

CAPÍTULO 3:

ESTIMA DE LA FRECUENCIA FUNDAMENTAL.

Para el cálculo de la frecuencia fundamental se han usado dos métodos: el método SIFT (Simple Inverse Filter Tracking) y el método AMDF (Average Magnitude Difference Function).

Como ya se ha dicho anteriormente la frecuencia fundamental media del llanto de los recién nacidos está comprendida entre los 400Hz y los 600Hz, sin embargo en el llanto hiperfonado puede llegar a superar los 700Hz. Por ello se ha tomado como rango para la frecuencia fundamental de 150 a los 900Hz.

3.1.- Método SIFT.

El método Sift está basado sobre modelo del tracto vocal descrito anteriormente llamado modelo de predicción lineal (LP) o modelo AR. En él el tracto volcal es dividido en N cavidades cilíndricas. Según este modelo la función de transferencia es una función todo-polo, con un Inverse Filter (IF).

El primer paso a dar sería calcular los parámetros AR con el algoritmo forward-backward. Para ello la entrada del filtro sería dada por la excitación de la glotis y la salida sería el sonido percibido, las ecuaciones matemáticas que describen este proceso son: [19]

- en el dominio del tiempo:

$$s(n) = \sum_{i=1}^N a_i s(n - i) + e(n)$$

- en el dominio de la frecuencia:

$$S(z) = E(z) \frac{1}{1 - \sum_{i=1}^N a_i z^{-i}} = E(z)H(z)$$

Según este modelo la muestra $s(n)$ se obtiene de la combinación lineal de las N muestras anteriores, donde los a_i son los pesos y $e(n)$ representa el error de predicción o residuo. Los pesos son los coeficientes del modelo AR de orden N. En el dominio de la frecuencia tenemos que $H(z)$ será un filtro todo-polo, considerado el inverso de la función todo ceros $A(z)$:

$$E(z) = S(z)A(z)$$

El residuo será pues la convolución entre la señal y el filtro $A(z)$, cuyos coeficientes a_i se obtienen del análisis de la señal. El problema que presenta

esta técnica es obtener el orden del filtro de manera que el ruido no afecte a la señal. Para ello se ha usado la técnica SVD (Singular Value Decomposition) que consiste en separar el subespacio de la señal del subespacio del ruido mediante una matriz A de rango k. Si no existe ruido, k será igual al rango del subespacio de la señal e igual al orden de los coeficientes. Sin embargo si hay ruido el valor de k será igual a la suma de ambos rangos: el de la señal y el del ruido. Por tanto, con la técnica SVD se obtiene una estima del menor orden para el filtro AR.[8].

Para calcular la dimensión de los dos subespacios se usa la técnica DME (Dynamic Mean Evaluation). Esta es aplicada a cada valor de la secuencia de manera que se obtienen órdenes diferentes para cada trozo de la señal usada consiguiendo la variabilidad propia de la señal.

Una vez obtenido los coeficientes del filtro, se pasa la señal por el filtro IF, obteniendo así los residuos. De la función de autocorrelación de este último se obtiene el valor estimado de la frecuencia fundamental correspondiente al máximo de secuencia dentro del intervalo de frecuencias fijado.

Sin embargo el llanto del niño se puede dividir en parte voiced y unvoiced, es decir, voiced es aquella producida directamente por la vibración de las cuerdas vocales, y la unvoiced sería creada por un flujo turbulento de aire al pasar por las zonas más estrechas del tracto vocal. La frecuencia fundamental nos interesa sólo la en las partes voiced.

Para localizar las partes voiced partimos de la secuencia de autocorrelación. Esta es una serie de picos con valores comprendidos entre -1 y 1 y con el máximo en el punto de origen. En la figura 3.1 se puede apreciar una secuencia de autocorrelación de ejemplo.

