TECNIACUSTICA
BARCELONA • 1996

Sistema de determinación de la frecuencia fundamental de la voz basado en funciones Wavelets moduladas

Juan José Bonet, Léonard Janer, Ignasi Esquerra

Dept. TSC Universitat Politècnica de Catalunya
c/ Sor Eulàlia d'Anzizu s/n 08034 España
Email: leonard@gps.tsc.ups.es

RESUMEN
En este artículo se presentan dos versiones mejoradas de un sistema de determinación de pitch ya existente. Como mejoras se han añadido en un caso un umbral a dos niveles y el el otro caso un umbral adaptativo y un detector de sonoridad para eliminar las zonas sordas y reducir así el número de errores. La eficiencia del sistema se ha testeado utilizando una base de datos segmentada semiautomáticamente con una referencia extraída a partir del laringograma de las señales grabado en el momento de la adquisición de la base.

INTRODUCCION

En este artículo se presenta un sistema de determinación de la frecuencia fundamental de la voz y una variación del mismo que mejora sus resultados de detección. Ambos sistemas se basan en la utilización de funciones wavelets moduladas para realizar el filtrado de la señal de voz [Janer95, Janer96]. Estos sistemas poseen 6 bandas distribuídas en frecuencia a lo largo de una escala de Bark (distribución logarítmica de frecuencias) que determinan el valor del período de pitch muestra a muestra de la señal de voz. En los últimos años se presentaron algunos sistemas de detección de la frecuencia fundamental de voz basados en funciones wavelets (todos con funciones diádicas) [González94, Kadambe91, Larreategui95] que demostraron la eficiencia de tales funciones en la tarea aquí tratada.

El artículo se divide en las siguientes secciones: en la siguiente sección presentamos el sistema inicial y la variación del mismo, en la tercera sección se presentan las evaluaciones de ambos sistemas en base a los errores cometidos y en la última sección detallamos las conclusiones extraídas del trabajo.

ALGORITMO DE DETECCION DE PITCH

Algoritmo inicial

En esta primera versión del algoritmo (ver figura 1) la señal de voz muestreada a una frecuencia de 20 kHz es filtrada por un banco de seis filtros construídos a partir de funciones wavelets gaussianas moduladas distribuídas en una escala logarítmica de Bark. La salida de este banco de filtros consiste en seis señales de diferentes escalas que pasan al siguiente módulo del sistema. Este módulo busca los máximos locales de cada una de las seis bandas y los procesa a través de un umbral a dos niveles.

El funcionamiento de este umbral es muy simple, consiste en calcular para cada máximo la media aritmética de los 5 máximos que se encuentran a su alrededor incluyendo el máximo en cuestión pero discriminando aquellos máximos que se encuentren a mayor distancia que un valor preestablecido. Una vez calculada la media para cada máximo el umbral binario toma valor máximo si la media es superior al valor de su máximo correspondiente y toma valor mínimo si la media es inferior al valor de dicho máximo. A partir de este momento sólo aquellos máximos que tengan un valor superior al del umbral en ese instante serán tomados en consideración, los restantes máximos son eliminados. A continuación evaluamos las distancias entre máximos y eliminamos aquellos cuyo valor de pitch no se encuentre entre los 50 Hz y los 400 Hz así como aquellos cuyo valor de pitch presente una variación muy brusca respecto a los anteriores valores de pitch de los máximos cercanos a él. Una vez depurados todos los máximos de todas las bandas se superponen las seis bandas en un único vector de marcas de pitch y se realizan tres tipos diferentes de postprocesado sobre dicho vector de marcas, el primero se basa en la moda, el segundo en la mediana y el tercero

