Uso del análisis de multirresolución para calcular el pitch de señales en presencia de ruido

José Romero y Salvador Cerdá
Lab Acústica. Dep. Física. U. Valencia

SUMMARY

Our aim was determinate the effect of noise presence at time to evaluate the pitch, and the study of possible improvements. Several algorithms (Terhard, Duifhuis, Cepstrum, simple FFT and visual determination), were studied and implementated. These algorithms were applicated to different vowels with a progresive augment of noise level; conclusions were that Terhardt method is very similar to the ear, Duifhuis implementation of Godstein´s model is too much sensible to the presence of noise and cepstrum and visual detection are very insensible to the presence of noise. In general, the noise presence involve a loss of accurancy because spurious components appear in the spectrum. At this point the multiresolution analysis (MRA) become a useful tool to eliminate the effect of elevate noise level. MRA were used to smooth the vowels analyzed. Then the algorithms were applicated to the news files obtained with the 10-Daudechies coefficients at differents levels of consecutive approximations. An accurancy augment in all methods was observed when this procedure was applicated.

INTRODUCCION

El pitch de un sonido simple (con una única componente en frecuencias), es una variable subjetiva que nos indica si el sonido es alto o bajo. Desde el punto de vista musical, nos indica dónde se situaria el sonido dentro de una escala musical. Como parámetro objetivo nos evalua la repetición del frente de ondas del sonido. Para una componente pura, el pitch nos proporciona como valor objetivo la frecuencia. Para sonidos complejos (con más de una componente en frecuencias) su valor corresponde normalmente con el de la frecuencia fundamental cuando el espectro es una serie de armónicos de una fo presente. Si el sonido complejo no tiene como espectro una serie de armónicos o la frecuencia fundamental no está presente, el oído extrae el denominado pitch virtual. Desde el punto de vista subjetivo se introduce el mel como unidad psicoacústica del pitch. Una escala de 0 a 2400 mels, describe el rango de audición de 20 a 16 000 Hz.

Para determinar el pitch se han elaborado muchos algoritmos. En esencia siguen dos planteamientos posibles correspondientes a los dos modelos básicos del mecanismo del oido para extraer el pitch. Uno consiste en un análisis en el dominio temporal. Es decir sobre la onda de sonido se produce el análisis (teoría de la periodicidad). El otro procedimiento extrae el pitch después de la descomposición en frecuencias que realiza la coclea (teoría de la frecuencia). La mayoría de los algoritmos desarrollados trabajan a partir del análisis espectral del sonido, habiendo evidencias a favor de una teoría de la frecuencia (Duifhuis 1982).

El propósito de este trabajo consiste en el estudio del efecto que produce la presencia de ruido en la evaluación del pitch en diversos algoritmos. Para ello se analizan diversas señales en las que se realiza un aumento progresivo del ruido. En general la presencia de ruido supone la no localización del pitch adecuado en ninguno de los algoritmos implementados, bien sean frecuenciales o temporales. Este resultado es debido a la difuminación de los armónicos en el espectro o de la propia frecuencia fundamental, en los métodos frecuenciales; y a la pérdida de periodicidad en la onda sonora, en los métodos temporales.

En los últimos años se ha desarrollado intensamente el análisis de wavelet. Diversas aplicaciones en el tratamiento de imágenes como en el análisis tempo-frecuencial han ido apareciendo convirtiéndose dicho análisis en una herramienta moderna de investigación. Una de estas aplicaciones consiste en un método de extracción del pitch haciendo uso de la transformada de wavelet. Qiu et al (1993) introducen un algoritmo que hace uso de la transformada de wavelet en el dominio temporal y en el de frecuencia para extraer el pitch en señales con presencia de ruido. Nuestra intención final consiste en estudiar la mejora que el análisis de multiresolución puede facilitar en un tratamiento anterior a la utiliza-

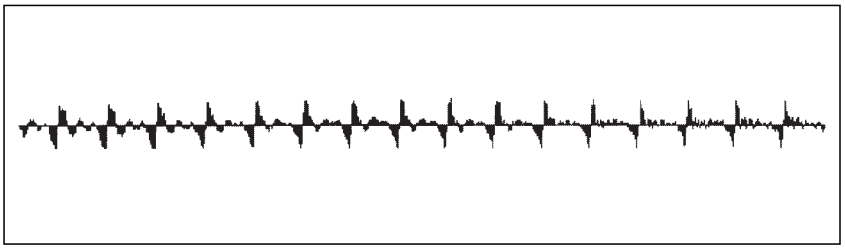


Figura 1. Dominio temporal de la vocal /o/ en “donde”.

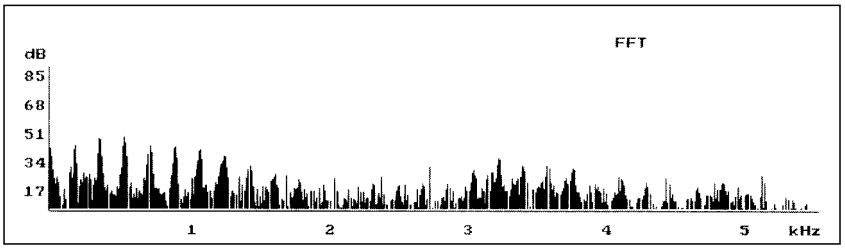


Figura 2. Módulo del FFTde la figura 1.

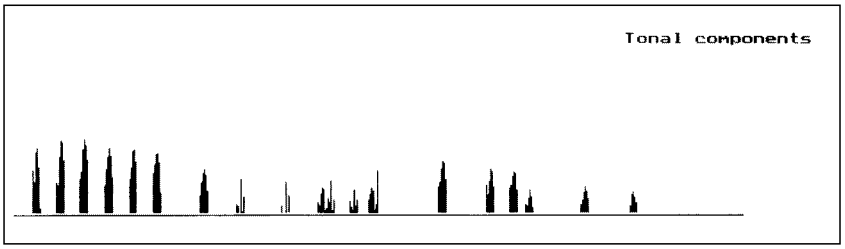


Figura 3. Componentes tonales de la figura 2.

