sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías

Gastón Schlotthauer[†], María E. Torres[†,§] y Hugo L. Rufiner[‡,§]

†Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería
Universidad Nacional de Entre Ríos, Oro Verde, Entre Ríos, metorres@bioingenieria.edu.ar
‡Laboratorio de Cibernética, Facultad de Ingeniería, UNER, Oro Verde, Entre Ríos
§SINC(i), Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Santa Fe.

***Resumen*— En este trabajo se propone un nuevo método para la extracción de la frecuencia fundamental (F_0) de la voz. Se estudia su comportamiento en el caso de voces patológicas, situación en la cual su determinación en vocales sostenidas es crucial para el diagnóstico. El método propuesto se basa en el algoritmo de descomposición modal empírica por conjuntos. Este es un algoritmo completamente guiado por los datos para la descomposición de señales en componentes AM - FM llamadas funciones modales intrínsecas o modos. Nuestros resultados indican que la frecuencia fundamental de la voz es capturada en un solo modo. Proponemos además un algoritmo basado en entropías para la selección automática del modo que contiene la frecuencia fundamental. Una vez seleccionado el modo, se obtiene la frecuencia instantánea mediante un método ya conocido. El comportamiento del algoritmo propuesto para la determinación de la frecuencia fundamental se compara con otros dos (algoritmo robusto para el seguimiento del pitch *-robust algorithm for pitch tracking*, RAPT- y algoritmo basado en autocorrelación -AC-), en vocales sostenidas normales y patológicas.**

***Palabras Clave*— Descomposición modal empírica por conjuntos, frecuencia fundamental, pitch, voces patológicas.**

1. INTRODUCCIÓN

El período fundamental T_0 de la señal de habla vocalizada puede definirse como el intervalo de tiempo que transcurre entre dos pulsos laríngeos consecutivos, siendo la frecuencia fundamental $F_0 = 1/T_0$ [1].

Aunque la frecuencia fundamental es útil en un amplio rango de aplicaciones, su estimación robusta y confiable sigue siendo aún hoy una tarea difícil. Esto resulta más evidente ante la presencia de ruido o en el caso de voces patológicas [1].

En el habla, las variaciones de la frecuencia fundamental contribuyen a la prosodia con diversas funciones, por ejemplo en la afirmación enfática, en los diferentes matices de un enunciado (afirmativo, interrogativo, etc.) y en las entonaciones regionales, entre otras. En lenguas tonales como el Chino Mandarín, el contorno de la frecuencia fundamental con el que se pronuncia cada sílaba sirve para crear contrastes fonológicos. Hubo varios intentos de utilizar la frecuencia fundamental en sistemas de reconocimiento automático del habla, con resultados diversos. Entre otros factores, esto podría ser una consecuencia de la ausencia de un método de estimación lo suficientemente robusto y confiable. Otras aplicaciones se relacionan con reconocimiento automático del hablante, clasificación de emociones basada en el habla, *morphing* de voz, canto y estudio de voces patológicas.

En la evaluación clínica de voces patológicas, el análisis de la perturbación de la frecuencia fundamental de vocales sostenidas es un procedimiento estándar para medir el grado de severidad de las patologías y para el seguimiento de la evolución del paciente [2]. En este tipo de aplicaciones, es esencial una estimación precisa y confiable de la frecuencia fundamental. Los algoritmos convencionales se basan en hipótesis de linealidad y estacionariedad, dos simplificaciones excesivas en el caso de voces patológicas.

El algoritmo de descomposición modal empírica (*Empirical Mode Decomposition*, EMD) propuesto por Huang [3] descompone, en forma adaptativa, señales no lineales y no estacionarias en una suma de componentes AM-FM llamadas funciones modales intrínsecas (*Intrinsic Mode Function*, IMF). Cada una de estas IMFs son monofrecuenciales, y pueden escribirse como $A(t)\cos(\phi(t))$, con amplitud instantánea $A(t)$ y frecuencia instantánea $f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}$. Esta nueva técnica consiste en la separación de una señal en oscilaciones rápidas y lentas de manera local y completamente guiada por los datos. En [4] se utilizaron seis filtros pasa banda para obtener un modelo AM-FM de la señal del habla; mientras que el algoritmo de EMD realiza la descomposición en una suma de componentes AM-FM de forma adaptativa.

