2, Pt.2, August 1969.

3.- Manuel Rodriguez, Juan C. Olabe, "Visión panorámica de la respuesta oral de las máquinas", Mundo Electrónico, No. 144, 1984, Madrid España.

4.- J.D. Markel, A.H. Gray, "Linear Prediction of Speech", Springen -Verlang, 1976.

5.- L.R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signals" Prentice Hall Inc., 1979.

6.- Luis F. Martínez, J.C. Moreno, "Reconocimiento de dígitos hablados con independencia del locutor", Universidad Politécnica de Madrid Octubre, 1980.

7.- D. O'Shaughnessy, "Automatic Speech Synthesis",IEEE Communications Magazine, December 1983.

8.- James L. Flanagan , M.R. Schoeder, B. Atal, "Speech Coding", IEEE Trans. on Communications, Vol. Com-27, No. 4, April 1979.

9.- Bruce Secrest, Masud Arjmand, "Speech Analysis and Synthesis become practical on µc chip", Electronic Design, May 27, 1982.

10.-Kevin McDonough, "Application Processors", Mini Micro Systems December 1983.

11.-Robert H. Cushman , "ICs and Semiconductors', EDN, July 16, 1982.

12.-TMS32010 User s Guide, Digital Signal Processor Products. Texas Instruments, November 1983.

13. TMS32010 Evaluation Module User's Guide. Digital Signal Processor Products. Texas Instruments, 1983.

14.-TMS32010 Analog Interface Board User s Guide. Digital Signal Products. Texas Instruments, 1983.

15.-TMS32010 Assembly Language Programmer s Guide. Digital Signal Products. Texas Instruments, 1983.