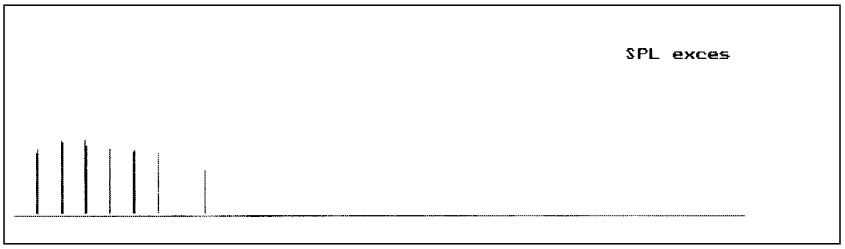


Figura 4. Exceso de S.P.L. de la figura 3.

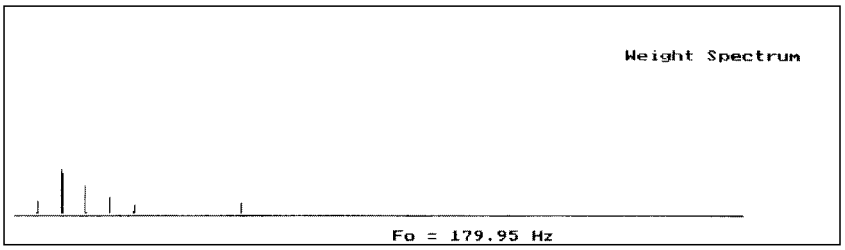


Figura 5. Espectro promedio de la figura 4.

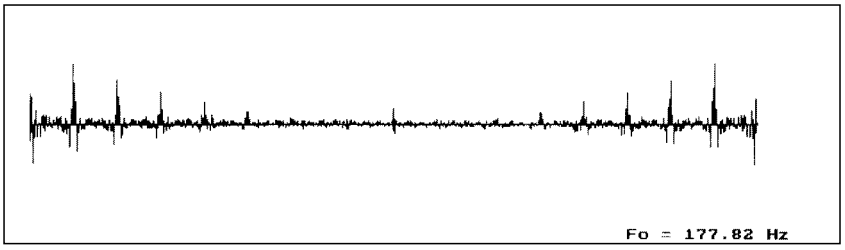


Figura 6. Cepstrum de la vocal /o/ de la figura 1.

ción de los algoritmos de extracción del pitch.

En el apartado I introducimos los diversos algoritmos estudiados describiendolos brevemente sin profundizar en los aspectos psicoacústicos. Para mayor información sobre los mismos les referimos a la bibliografía presentada. En el apartado I, introducimos los conceptos básicos relacionados con el análisis de multiresolución. En el apartado I I I describimos el método experimental empleado, su objetivo y su evaluación. Finalmente en el apartado I V, presentamos los resultados y las conclusiones.

I. ALGORITMOS DE CALCULO DEL PITCH.

Se introducen a continuación los algoritmos de determinación del pitch . Muchos de ellos incluyen diversas consideraciones psicoacústicas. Para profundizar en ellas remítase a la bibliografía presentada. La mayoría de los algoritmos utilizan procedimientos largos que se evitan desarrollar extensamente. Se describen sucintamente dando la bliografía en donde se pueden encontrar correctamente descritos.

1. FFT- método.

Este método es el utilizado en Hiraoka et al (1984). En el se realiza un estudio directo del espectro de la muestra dentro del rango de frecuencias comprendido entre 95 y 500 Hz El mecanismo consiste en los siguientes pasos:

1. Se selecciona la frecuencia F que proporciona el máximo del espectro dentro del intervalo [95,500] Hz
2. Se seleccionan todos los máximos locales en el intervalo anterior cuya intensidad sea almenos un 10% de la intensidad de la frecuencia máxima anterior en escala lineal.
3. De los candidatos anteriores , se seleccionan aquellos que son division entera de F.
4. La fo seleccionada es el menor divisor entero encontrado .

Para calcular la transformada de Fourier se utiliza un algoritmo de FFT, con una ventana Hamming. Se analiza

una muestra de 93 ms con frecuencia de muestreo de 11025 Hz, que nos proporciona una precisión en frecuencia de 10 Hz. En todo el algoritmo se trabaja con el módulo de la transformada de Fourier en escala lineal .

2. Modelo de Goldstein.

Utilizamos la implementación del modelo de Goldstein de la percepción del pitch debida a Duifhuis et al. (1982). Este mecanismo propuesto por Goldstein, pretende determinar la frecuencia fundamental a partir de consideraciones psicoacústicas. En términos generales, a partir de un sonido complejo, con más de una componente en frecuencias, se intenta obtener el mejor patrón de armónicos que aparece en el espectro de la señal.

El algoritmo consta de los siguientes pasos :

1. Análisis mediante FFT de una muestra de 93 ms .

2. Selección de los picos presentes en el espectro bajo una serie de requisitos

3. Procesamiento de estos picos en busca del mejor patrón de armónicos que describe los picos seleccionados .

4. Se selecciona la fo a partir de un proceso de búsqueda del estimador de máxima verosimilitud.

2.1 Análisis espectral.

A partir de 1024 muestras correspondientes a 93 ms (11025 Hz de frecuencia de muestreo), se realiza un análisis de Fourier mediante un algoritmo de FFT para funciones reales y con una ventana Hamming. Este algoritmo nos proporciona una precisión aproximada de 10 Hz. Y se calcula el módulo del espectro. En el resto del algoritmo se trabaja a partir de dicho módulo, siempre en escala lineal.