Si bien se han propuesto otros algoritmos basados en EMD para la estimación de la F_0 [5, 6], estos presentan el problema conocido como "mezcla de modos". Para

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

intentar atenuarlo, utilizan un conjunto de reglas en una etapa de post procesamiento.

La mezcla de modos es quizás el mayor inconveniente del EMD original. Se habla de este efecto cuando aparecen oscilaciones de escalas (o energías) muy diferentes en cierta IMF, o ante la presencia de escalas similares en diferentes modos [7]. Wu y Huang [7] propusieron una modificación al algoritmo de EMD original, llamada descomposición modal empírica por conjuntos (*ensemble empirical mode decomposition*, EEMD), que alivia en gran medida dicho problema.

En este artículo presentamos un nuevo método basado en EEMD que permite extraer la frecuencia fundamental en vocales sostenidas normales y patológicas.

2. MATERIALES Y MÉTODOS

2. 1. Base de datos

Se utilizó una base de datos de vocales con señales procedentes de 710 personas de ambos sexos [8], con una frecuencia de muestreo $f_s = 22050$ Hz. La misma incluye fonaciones sostenidas de la vocal /a/ de individuos sanos y de pacientes con una amplia variedad de desórdenes de la voz (orgánicos, neurológicos, traumáticos y psicogénicos). En esta base de datos, la frecuencia fundamental promedio de las voces normales se encuentra entre 120,39 y 316,50 Hz.

2. 2. Descomposición modal empírica por conjuntos

Como se indicó en la Sec. 1., la EMD descompone una señal $x(t)$ en un número pequeño de IMFs. Cada una de ellas debe satisfacer dos condiciones: (i) el número de extremos debe ser igual al de cruces por cero, o bien pueden diferir en uno; y (ii) en cada punto, el valor medio entre las envolventes superior e inferior debe ser cero. Dada una señal $x(t)$, el algoritmo EMD es el siguiente [3]:

1. encontrar todos los extremos de $x(t)$,
2. interpolar entre los mínimos (máximos), para obtener la envolvente $e_{min}(t)$ ($e_{max}(t)$),
3. calcular la media local $m(t) = (e_{min}(t) + e_{max}(t))/2$,
4. extraer la candidata a IMF $d(t) = x(t) - m(t)$,
5. verificar las propiedades de $d(t)$:
   - si $d(t)$ no es una IMF, reemplazar $x(t)$ con $d(t)$ e ir al paso 1,
   - si $d(t)$ es una IMF, evaluar $r(t) = x(t) - d(t)$,
6. repetir los pasos 1 al 5 sobre la señal residuo $r(t)$ hasta satisfacer un criterio predefinido [9].

Como ya se señaló, uno de los inconvenientes más relevantes del algoritmo EMD es la mezcla de modos. Este problema se ilustra en la columna izquierda de la Fig. 1, donde se muestran 60 ms de una vocal sostenida /a/ procesados mediante EMD y las cuatro IMFs con mayor energía. La aparición de oscilaciones de diferentes escalas resulta clara en la IMF 2. Otro ejemplo

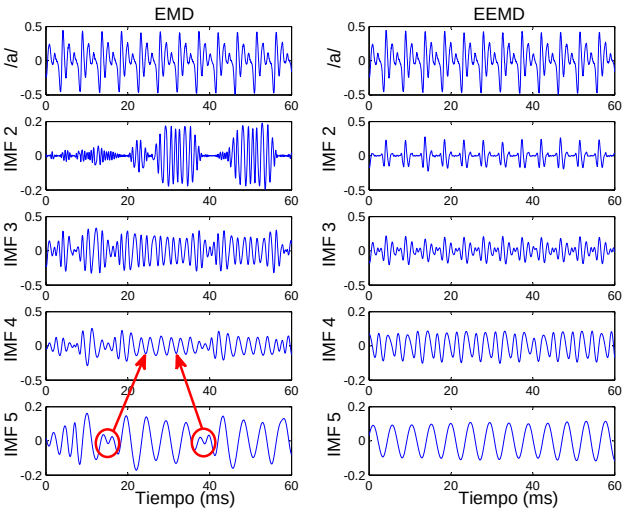

Figura 1: Vocal sostenida /a/ analizada con EMD (izquierda) y EEMD (derecha). Se muestran las IMFs 2 a 5. En la IMF 5 obtenida con EMD, se indican con círculos segmentos donde ocurre "mezcla de modos".

puede apreciarse en la IMF 5, donde dos segmentos con oscilaciones muy similares a las que aparecen en la IMF 4 se señalan con círculos.