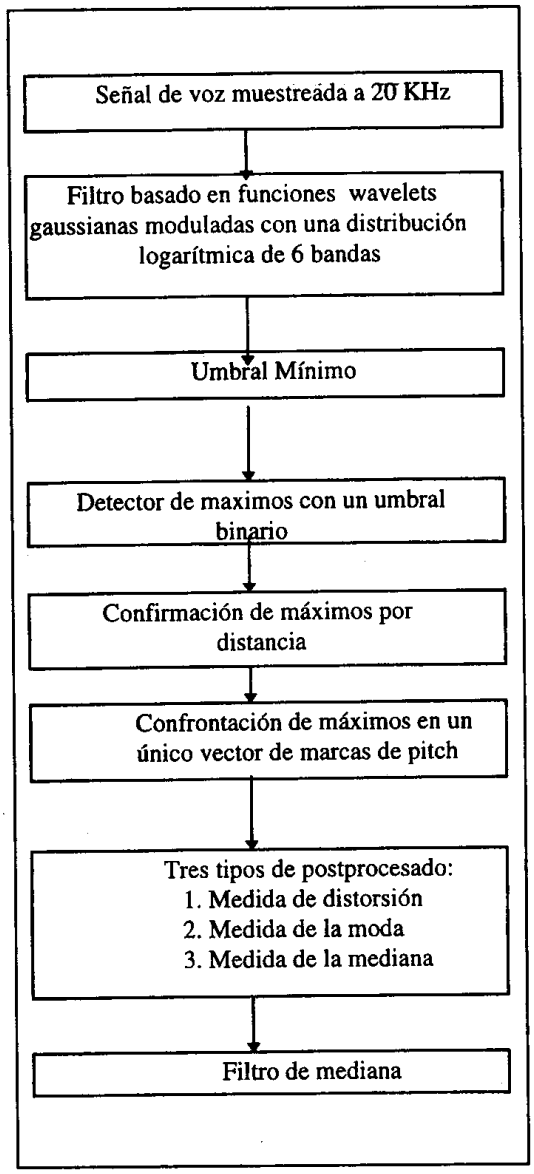

Figura 1: Diagrama de bloques para el sistema inicial.

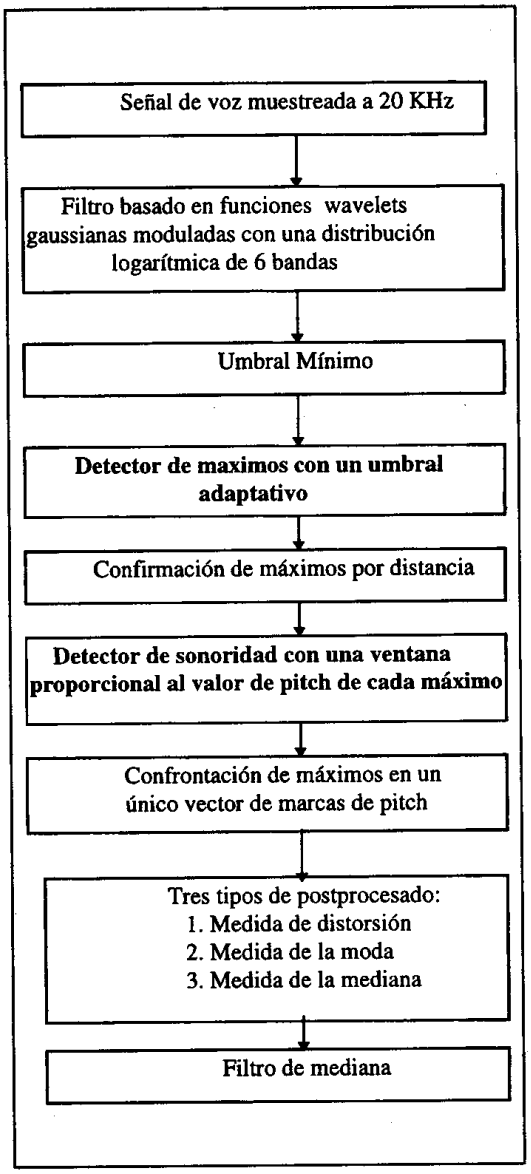

Figura 2: Diagrama de bloques para el sitema variado

es una medida de la distorsión de las marcas de pitch. Finalmente, el último bloque del sistema consiste en un filtro a la mediana .

Variación del algoritmo

En la figura 2 se presenta el esquema de este nuevo algoritmo. Como puede observarse es básicamente como el primero con sólo dos diferencias, se ha cambiado el umbral a dos niveles por un umbral adaptativo y se ha añadido un detector de sonoridad tras la etapa de confirmación de máximos por distancia para mejorar los errores en las zonas sordas de la señal de voz donde no se debería detectar el pitch.