2.2 Selección de los picos.

Se calcula el máximo global del módulo del espectro. Seguidamente se determinan los máximos locales que no sean inferiores al 15 % del nivel del máximo global en escala lineal. Se seleccionan como máximo 15 picos. A partir del máximo global se seleccionan los restantes picos que aparecen en el espectro dentro de los límites anterio-

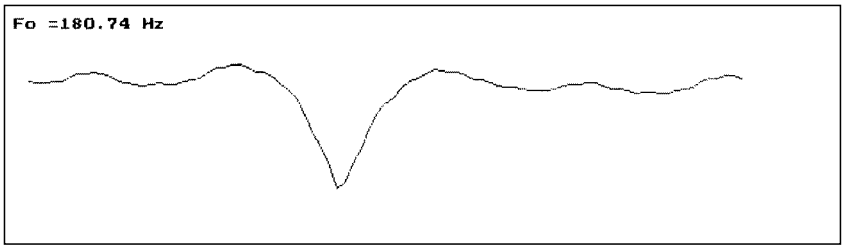


Figura 7. Autopitch de la vocal /o/ de la figura 1.

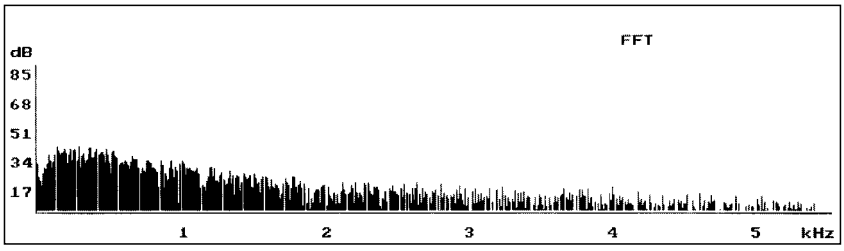


Figura 8. Módulo del FFTdel ruido utilizado

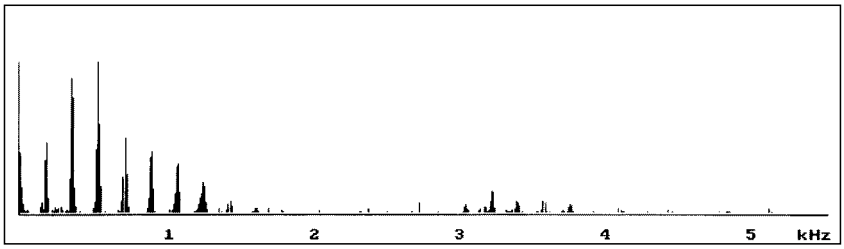


Figura 9. Módulo del FFT en escala lineal de la vocal /o/ de la figura 1.

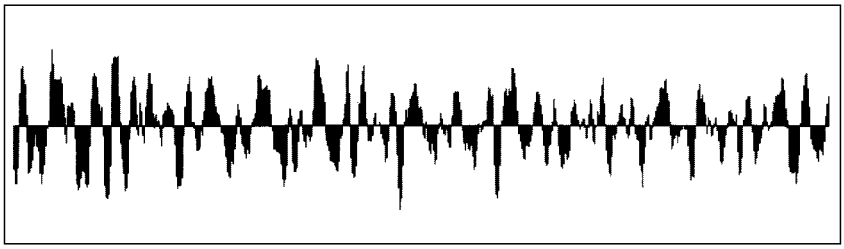


Figura 10. Vocal /o/ de la figura 1 mezclada con ruido.

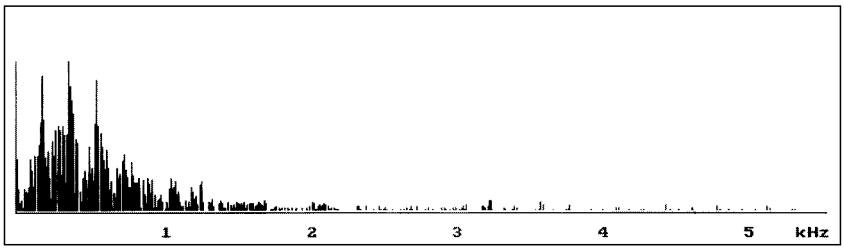


Figura 11. Módulo del FFT en escala lineal de la figura 10.

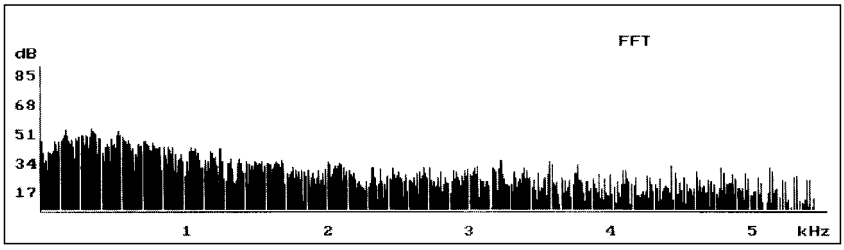


Figura 12. Módulo del FFT en dB de la figura 10.