El algoritmo EEMD[1] es una extensión del EMD, previamente descripto, y define a la *verdadera* IMF como la media de las correspondientes IMFs obtenidas a partir de cierto número (N_e) de nuevas realizaciones generadas sumando ruido blanco gaussiano a la señal original. Este método provee una mejora sustancial al algoritmo de EMD y reduce la mezcla de modos [7]. En la columna derecha de la Fig. 1 se presenta un ejemplo de sus capacidades. Se utilizó un conjunto con $N_e = 5000$ elementos, y el ruido blanco agregado a cada miembro del conjunto tiene desvío estándar $\epsilon = 0{,}2$. En general unos pocos cientos de elementos permiten lograr buenos resultados [7], aunque con una carga computacional sensiblemente mayor que EMD. El ruido remanente, definido como la diferencia entre la señal y la suma de las IMFs obtenidas por EEMD, tiene un desvío estándar $\epsilon_r = \epsilon/N_e$. Las IMFs 2 a 5 se muestran en la columna derecha de la figura 1, debajo de la vocal sostenida /a/. Puede apreciarse que las IMFs obtenidas por EEMD son mucho más regulares que aquellas halladas por EMD. Además, se observa que las oscilaciones de la IMF 5 capturan el período fundamental de la vocal /a/.

Este hecho se destaca en la Fig. 2(a), donde se presenta la vocal /a/ en azul y la IMF 5 obtenida con EEMD se superpone en rojo. En la Fig. 2(b) se presentan las densidades espectrales de potencia de la vocal /a/ y de la IMF 5. Esta última muestra un pico bien definido en la frecuencia $F = 210$ Hz, la que puede interpretarse como la frecuencia fundamental promedio.

[1] Disponible en `http://rcada.ncu.edu.tw/`.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

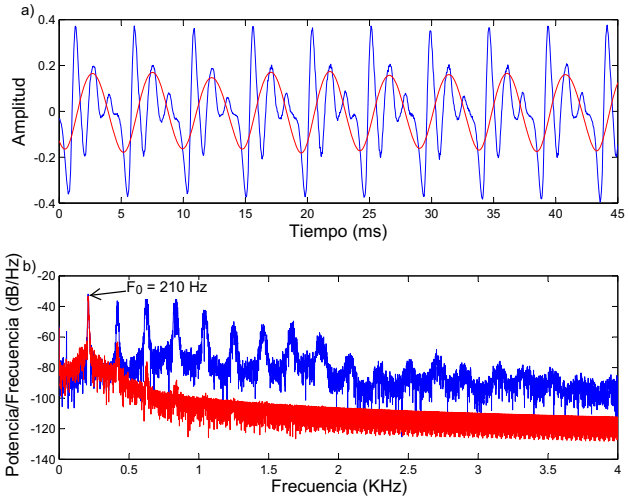

Figura 2: a) Vocal sostenida /a/ (azul) e IMF 5, obtenida con EEMD (rojo). b) Espectro de la vocal sostenida /a/ (azul) y su IMF 5 obtenida mediante EEMD (rojo). El pico del espectro de la IMF 5 se indica con $F_0 = 210$ Hz.

2. 3. Algoritmo discreto de separación de energía (DESA-1)

Habitualmente se utilizan técnicas basadas en la transformada de Hilbert (HT) para extraer la frecuencia instantánea en señales mono-frecuenciales. Sin embargo, el algoritmo discreto de separación de energía (*Discrete Energy Separation Algorithm*, DESA-1) da resultados superiores a los métodos HT cuando se consideran señales del mundo real [10].

Sea $d^m(n)$ una versión muestreada de una IMF continua, con $n = 1, \ldots, N$, para $m = 1, \ldots, M_x$, donde M_x indica el número de modos en los que se descompone $x(t)$. Se define entonces el operador discreto de energía Teager mediante $\Psi\left[d^m(n)\right] = \left(d^m(n)\right)^2 - d^m(n-1)d^m(n+1)$, para $n = 2, \ldots, N-1$.

