



Decodificador DTMF y síntesis de voz en tiempo real sobre la línea telefónica



L. Escobar Salguero, *Departamento de Procesamiento de Señales, Facultad de Ingeniería, UNAM.*
 C. Maya León, *Departamento de Procesamiento de Señales, Facultad de Ingeniería, UNAM.*

RESUMEN: A través de líneas telefónicas que se pueden enviar voz, música, imágenes y datos. Cuando se pulsa una tecla se envía un tono de doble frecuencia (DTMF) y se puede utilizar para requerimientos específicos. Nuestro sistema está conectado a una interfaz y un procesador digital de señales (DSP) el cual recibe una llamada, realiza el descuelgue contestando con voz sintética, la persona que llama escucha a la información y elige una opción a través del teclado numérico, la línea transmite una señal DTMF, ésta es detectada por nuestro sistema y la decodifica dando otro menú como consecuencia. Los resultados muestran que el sistema implementado funciona en tiempo real aun en entornos ruidosos.

Implementación del sistema (Figura 1)

Consiste en una interfaz que se comunica con la línea telefónica por medio de un conector R-J11 contenido en la tarjeta del DSP, cuando el usuario presiona una tecla del teléfono produce un código DTMF correspondiente con la tecla, este código viaja por la línea telefónica hasta la tarjeta del DSP el cual detecta el ring, descuelga virtualmente, decodifica la señal DTMF y responde con voz sintética dependiendo de la tecla detectada.

Codificador DTMF (Figura 2)

El codificador consiste en generar dos tonos senoidales, uno alto y uno bajo, que están dentro del ancho de banda de la voz, estos dos tonos se suman y son enviados a través de la línea telefónica.

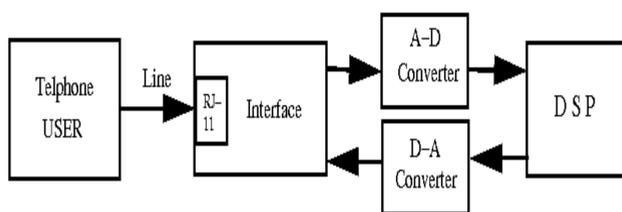


Figura 1: Diagrama general del sistema

	F_H			
	1209	1336	1477	1633
697 Hz	1	2	3	A
770 Hz	4	5	6	B
852 Hz	7	8	9	C
941 Hz	*	0	#	D

Figura 2: Matriz de frecuencias del teclado telefónico

Decodificador DTMF

El proceso de detección de un tono DTMF es posible usar un banco de filtros, la Transformada Discreta de Fourier (DFT) o bien la Transformada Rápida de Fourier (FFT). De acuerdo al estándar solo se tienen ocho frecuencias de interés por lo que un banco de filtros resulta más conveniente. Para este caso es más eficiente calcular los coeficientes DFT utilizando el algoritmo de Goertzel donde cada filtro se sintoniza con cada frecuencia k . Dicho filtro tiene un polo sobre el círculo unitario en la frecuencia $\omega_k = 2\pi k/N$.

El algoritmo de Goertzel necesita calcular de forma recursiva los filtros en $(N-1)$ iteraciones:

$$v(n) = x(n) + 2 \cos(2\pi k/N) v(n-1) - v(n-2)$$

Y de forma no recursiva:

$$X(k) = v(N-1) - e^{j2\pi k/N} v(n-2)$$

Para obtener la magnitud se puede utilizar:

$$|X(k)|^2 = v^2(N-1) - 2 \cos(2\pi k/N) v(N-1)v(N-2) + v^2(N-2)$$

Generación de voz sintética

Para generar la voz sintética es necesario contar con ciertos parámetros correspondientes a cada ventana o "frame" que se obtiene al realizar un análisis previo de una muestra de voz real. El método utilizado es Codificación de Predicción Lineal (LPC).