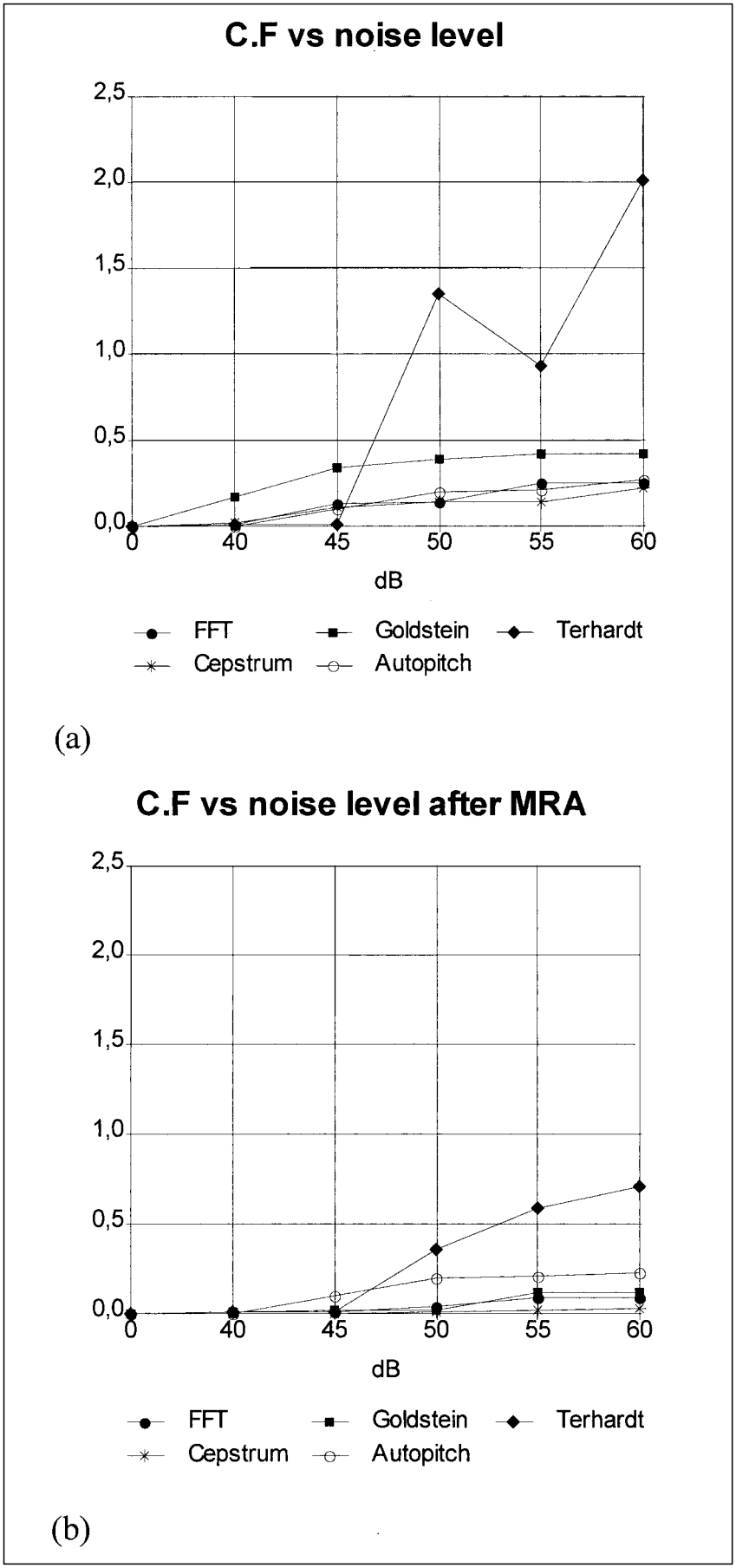


Figura 13. a) Coeficiente C.F inicial a diferentes niveles de ruido; b) Coeficiente C.F tras el uso del MRA.

res. Además, la presencia de un máximo local produce un efecto de enmascaramiento en los máximos próximos a el. Esto se tiene en cuenta, aceptando como nuevos máximos locales, aquellos que esten como muy cerca a 50 Hz de un pico ya seleccionado.

2.3 Patrón de armónicos.

Se pretende ahora encontrar el mejor patrón de armónicos que describe los picos seleccionados en el apartado anterior. Esto consiste en determinar la frecuencia fundamental y el número de armónico que un pico seleccionado ocupa para dicha frecuencia fundamental. Dentro del intervalo de frecuencias [95,500], en pasos de 10 Hz aproximadamente, se verifica si para una frecuencia dada, los picos que se han seleccionados coinciden con uno de los primeros 15 armónicos, con una precisión del 8 %. En caso que así sea se considera el pico como un armónico presente de la frecuencia dada y se indica qué número de armónico resulta ser. Para cada frecuencia del intervalo, en pasos de 10 Hz , se determinan los siguientes parámetros :

1. NL (f), número efectivo de armónicos que pueden estar presentes. Es decir el número total de picos seleccionados menos aquellos que están por encima del máximo armónico permitido dada una frecuencia (es decir inferiores a 15 f).
2. KL (f), número total de picos clasificados. Que consiste en los picos que resultan ser posibles armónicos de una frecuencia f .
3. ML (f), el mayor número de armónico que se determina con los picos para una frecuencia dada f.

A partir de estos valores se determina la cantidad CL (f) = (ML + NL)/ / KL. La f que nos hace mínimo el valor de CL se considera la que nos proporciona el mejor patrón de armónicos .

2.4 Estimación de fo.

Una vez determinada la f que nos minimiza la cantidad CL, podríamos tomar dicha f como la fo buscada . Sin embargo, consideramos como valor más adecuado

$$fo = \sum_{i=1}^{k} x(i)\, n(i) \Big/ \sum_{i=1}^{k} n(i)^2 \quad (1)$$

Donde x(i) representa los picos seleccionados y n(i) el número de armónico que representa dicho pico para la frecuencia f que minimiza CL. En caso de que un pico seleccionado no se considere armónico de dicha f n(i) se considera cero .

3. Al goritmo de Terhardt.

Se utiliza el algoritmo de extracción del pitch propuesto por Terhardt et al. (1982). El algoritmo se fundamenta en aspectos psicoacústicos estudiados por Terhardt y otros. Se puede resumir en los siguientes apartados, a partir de la señal temporal (Fig 1) :

1. Análisis espectral (Fig 2).
2. Extracción de las componentes tonales (Fig 3) .
3. Evaluación de los efectos de enmascaramiento (Fig 4).
4. Peso de las componentes (Fig 5).
5. Extracción del pitch virtual .

3.1. Análisis espectral.

Se realiza un anáilisis de Fourier con algoritmo FFT y una ventana Hamming . Se calcula el módulo del espectro en dB . Se considera como potencial de referencia el nivel de saturación de la tarjeta de sonido (85 dB). Se mantiene una frecuencia de muestreo de 11025 Hz. Esto nos proporciona una precisión aproximada de 10 Hz (Fig 2).

3.2. Componentes tonales.

Una vez calculado el módulo del espectro en escala de dB, extraemos lo que Terhardt denomina componentes tonales. Para ello determinamos los máximos locales que aparecen en todo el espectro. Es decir aquellos que satisfacen:

$$N_{i-1} < N_i \geq N_{i+1} \quad (2)$$

Para que los máximos locales se consideren componentes tonales ha de satisfacerse

$$N_i - N_{i+j} \geq 7 \text{ dB} \quad \text{para } j = -3, -2, +2, +3 \quad (3)$$

En este caso se consideran las siete muestras del espectro (i - 3 , i - 2 ,. . . , i +3) como una componente tonal (Fig 3) .