La misión del umbral adaptativo es la de ajustar al máximo los inicios y finales de tramos sonoros de voz y eliminar la mayor parte de máximos pertenecientes a las zonas sordas (ver figura 3,4). Para obtener el umbral adaptativo se ha calculado la media de los dos máximos adyacentes a cada máximo junto con dicho máximo, sólo que la contribución de los dos máximos adyacentes se ha ponderado en función de la distancia al máximo central disminuyendo dicha contribución cuanto más lejos se encuentren del mismo. Mientras el valor de los máximos se mantenga por debajo de la media calculada el umbral toma el valor de dicha media siempre y cuando ésta sea superior a un valor mínimo, en caso contrario el umbral toma dicho valor mínimo. Cuando la señal de máximos supera el valor de la media el umbral toma el valor de

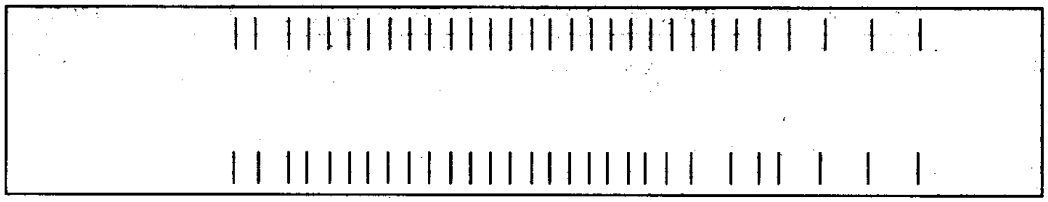

Figura 3. Marcas de pitch de la referencia (arriba) y marcas de pitch del sistema (abajo).

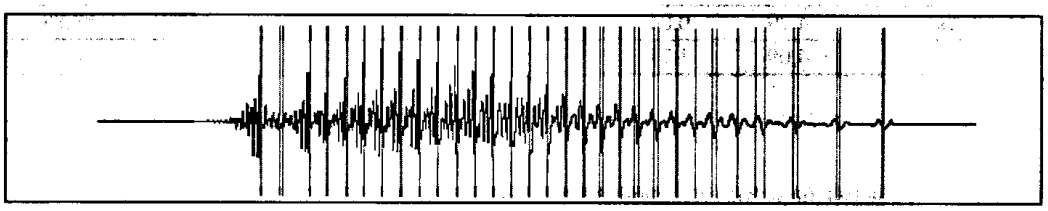

Figura 4. Trozo sonoro de la señal de voz de una mujer.

la media en el instante en que los máximos superan a la media siempre que dicho valor sea inferior a un cierto valor máximo, en caso contrario el umbral toma dicho valor máximo. El umbral se mantiene constante hasta que la señal de máximos es inferior a dicho valor fijo del umbral y entonces se vuelve a retomar el valor de la media.

El detector de sonoridad se encarga de limpiar el resto de máximos de las zonas sordas que el umbral adaptativo no ha sido capaz de eliminar y procura dejar intactos los máximos que pertenecen a zonas sonoras. El funcionamiento del detector de sonoridad es el siguiente: para cada máximo de cada banda se coge una ventana proporcional al valor del pitch de ese máximo centrada en la posición del mismo y se realiza un conteo de todos los máximos encontrados en todas las bandas del sistema dentro de dicha ventana, si el número de máximos supera un determinado valor se etiqueta el máximo en cuestión como sonoro, en caso contrario se etiqueta como sordo y se elimina.

RESULTADOS

En las tablas 1 y 2 se muestran los resultados obtenidos para ambos sistemas. La base de datos utilizada está formada por 5 locutores y 5 locutoras que pronuncian frases durante 40 segundos cada uno/a y que son muestreadas a 20 KHz. La señal de referencia del período de pitch se ha obtenido a partir de un sistema semiautomático basado en el laringograma de la señal de voz [Meyer95, Navarro95].