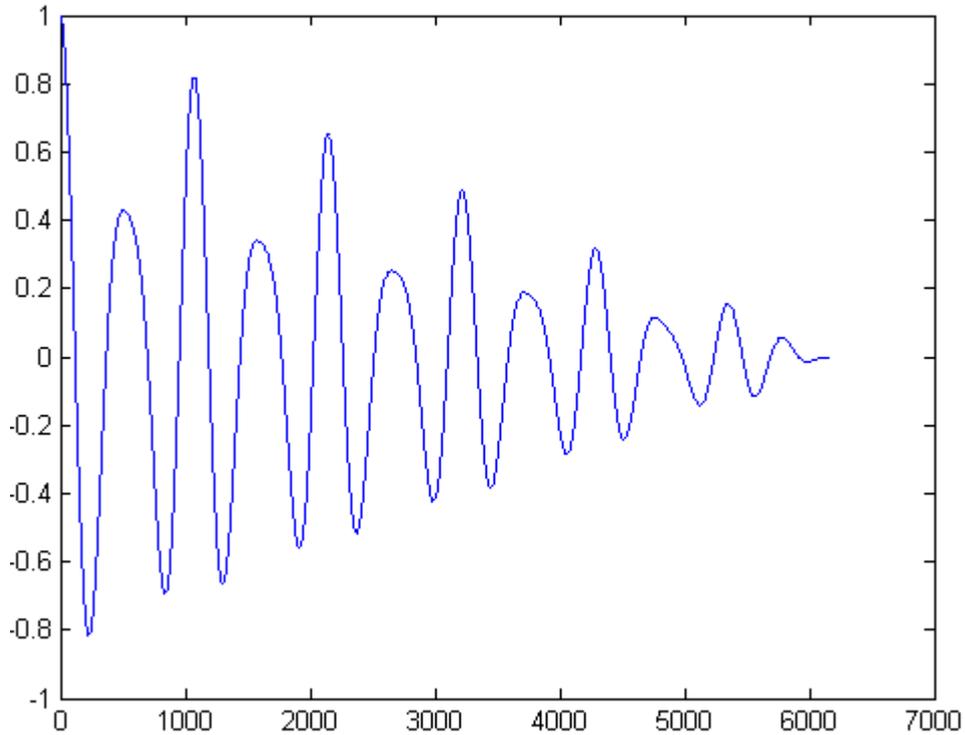


Figura 3.1: Autocorrelación de uno de los bloques.

Las partes voiced muestran un pico en correspondencia con la frecuencia del periodo fundamental mientras que las partes unvoiced están ausentes. Por ello se ha definido el periodo fundamental como el pico más alto en el intervalo deseado [5,19]

Suponiendo el ruido de naturaleza gaussiana se puede definir un valor umbral de manera que todo pico de valor inferior a este sea considerado ruido. La autocorrelación que se obtiene, debido a la aproximación con el tracto vocal, presenta una forma oscilatoria amortiguada, en base a esto se puede definir el pico $\hat{\gamma}$ (valor umbral) para un intervalo P como:

$$\hat{\gamma} = -0.003P + 0.9$$

donde P representa el intervalo del periodo en ms. En este caso como la frecuencia fundamental está comprendida entre 150 y 900Hz los valores de P

serán $P=1.1\text{ms}$ y $P=6.7\text{ms}$ dando lugar al límite inferior de $\hat{\gamma}_{inf} = 0.7$ y $\hat{\gamma}_{sup} = 0.867$. En un principio bastaría decir que si el pico es inferior a 0.7 estamos en la parte unvoiced y si es mayor de 0.7 es parte voiced. Sin embargo este sistema puede presentar errores ya que a veces se encuentran picos voiced por debajo de los 0.7, esto es debido a que rápidos cambios en la frecuencia fundamental pueden llevar a un decremento de 15-20% en el valor de los picos. Para solucionar este hecho cuando se hace la autocorrelación se interpola por un valor de 7 que aumenta significativamente los valores de pico.

Algoritmo realizado

Toda la parte del algoritmo definida a continuación se haya implementada en la función *Sift2*.

En un principio se calcula el número de puntos que tendrá cada ventana. Este es igual a $m = 3F_s/\text{lim_inf}$, donde F_s es la frecuencia de muestreo y el lim_inf será el valor mínimo de la frecuencia fundamental, es decir, para el caso del llanto infantil, 150Hz.

Una vez hallado el número de puntos que tiene cada ventana o bloque y, como el número de puntos totales que componen la señal es conocido, se puede calcular el número de bloques en los que hay que dividir la señal. A continuación se divide la señal en estos y para cada uno de ellos se hace que la señal tenga media cero, ya que una media distinta de cero mete un error sistemático en los coeficientes del modelo AR. Para cada bloque se realizan las operaciones descritas anteriormente, es decir, se halla el orden del Inverse Filter usando los métodos SVD y DME; se calculan los coeficientes usando la técnica forward-backward y por último el residuo realizando la convolución entre la señal que comprende la ventana y los coeficientes calculados anteriormente. El resultado obtenido se encuentra en altas frecuencias debido a la aproximación del tracto vocal con el modelo AR, dejando la información

fundamental a bajas frecuencias. Para recuperar el residuo se filtra con un filtro pasa banda con la banda pasante entre los 50 y los 1000Hz.