La frecuencia que tomamos para el máximo local determinado por las tres muestras del espectro se calcula como:

$$f = f(i) + 0.46\ (\text{Hz} / \text{dB})\ (N_{i+1} - N_{i-1}) \quad (4)$$

3.3 Efectos de enmascaramiento .

Una vez determinadas las componentes tonales, se intenta determinar cuales de estas son consideradas audibles por el oido. El procedimiento es bastante elaborado. Consiste en determinar lo que Terhardt denomina el exceso de SPL. Se tiene en cuenta en ello el nivel de la frecuencia central de la componente tonal dado por el espectro. Por otro lado, se tiene en cuenta que la presencia de otras componentes tonales, hace que este nivel aumente o disminuya

Además las frecuencias presentes en una componente tonal junto a la frecuencia tonal, introducen también un enmascaramiento en la frecuencia central. Finalmente se considera el unbral de audibilidad para cada frecuencia presente en una componente tonal. Bajo todas estas consideraciones se determinan las componentes tonales que participan en la percepción del pitch (Fig 4).

Además, la presencia de distintas componentes tonales hace que el pitch asociado a una frecuencia sufra cierto desplazamiento. Dicho desplazamiento se calcula a partir del exceso de SPL de cada componente tonal seleccionada .

3.4 Peso de las componentes.

Una vez determinadas las componentes tonales que intervienen en la percepción del pitch mediante el cálculo del exceso de SPL, se introduce el peso que cada frecuencia tiene en la percepción del pitch. En esto se considera que dicho peso depende del exceso de SPL y de la frecuencia de la componente tonal. Así, se determina la competitividad de cada componente tonal seleccionada hasta este paso, determinando un patrón de posibles pitch (SP) y evaluando en qué medida coinciden con el pitch real observado (WS) (Fig 5)

3.5 Extracción del pitch virtual.

En estos momentos tenemos el SP que nos facilita los posibles pitch y un peso WS para cada uno de ellos dando cuenta de su competitividad. Se clasifican entonces de mayor a menor peso todos los posibles pitch hasta aquí seleccionados. Ahora se ha de tener en cuenta la posible armonicidad o inarmonicidad del espectro. Es decir puede que se hallan seleccionado pitch que sean armónicos unos de otros, o puede que no se correspondan a un patrón de armónicos. Para tener en cuenta todo esto se comparan de dos en dos los pitch seleccionados y se determina un nuevo peso W para cada picth. Se selecciona finalmente como pitch virtual aquel que ofrece mayor peso W .

4. Algoritmo de cepstrum.

Este algoritmo consiste en un análisis inicial de una muestra de 93 ms mediante un algoritmo de FFT con una ventana Hamming (Michael 1976). La frecuencia de muestreo de 11025 Hz nos permite una precisión de 10 Hz . Se calcula el módulo de la FFT resultante en escala lineal. A partir de este módulo se construye un vector de datos complejos con parte real el módulo anterior y de parte imaginaria nula. Se realiza la transformada de Fourier inversa. A partir del resultado obtenido se selecciona el máximo que aparece dentro del intervalo temporal de 2 ms a 10 ms (Fig 6). Se escoje como frecuencia fundamental el inverso del valor temporal que produce el máximo .

5. Autopitch.

Este método consiste en la búsqueda de un patrón de periodicidad por mecanismos automáticos. En este sentido se simula el simple contaje visual que en una pantalla gráfica se puede realizar. Un método semejante, incluyendo inicialmente un filtraje de la señal original, se propone en Yumoyo et al (1982). En nuestro caso, se toma un intervalo de longitud determinada t, y se va comparando con todos los intervalos de igual longitud presente en la muestra. De esta comparación se obtiene un valor P (Fig 7), resultado de la suma del valor absoluto de la resta del intervalo original con todos

los otros intervalos presentes en la muestra. El proceso se repite para intervalos de distinta duración dentro del intervalo {1/500, 1/95}, es decir buscando un patron de periodicidad definido entre 95 y 500 Hz. La frecuencia fundamental escojida es el inverso de la duración que nos da un valor de P mínimo , pues dicho patrón es el que más se repite en toda la muestra .

II . ANALISIS DE MULTIRESOLUCION

La transformada de wavelet se define como :

$$WT(a,b)(f) = |a|^{-1/2} \int_R f(t)\, \varphi^*(t-b/a)\, dt \qquad (5)$$

La función φ ha de estar localizada en el tiempo y en frecuencias. Además se verifica:

$$\int_R \varphi(t)\, dt = 0 \qquad (6)$$

Por otro lado la función $\varphi(t)$ ha de satisfacer otra serie de requisitos para que la transformada de wavelet este bien definida y sea invertible (Chui 1992, Daubechies 1992).

A la hora de construir algoritmos eficientes para calcular WT(a,b)(f), se suele trabajar con la transformada discreta de wavelet (DWT), que consiste en la discretización del parámetro a (parámetro de traslaciones), y del parámetro b (parámetro de dilataciones). Se considera entonces la transformada diádica de wavelet como $WT(2^{-j}, k/2^j)(f)$. En estos términos se observa que :

$$WT(2^{-j}, k/2^j)(f) = \int_R f(t)\, \{2^{j/2}\, \varphi^*(2^j - k)\}\, dt = [f, \varphi_{j,k}] \qquad (7)$$

donde se introduce la notación:

$$\varphi_{j,k} = 2^{j/2}\, \varphi^*(2^j - k)\} \qquad (8)$$

simbolizando el asterisco la conjugación compleja y los corchetes el producto escalar de $L^2(R)$.