Si $d^m(n)$ es un coseno de tiempo discreto con amplitud constante A y frecuencia ω, $d^m(n) = A\cos\left(\Omega n + \theta\right)$, con $\Omega = \omega T$ y T es el período de muestreo, entonces:

$$\Psi\left[d^m(n)\right] \quad = \quad A^2\omega^2\left(\frac{\sin\Omega}{\Omega}\right)^2.$$

A partir de estas relaciones, aplicamos el algoritmo DESA-1 para la separación AM-FM [11]. Este algoritmo estima la frecuencia instantánea $\Omega(n)$ y la envolvente instantánea $a(n)$ como:

$$\Omega(n) \quad = \quad \arccos\left(1 - \frac{\Psi\left[y(n)\right] + \Psi\left[y(n+1)\right]}{4\,\Psi\left[d^m(n)\right]}\right),$$

$$|a(n)| \quad = \quad \sqrt{\frac{\Psi\left[d^m(n)\right]}{1 - \left(1 - \frac{\Psi[y(n)]+\Psi[y(n+1)]}{4\,\Psi[d^m(n)]}\right)^2}},$$

donde $y(n) = d^m(n) - d^m(n-1)$ para $n = 2, \ldots, N$.

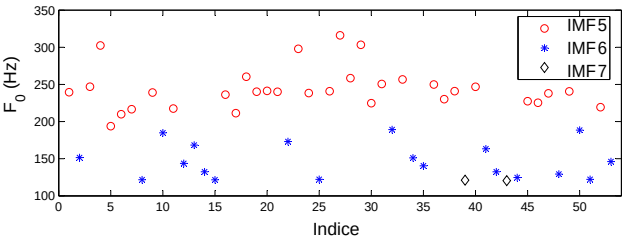

Figura 3: F_0 promedio de las 53 vocales sostenidas analizadas. Círculos (rojo), asteriscos (azul) y diamantes (negro) indican los registros donde la F_0 se halló en los modos 5, 6 y 7 respectivamente.

2. 4. Extracción de F_0 basada en EEMD

En esta sección se presentan y discuten las ideas principales del algoritmo de extracción de F_0 basado en EEMD.

Luego de aplicar la EEMD, cabe preguntarse en qué modo se encuentra *oculta* la F_0. Con esto en mente, realizamos una inspección visual de las IMFs obtenidas de cada una de las vocales sostenidas normales en nuestra base de datos y, como en el ejemplo de la Fig. 2, identificamos el modo candidato.

Para nuestras 53 voces normales, la F_0 se encontró en las IMFs 5, 6 y 7. Sólo en dos ocasiones se halló F_0 en la IMF 7, con promedios 120,394 Hz y 121,102 Hz, mientras que diecinueve veces apareció en IMF 6 con promedios entre 121,652 Hz y 189,295 Hz, y finalmente se encontró en la IMF 5 en las 32 voces restantes en un rango de F_0 entre 193,934 Hz y 316,504 Hz. En la Fig. 3 se muestra el valor promedio de la frecuencia instantánea extraída del modo así identificado para cada una de las voces sanas de la base de datos. Las voces donde se encontró la F_0 en el modo 5 se representan con círculos rojos, mientras que los asteriscos azules y los diamantes negros indican las voces donde la F_0 se halló respectivamente en los modos 6 y 7. Puede apreciarse una relación entre el valor promedio de la F_0 y el modo donde se encuentra. Este hecho concuerda con los estudios de Flandrin y col., quienes mostraron que la EMD actúa como un banco de filtros diádico y adaptativo cuando se aplica a ruido blanco [12].

A los efectos de obtener un método automático para elegir el modo donde se presenta la F_0, en este trabajo exploramos la capacidad de la entropía como una medida de información que permita discriminar en este contexto.