El siguiente paso a dar es hallar la secuencia de correlación mediante la técnica Biased e interpolar con un valor de 7, útil para el cálculo de la parte voiced. Una vez hecho esto se normaliza la función de autocorrelación y se busca el máximo existente entre los 150Hz y lo 900Hz. Una vez obtenido el pico de mayor tamaño se toma este para definir si el bloque es voiced o unvoiced. Para ello la primera operación a realizar es ver si el pico es mayor o menor al valor umbral. Si es mayor se verifican los bloques anterior y posterior. Si uno de ellos o los dos es de naturaleza unvoiced, se cambia el valor umbral, este se aumenta un 10% para comprobar si realmente el bloque es voiced o se trata de una anomalía. Si el pico supera este nuevo valor umbral o el bloque siguiente es de naturaleza voiced, el bloque actual se considera voiced. Si los dos son de naturaleza voiced el bloque es voiced.

Si por el contrario el pico no supera el valor umbral, se controlan nuevamente los bloques adyacentes, si los dos son de tipo voiced, el valor umbral se baja un 25% si el pico supera este valor es considerado voiced. Sino es considerado unvoiced.

Para el primer y último bloque se usa un comportamiento distinto debido a la imposibilidad de comparar con los bloques adyacentes. El primer bloque será de la misma naturaleza que el segundo y el último igual al penúltimo.

Una vez que se sale del bucle, se tiene la señal dividida en partes voiced y unvoiced. Ya sólo resta calcular la frecuencia fundamental. La relación existente entre este máximo y F_0 viene definida como:

$$F_0 = \frac{F_s}{\tau}$$

donde F_s es la frecuencia de muestreo y τ el índice correspondiente al máximo de la función de autocorrelación.

3.2.- Método AMDF.

El método AMDF es ideal para los cambios rápidos y lentos de la frecuencia fundamental. Para una señal $s(n)$ con $n=1,2,\dots,N$ dividida en k bloques contiguos de una duración $M=N/k$, la función AMDF para cada bloque viene definida como [8]:

$$AMDF(\eta) = \sum_{i=1}^M |s(i) - s(i + \eta)|, \quad \text{con } \eta = 0, \dots, M - 1$$

Según esta notación la señal se extiende fuera de la ventana de observación. Esto presenta problemas para las funciones estacionarias como es el caso del llanto. Por ello se usa la siguiente fórmula:

$$AMDF(\eta) = \frac{1}{M - \eta} \sum_{i=1}^{M-\eta} |s(i) - s(i + \eta)|, \quad \text{con } \eta = 0, \dots, M - 1$$

Si la señal es periódica con periodo T , entonces se verifica que $AMDF(0)=AMDF(T)=0$. Sin embargo para las señales no estacionarias el mínimo es mayor de cero. Para encontrar F_0 habrá que encontrar el valor η que da lugar al mínimo de la AMDF. Después bastará aplicar la relación:

$$F_0 = \frac{F_s}{\eta_{min}}$$

Algoritmo realizado

El algoritmo que se describirá a continuación se haya en la función $F0_noise$.

Para usar el método AMDF primero es necesario haber realizado el método SIFT. Los bloques usado con el método AMDF son de menor tamaño que los usado con el método SIFT, en general cada bloque SIFT será dividido en un número variable de bloques $k>1$. Para cada uno de estos k bloques se

tomará un límite inferior y otro superior, estos límites vienen definidos por las frecuencias fundamentales de los bloques SIFT. Así el límite inferior será la mínima frecuencia fundamental del bloque SIFT correspondiente, el anterior y el posterior, y lo mismo para la frecuencia superior que será la máxima de estas tres. A estos límites se le resta, al inferior, o se la suma, al superior, una tolerancia de 10Hz para obtener una dinámica más amplia.

Si el bloque SIFT a dividir tiene frecuencia fundamental cero o es unvoiced, el número de bloques en que será dividido es uno, es decir, el bloque AMDF tendrá el mismo tamaño que el SIFT y su frecuencia fundamental será definida como la misma que la del bloque SIFT.