Como dijimos anteriormente la función $\varphi(t)$ que define la transformada de wavelet debe satisfacer algunos requisitos. Estos requisitos en esencia permiten que a partir de el conjunto de funciones $\varphi_{j,k}$, resultado de dilataciones y traslaciones de la función φ (wavelet madre), se obtenga una descomposición de $L^2(R)$ (el espacio matemático de las señales de energía finita). Así, dada una señal f(t), podemos obtener su descomposición en términos de las funciones $\varphi_{j,k}$.,calculando los productos $[f, \varphi_{j,k}]$. Para ello podemos utilizar fórmulas de cuadratura. Esto supone muchos cálculos cuando la función $\varphi_{j,k}$ es distinta de cero en un intervalo temporal muy amplio. Sin embargo cabe esperar que si existieran relaciones entre las distintas funciones $\varphi_{j,k}$ se podrían calcular los productos $[f, \varphi_{j,k}]$ unos en función de otros. Normalmente no existen este tipo de relaciones. Lo que si que ocurre es que existan relaciones con otra función distinta a las $\varphi_{j,k}$. Esto nos lleva a las funciones escala (Chui 1992) asociadas a una función de wavelet. Entre una función escala y su wavelet se verifica la siguiente relación :

$$\varphi(x) = \sum q_k\, \phi(2x - k) \qquad (9)$$

siendo $\phi(x)$ la función escala mencionada. Ademas la función escala goza de la siguiente propiedad :

$$\phi(x) = \sum p_k\, \phi(2x - k) \qquad (10)$$

Junto a estas dos relaciones existen otras relaciones entre la función escala y la función wavelet que permitiran el cálculo de los productos $[f, \varphi_{j,k}]$:

$$\phi(x) = \sum \{ a_{-2k}\, \phi(x-k) + b_{-2k}\, \varphi(x-k) \} \qquad (11)$$

$$\phi(2x - 1) = \sum \{ a_{1-2k}\, \phi(x-k) + b_{1-2k}\, \varphi(x-k) \} \qquad (12)$$

Estas dos relaciones permiten encontrar los productos $[f, \varphi_{j,k}]$ a partir de un producto inicial $[f, \phi_{0,0}]$. Consideremos $c^j k = [f, \phi_{j,k}]$ y $d^j k = [f, \varphi_{j,k}]$. Las igualdades anteriores permiten introducir el algoritmo de descomposición:

$$c^{j-1} k = \sum_k a_{l-2k}\, c^j l \qquad (13)$$

$$c^{j-1} k = \sum_k b_{l-2k}\, c^j l \qquad (14)$$

Los coeficientes $c^j k$ proporcionan la descomposición de la señal original f(t) en términos de la familia de funciones $\varphi_{j,k}$ y los coeficientes $d^j k$ son los coeficientes de wavelet diádicos.

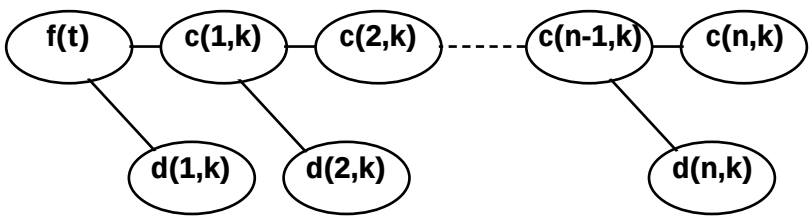
Los coeficientes $c^j k$ proporcionan diferentes aproximaciones a la señal original. Por otro lado se observa que los coeficientes $d^j k$ no son mas que la diferencia entre dos coeficientes $c^j k$. Este algoritmo se esquematiza como en la figura 1.

Como se observa en cada paso se calcula una aproximación a la señal y un coeficiente $d^j k$.

En cada aproximación la señal va perdiendo energía debido a que la información que se extrae en los coeficientes $d^j k$ ya no aparece en las siguientes aproximaciones $c^j k$.

Repasemos el significado de cada término:

1. El parámetro k sitúa la localización del análisis temporalmente (traslación temporal).
2. El parámetro j fija el tamaño de la ventana de análisis (dilatación temporal). Esto equivale a fijar la ventana de frecuencias.
3. Los coeficientes W(j,k) (f) nos proporcionan la descomposición de la señal en términos de una ventana de anchura fijada por j y situada en el tiempo en la posición k. Cada fijación de anchura corresponde a una fijación en frecuencias diferente.



Proceso de cálculo del algoritmo de multirresolución MRA.

III EXPERIMENTO

1. Método.

En general la presencia de ruido supone una pérdida de claridad a la hora de percibir distintos sonidos. Así , un elevado nivel de ruido puede incluso hacernos imposible distinguir un sonido inicial o hacerlo pasar desapercibido. La presencia de ruido introduce modificaciones en nuestra percepción del pitch de un sonido. Un ruido a bajas frecuencias hace que el picth percibido tienda a desplazarse a frecuencias superiores. Nuestra intención es el estudio de los efectos de la presencia de ruido en la determinación del pitch para una serie de algoritmos. Y el estudio de la utilización del MRA para mejorar los resultados.

Para observar el efecto del ruido en los algoritmos, se analizan varias vocales españolas. Las muestras se han tomado mediante una tarjeta de sonido Sound Blaster de 16 bits. Como fuente de grabación se ha tomado un CD editado por la UNED para un curso de Logoaudiometría (1994). Para la calibración de los algoritmos se han utilizado las grabaciones que aparecen en el CD editado por la ASA "Auditory Demonstrations" (1987). El formato de grabación consiste en archivos mono en una frecuencia de muestreo de 11025 Hz a 8 bits. Como ruido enmascarante se utiliza ruido blanco de banda estrecha centrado en los 1000 Hz. (Fig 8). Este ruido tiene componentes en frecuencia principalmente dentro de los 0 Hz a 1000 Hz que incluye precisamente el intervalo de determinación de pitch de todos los algoritmos .

Los ejemplos estudiados son:

la vocal /o/ de la palabra "donde" (E1), la vocal /i/ de la palabra "tinte" (E2); la vocal /a/ de la palabra "al" (E3); la vocal /u/ de la palabra "usen" (E3); la vocal /o/ de la palabra "noche" (E5). Se analiza primeramente las señales originales sin ruido en los diversos algoritmos presentados (E1: 52 dB;E2: 55 dB; E3: 46 dB; E4: 58 dB; E5: 57 dB). Seguidamente se introduce un ruido blanco de pasa -baja en los 1000 Hz (Fig 9) a diversos niveles (40 dB, 45 dB, 50 dB, 55 dB, 60 dB). Se aplican los algoritmos en cada nivel de ruido observando los nuevos valores para el pitch.