En la Fig. 4(a) se muestran los diagramas de cajas de la entropía discreta de Shannon (H) [13] estimada a partir del histograma con 500 particiones, para los diez primeros modos de las vocales sostenidas en las que la F_0 se encontró en el modo 5. En la Fig. 4(b) se presentan las entropías de las voces donde la frecuencia fundamental aparece en el modo 6. Se observa que el primer modo posee una entropía promedio menor que la de los siguientes cuatro o cinco modos (Figs.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

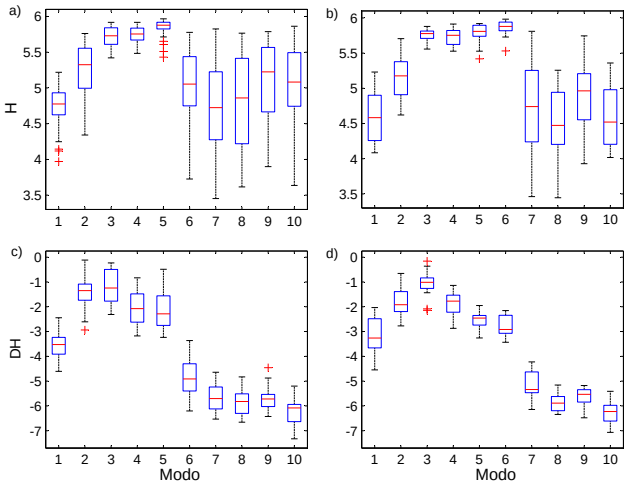

Figura 4: a, b) Entropías discretas de los modos 1 a 10 de vocales normales sostenidas /a/ en las cuales F_0 está presente en los modos 5 y 6, respectivamente. c, d) Entropías diferenciales de los modos 1 a 10 de las vocales normales sostenidas /a/ en las que F_0 se encuentra en los modos 5 y 6 respectivamente.

4(a) y 4(b), respectivamente). Esto es coherente con el hecho de que el primer modo contiene principalmente ruido de alta frecuencia: el que ya poseían los registros sumado al ruido gaussiano que se agregó en el proceso de EEMD. Es conocido que el ruido gaussiano tiene menor entropía que una senoide. Puede observarse que en el caso de las voces donde la F_0 está en la IMF 5, la entropía presenta un salto en este modo, mientras que existe un salto similar en el modo 6 para las voces donde la frecuencia fundamental se halló en la IMF 6. Existe sin embargo un solapamiento, que no aparece si utilizamos una estimación de la entropía diferencial (DH) [13] en lugar de la discreta, también utilizando 500 particiones. Los resultados mostrados en las Figs. 4(c) y 4(d) corresponden respectivamente a voces donde la frecuencia fundamental apareció en los modos 5 y 6. Debemos señalar aquí que las IMFs obtenidas a partir de EEMD en el caso de voces normales, tienen una morfología sinusoidal. Más aún, la función de densidad de probabilidades estimada para estas IMFs son también similares a las correspondientes a sinusoidales. La entropía diferencial de una función senoidal con amplitud A viene dada por $DH = \ln(\pi A/2)$. Es razonable por lo tanto pensar que el logaritmo de la potencia de las IMFs pueda usarse como un buen índice para encontrar el modo donde se encuentra la F_0. Esta línea será abordada en futuros trabajos.

Tomando en cuenta estos resultados, para los modos $m = 5, 6$ y 7, podemos proponer los umbrales T_5, T_6 y T_7 de la siguiente manera: $-3{,}365 < T_5 < -3{,}234$, $-4{,}224 < T_6 < -3{,}433$ y $-5{,}762 < T_7 < -4{,}172$. Es así que, si la DH del modo 5 es mayor que T_5 mientras que la DH del modo 6 es menor que T_5, se espera que F_0 esté en el modo 5. Si esto no ocurre, debemos

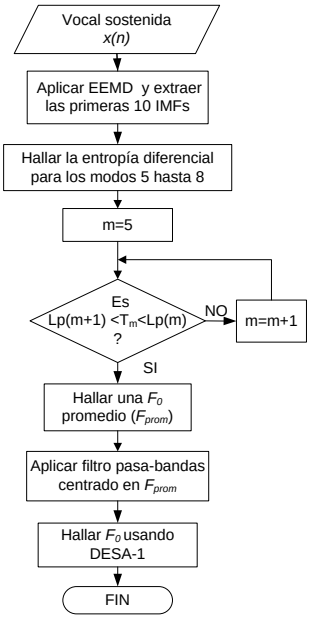

Figura 5: Extracción de F_0 basada en EEMD.