La frecuencia de cada bloque se está almacenando en dos vectores distintos: *Pitch* y *Pitch_2*. En el primero se almacena la frecuencia tal cual y en el segundo se almacena con una peculiaridad. Cuando estamos en una parte unvoiced la frecuencia fundamental se hace cero. Esta parte ha sido añadida de manera que cuando se esté visualizando la frecuencia fundamental se usará el vector *Pitch* y sí se verán las frecuencias unvoiced, pero cuando se estén calculando las frecuencias de resonancia se usará el vector *Pitch_2* y sólo se tendrán en cuenta las partes con F_0 distinta de cero, es decir, las voiced.

En otro caso los bloques AMDF tendrán un tamaño [9]:

$$M = \frac{4N}{\lim_inf \frac{4N}{FS} - (\delta - 1)}$$

Donde:

- Lim_inf es la mínima frecuencia calculada como se dijo anteriormente.

- N es el número de la transformada discreta de Fourier, en este caso se toma $N=32768$
- F_s es la frecuencia de muestreo
- δ número de puntos del espectro que no tiene componente armónica, en este caso $\delta=5$

Una vez que se tiene dividida la señal en los bloques de tamaño deseado hay que filtrar la señal ya que para realizar el método AMDF es necesario que la señal se encuentre en el dominio del tiempo. Este filtrado se realizará con la transformada wavelet continua, para ello se usa la wavelet madre "Mexican hat".

Por último se calcula la secuencia AMDF con la ecuación para ello se llama a la función `Amdf2` y, con la secuencia interpolada con un factor de 7, se calcula la frecuencia fundamental correspondiente al mínimo de la técnica AMDF.

En la función `F0_noise` también se calcula el ruido de la señal, sin embargo esta parte no será explicada ya que cuando se utiliza el programa para analizar el llanto infantil este no es necesario.

Una vez calculada la frecuencia fundamental para cada una de las ventanas se calcula la frecuencia fundamental media para toda la señal.

3.3.-Evaluación de los parámetros referentes a la frecuencia fundamental.

Como ya se ha dicho anteriormente la frecuencia fundamental es uno de los valores más importante para la evaluación del llanto de los neonatos. Pero de ella no sólo importa su valor numérico sino también su evolución en el tiempo.

En este proyecto se ha buscado presentar de manera lo más clara posible los resultados obtenidos hasta ahora, así como añadir otros, de manera que los profesionales de la medicina puedan sacar las mejores conclusiones sin necesidad de un gran dominio sobre el PC.

La función *plotsingolo* se encarga de dibujar todas las gráficas y valores necesarios para la frecuencia fundamental. Antes de dibujar el cambio que sufre la frecuencia con respecto al tiempo, es necesario calcular dónde empieza y comienza cada parte voiced (cómo se realiza este cálculo será explicado más adelante). Estos datos se usarán en el dibujo de la gráfica ya que la parte voiced y unvoiced serán dibujadas de distinto color y con distinto símbolo. Para ello el primer cambio que se ha debido realizar es que la función reciba la matriz *voicedT*, en la primera fila de esta viene almacenada si la frecuencia es voiced o unvoiced (se pasará al vector *voiced*) y en la segunda fila el tiempo correspondiente a esa frecuencia (se pasará al vector *TT*). Debido a que el vector *voiced* ha sido calculado cuando se calculaba la frecuencia según el método SIFT y el vector que contiene todas las frecuencias ha sido calculado con el método AMDF este último es de mayor longitud, por tanto no se pueden comparar directamente entre si.

Una vez que se tiene almacenada en la matriz *tablavoiced* donde comienza, termina y la duración total de cada una de ellas, se realiza lo siguiente para cada una de las frecuencias.

Se comprueba si la frecuencia se encuentra en un tiempo menor que el valor inicial de la parte voiced que está analizando o si este es mayor que el final de la última parte voiced, en este caso estaremos en una parte unvoiced y la frecuencia y su tiempo correspondiente serán almacenados en los vectores *valorUV* y *tiempoUV*.

Sin embargo si el valor de tiempo se encuentra comprendido entre el valor inicial y final de la parte voiced que se está analizando, la frecuencia y su tiempo correspondiente serán almacenados en los vectores *valorV* y *tiempoV*. En este caso se pone la bandera *dentro=1*. Esta bandera indica si se está dentro de una parte voiced.