TABLA I: Pitch (in Hz) de E1 a diferentes niveles de ruido.

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	183	183	183	172	172	172
Goldstein	180	90	102	90	90	92
Terhardt	180	183	181	547	410	700
Cepstrum	178	178	178	178	178	178
Autopitch	181	181	181	181	181	181

TABLA II: Pitch (in Hz) de E2 a diferentes niveles de ruido.

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	215	215	97	97	129	129
Goldstein	211	133	136	135	102	92
Terhardt	211	211	211	771	385	770
Cepstrum	208	221	221	225	225	95
Autopitch	208	208	208	99	97	95

TABLA III: Pitch (in Hz) de E3 a diferentes niveles de ruido.

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	118	118	118	118	194	194
Goldstein	113	113	114	91	92	92
Terhardt	114	114	114	229	229	229
Cepstrum	114	114	114	114	114	114
Autopitch	115	115	114	114	114	145

TABLA IV: Pitch (in Hz) de E4 a diferentes niveles de ruido.

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	205	205	205	194	194	194
Goldstein	200	200	102	97	101	104
Terhardt	199	200	200	399	498	—
Cepstrum	197	197	99	99	99	99
Autopitch	201	201	100	100	101	100

TABLA V: Pitch (in Hz) de E5 a diferentes niveles de ruido.

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	151	140	140	140	140	140
Goldstein	147	146	93	92	92	95
Terhardt	147	147	146	144	144	368
Cepstrum	145	145	145	133	133	133
Autopitch	147	147	147	145	145	145

TABLA VI: coeficiente C.F a diferentes niveles de ruido de las vocales analizadas

Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	0	0	0.14	0.16	0.29	0.29
Goldstein	0	0.21	0.33	0.39	0.43	0.44
Terhardt	0	0.01	0.01	1.68	1.16	2.18
Cepstrum	0	0.02	0.14	0.15	0.15	0.26
Autopitch	0	0	0.13	0.25	0.26	0.33

TABLA VII: Pitch (in Hz) de E1 a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	183	183	183	172	172	172
Goldstein	180	177	177	177	177	176
Terhardt	180	183	181	168	57	35
Cepstrum	178	178	178	178	178	178
Autopitch	181	181	181	181	181	181

TABLA VIII: Pitch (in Hz) de E2 a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	215	215	215	215	215	215
Goldstein	211	215	222	220	221	220
Terhardt	211	211	211	—	—	420
Cepstrum	208	208	208	212	212	221
Autopitch	208	208	208	108	95	95

TABLA IX: Pitch (in Hz) de E3 a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	118	118	118	118	151	151
Goldstein	113	113	114	113	142	142
Terhardt	114	114	114	—	—	—
Cepstrum	114	114	114	114	114	114
Autopitch	115	115	114	114	114	125

TABLA X: Pitch (in Hz) de E4 a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	205	205	205	194	194	194
Goldstein	200	200	193	193	256	256
Terhardt	199	200	200	394	408	408
Cepstrum	197	197	193	193	187	187
Autopitch	201	201	99	98	98	98

TABLA XI: Pitch (in Hz) de E5 a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	151	140	140	140	140	140
Goldstein	147	146	146	149	147	147
Terhardt	147	147	146	144	147	147
Cepstrum	145	145	145	149	149	149
Autopitch	147	147	147	145	145	145

TABLA XII: Coeficiente C.F a diferentes niveles de ruido tras MRA.						
Metod \Nivel ruido	0 dB	40 dB	45dB	50 dB	55 dB	60 dB
FFT	0	0	0.01	0.04	0.09	0.09
Goldstein	0	0.01	0.02	0.02	0.12	0.12
Terhardt	0	0.01	0.01	0.36	0.59	0.71
Cepstrum	0	0	0	0.01	0.02	0.03
Autopitch	0	0	0.1	0.2	0.21	0.23

Cuando la presencia de ruido conlleva valores diferentes a los proporcionados para la señal sin ruido, se utiliza el algoritmo de descomposición para obtener diversas aproximaciones a la muestra original. Dicho algoritmo se utiliza con 2-coeficientes de Daubechies de orden 10 [Chui 1992]. En cada nivel de aproximación del MRA, se vuelve a aplicar el algoritmo que nos daba un valor del pitch diferente. Se observa el nuevo valor del pitch que proporcionan los algoritmos sobre las consecutivas aproximaciones del MRA

Para evaluar el efecto que produce la presencia de ruido en la determinación del pitch introducimos un coeficiente C.F que nos indique la desviación del valor del pitch en presencia de ruido con el valor obtenido para la señal original limpia. Sea Fo(i) el valor del pitch para la señal inicial y sea Fo(n) el valor del pitch obtenido con un nivel particular de ruido. Consideramos el cociente

$$Fo(n) / Fo(i) \qquad (15)$$

Cuando ambos valores sean iguales dicho cociente valdra 1. En caso que sean diferentes el cociente se alejará de la unidad. Tomamos como medida de la incorrección del valor Fo(n) del picth en presencia de ruido como la diferencia del coeficiente con la unidad en valor absoluto:

$$c.f\ (n) = |1 - F0(n) / F0(i)| \qquad (16)$$

En el caso de tener N ejemplos , para cada método y nivel de ruido añadido se introduce el coeficiente

$$C.F = N^{-1} \sum_{n=1}^{N} c.f\ (n) \qquad (17)$$

Este coeficiente nos da cuenta exclusivamente del alejamiento medio del pitch obtenido con ruido del pitch de la señal limpia. Obviamente $C.F \geq 0$, tomando el valor cero cuando el picth obtenido en presencia de ruido sea igual al obtenido en la señal original.