comprobar la existencia de un salto entre los modos 6 y 7 utilizando el umbral T_6, y luego entre los modos 7 y 8 mediante el umbral T_7. Esta hipótesis deberá ser confirmada con un estudio sobre una base de datos más extensa, lo que permitirá además ajustar los umbrales. Una vez seleccionado el modo donde se espera que se encuentre la F_0 , y con el propósito de eliminar componentes espurios, se aplica un filtro Chebyshev Tipo II, centrado en la frecuencia correspondiente al máximo del espectro de dicho modo. Como se muestra en la Fig. 2(b) esta frecuencia es una buena aproximación a la media de F_0. El ancho de banda del filtro es de 150 Hz. Posteriormente se aplica un algoritmo de separación AM-FM. En nuestro caso utilizamos el DESA-1 descripto en la sección anterior, como último paso de nuestro algoritmo propuesto para la extracción de F_0. En la Fig. 5 se presenta el diagrama de flujo del algoritmo completo.

3. RESULTADOS

Con objetivo de ilustrar el funcionamiento del algoritmo propuesto, en la Fig. 6 se presenta en rojo la frecuencia fundamental estimada para dos vocales sostenidas de individuos con voces sanas de la base de datos [8]. La Fig. 6(a) corresponde al registro EDC1NAL y la Fig. 6(b) al registro JTH1NAL. Se muestran también los resultados obtenidos con dos métodos conocidos a los efectos de realizar una comparación. Estos métodos son RAPT (negro)[14], con la implementación del Toolkit para Matlab VOICEBOX[2] y un método basado en AC (azul) [15][3]. Se utilizaron los parámetros por defecto en estos dos últimos algoritmos. Puede verse que los resultados son similares.

[2]VOICEBOX toolkit v. 1.18 (2008), disponible en `http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html`.

[3]PRAAT v. 5.0.32, disponible en `http://www.praat.org`.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

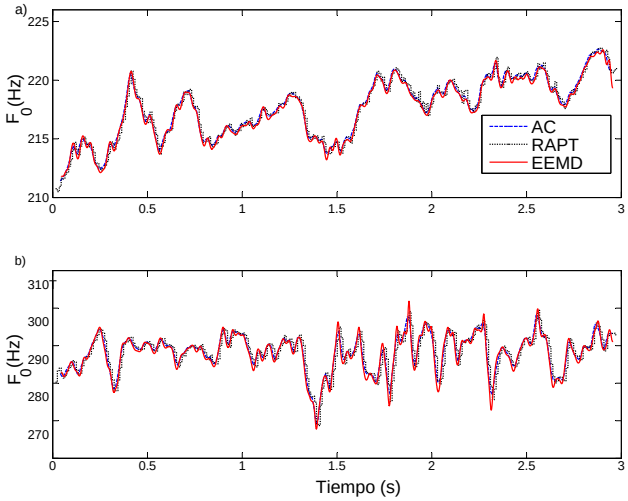

Figura 6: F_0 de dos vocales sostenidas /a/ de individuos con voces sanas: (a) EDC1NAL y (b) JTH1NAL. Método basado en EEMD (rojo), RAPT (negro) y AC (azul).

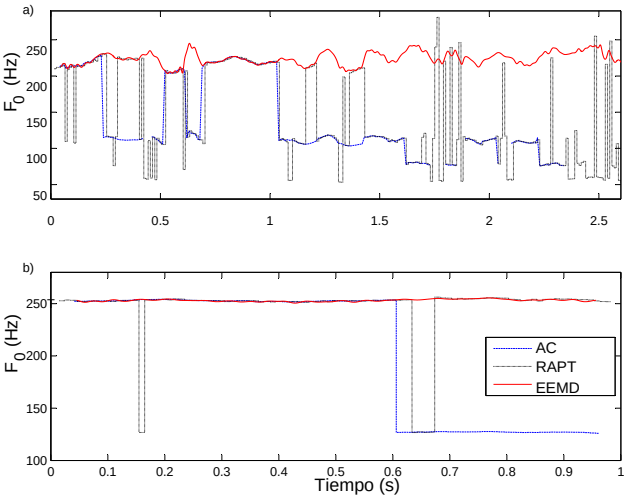

Figura 7: F_0 de dos vocales sostenidas correspondientes a pacientes con a) disfonía por tensión muscular y b) disfonía espasmódica. Método basado en EEMD (rojo), RAPT (negro) y AC (azul).