La siguiente comprobación que se hace es ver si la frecuencia es la primera frecuencia de una parte unvoiced, para ello se mira que, además de pertenecer a la parte unvoiced, la bandera *dentro* se encuentre a 1. Si se cumplen estos requisitos la frecuencia y su tiempo correspondiente serán almacenados en los vectores *valorUV* y *tiempoUV*. Además se pasa a dibujar la parte voiced anterior en rojo. Como la comparación se hace por los tiempos puede suceder que una frecuencia que sea cero, unvoiced, se considere voiced al estar muy cerca del tiempo en que comienza la parte voiced. Si esto sucede, es decir, es considerada como voiced por error, la frecuencia será dibujada en azul, color de las zonas unvoiced, y su valor, después de dibujada, pasará a ser el de la frecuencia media. El que se cambie su valor al de la frecuencia media es para que, en el paso siguiente, cuando se calcula la mínima y la máxima frecuencia de esa parte voiced no influya su valor. Por último se pone la bandera *dentro* a 0 y se aumenta el valor de n. Este valor indica la parte voiced con la que se está comparando.

La última comprobación que se realiza es ver si la frecuencia a considerar es la última y aún estamos en parte voiced. En ese caso se almacena el valor correspondiente como voiced, se dibuja toda la parte voiced y se calcula el máximo y mínimo correspondiente a esa parte. Como se explico anteriormente si por error una frecuencia que es igual a cero es considerada voiced se dibuja como unvoiced y su valor se igual a la frecuencia media.

Hasta ahora sólo se han ido dibujando las partes voiced según llegaban a su fin. Cuando se termina de comprobar todas las frecuencias se pasan a dibujar las frecuencias unvoiced, en azul, en la misma gráfica. Para que no se borren las partes voiced dibujadas con anterioridad se debe dar la orden *hold on* cuando estas son dibujadas.

A título de ejemplo se muestra a continuación una gráfica de la frecuencia fundamental en función del tiempo de un niño sano.

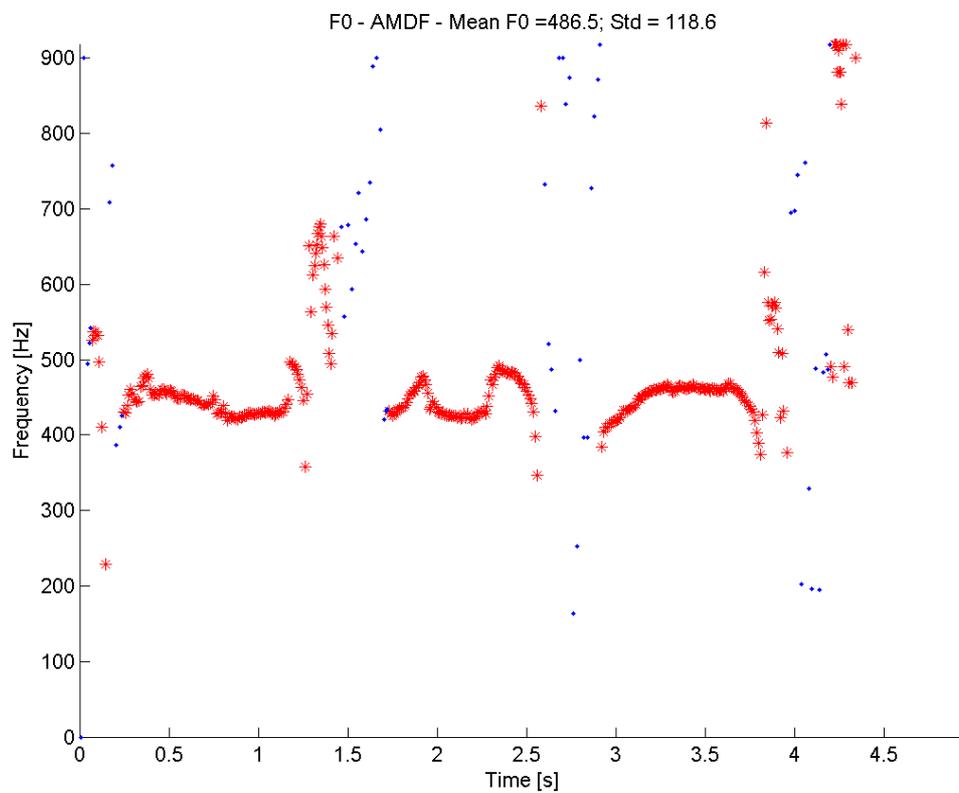


Figura 3.2: Frecuencia fundamental media de un niño sano.