Lo que se realiza es a partir de una muestra de voz real seccionarlo en ventanas de 30 ms, identificar si la ventana no se trata de silencio entonces extraer sus parámetros que son el Pitch, coeficientes LP, Ganancia y si se trata de un segmento voceado o no voceado. Los coeficientes LP se obtienen al resolver la ecuación normal:

$$R_p a_i = r_p$$

Donde R_p es la matriz de autocorrelación, a_i son los coeficientes LP y r_p el vector de autocorrelación.

Con muestreo de 8KHz, ventanas de 30 ms y un orden de predicción de $P=10$, se puede comprimir la información de hasta un 5.42% del original.

Con estos parámetros se alimenta el modelo del filtro todo polo mostrado en la figura 3 en el cual si el segmento es voceado se mandará un tren de impulsos al filtro, de lo contrario se mandará ruido blanco.

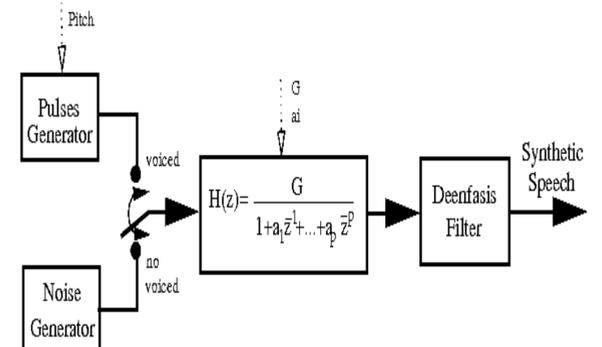


Figura 3: Modelo de producción de voz sintética

Resultados

En la figura 4 se muestra las aproximaciones que se tienen a diferentes órdenes de predicción donde la más cercana a la original es cuando $P = 15$ y a partir de un orden de predicción $P = 10$ la señal de voz sintética es inteligible.

En la figura 5 de igual manera se evalúa el desempeño del sistema a diferentes niveles de ruido que distorsionan la señal DTMF donde responde correctamente para señales SRN mayores a 3 dBs.

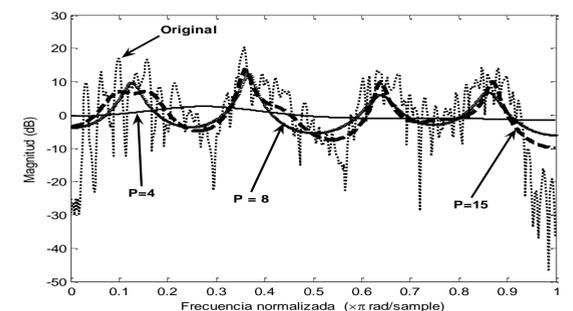


Figura 4: Respuesta a diferentes órdenes de P

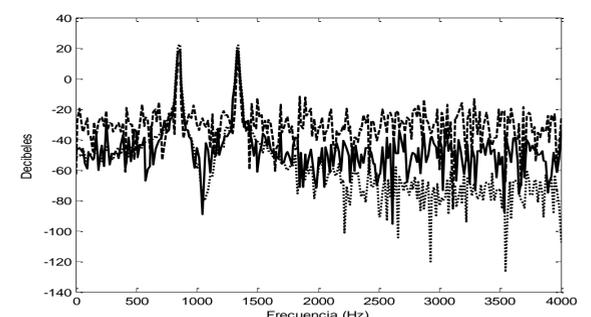


Figura 5: Señales DTMF a diferentes órdenes de SNR

CONCLUSIONES

La decodificación en tiempo real y la síntesis de voz se ha implementado en una plataforma DSP en una línea telefónica tanto residencial como móvil, el sistema soporta hasta 5 dBs de SNR. La calidad de la señal de voz es buena e inteligible, los mejores resultados se tuvieron con un orden de predicción de 10 y 400 muestras por ventana. El performance del DSP es bueno ya que se utilizan 66 MIPS de sus 100 MIPS pudiendo atender tareas adicionales.