2 Resultados.

En las Figuras 1,2,9 se muestra la vocal /o/, en dominio temporal y dominio de frecuencia (escala dB y escala lineal). Dicha vocal se analiza utilizando los diversos algoritmos implementados

obteniendo como Fo los valores que se muestran en la Tabla I. A estos ejemplos de vocales españolas, se añade ruido blanco de banda estrecha centrado en los 1000 Hz (Fig. 8), a distintos niveles (45, 50, 55, 60 dB). El efecto de la presencia de ruido tiene resultados visibles en el dominio temporal como se observa en la Fig 10. El efecto en el dominio frecuencial se puede apreciar en las figuras 11 (escala lineal) y 12 (escala dB). Los valores del pitch en los distintos niveles de ruido se muestran en las Tabla I-V para los cinco ejemplos estudiados.

Como se aprecia en las tablas I-V , la presencia de ruido en general motiva valores diferentes del pitch para la señal original y las señales con ruido. Como indicamos anteriormente, el coeficiente C.F nos permite cuantificar estas diferencias en valor absoluto. Los valores obtenidos para este coeficiente se presentan en la Tabla VI. Dichos valores se comparan graficamente para los algoritmos estudiados en la Fig 13 (a). Se observa en general que con el aumento del nivel de ruido, el coeficiente C.F aumenta. Se destaca el aumento claro de C.F para el algoritmo de Terhardt una vez el nivel de ruido pasa los 45 dB.

Uso del MRA

Una vez observado el efecto que el ruido produce en los algoritmos, pasamos a estudiar la mejora que produce en la determinación del pitch el uso del MRA. En las tabla VII - XI, se presentan los valores mejorados al aplicar los algoritmos a las distintas aproximaciones sobre la señal con ruido que proporciona el MRA. De nuevo estudiamos el coeficiente C.F. Los nuevos valores se presentan en la tabla XII. La comparación gráfica de los diversos algoritmos se observa en la Fig 13 (b) .

IV. CONCLUSIONES.

Inicialmente hay que destacar que los algoritmos cualitativamente se comportan de forma muy semejante. Como cabe esperar, la presencia de ruido modifica los valores obtenidos del pitch dependiendo de la relación existente entre el nivel de ruido (en dB) y el nivel de la señal (en dB). Con el aumento del nivel de ruido los algoritmos empiezan a dar valores más alejados del pitch de la señal limpia. El algoritmo de Duifhuis que implementa el modelo de Goldstein, es el más sensible a la presencia de ruido. Para el nivel más bajo de 40 dB, ya se ve afectado. Los algoritmos de FFT, Cepstrum y Autopitch, siempre mantienen un coeficiente C.F inferior al coeficiente C.F de Goldstein. Destacable entre todos los demás, el algoritmo de Terhardt mantiene el comportamiento más semejante al del oido: a nivel bajo de ruido aprecia bien el pitch original y a niveles altos de ruido proporciona valores altos del coeficiente C.F. El algoritmo de Cepstrum es el que muestra menor influencia por la presencia de ruido.

En la Tabla XII se aprecia cómo, tras hacer uso del MRA, los valores del coeficiente C.F disminuyen en todos los algoritmos con respecto a los valores del C.F de la Tabla VI. Esto indica que el uso del MRA supone en términos generales una mejora a la hora de determinar el pitch de la señal original enmascarada por ruido. Comparando las Fig 13 (a) y 13 (b) la reducción del coeficiente C.F se hace patente a simple vista. Se observa que el algoritmo de Autopitch es el que se ve menormente afectado por el uso del MRA; efecto causado sin duda por tratarse de un método que trabaja en el dominio temporal de la señal. El algoritmo de Terhardt sigue manteniendo un comportamiento muy bueno con el nivel de ruido. La mejora más importante se produce en el algoritmo de Goldstein. Por tanto este algoritmo que era el más sensible a la presencia de ruido, es también el que admite una mayor mejora con la utilización del MRA. Se observa también que el algoritmo de Cepstrum que tenía un coeficiente C.F bajo para cualquier presencia de ruido, reduce casi completamente dicho coeficiente. Lo cual significa que el MRA permite con su uso reducir casi totalmente los efectos de la presencia de ruido a la hora de determinar el pitch con el algoritmo de Cepstrum, mejorar plenamente los resultados en el algoritmo de Duifhuis y del método utilizado por Hiraoka así como proporcionar resultados sensiblemente mejores en los otros algoritmos estudiados.

BIBLIOGRAFÏA

H. Duifhuis, L.F Wilems and R.J Sluyter , "Measurements of pitch in speech: An implementation of Goldstein´s theory of pitch perception", J.Acoust. Soc. Am. 71, 1568-1579. 1982

N. Hiraoka, Y. Kitazoe, H. Ueta, S. Tanaka and M Tanabe "Harmonic-intensity analysis of normal and hoarnes voices", J. Acoust. Soc. Am. 76 ,1648-1651. 1984.

A. Michael "Cepstrum pitch determination", J. Acoust. Soc. Am. 41 ,293-309 . 1978.

E. Yumoto , W. J. Gould and T. Baer "Harmonic-to-noise ratio as an index of the degree of hoarseness" , J. Acoust. Soc. Am. 71 , 1544-1549 .1982

E. Terhardt, G Stoll and M. Seewann "Algorithm for extraction of pitch salience from complex tonal signals" , J. Acoust. Soc. Am. 71 , 679-688 . 1982.

L Qiu, S. Kog and H. Yang "Pitch determination of noisy speech using wavelet transform in time and frequency domains" , IEEE TENCON´93 / Beijing 337-340. 1993.

I. Daubechies, Ten Lectures on Wavelets (SIAM, Philadelphia,1992)

C.K. Chui, An Introduction to Wavelets (Academic, San Diego, 1992).

A.J.M. Houtsma T.D. Rossing, and W.M. Wagenaars "Auditory Demostrations" , IPO NIU ASA . 1987.

M.R. de Cárdenas and V. Marrero "Cuaderno de logoaudiometría" edited by U.N.E.D Madrid 1994.