El coeficiente de correlación de Pearson entre la F_0 promedio de cada una de las 53 vocales sostenidas de pacientes sanos que indica la base de datos [8] y la frecuencia fundamental promedio hallada con el método propuesto fue $r = 0{,}999995$.

En la evaluación clínica de patologías vocales, es común el estudio de una gran cantidad de parámetros que son extraídos a partir de una estimación de la frecuencia fundamental [2]. Resulta entonces de gran importancia la estimación de F_0 de forma robusta y confiable. Desafortunadamente, no existen métodos para esta tarea que operen de forma consistente en el caso de voces patológicas. Esto se debe en parte a las complejas irregularidades de la vibración de las cuerdas vocales, mucho más pronunciadas en el caso de individuos con voces patológicas que en aquellos con voces sanas. Son varias las dificultades que surgen al estimar la F_0, pero los errores más notorios se manifiestan como una estimación de la longitud del período fundamental del doble o de la mitad del valor real. Estos problemas se intensifican de manera importante en los casos patológicos.

En la Fig. 7 se muestra la F_0 correspondiente a dos voces patológicas. En la Fig. 7(a) se analiza la frecuencia fundamental de la vocal sostenida /a/ de un paciente que padece disfonía por tensión muscular. Por otro lado, la Fig. 7(b) muestra la F_0 de un paciente con disfonía espasmódica de aducción. Como en la Fig. 6, la frecuencia fundamental hallada por nuestro método se presenta en rojo, mientras que con negro y azul se indican respectivamente los resultados hallados con RAPT y AC. Aunque los métodos basados en la autocorrelación han sido reportados como las mejores técnicas para el análisis de vocales /a/ sostenidas en casos patológicos [16], puede apreciarse fácilmente en la Fig. 7 que éste falla en varias ocasiones. También falla el algoritmo RAPT, mientras que el aquí propuesto exhibe claramente un mejor desempeño.

En un estudio con 35 vocales sostenidas de pacientes con desórdenes de la voz (15 con disfonía por tensión muscular y 20 con disfonía espasmódica de aducción), se observó que en la tarea de extraer correctamente la F_0 el algoritmo RAPT y el basado en autocorrelación fallaron en 22 voces (62,86 %). El método propuesto en el presente trabajo redujo el número de fallas a 10 voces (28,57 %). La estimación de F_0 se consideró fallida cuando se observó al menos un evento de duplicación o reducción a la mitad en la estimación del período fundamental, o cuando se detectaron artefactos en forma de espigas (como el que se muestra como ejemplo en la Fig. 8). Este último tipo de artefactos se presentaron en el método aquí propuesto, y fueron coincidentes con segmentos de voces patológicas de muy baja energía. Con el objetivo de detectar estos segmentos y para prevenir este tipo de errores en la estimación de la F_0, consideramos que podría aplicarse un método de detección de actividad vocal (*voice activity detection*, VAD) como una etapa previa de procesamiento. Sin embargo, las fallas de los otros algoritmos han resultado mucho más evidentes y no sólo asociadas a variaciones de energía. Es importante resaltar que la longitud total de los segmentos donde los métodos RAPT y de AC fallaron, exceden ampliamente la de los segmentos donde falló el método basado en EEMD. Por tal motivo, su ventaja sería mucho más notoria si se utilizara como medida de comparación el porcentaje de longitud de señal donde las estimaciones de F_0 son satisfactorias. Este enfoque se explorará en futuros trabajos.

4. DISCUSIÓN Y CONCLUSIONES

En este trabajo presentamos las capacidades de la EEMD para extraer la F_0 de vocales sostenidas /a/,

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
G. Schlotthauer, M. E. Torres & H. L. Rufiner; "Determinación de la frecuencia fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías"
XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009), sep, 2009.