Para el primer algoritmo presentamos las siguientes estadísticas: GPER (Errores Graves o errores mayores que 1ms), PDER (Errores dobles o errores en que el pitch de una mujer se detecta como el pitch de un hombre), PHER (Errores mitad o errores en que el pitch de un hombre es detectado como el de una mujer), FPER (Errores débiles o errores inferiores a 1ms) y PR (Porcentaje de errores débiles más aciertos en la estimación del periodo de pitch). Para el segundo algoritmo se han presentado los errores en la detección de sonoridad: UVER(%) (Errores de falsa sonoridad: la referencia indica que el tramo de señal es sordo y

	GPER(%)	PDER(%)	PHER(%)	VuVER(%)	FPER(%)	PR(%)
Resultados para la MODA						
Hombres	7.55	0.00	0.97	1.88	78.08	89.59
Mujeres	5.06	3.45	0.33	5.84	74.55	85.33
Resultados para la MEDIANA						
Hombres	10.07	0.03	0.58	1.88	77.57	87.44
Mujeres	7.87	2.77	0.04	10.19	68.09	79.14
Resultados para la DISTANCIA						
Hombres	9.27	1.12	1.00	4.22	75.74	84.4
Mujeres	8.01	4.43	0.25	5.88	73.65	81.43

Tabla 1: *Resultados de la detección de pitch para el sistema inicial.*

	GPER(%)	PDER(%)	PHER(%)	UVER(%)	VuVER(%)	FPER(%)	PR(%)
Resultados para la MODA							
Hombres	3.06	0.00	0.45	1.14	3.53	43.75	90.36
Mujeres	2.13	1.07	0.06	6.04	5.7	34.16	83.93
Resultados para la MEDIANA							
Hombres	3.85	0.00	0.47	1.14	3.53	43.05	89.61
Mujeres	6.15	1.23	0.06	6.04	5.7	30.53	79.75
Resultados para la DISTANCIA							
Hombres	4.29	0.04	0.72	1.14	3.53	42.82	88.88
Mujeres	4.09	1.71	0.06	6.04	5.7	32.8	81.33

Tabla 2: *Resultados de la detección de pitch para el nuevo sistema*

el algoritmo de detección determina sonoridad) y VER(%) (Errores de falta de sonoridad: El sistema de referencia determina que el tramo es sonoro, y el algoritmo decide que el tramo es sordo).

CONCLUSIONES

En este trabajo se ha presentado una variante de un sistema de detección de la frecuencia fundamental de señales de voz, que funcionando muestra a muestra, determina la presencia de sonoridad y el período de pitch. Esto representa una mejora respecto a un sistema existente que nos da la periodicidad de las señales pero no nos determina su sonoridad.

REFERENCIAS

(González94) N. González and D. Docampo. *"Application of singularity detection with wavelets for pitch estimation of speech signals"*. In Proceedings EUSIPCO, volumen 3, pag. 7P.8 1657-1680, 1994.

(Janer95) Léonard Janer. *"Modulated Gaussian Wavelet Transform based Speech Analyser (MGWTSA) Pitch Detection Algorithm (PDA)"*. In Proceedings EUROSPEECH, volumen 1, pag. 401-404, 1995.

(Janer96) Léonard Janer, Juan José Bonet, Eduardo Lleida-Solano. *"Pitch Detection and Voiced/unvoiced Decision Algorithm based on Wavelet Transforms"*. In Proceedings ICSLP96, Philadelphia, October 1996.

(Kadambe91) S. Kadambe and G.F. Boudreaux-Bartels. *"A Comparison of Wavelet Functions for Pitch Detection of Speech Signals"*. In Proceedings ICASSP, 1991.

(Larreategui95) Mikel L. Larreategui, F.J. Ancin and Rolando A. Carrasco. *"An Improved Epoch Detection Algorithm Based on Sinusoidal Modelling of Speech"*. In Proceedings EUROSPEECH, volumen 1, pag. 409-412, 1995.

(Meyer95) G.F. Meyer, Plante F. y Ainsworth W.A. *"A pitch extraction reference database"*. In Proceedings EUROSPEECH, volumen 1, pag. 837-840, 1995.

(Navarro95) Juan Luis Navarro-Mesa y Ignasi Esquerra-Lluciá. *"A Time-Frequency Approach to Epoch Detection"*. In Proceedings EUROSPEECH, volumen 1, pag. 405-408, 1995.