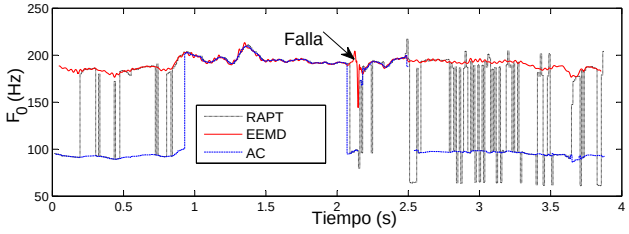

Figura 8: F_0 una vocal sostenida patológica /a/ estimada con EEMD (rojo), RAPT (negro) y AC (azul). Pese a observarse una falla en el método basado en EEMD alrededor de $t = 2{,}1$ s, los otros dos métodos fallan de manera más notoria (estimación del doble o mitad del período).

en combinación con un algoritmo de estimación de la frecuencia instantánea (DESA-1). Además, se propuso una técnica basada en entropía diferencial para la selección automática del modo a partir del cual se puede extraer la frecuencia fundamental. El nuevo método se probó con éxito sobre voces normales y patológicas y se comparó con otros algoritmos. El algoritmo basado en EEMD tiene la ventaja de no requerir el ajuste de parámetros, propiedad interesante para operadores que no sean expertos en el área, tales como los foniatras, potenciales usuarios de una aplicación. Estos resultados preliminares sugieren que la técnica propuesta provee mejoras de importancia en el área y nos impulsan a continuar investigando estas ideas. Aunque resultan prometedoras, las conclusiones de este trabajo necesitan una prueba estadística sobre una base de datos más extensa. En un futuro se abordarán también señales de habla continua y con ruido.

AGRADECIMIENTOS

Este trabajo fue realizado con el apoyo de la UNER, la UNL, CONICET y la ANPCyT. Los autores agradecen a la Dra. María C. Jackson Menaldi del Lakeshore Ear, Nose and Throat Center, St. Clair Shores (USA), y de Wayne State University, Detroit (USA) por sus valiosos comentarios y a Kay Elemetrics Corp.

Referencias

[1] WJ Hess. *Springer Handbook of Speech Processing*, Pitch and Voicing Determination of Speech with an Extension Toward Music Signals, 181–208. Springer, 2008.

[2] G Schlotthauer, ME Torres, M Jackson-Menaldi. A pattern recognition approach to spasmodic dysphonia and muscle tension dysphonia automatic classification. *J Voice*, 2009. En Prensa.

[3] NE Huang y col. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Royal Society A*, 454:903–995, 1998.

[4] D Dimitriadis, P Maragos. Continuous energy demodulation methods and application to speech analysis. *Speech Comm*, 48(7):819–837, 2006.

[5] H Huang, J Pan. Speech pitch determination based on Hilbert-Huang transform. *Signal Proc*, 86(4):792–803, 2006.

[6] H Weiping y col. A novel pitch period detection algorithm bases on HHT with application to normal and pathological voice. En *Proc 27th Annual Int Conf Eng Med & Biol Soc IEEE-EMBS*, 4541–4544, Shangai, 2005.

[7] Z Wu, NE Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1–41, 2009.

[8] Kay Elemetrics Corp. Disordered voice database 1.03. Massachusetts Eye and Ear Infirmary, Voice and Speech Lab, Boston, 1994.

[9] G Rilling, P Flandrin, P Gonçalvès. On empirical mode decomposition and its algorithms. En *Proc IEEE-EURASIP Workshop NSIP-03*, Grado, Italia, 2003.

[10] M Diaz, R Esteller. Comparison of the non linear energy operator and the hilbert transform in the estimation of the instantaneous amplitude and frequency. *IEEE Latin America Trans*, 5(1):1–8, 2007.

[11] P Maragos, JF Kaiser, TF Quatieri. Energy separation in signal modulations with application to speech analysis. *IEEE Trans Signal Proc*, 41(10):3024–3051, 1993.

[12] P Flandrin, G Rilling, P Gonçalvès. Empirical mode decomposition as a filter bank. *IEEE Signal Proc Letters*, 11(2):112–114, 2004.

[13] A Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3ra ed, 1991.

[14] D Talkin. *Speech Coding and Synthesis*, A robust algorithm for pitch tracking (RAPT), 121–173. Elsevier, 1995.

[15] P Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. En *Proc Institute of Phonetic Sciences*, (17):97–110, 1993.

[16] SJ Jang y col. Evaluation of performance of several established pitch detection algorithms in pathological voices. En *Proc 29th Annual Int Conf IEEE Eng Med & Biol Soc*, 620–623, Lyon, 2007.